



Intelligent Fault Diagnosis Method of Wind Turbines Planetary Gearboxes Based on a Multi-Scale Dense Fusion Network

Xinghua Huang¹, Yuanyuan Li² and Yi Chai^{1*}

¹College of Automation, Chongqing University, Chongqing, China, ²College of Automation, Chongqing University of Posts and Telecommunications, Chongqing, China

OPEN ACCESS

Edited by:

Yolanda Vidal,
Universitat Politècnica de Catalunya,
Spain

Reviewed by:

Davide Astolfi,
University of Perugia, Italy
Kang Ding,
South China University of Technology,
China
Yongbo Li,
Northwestern Polytechnical
University, China

*Correspondence:

Yi Chai
chaiyi@cqu.edu.cn

Specialty section:

This article was submitted to
Wind Energy,
a section of the journal
Frontiers in Energy Research

Received: 26 July 2021

Accepted: 25 October 2021

Published: 29 November 2021

Citation:

Huang X, Li Y and Chai Y (2021)
Intelligent Fault Diagnosis Method of
Wind Turbines Planetary Gearboxes
Based on a Multi-Scale Dense
Fusion Network.
Front. Energy Res. 9:747622.
doi: 10.3389/fenrg.2021.747622

Due to the powerful capability of feature extraction, convolutional neural network (CNN) is increasingly applied to the fault diagnosis of key components of rotating machineries. Due to the shortcomings of traditional CNN-based fault diagnosis methods, the continuous convolution and pooling operations result in the constant decrease of feature resolution, which may cause the loss of some subtle fault information in the samples. This paper proposes a CNN-based model with improved structure multi-scale dense fusion network (MSDFN) to realize the fault diagnosis of wind turbines planetary gearboxes under complicated working conditions. First, the continuous wavelet transform is applied to preprocess the vibration signals, and the two-dimensional wavelet time-frequency diagrams are used as the network input. Then, the multi-scale feature fusion (MSFF) module and a feature of maximum (FoM) module are used in the extraction and classification stages of fault features, respectively. Next, the multi-scale features of each network layer are fused to enhance the fault features. Finally, the high fault diagnosis accuracy is achieved by extracting the separable fusion result of fault features. The proposed method achieves more than 99% fault diagnosis average accuracy on a planetary gearbox dataset. The comparative experimental results verify the effectiveness of the proposed method and its superiority to some mainstream approaches. The ablation study further confirms that MSFF module and FoM module play the positive role in fault diagnosis.

Keywords: wind turbines planetary gearbox, fault diagnosis, convolutional neural network, feature fusion, wavelet transform

1 INTRODUCTION

Planetary gearbox is a key component in the transmission system of wind turbines (WT) (Feng and Liang, 2014; Wang et al., 2019). Due to the harsh working environment and complex structure, the key gear components in the wind turbine planetary gearboxes are prone to damage, which adversely affect the entire transmission system. Since wind turbines are often installed in places with inconvenient transportation (Lu et al., 2020; Sun et al., 2021), any gear fault of planetary gearboxes may cause the long downtime of the corresponding wind turbine and the high cost of the related operation, maintenance, and reparation (Cao et al., 2019; Sun et al., 2019). During the service life of a wind turbine, the cost of the related maintenance and operation account for about

75% of the total investment (Lin et al., 2018; Zhu et al., 2021). The monitoring of the health status of each planetary gearbox plays an important role in the normal operation of a wind turbine. The gear fault types of wind turbine planetary gearboxes mainly include chipped tooth, missing tooth, crack tooth, and surface wear (Wang et al., 2018; Liang et al., 2020). At present, most of the fault diagnosis research on the gears of planetary gearboxes are based on vibration signals (Lei et al., 2020).

Due to no-stationary working condition and complex structure, it is extremely difficult to establish a general mathematical model for the vibration signals of planetary gearboxes (Feng and Zuo, 2012). Since the vibration signals of planetary gearboxes have three main features, composite signals, pass-through effect, and nonlinearity, it is difficult to directly extract fault features by observing vibration responses (LI et al., 2017). In addition, traditional signal processing methods are difficult to process the monitoring data with massive states in time in modern industry (Pan et al., 2019). Therefore, deep learning-based methods with powerful feature extraction capabilities are increasingly applied to the fault diagnosis of key components of rotating machinery. They usually have three main steps, data preprocessing, fault feature extraction, and fault classification (Liu et al., 2018; Ma et al., 2019; Liang et al., 2020).

Deep learning-based fault diagnosis methods often preprocess the original vibration signals first. Wavelet transform, which has excellent time-frequency analysis ability for non-stationary vibration signals, is often applied to fault diagnosis of rotating machinery. The optimized Morlet wavelet transform is used to process the vibration signals to obtain better time- and frequency-domain statistical feature sets (Wang et al., 2018). Fault diagnosis can also be well achieved by wavelet packet coefficient matrix of vibration signals (Zhao et al., 2017) or time-frequency images (Liang et al., 2020; Zhao et al., 2020c; Cheng et al., 2021).

Since deep learning can automatically extract abstract features from the original data (Wen et al., 2017), various deep learning-based models are often applied to extract and classify fault features. These methods can effectively avoid using complex signal processing methods to calculate feature parameters for the expression of fault information. According to vibration signals or statistical feature sets, frequency spectrum, and time-frequency spectrum of vibration signals, DBN (Wang et al., 2018; Kang et al., 2020), LSTM (Cao et al., 2019), SAE (Jiang et al., 2017), RNN (Miao et al., 2020), and CNN (Jiang et al., 2018; Wang et al., 2020a) can obtain relatively high fault diagnosis accuracy.

Considering the powerful feature extraction capabilities of deep learning, this paper proposes a CNN-based fault diagnosis model for wind turbine planetary gearboxes. The proposed model aims to use CNN to extract fault features from the time-frequency images of vibration signals and achieve fault classification. However, in actual wind turbine applications, due to noise interference and difficulty in determining the impact of faults, etc., the fault information of the vibration signals of wind turbine planetary gearboxes mapped on time-frequency images may be extremely subtle, especially in the early-stage faults (Wei et al., 2019). Following the continual

decrease of feature resolution, some important fault information may be lost to cause adverse impact (Wang et al., 2020b), when CNN is used to extract fault features. In the traditional CNN methods, the learning of each network layer is always only based on the features learned in the previous layer. If partial information of the learned features is lost, the features learned later is adversely affected. To solve the above issue, this paper proposes an intelligent fault diagnosis method based on multi-scale dense fusion network (MSDFN).

MSDFN consists of two parts, feature extraction network and classifier. A dense feature fusion structure Huang et al. (2017) based on projection and back-projection operators Irani and Peleg (1991). Dai et al. (2007) is used to optimize the traditional CNN network, and the output of each layer of the feature extraction network is fused as the input of the classifier. Through the dense fusion of multi-scale features to supplement the fault information, the fault diagnosis of wind turbine planetary gearboxes under complicated working conditions is realized. The main contributions of this paper can be summarized as follows.

- 1) A CNN-variant MSDFN is proposed for fault diagnosis of wind turbine planetary gearboxes under complex working conditions. It could extract enhanced fusion results of fault features from the time-frequency images of vibration signal to improve diagnosis accuracy.
- 2) The MSFF module designed by projection and back-projection operators are embedded in each network layer. And the fault feature enhancement algorithm based on multi-scale feature fusion is used to supplement the missing information of every layer in time. The fused features can express fault information more effectively and have stronger separability.
- 3) A FoM module is designed to fuse the output fault features of each feature extraction layer. Specifically, this module uses adaptive maximum pooling to convert the features of each layer to the same resolution and concatenate them. It makes the input features of the classifier have more complete fault information, and the corresponding diagnosis accuracy is improved.

The rest of this paper is organized as follows. **Section 2** reviews the CNN-based fault diagnosis research of key components of rotating machinery in recent years; **Section 3** describes the proposed fault diagnosis method based on MSDFN; **Section 4** compares the proposed method with several existing fault diagnosis networks to verify its effectiveness and also conducts the corresponding ablation study to test the performance of each module; and **Section 5** concludes this paper.

2 RELATED WORK

In existing deep learning-based fault diagnosis research of rotating machinery, CNN is one of the most commonly used deep learning models for fault diagnosis. Compared with SAE, DBN, and other models, CNN and its variants such as Deep

residual network (DRN), deep convolutional neural network (DCN), etc. are more convenient for training (Wen et al., 2017; Zhao et al., 2017), when vibration signals or their time-frequency features are used as input. Because their local receptive fields and weight sharing strategy usually only needs a smaller amount of parameters. In recent years, many CNN-based fault diagnosis solutions of rotating machinery have been published.

CNN was applied to the fault diagnosis of rotating machinery and fault features were extracted from the spectrum of vibration signals, which achieved better performance than classical classifiers such as random forest and SVM (Janssens et al., 2016). Two-dimensional DCN was used to extract fault features from the wavelet packet energy images of vibration signals, and the corresponding method achieved high fault diagnosis accuracy (Ding and He, 2017). As the input time-frequency information composed of the original vibration signals and its frequency spectrum, a one-dimensional CNN was applied to the fault diagnosis of planetary gearboxes (Jing et al., 2017). Fault features were extracted from multi-sensor data by a CNN with the multi-input branch structure to realize the fault diagnosis of rotating machinery (Xia et al., 2017). A one-dimensional CNN-based method realized the end-to-end fault diagnosis of rotating machinery (Wu et al., 2019). Discrete wavelet transform was used to obtain the time-frequency matrix of the vibration signals, and CNN was applied to extract the fault features of planetary gearboxes (Chen et al., 2019). Vibration signals were first analyzed by recursive graphs, and then CNN was used to achieve the fault diagnosis of rotating machinery according to the obtained recursive matrix (Wang et al., 2020a). Since the second-order cyclostationary behavior of vibration signals can reveal valuable health information, cyclic spectral coherence (CSCoh) analysis of vibration signals was used to preprocess the original data, which reduced the difficulty of feature learning in deep diagnosis model and improved diagnosis accuracy (Chen et al., 2020).

The above work focuses on the preprocessing methods of raw data, and studies the effects of various data processing methods on fault diagnosis. However, the related research not only relies on rich knowledge of vibration signal processing, but also increases the workload of data processing. The following research proves that the improvement of network structure is conducive to extracting detailed fault features and achieving high diagnostic accuracy. Zhao proved that deep residual network (DRN) can efficiently extract the high-level fault features contained in the wavelet packet coefficients (Zhao et al., 2017; Zhao et al., 2020b). On the basis of DRN, the dynamic weight module was introduced to weight the fault features of different frequency bands in the time-frequency images, which improved the diagnosis accuracy (Zhao et al., 2017). A multi-branch CNN network structure was used to extract features from different scales and improve the diagnostic ability of the model (Pan et al., 2019; Peng et al., 2020). MSCNN used multiple CNN branches to process vibration signals from multiple scales and the fault diagnosis accuracy of wind turbine planetary gearboxes was improved (Jiang et al., 2018). An SE-Res module (Hu et al., 2018) was added to the ordinary CNN network to reduce the interference of redundant information and enhance fault features

(Cao et al., 2020). A CNN model using hollow convolutions was applied to increase the receptive fields to improve the gear fault diagnosis accuracy of planetary gearboxes (Han et al., 2019). To accurately and automatically identify the health status of rotating machinery, a normalized convolutional neural network was proposed for the fault type diagnosis of rotating machinery under variable operating conditions (Zhao et al., 2020a). A multi-core cascade structure of CNN was used to substitute a single core for fault diagnosis (Wang et al., 2020b). Xu developed a new method VMD-DCNNs that integrated convolutional neural networks with variational mode decomposition (VMD) algorithms (Xu et al., 2020). This method used CNN to extract features from each intrinsic mode function (IMF) and directly processed the original vibration signals in an end-to-end manner without any manual experiences and intervention to realize the fault diagnosis of the key components of wind turbines.

All the above methods use the CNN model as the fault feature extractor and classifier, and achieve good performance in fault diagnosis of the key components of rotating machinery. However, these methods ignore that when CNN performs feature extraction, the reduction of feature resolution may cause the loss of partial fault information and even the decrease of fault diagnosis accuracy (Wang et al., 2020b). The dense feature fusion structure makes the input passed to the next network layer come from all the extracted features, which can effectively supplement the fault information. Therefore, according to the inspiration and guidance of the above-mentioned work, this paper focuses on improving the network by using the feature fusion structure, and develops a multi-scale dense fusion network (MSDFN). The proposed model can extract the enhanced fusion results of fault features from the time-frequency images of vibration signals of wind turbine planetary gearboxes under complicated working conditions, so the relatively complete fault information can be obtained, and the fault diagnosis accuracy can be improved. **Section 3** specifies the overall structure and working principle of MSDFN.

3 MSDFN-BASED FAULT DIAGNOSIS METHOD

The fault diagnosis of wind turbine planetary gearbox is really important. However, the traditional CNN-based fault diagnosis methods may cause the loss of fault information. So, this paper proposes a MSDFN-based intelligent fault diagnosis method for wind turbine planetary gearboxes. **Figure 1** shows the diagnosis process. The proposed method uses continuous wavelet transform to preprocess the original vibration signal data. A multi-scale feature fusion (MSFF) module is embedded into each feature extraction network layer. An FoM module is used to fuse the output of each feature extraction network layer to obtain the classifier input. So, the information loss is reduced during the fault extraction, and the corresponding fault diagnosis accuracy is improved. The feature extraction network contains five layers, which can fuse features of five scales at most. A structure that is too shallow will result in poor accuracy, and a structure that is too deep will result in a significant increase in the

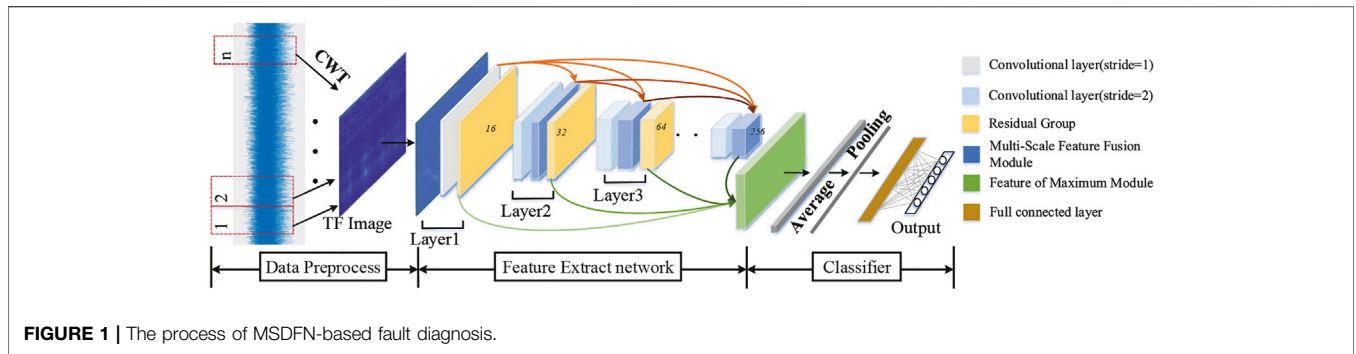


TABLE 1 | The similarity of five commonly used wavelet basis functions.

Wavelet basis function	Coif5	Meyr	Morlet	Db10	Cmor3-3
similarity	6.3298	6.6082	7.2953	7.4970	40.0507

amount of calculation and cannot improve the accuracy. This is verified in the Section 4.

3.1 Data Preprocessing

3.1.1 The Generation of Time-Frequency Images

Wavelet transform is widely used in vibration signal processing of rotating machinery due to its excellent time-frequency analysis ability for unsteady signals. The commonly used expression of wavelet transform is shown in Eq. 1.

$$WT(a, \tau) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t)\psi\left(\frac{t-\tau}{a}\right)dt \quad (1)$$

where $WT(a, \tau)$ is the wavelet coefficient, $f(t)$ is the input signal, $\psi(t)$ is the wavelet basis function, a is the scale factor that controls the expansion and contraction of the wavelet basis function, and τ is the translation amount that controls the translation of the wavelet basis function. The scale corresponds to frequency, and the amount of translation corresponds to time. When the wavelet function is translated on the time axis at each scale, it is multiplied by the input signals respectively. So, the frequency components of the signals in each time period can be obtained.

The selection of wavelet basis function is an important step in using wavelet transform. Similarity coefficients (Mao et al., 2019; Liang et al., 2020) are applied to select the wavelet basis function for quantitative analysis. The expression of similarity coefficients is shown in Eq. 2.

$$\delta = \sum_{i=1}^k \alpha_i \frac{m_i^2}{s_i} \quad (2)$$

where δ is the dimensionless similarity coefficient, k and s_i, m_i, α_i is the number of peaks, area of each peak, maximum of each peak, weighted coefficient of each peak after making absolute value for wavelet basis function.

Table 1 lists the similarity coefficients of several commonly used wavelets for fault diagnosis. When the similarity coefficient

of a wavelet increases, the wavelet gets close to the original signals, which means the wavelet contains more fault information to facilitate diagnosis. Therefore, the cmor wavelet defined in Eq. 3 is used. The complex Morlet wavelet is a complex sinusoid modulated by a Gaussian envelope defined by:

$$cmor(x) = (\pi F_b)^{-\frac{1}{2}} \exp(2i\pi F_c x) \exp\left(\frac{-x^2}{F_b}\right) \quad (3)$$

where F_b is the bandwidth of the wavelet, F_c is the central frequency of the wavelet function, and i is the imaginary unit. This paper takes $F_c = F_b = 3$.

3.1.2 Time-Frequency Image Preprocessing

Time-frequency images need to be preprocessed to improve the training performance before inputting into the network. For image datasets, the commonly used data enhancement methods include rotation, flipping, and random cropping. As the purpose of data enhancement, diverse training samples enable the network to extract key features from the samples undergoing various changes for improving the generalization ability of the network. However, since the vibration signal sequence is time-dependent, it can be regarded as a discrete time function. Flipping, rotation, and random cropping disrupt the relationship between time and frequency features, resulting in poor training performance. Therefore, the proposed image processing method first transforms the size of input images, and then normalizes them to improve the training speed.

Bilinear interpolation is used to convert the resolution of time-frequency images. Bilinear interpolation is the expansion of two-variable linear interpolation, which performs linear interpolation in two directions to obtain the value of the unknown function $f(x, y)$ at the point $P = (x, y)$. The transformation expression of bilinear interpolation is shown in Eq. 4.

$$f(x, y) = \sum_{i=1}^2 \sum_{j=1}^2 \left| \frac{f(Q_{ij})(x_i - x)(y_j - y)}{(x_2 - x_1)(y_2 - y_1)} \right| \quad (4)$$

where $Q_{11}(x_1, y_1), Q_{21}(x_2, y_1), Q_{12}(x_1, y_2),$ and $Q_{22}(x_2, y_2)$ are the four points closest to the target point $f(x, y)$, and $f(Q_{ij})$ represents the value of the point $Q_{ij}, x_{i,j}$, and $y_{i,j}$ represent the abscissa and ordinate of the point Q_{ij} , respectively.

Image normalization is used to scale the value of each image pixel to a small specific interval, remove the data unit limit, and

TABLE 2 | Detailed structure of MSDFN. “Conv”, convolutional layer; “RG”, residual group; “AMP”, adaptive maximum pooling; “AGAP”, adaptive global average pooling; “Fc”, fully connected layer; “numclass”, the number of fault categories.

Layer name	Structure	Input channels, output channels	Output size (Input size = 256)
Layer1	Conv1	3, 16	256 × 256
	RG1	16, 16	256 × 256
Layer2	Conv2	16, 32	128 × 128
	MSFF1	(16,32), 32	128 × 128
	RG2	32, 32	128 × 128
Layer3	Conv3	32, 64	64 × 64
	MSFF2	(16,32,64), 64	64 × 64
	RG3	64, 64	64 × 64
Layer4	Conv4	64, 128	32 × 32
	MSFF3	(16,32,64,128), 128	32 × 32
	RG4	128, 128	32 × 32
Layer5	Conv5	128, 256	16 × 16
	MSFF4	(16,32,64,128,256), 256	16 × 16
FoM	AMP	(16,32,64,128,256), 496	16 × 16
GAP	AGAP	496, 496	1 × 1
Classifier	Fc	–	1 × numclass

convert it into a dimensionless pure value, which facilitates the comparison, and weighting of the indicators of different units or magnitudes. The normalization formula used in this paper is shown in Eq. 5.

$$\bar{x} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (5)$$

where the pixel value of a point is converted from x to \bar{x} , x_{\max} and x_{\min} are the maximum and minimum values of the sample image pixel values, respectively. The values of all points can be converted to the interval of 0–1 by normalization, which improves the convergence speed and accuracy of the model.

3.2 MSDFN

As shown in Figure 1, the MSDFN consists of two parts: a feature extraction network and a classifier. Table 2 shows the specific composition of MSDFN. In Table 2, “Conv” refers to convolutional layer; “RG” refers to residual group; “AMP” refers to adaptive maximum pooling; “AGAP” refers to adaptive global average pooling; “Fc” refers to fully connected

layer; “numclass” refers to the number of fault categories. The convolutional layer of the feature extraction network layer is used to transform the channel and reduce the feature resolution, the Residual Group (RG) is used for fault feature extraction, and the MSFF module is used for multi-scale feature fusion. FoM merges the output features of each layer of the feature extraction network again, and inputs it into the classifier constructed by a fully connected layer to obtain a prediction vector. The residual group as shown in Figure 2A is composed of three residual blocks as shown in Figure 2B. ReLU function is be used as the activation function. Although multi-scale feature fusion creates redundancy, it supplements more complete fault information, so that the subsequent network layer can extract more comprehensive fault information, which enhancing the extracted fault features and improving diagnosis accuracy.

3.2.1 MSFF Module

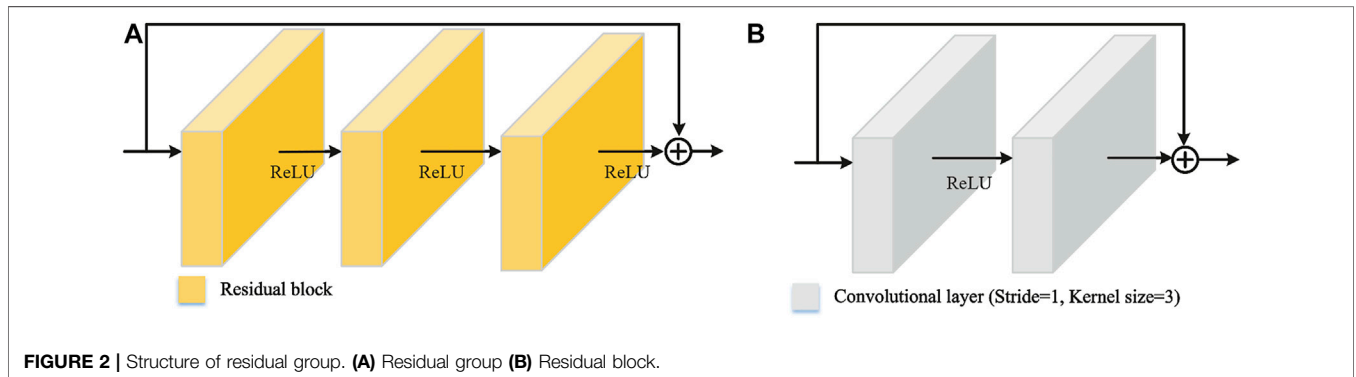
According to the feature fusion method proposed in (Dong et al., 2020), the MSFF module is proposed to apply projection and back projection operators to the fault diagnosis of planetary gearboxes. As shown in Figure 1, the MSFF module fuses the features of all layers to supplement important time-frequency and fault information. Figure 3 shows the structure of the MSFF module. The MSFF module of the n -th network layer is defined as Eq. 6.

$$\tilde{j}^n = D^n(j^n, \{\tilde{j}^1, \tilde{j}^2, \dots, \tilde{j}^{n-1}\}) \quad (6)$$

where j^n is the latent feature of the n -th feature extraction network layer, \tilde{j}^n is the enhanced features obtained through dense fusion, and $\{\tilde{j}^1, \tilde{j}^2, \dots, \tilde{j}^{n-1}\}$ are the enhanced fusion features from n MSFF modules before this layer in the network. This paper uses the enhanced features $\tilde{j}^{n-t}, t \in \{1, 2, \dots, n-1\}$, t times to gradually improve the enhanced features j^n . The specific update and improvement process is shown as follows.

- 1) As shown in Eq. 7, the difference e_t^n between the enhanced feature \tilde{j}_t^n of the t -th iteration and the t -th enhanced feature \tilde{j}^t is calculated.

$$e_t^n = q_t^n(j_t^n) - \tilde{j}^t \quad (7)$$



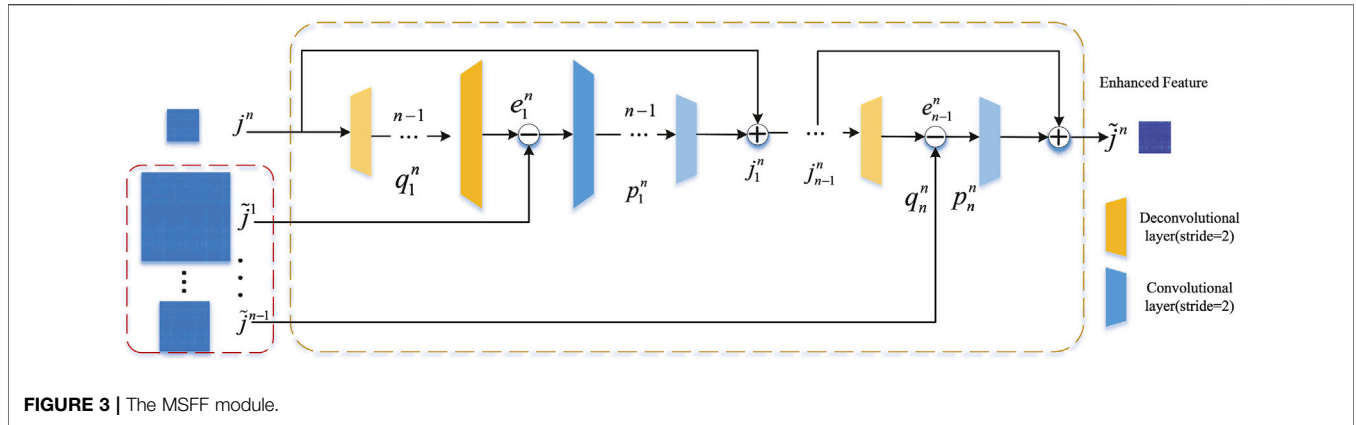


FIGURE 3 | The MSFF module.

where q_t^n represents the back projection operator, which upsamples the promoted feature j_t^n to the same dimension as \tilde{j}^t .

- 2) The enhanced feature \tilde{j}_t^n is updated through back projection for the difference calculated by Eq. 7, as shown in Eq. 8.

$$\tilde{j}_{t+1}^n = p_t^n(e_t^n) + \tilde{j}_t^n \quad (8)$$

where p_t^n is the projection operator, which downsamples the difference e_t^n in the t -th iteration to the same dimension of \tilde{j}_t^n .

- 3) After the iteration of all previous enhanced features, the enhanced feature \tilde{j}^n is finally obtained.

Unlike the traditional back projection techniques, the sampling operators q_t^n and p_t^n in the network are unknown. The proposed method uses strided convolution (deconvolution layer) to learn the downsampling (upsampling) operator in an end-to-end manner. In order to avoid introducing too many parameters, this paper uses 2 as the stride size and stacks the convolutional and deconvolutional layers of $n - t$ -th layer together to achieve downsampling and upsampling learning in q_t^n and p_t^n .

In summary, the multi-scale feature fusion (MSFF) algorithm used to enhance fault features is described in Algorithm 1.

Algorithm 1. Fault feature enhancement algorithm based on multi-scale feature fusion

```

Input: feature extraction network  $N$ , time-frequency images, back projection operator  $q_t^n$ , projection operator  $p_t^n$ , number of network layers  $L$ 
Output: feature extraction network  $N'$ , optimized back projection operator  $q_t^n$ , projection operator  $p_t^n$ , fused feature  $\tilde{j}^n$ 
if  $n = 0$  then
  Extract the hidden feature at 0-th network layer
end if
for  $n = 0$  to  $L$  do
  Convolution feature extraction:
  Extract the hidden feature  $j^n$  at 0-th network layer
  MSFF fusion feature extraction:
  for  $t = 0$  to  $t = n - 1$  do
     $\tilde{j}_t^n = j^n$ 
    Calculate the difference  $e_t^n$  between the enhanced feature  $\tilde{j}_t^n$  of the  $t$ -th iteration and the  $t$ -th enhanced feature  $j^t$  according to Eq. 7
    According to Eq. 8, the enhanced features  $\tilde{j}_t^n$  and  $\tilde{j}^t$  are updated through the back projection for the difference calculated by Eq. 7
  end for
end for
  
```

3.2.2 FoM

The MSFF module achieves feature fusion in the feature extraction process, but the issue of incomplete information still exists when the network uses the extracted features to classify faults. Therefore, the FoM module is designed before the classifier part, and its definition fom is shown in Eq. 9. As shown in Figure 4A, the FoM module is used to convert the output features of each network layer to a uniform size and concatenate on the channel dimensions to achieve the fusion of predicted features.

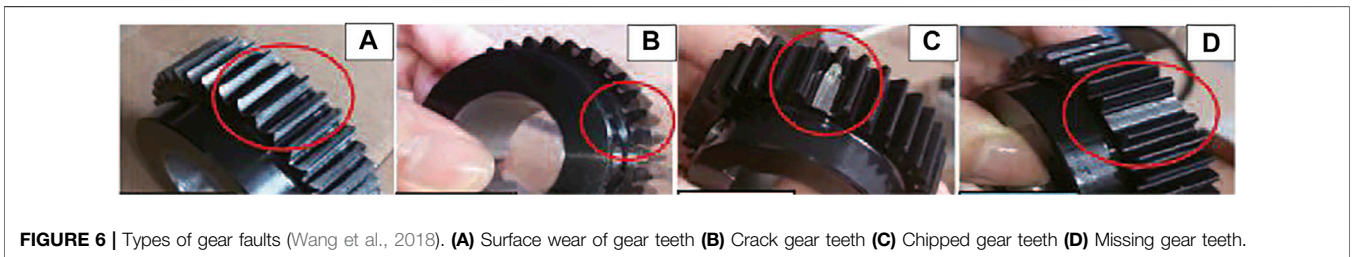
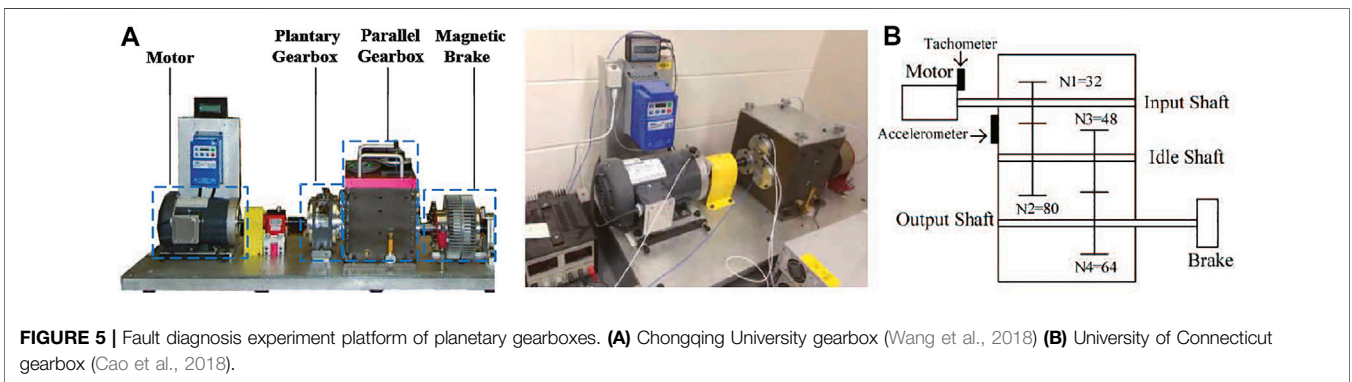
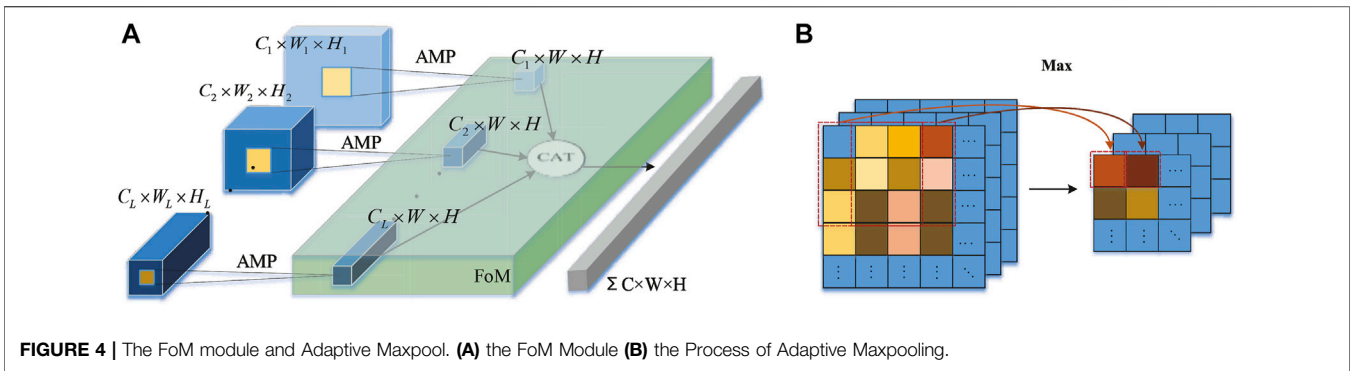
$$fom(y_1, y_2 \dots y_L) = cat(v_1, v_2 \dots v_L) \quad (9)$$

$$v_i = amp_{W \times H}(y_i) \quad (10)$$

where cat represents the concatenation of each input feature on the channel dimensions, and $amp_{W \times H}$ represents the adaptive maximum pooling operation with an output size of $W \times H$. As shown in Figure 4B, the features of each scale are converted into feature vectors with the same resolution and different channel numbers to represent the diagnosis result obtained by each layer. $\{y_1, y_2 \dots y_L\}$ are the output features of the first to L -th ($L = 5$ in this paper) network layers. The FoM module performs the second-time fusion of fault features to prevent the feature vectors used for classification from affecting the diagnosis result due to insufficient or missing information.

4 EXPERIMENT

In order to verify the effectiveness of the proposed MSDFN-based wind turbine planetary gearbox fault diagnosis method and the performance of the feature fusion module MSFF and FoM modules, based on the planetary gearbox dataset of the State Key Transmission Laboratory of Chongqing University (Wang et al., 2018) and gearbox dataset of the University of Connecticut (Cao et al., 2018), comparative experiments and ablation experiments were carried out. The aim is to use the multi-stage gearbox of the experimental platform to simulate the gearbox of the wind turbine for fault diagnosis research.



4.1 An Introduction of the Dataset

The vibration signals datasets of the planetary gearboxes used in this research come from the University of Connecticut gearbox data (Cao et al., 2018) and actual experimental data of the State Key Laboratory of Mechanical Transmission of Chongqing University (Wang et al., 2018). The following describes the basic situation of the two data sets respectively.

1) Chongqing University gearbox dataset. As shown in Figure 5A, the fault diagnosis experiment platform of planetary gearboxes is mainly composed of a motor, a planetary gearbox, a parallel shaft gearbox, and an electromagnetic brake. The multi-axis accelerometer sensor is installed at the housing directly above the second-stage Sun gear of planetary gearboxes to collect the original vibration signals; and the rotation frequency of the second-stage Sun gear is set to 4.17 Hz. As shown in Figure 6, the types of gear faults in the experiments contain 1) surface wear of gear teeth, 2) cracked gear teeth, 3) chipped gear teeth, and 4) missing gear teeth. For the

TABLE 3 | Accuracy under different ratio of training and testing sets.

Train set: Test set	5: 5	6: 4	7: 3	8: 2
Train set	2,000	2,400	2,800	3,200
Test set	2,000	1,600	1,200	800
Accuracy	99.73%	99.72%	99.75%	99.74%

collection of vibration signals, the sampling frequency is set to 5,120Hz, and each group of sampling time lasts for 200 s. So, each type of fault includes 1,024,000 sampling points under each load condition. The working conditions are often unmeasured and easy to fluctuate under actual engineering conditions. To discuss the fault diagnosis performance of network models on planetary gearboxes under different load conditions, four load conditions are set to 0 N.m, 1.4 N.m, 2.8 N.m, and 25.2 N.m for each type of Sun gear fault, respectively. The collected data is processed by high-pass filtering to remove some low-frequency noise

TABLE 4 | time-frequency images datasets of Chongqing University gearboxes.

Dataset	Chipped	Crack	Missing	Normal	Surfacewear	Loadcondition	Train set	Test set
Label	0	1	2	3	4			
DatasetA	200	200	200	200	200	L1	700	300
DatasetB	200	200	200	200	200	L2	700	300
DatasetC	200	200	200	200	200	L3	700	300
DatasetD	200	200	200	200	200	L4	700	300
DatasetE	200	200	200	200	200	L1,L2,L3,L4	700	300

TABLE 5 | Time-frequency images dataset of University of Connecticut gearbox.

Dataset	Chipping1a	Chipping2a	Chipping3a	Chipping4a	Chipping5a	Crack	Healthy	Missing	Spall
Label	0	1	2	3	4	5	6	7	8
Dataset F	104	104	104	104	104	104	104	104	104

interference in the vibration signals. All the experimental data used in this paper are the time sequence of filtered vibration signals (Wang et al., 2018).

2) University of Connecticut gearbox data, the experimental data are collected from a benchmark two-stage gearbox with replaceable gears as shown in **Figure 5B**. The gear speed is controlled by a motor. The torque is supplied by a magnetic brake which can be adjusted by changing its input voltage. A 32-tooth pinion and an 80-tooth gear are installed on the first stage input shaft. The second stage consists of a 48-tooth pinion and 64-tooth gear. The input shaft speed is measured by a tachometer, and gear vibration signals are measured by an accelerometer. The signals are recorded through a dSPACE system (DS1006 processor board, dSPACE Inc.) with sampling frequency of 20 KHz. Nine different gear conditions are introduced to the pinion on the input shaft, including healthy condition, missing tooth, root crack, spalling, and chipping tip with five different levels of severity (Cao et al., 2018). This dataset contains 104 samples every class and every sample contain 3,600 points.

The gearbox data set of Chongqing University contains usual types of gear failures, including data on a wide range of working conditions, which can better simulate the variable characteristics of wind turbine gearbox load conditions Cao et al. (2019). At the same time, the gearbox data shared by the team of Professor Jiong Tang from the University of Connecticut was selected to further verify the effectiveness of the proposed method. This data set contains nine types of failures, which can better test the model's ability to distinguish samples of the same type but with different failure levels. The experimental platform configuration of these two datasets and the structure of the gearbox are relatively reasonable, which has been proved in the related wind turbine gearbox fault diagnosis researches (Jiang et al., 2018; Lu et al., 2020).

4.2 Data Preprocessing

According to the method proposed in **Section 3**, the original vibration signals are preprocessed to obtain the dataset of time-frequency images. 1) Chongqing University gearbox dataset, since

TABLE 6 | Diagnostic accuracy of MSDFN of different depth feature extraction networks.

Depth of the feature extraction network L	3	4	5	6
Accuracy	96.53%	98.12%	99.75%	99.72%
Parameters	1.53M	6.15M	24.66M	98.68M

the rotation frequency of faulty gears is 4.17 Hz, the data collected per second can contain the vibration information of multiple rotation periods. So, a sample is composed of the data per second. 200 time-frequency images are obtained for each fault type. Four load conditions are marked as L1, L2, L3, and L4. Each of them has 1,000 image samples. 2) University of Connecticut gearbox data, according to the data set structure and sampling frequency, 104 samples of each class are obtained.

Table 3 shows the average accuracy of MSDFN with different ratio of training and testing sets from the Chongqing University gearboxes dataset. It can be found that the accuracy has not changed significantly, nor has it shown an obvious monotonic trend, but fluctuates slightly as the ratio changes. Considering the number of samples, in order to obtain a higher diagnostic accuracy, all image samples are randomly divided into the training and testing sets at a ratio of 7:3. The data collected from different load conditions is combined to obtain a mixed dataset. The sample size of the obtained time-frequency images is 256×256 . **Tables 4, 5** shows the details of the obtained datasets.

4.3 Model Training

4.3.1 Hyperparameter Setting

In the training process of deep learning networks, the hyperparameter settings have considerable impact on the training performance. In the experiments, the batch size of the network input (the number of samples input to the network each time) affects the testing accuracy of the model. If the batch size is too large, the model is difficult to fit or the fitting performance is poor. If the batch size is too small, the model is difficult to

TABLE 7 | Accuracy of different methods on different dataset.

Network		Dataset A (%)	Dataset B (%)	Dataset C (%)	Dataset D (%)	Dataset E (%)	Dataset F (%)
DenseNet121, Huang et al. (2017)	Train	99.80	100.00	99.80	100.00	100.00	100.00
	Test	98.40	98.60	98.60	98.60	98.35	99.65
VGG11, Simonyan and Zisserman (2017)	Train	96.80	98.80	98.60	95.40	99.70	100.00
	Test	96.60	97.60	97.00	93.60	97.05	99.54
ResNet18, He et al. (2016)	Train	99.80	99.80	100.00	100.00	100.00	100.00
	Test	99.20	98.82	99.20	99.60	99.50	100.00
DBN, Wang et al. (2018)	Train	99.92	99.94	99.33	98.28	98.22	100.00
	Test	99.52	98.82	98.65	95.82	95.72	98.92
WT-CNN, Liang et al. (2020)	Train	100.00	100.00	100.00	99.90	99.90	100.00
	Test	99.40	98.20	98.40	98.80	97.60	99.23
DWWC + DRN, Zhao et al. (2017)	Train	99.40	100.00	100.00	99.90	99.90	100.00
	Test	98.80	98.20	98.60	99.20	98.90	99.85
MSDFN	Train	100.00	100.00	100.00	100.00	100.00	100.00
	Test	98.62	99.27	99.03	99.35	99.75	100.00

converge, and the accuracy rate oscillates unevenly. Considering the number of samples and the testing results, 16 is chosen.

The learning rate is another important hyperparameter. The learning rate represents the updated stride size of network parameters. If the learning rate is too small, it causes too low training efficiency and too much time is spent on training. If the learning rate is too large, it leads to local optimization, loss of oscillation, and model failure to converge. The dynamic adjustment mechanism of learning rate is adopted. The initial learning rate is 0.001, the learning rate transformation coefficient is 0.9. For instance, the learning rate is multiplied by 0.9 for every 10 training cycles. It ensures a fast training speed in the initial stage of training. When the convergence speed of the model slows down, the learning rate is reduced to gradually approach the optimal network parameter values.

4.3.2 Other Training Details

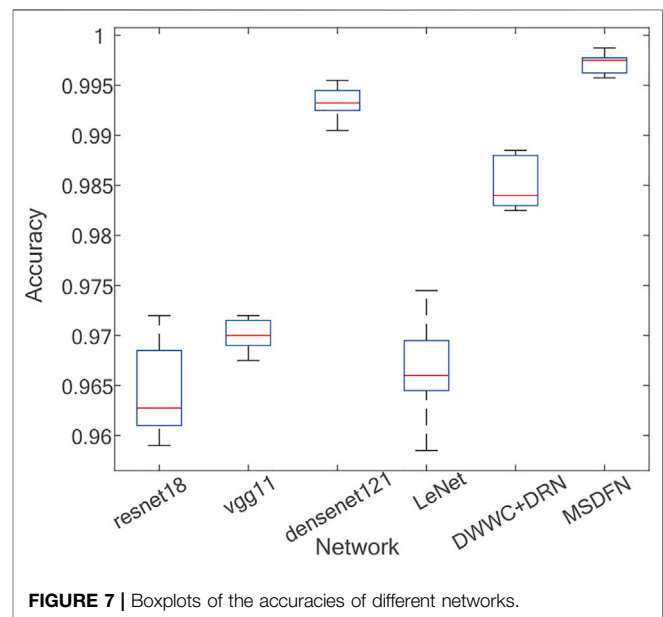
All experiments are implemented on a computer with a RTX3090 GPU, 16 GB RAM, and an Intel i710700 CPU. The implementation, training and testing of the network model are executed on the Pytorch1.7.0 deep learning framework. Matlab 2018a is used to divide the original vibration signals into sample sequences and generate time-frequency image samples.

This paper uses the softmax cross-entropy loss function. The corresponding loss calculation formula for a single sample is shown in Eq. 11.

$$lo(x, c) = -\log\left(\frac{\exp(x_c)}{\sum_j \exp(x_j)}\right) = -x_c + \log\left(\sum_j \exp(x_j)\right) \quad (11)$$

where c is the label (the category index), the vector x is the prediction result of each category (the network output), and x_c represents the c -th element of x . As the network trains, x_c approaches 1, so the loss approaches 0. The loss of each batch is shown in Eq. 12.

$$loss = \frac{\sum_{i=1}^N lo(x_i, C_i)}{N} \quad (12)$$

**FIGURE 7** | Boxplots of the accuracies of different networks.

The vectors x_i and C_i are the network output and label category of each sample respectively, and N is the batch size. The loss of each batch is the average loss of each sample.

In order to speed up the convergence speed and avoid falling into the local optimization, the momentum-driven stochastic gradient descent (Momentum-SGD) algorithm is used to optimize the network model, and the momentum parameter is set to 0.9.

4.4 Experimental Results and Analysis

The depth of the feature extraction network in MSDFN will affect feature fusion and the accuracy of fault diagnosis. Therefore, the best depth is selected through experiments. Table 6 shows the diagnostic accuracy of MSDFN of different depth feature extraction networks on Dataset E. It can be seen that if the structure is shallow, the diagnostic accuracy will decrease, and if

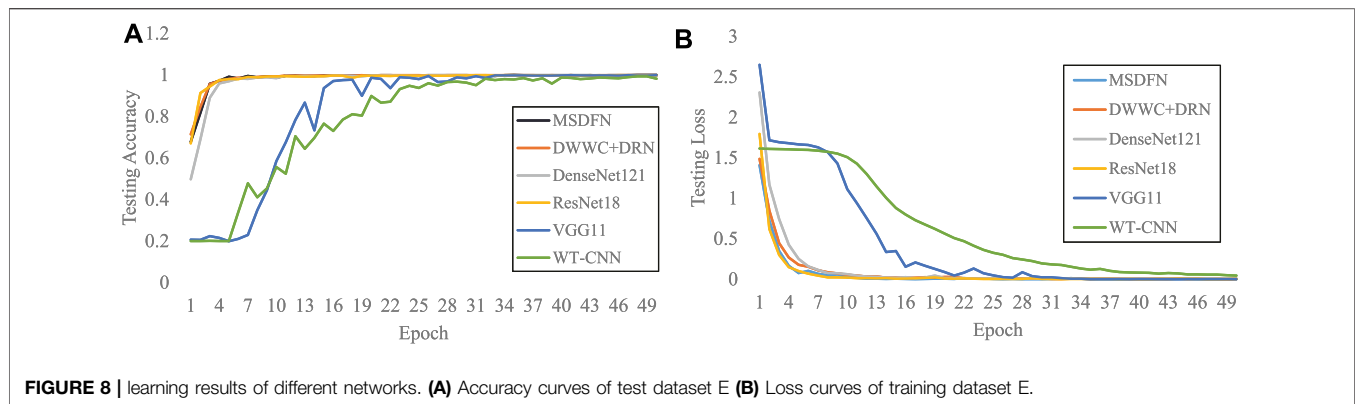


FIGURE 8 | learning results of different networks. **(A)** Accuracy curves of test dataset E **(B)** Loss curves of training dataset E.

the structure is too deep, it does not improve the accuracy. Adding one layer will increase the amount of model parameters several times. Therefore, we set the feature extraction network to five layers.

4.4.1 Performance Comparison of Different Network Models

In order to verify the effectiveness of the proposed method, the proposed MSDFN is compared with DWWC + DRN (Zhao et al., 2017), WT-CNN (Liang et al., 2020), and three traditional image classification networks: ResNet18, DenseNet201, VGG11. There are the following explanations about the comparative experiment.

Since the innovations of the proposed method focus on the fault diagnosis network, the networks proposed by existing methods are used in the comparative experiments. In comparative experiments, the data preprocessing methods of the original literature are not included, but the same dataset of time-frequency images is used. Therefore, this paper selects the networks proposed in Zhao et al. (2017) (the original literature uses wavelet packet coefficient matrix) and Liang et al. (2020) (the original literature uses time-frequency images) for comparison.

In addition, several traditional image classification networks are applied to the performance comparison of network models. DBN (Wang et al., 2018) is the source of the dataset used in this article. The listed results are from the original literature. Since the original network structure is not convenient to process time-frequency images and its training method is quite different from others, this paper does not test DBN. All the related comparative results are from the original literature (Wang et al., 2018). Comparative experimental results. **Table 7** shows the accuracy of fault diagnosis on each dataset. Compared with DBN and CNN using the traditional structures, the residual network and MSDFN have obvious advantages in accuracy. Compared with the residual network, the proposed network uses fewer network layers to achieve a certain improvement in fault diagnosis accuracy. The accuracy of each network on Dataset F, the University of Connecticut dataset proves the effectiveness of the proposed gearbox fault diagnosis method. In addition, all networks perform better on Dataset F, which also reflects the different complexity of the dataset distribution. **Figure 7** shows the accuracies of several CNN-based networks obtained in 10 times repeated experiments. MSDFN has a stabler and

TABLE 8 | The ablation networks.

Network	A	B	C	D
Backbone	✓	✓	✓	✓
MSFF		✓		✓
FoM			✓	✓

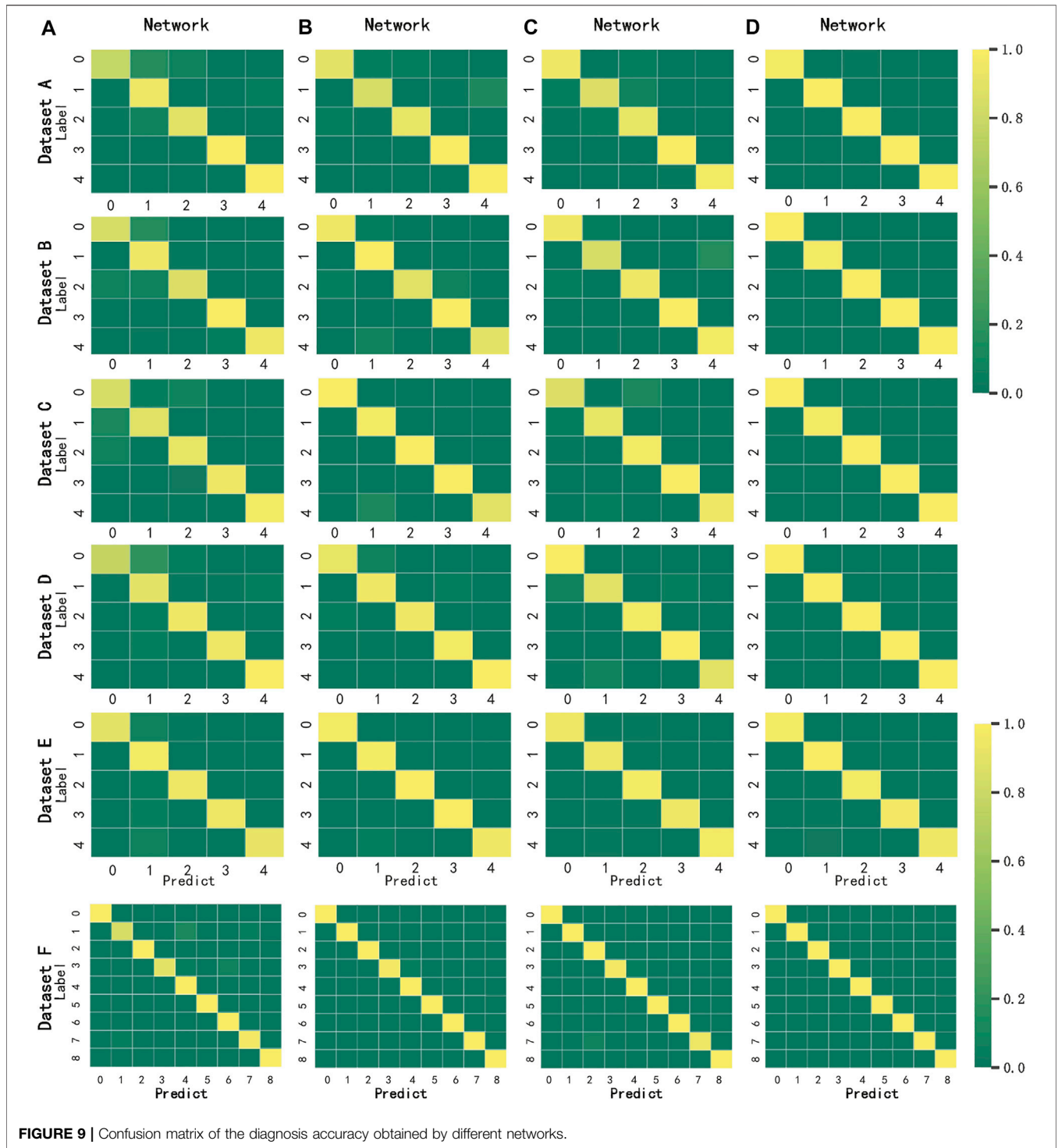
TABLE 9 | Ablation experiment results.

Network	A (%)	B (%)	C (%)	D (%)
Dataset A	92.67	95.00	94.00	98.62
Dataset B	92.67	95.30	95.00	99.27
Dataset C	92.33	97.33	95.00	99.03
Dataset D	92.00	97.00	95.33	99.35
Dataset E	92.58	97.08	96.00	99.75
Dataset F	96.88	99.65	98.96	100.00

higher accuracy rate on randomly allocated datasets, which means that the proposed network has stronger robustness. **Figures 8A,B** respectively compare the accuracy and loss curves of the Dataset E during the training process of each network. The convergence speed of MSDFN is close to ResNet, and its accuracy fluctuation is small.

4.4.2 The Role of Dense Feature Fusion

In this paper, the feature fusion module first is removed and then ablation study is performed on the complete model to verify the impact of the proposed multi-scale dense fusion mechanism on diagnosis performance. The different degrees of ablation models are shown in **Table 8**. After removing all the remaining parts of the feature fusion structure, the diagnosis network shown in **Figure 1** is used as backbone, which is mainly composed of a convolutional layer, a residual block, and a fully connected layer. The complete network D is MSDFN. The diagnostic accuracy of different degrees of ablation models on each dataset is shown in **Table 9**. It can be seen that the feature fusion module has a positive effect on improving the accuracy of fault diagnosis. The diagnostic accuracy of different degrees of ablation models does not fluctuate greatly under different working conditions, but they are still affected by the working conditions. Different models show different diagnostic



performance on different datasets. As the load increases, the diagnostic accuracy of models B, C, and D is also improving, but model A does not have this rule, which reflects that the feature fusion module increases the model sensitivity to changes in load conditions.

Figure 9 shows the confusion matrices of the diagnosis results obtained by each ablated network model on the testing set of Dataset E containing 1,200 images. After adding the feature fusion

structure to the network, the accuracy rate is significantly improved. Additionally, due to the dense feature fusion at each network layer, the MSFF module plays a more important role than the FoM module. Although the projection and back projection operations of each layer increase the amount of calculation, the fault diagnosis accuracy is effectively improved. Three images incorrectly diagnosed by the complete network D belong to different categories and are

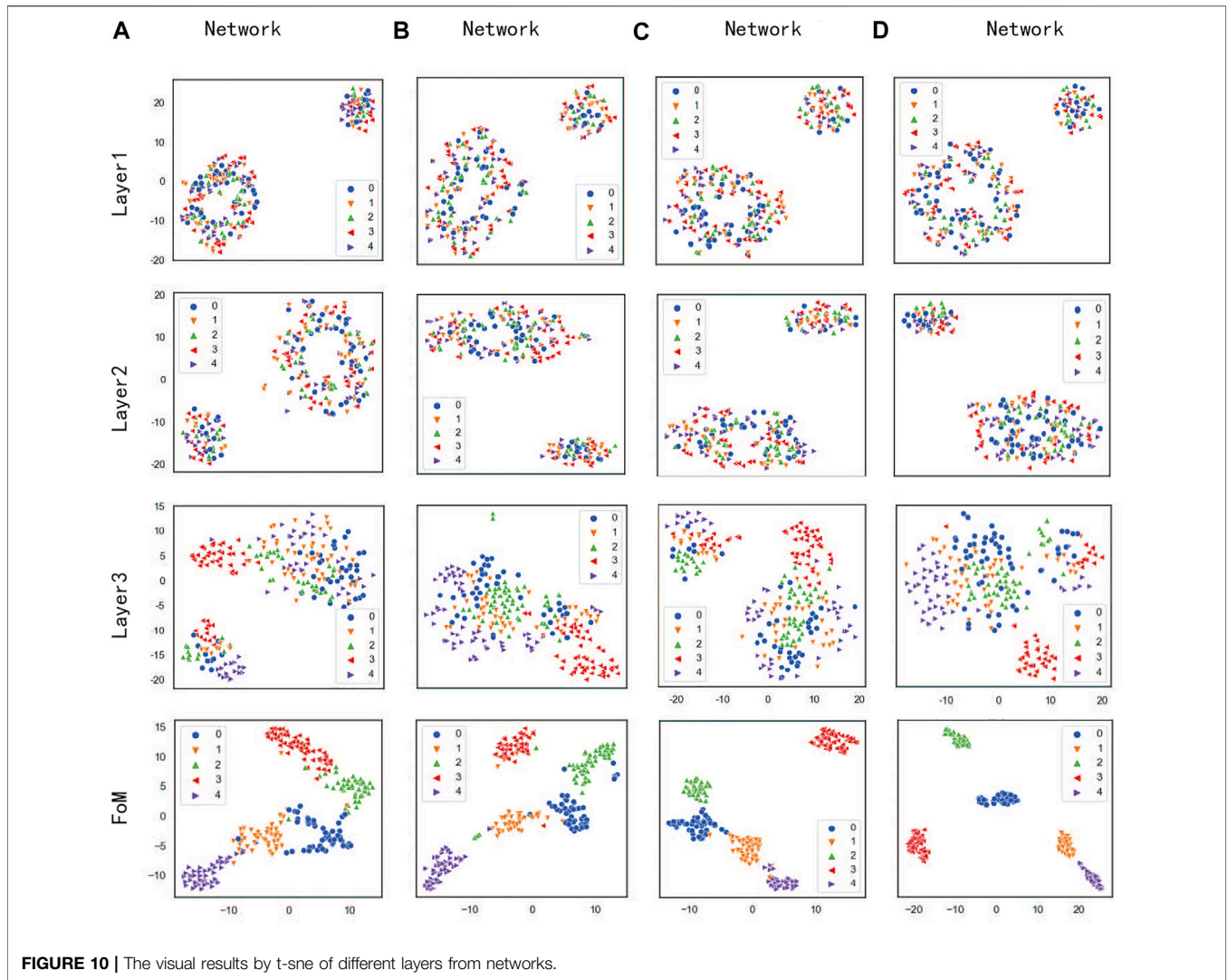


FIGURE 10 | The visual results by t-sne of different layers from networks.

predicted to be three different categories, which indicates that MSDFN has no obvious deviation in the diagnosis ability of different fault types. However, the obvious deviation in the diagnosis ability exists in networks A, B, and C. The integration of the two feature fusion modules, MSFF and FoM, effectively improves the unstable recognition ability of the network for various fault types. 240 images are randomly selected from the testing set of Dataset E to compose a dataset. According to the network with various degrees of ablation, **Figure 10** shows the distribution of output features after t-sne dimensionality reduction. The distribution of output features intuitively shows the separability of output features. The fault feature extraction performance of each ablated network can be evaluated.

The output features of the first two feature extraction network layers of each network model show two parts, a large part and a small part, because the separability of features is low at this time. Since testing is carried out on a mixed dataset, there are obvious data distribution differences between the data of the mixed working conditions. This mainly reflects the load among L1 (0 N.m), L2 (1.4 N.m), L3 (2.8 N.m), and L4 (25.2 N.m), so the

two parts in the scatter diagram show an approximate ratio of 3:1. The performance of the two feature fusion modules on fault diagnosis can be analyzed by comparing the feature distribution of networks A, B, C, and D. As the network depth increases, the differences caused by working conditions gradually decrease, so the features of the same type of faults gradually become concentrated. Different ablated networks show different recognition capabilities. Compare with the third-layer output of different networks, the introduction of the MSFF module enhances the ability to distinguish features. According to the output comparison of each layer of each model, the network D achieves the best performance.

The fault recognition capability of each network can be determined by observing the output of FoM module in each model. Compared with other models, the output features of FoM module in model A (backbone network) have many errors, and the features of each fault category are relatively close. As the complete network, the output features of FoM module in model D show high separability, and 100% accuracy is achieved on the set of 240 random samples.

5 CONCLUSION

Monitoring the health status of each wind turbine planetary gearbox is of great significance for reducing the operation and maintenance cost of wind turbines. In order to solve the loss of fault information in the diagnosis process and improve the model diagnosis ability, this paper proposes an intelligent fault diagnosis method for wind turbine planetary gearboxes based on MSDFN. Both MSFF and FoM modules are used to perform feature fusion in the feature extraction and fault classification, respectively. The loss of fault information caused by continuous convolution and pooling in CNN is effectively mitigated. The proposed method is compared with two mainstream methods and three traditional image classification networks to verify its effectiveness. Compared with traditional CNN-based networks, MSDFN uses feature fusion twice to improve the accuracy of fault diagnosis under both single and mixed loads. In repeated experiments, the proposed method achieves more than 99.5% accuracy rate. The ablation study verifies that feature fusion is conducive to the gear fault diagnosis of planetary gearboxes. On the mixed dataset, MSFF and FoM modules increase the accuracy rate by 4.5% and 3.42%, respectively. To enable the proposed model to maintain high fault diagnosis accuracy in the case of large changes in working conditions, future work will focus on how to use transfer learning to improve the model's adaptability to working conditions.

REFERENCES

- Cao, L., Qian, Z., Zareipour, H., Huang, Z., and Zhang, F. (2019). Fault Diagnosis of Wind Turbine Gearbox Based on Deep Bi-directional Long Short-Term Memory under Time-Varying Non-stationary Operating Conditions. *IEEE Access* 7, 155219–155228. doi:10.1109/access.2019.2947501
- Cao, P., Zhang, S., and Tang, J. (2018). Preprocessing-free Gear Fault Diagnosis Using Small Datasets with Deep Convolutional Neural Network-Based Transfer Learning. *Ieee Access* 6, 26241–26253. doi:10.1109/access.2018.2837621
- Cao, X., Chen, B., and Zeng, N. (2020). A Deep Domain Adaption Model with Multi-Task Networks for Planetary Gearbox Fault Diagnosis. *Neurocomputing* 409, 173–190. doi:10.1016/j.neucom.2020.05.064
- Chen, R., Huang, X., Yang, L., Xu, X., Zhang, X., and Zhang, Y. (2019). Intelligent Fault Diagnosis Method of Planetary Gearboxes Based on Convolution Neural Network and Discrete Wavelet Transform. *Comput. Industry* 106, 48–59. doi:10.1016/j.compind.2018.11.003
- Chen, Z., Mauricio, A., Li, W., and Gryllias, K. (2020). A Deep Learning Method for Bearing Fault Diagnosis Based on Cyclic Spectral Coherence and Convolutional Neural Networks. *Mech. Syst. Signal Process.* 140, 106683. doi:10.1016/j.ymssp.2020.106683
- Cheng, Y., Lin, M., Wu, J., Zhu, H., and Shao, X. (2021). Intelligent Fault Diagnosis of Rotating Machinery Based on Continuous Wavelet Transform-Local Binary Convolutional Neural Network. *Knowledge-Based Syst.* 216, 106796. doi:10.1016/j.knsys.2021.106796
- Dai, S., Han, M., Wu, Y., and Gong, Y. (2007). “Bilateral Back-Projection for Single Image Super Resolution,” in 2007 IEEE International Conference on Multimedia and Expo (IEEE), 1039–1042. doi:10.1109/icme.2007.4284831
- Ding, X., and He, Q. (2017). Energy-fluctuated Multiscale Feature Learning with Deep Convnet for Intelligent Spindle Bearing Fault Diagnosis. *IEEE Trans. Instrum. Meas.* 66, 1926–1935. doi:10.1109/tim.2017.2674738
- Dong, H., Pan, J., Xiang, L., Hu, Z., Zhang, X., Wang, F., et al. (2020). “Multi-scale Boosted Dehazing Network with Dense Feature Fusion,” in Proceedings of the

DATA AVAILABILITY STATEMENT

The data analyzed in this study is subject to the following licenses/restrictions: The dataset is provided by the author of paper “<https://content.iospress.com/articles/journal-of-intelligent-and-fuzzy-systems/ifs169538>”. Requests to access these datasets should be directed to <https://content.iospress.com/articles/journal-of-intelligent-and-fuzzy-systems/ifs169538>.

AUTHOR CONTRIBUTIONS

Conceptualization, YC and XH; Methodology, YL and XH; Software, XH; Writing XH and YC.

FUNDING

This work is jointly supported by the National Natural Science Foundation of China under Grant No. 61906026; Innovation research group of universities in Chongqing; Special key project of Chongqing technology innovation and application development: cstc2019jscx-zdztzx0068; The Chongqing Natural Science Foundation under Grant cstc2020jcyj-msxmX0577, cstc2020jcyj-msxmX0634; “Chengdu-Chongqing Economic Circle” innovation funding of Chongqing Municipal Education Commission KJCXZD2020028.

- IEEE/CVF Conference on Computer Vision and Pattern Recognition (IEEE), 2157–2167. doi:10.1109/cvpr42600.2020.00223
- Feng, Z., and Liang, M. (2014). Complex Signal Analysis for Wind Turbine Planetary Gearbox Fault Diagnosis via Iterative Atomic Decomposition Thresholding. *J. Sound Vibration* 333, 5196–5211. doi:10.1016/j.jsv.2014.05.029
- Feng, Z., and Zuo, M. J. (2012). Vibration Signal Models for Fault Diagnosis of Planetary Gearboxes. *J. Sound Vibration* 331, 4919–4939. doi:10.1016/j.jsv.2012.05.039
- Han, Y., Tang, B., and Deng, L. (2019). An Enhanced Convolutional Neural Network with Enlarged Receptive fields for Fault Diagnosis of Planetary Gearboxes. *Comput. Industry* 107, 50–58. doi:10.1016/j.compind.2019.01.012
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). “Deep Residual Learning for Image Recognition,” in Proceedings of the IEEE conference on computer vision and pattern recognition (IEEE), 770–778. doi:10.1109/cvpr.2016.90
- Hu, J., Shen, L., and Sun, G. (2018). “Squeeze-and-excitation Networks,” in Proceedings of the IEEE conference on computer vision and pattern recognition (IEEE), 7132–7141. doi:10.1109/cvpr.2018.00745
- Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2017). “Densely Connected Convolutional Networks,” in Proceedings of the IEEE conference on computer vision and pattern recognition (IEEE), 4700–4708. doi:10.1109/cvpr.2017.243
- Irani, M., and Peleg, S. (1991). Improving Resolution by Image Registration. *CVGIP: Graphical models image Process.* 53, 231–239. doi:10.1016/1049-9652(91)90045-1
- Janssens, O., Slavkovikj, V., Vervisch, B., Stockman, K., Loccufier, M., Verstockt, S., et al. (2016). Convolutional Neural Network Based Fault Detection for Rotating Machinery. *J. Sound Vibration* 377, 331–345. doi:10.1016/j.jsv.2016.05.027
- Jiang, G., He, H., Xie, P., and Tang, Y. (2017). Stacked Multilevel-Denoising Autoencoders: A New Representation Learning Approach for Wind Turbine Gearbox Fault Diagnosis. *IEEE Trans. Instrum. Meas.* 66, 2391–2402. doi:10.1109/tim.2017.2698738
- Jiang, G., He, H., Yan, J., and Xie, P. (2018). Multiscale Convolutional Neural Networks for Fault Diagnosis of Wind Turbine Gearbox. *IEEE Trans. Ind. Elect.* 66 (4), 3196–3207. doi:10.1109/TIE.2018.2844805

- Jing, L., Zhao, M., Li, P., and Xu, X. (2017). A Convolutional Neural Network Based Feature Learning and Fault Diagnosis Method for the Condition Monitoring of Gearbox. *Measurement* 111, 1–10. doi:10.1016/j.measurement.2017.07.017
- Kang, S., Chen, W., Wang, Y., Na, X., Wang, Q., and Mikulovich, V. I. (2020). Method of State Identification of Rolling Bearings Based on Deep Domain Adaptation under Varying Loads. *IET Sci. Meas. Techn.* 14, 303–313. doi:10.1049/iet-smt.2019.0043
- Lei, Y., Yang, B., Jiang, X., Jia, F., Li, N., and Nandi, A. K. (2020). Applications of Machine Learning to Machine Fault Diagnosis: A Review and Roadmap. *Mech. Syst. Signal Process.* 138, 106587. doi:10.1016/j.ymsp.2019.106587
- Li, W., Wang, F.-g., Liu, C., and Xu, P.-m. (2017). A New Method for Feature Extraction of Vibration Signals of Planetary Gearboxes. *Noise and Vibration Control*.
- Liang, P., Deng, C., Wu, J., and Yang, Z. (2020). Intelligent Fault Diagnosis of Rotating Machinery via Wavelet Transform, Generative Adversarial Nets and Convolutional Neural Network. *Measurement* 159, 107768. doi:10.1016/j.measurement.2020.107768
- Lin, Z., Chen, Z., Wu, Q., Yang, S., and Meng, H. (2018). Coordinated Pitch & Torque Control of Large-Scale Wind Turbine Based on Pareto Efficiency Analysis. *Energy* 147, 812–825. doi:10.1016/j.energy.2018.01.055
- Liu, R., Yang, B., Zio, E., and Chen, X. (2018). Artificial Intelligence for Fault Diagnosis of Rotating Machinery: A Review. *Mech. Syst. Signal Process.* 108, 33–47. doi:10.1016/j.ymsp.2018.02.016
- Lu, L., He, Y., Ruan, Y., and Yuan, W. (2020). Wind Turbine Planetary Gearbox Condition Monitoring Method Based on Wireless Sensor and Deep Learning Approach. *IEEE Trans. Instrumentation Meas.* 70, 1–16.
- Ma, S., Chu, F., and Han, Q. (2019). Deep Residual Learning with Demodulated Time-Frequency Features for Fault Diagnosis of Planetary Gearbox under Nonstationary Running Conditions. *Mech. Syst. Signal Process.* 127, 190–201. doi:10.1016/j.ymsp.2019.02.055
- Mao, W., He, J., and Zuo, M. J. (2019). Predicting Remaining Useful Life of Rolling Bearings Based on Deep Feature Representation and Transfer Learning. *IEEE Trans. Instrumentation Meas.* 69 (4), 1594–1608. doi:10.1109/TIM.2019.2917735
- Miao, X., Li, S., Zhu, Y., and An, Z. (2020). A Novel Real-Time Fault Diagnosis Method for Planetary Gearbox Using Transferable Hidden Layer. *IEEE Sensors J.* 20, 8403–8412. doi:10.1109/jsen.2020.2965988
- Pan, T., Chen, J., Zhou, Z., Wang, C., and He, S. (2019). A Novel Deep Learning Network via Multiscale Inner Product with Locally Connected Feature Extraction for Intelligent Fault Detection. *IEEE Trans. Ind. Inf.* 15, 5119–5128. doi:10.1109/tii.2019.2896665
- Peng, D., Wang, H., Liu, Z., Zhang, W., Zuo, M. J., and Chen, J. (2020). Multibranch and Multiscale Cnn for Fault Diagnosis of Wheelset Bearings under strong Noise and Variable Load Condition. *IEEE Trans. Ind. Inf.* 16, 4949–4960. doi:10.1109/tii.2020.2967557
- Simonyan, K., and Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv preprint arXiv:1409.1556
- Sun, J., Qi, G., Mazur, N., and Zhu, Z. (2021). Structural Scheduling of Transient Control under Energy Storage Systems by Sparse-Promoting Reinforcement Learning. *IEEE Trans. Ind. Inform.* 18 (2), 744–756. doi:10.1109/TII.2021.3084139
- Sun, J., Zhu, Z., Li, H., Chai, Y., Qi, G., Wang, H., et al. (2019). An Integrated Critic-Actor Neural Network for Reinforcement Learning with Application of Ders Control in Grid Frequency Regulation. *Int. J. Electr. Power Energy Syst.* 111, 286–299. doi:10.1016/j.ijepes.2019.04.011
- Wang, D.-F., Guo, Y., Wu, X., Na, J., and Litak, G. (2020a). Planetary-gearbox Fault Classification by Convolutional Neural Network and Recurrence Plot. *Appl. Sci.* 10, 932. doi:10.3390/app10030932
- Wang, F., Liu, R., Hu, Q., and Chen, X. (2020b). Cascade Convolutional Neural Network with Progressive Optimization for Motor Fault Diagnosis under Nonstationary Conditions. *IEEE Trans. Ind. Inform.* 17 (4), 2511–2521. doi:10.1109/TII.2020.3003353
- Wang, X., Qin, Y., and Zhang, A. (2018). An Intelligent Fault Diagnosis Approach for Planetary Gearboxes Based on Deep Belief Networks and Uniformed Features. *Ijs* 34, 3619–3634. doi:10.3233/jifs-169538
- Wang, Y.-r., Jin, Q., Sun, G.-d., and Sun, C.-f. (2019). Planetary Gearbox Fault Feature Learning Using Conditional Variational Neural Networks under Noise Environment. *Knowledge-Based Syst.* 163, 438–449. doi:10.1016/j.knosys.2018.09.005
- Wei, Y., Li, Y., Xu, M., and Huang, W. (2019). A Review of Early Fault Diagnosis Approaches and Their Applications in Rotating Machinery. *Entropy* 21, 409. doi:10.3390/e21040409
- Wen, L., Li, X., Gao, L., and Zhang, Y. (2017). A New Convolutional Neural Network-Based Data-Driven Fault Diagnosis Method. *IEEE Trans. Ind. Elect.* 65 (7), 5990–5998. doi:10.1109/TIE.2017.2774777
- Wu, C., Jiang, P., Ding, C., Feng, F., and Chen, T. (2019). Intelligent Fault Diagnosis of Rotating Machinery Based on One-Dimensional Convolutional Neural Network. *Comput. Industry* 108, 53–61. doi:10.1016/j.compind.2018.12.001
- Xia, M., Li, T., Xu, L., Liu, L., and De Silva, C. W. (2017). Fault Diagnosis for Rotating Machinery Using Multiple Sensors and Convolutional Neural Networks. *IEEE/ASME Trans. mechatronics* 23 (1), 101–110. doi:10.1109/TMECH.2017.2728371
- Xu, Z., Li, C., and Yang, Y. (2020). Fault Diagnosis of Rolling Bearing of Wind Turbines Based on the Variational Mode Decomposition and Deep Convolutional Neural Networks. *Appl. Soft Comput.* 95, 106515. doi:10.1016/j.asoc.2020.106515
- Zhao, B., Zhang, X., Li, H., and Yang, Z. (2020a). Intelligent Fault Diagnosis of Rolling Bearings Based on Normalized Cnn Considering Data Imbalance and Variable Working Conditions. *Knowledge-Based Syst.* 199, 105971. doi:10.1016/j.knosys.2020.105971
- Zhao, M., Kang, M., Tang, B., and Pecht, M. (2017). Deep Residual Networks with Dynamically Weighted Wavelet Coefficients for Fault Diagnosis of Planetary Gearboxes. *IEEE Trans. Ind. Elect.* 65 (5), 4290–4300. doi:10.1109/TIE.2017.2762639
- Zhao, M., Zhong, S., Fu, X., Tang, B., Dong, S., and Pecht, M. (2020b). Deep Residual Networks with Adaptively Parametric Rectifier Linear Units for Fault Diagnosis. *IEEE Trans. Ind. Elect.* 68 (3), 2587–2597. doi:10.1109/TIE.2020.2972458
- Zhao, Z., Li, T., Wu, J., Sun, C., Wang, S., Yan, R., et al. (2020c). Deep Learning Algorithms for Rotating Machinery Intelligent Diagnosis: An Open Source Benchmark Study. *ISA Trans.* 107, 224–255. doi:10.1016/j.isatra.2020.08.010
- Zhu, Z., Zeng, F., Qi, G., Li, Y., Jie, H., and Mazur, N. (2021). Power System Structure Optimization Based on Reinforcement Learning and Sparse Constraints under Dos Attacks in Cloud Environments. *Simulation Model. Pract. Theor.* 110, 102272. doi:10.1016/j.simpat.2021.102272

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Huang, Li and Chai. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.