



Research article

Fault diagnosis of printing press bearing based on deformable convolution residual neural network

Qiumin Wu^{1,*}, Ziqi Zhu¹, Jiahui Tang² and Yukang Xia¹

¹ Faculty of Printing, Packaging Engineering and Digital Media Technology, Xi'an University of Technology, Xi'an 710048, China

² School of Mechanical and Precision Instrument Engineering, Xi'an University of Technology, Xi'an 710048, China

* **Correspondence:** Email: wuqiumin@xaut.edu.cn; Tel: 86 15309215069.

Abstract: Rolling bear is a major critical component of rotating machinery, as its working condition affects the performance of the equipment. As a result, the condition monitoring and fault diagnosis of bearings get more and more attention. However, the strong background noise makes it difficult to extract the bearing fault features exactly. Furthermore, regular gradient disappearance and overfit appear in traditional network model training. Therefore, taking the printing press bearings as the research object, an intelligent fault diagnosis method based on strong background noise is proposed. This method integrates frequency slice wavelet transform (FSWT), deformable convolution and residual neural network together, and realizes the high-precision fault diagnosis of the printing press bearings. First, FSWT is used to preprocess the original vibration signal to obtain bearing fault features in the time and frequency domain, reconstruct the signal in any frequency band and describe local features accurately. Second, the ResNet is selected as the base network, and the two-dimensional time-frequency diagrams (TFD) obtained by preprocessing are used as input. For the model that has a poor ability to extract subtle features under strong background noise, the deformable convolution layer is introduced to reconstruct the convolution layer of ResNet, called deformable convolution residual neural network (DC-ResNet). Finally, the effectiveness of this method is verified by using the data sets collected under experimental conditions and actual working conditions for fault diagnosis of the printing press. The results show that the DC-ResNet can classify different bearing faults under strong background noise, and the accuracy and stability are greatly improved, which the accuracy meets 93.90%. The intelligent fault diagnosis with high-precision of printing press bearings under complex working

conditions is realized by the proposed method.

Keywords: frequency slice wavelet transform; residual neural network; deformable convolution; rolling bear; fault diagnosis

1. Introduction

In recent years, the printing press has developed in the direction of green, high speed, automation and intelligence in modern industry. Rolling bearings are one of the most important components in rotating machinery, and play a role in reducing the friction coefficient during motion, and ensuring rotary accuracy in the printing press [1–4]. Therefore, it's health status has an important influence on the printing press's overall performance, stability and life [5]. Due to the corrosion of the rolling bears by chemical substances, such as fountain solution, ink, or bearing failure, may be due to insufficient lubrication, water or foreign matter intrusion, improper bearing assembly, long time operation of the machine, etc. It is easy to cause failure and abnormal vibration of the printing press bearings, then cause faults in the printing press, such as paper gripper being gripped, inaccurate overprinting, ink bar and poor overprinting [1]. Also, the printing presses usually work in the poor environment, so fault is not easy to be found. This strong background noise environment can lead to low fault diagnosis accuracy, which can lead to safety problems. Therefore, monitoring the operating status of the printing press bearings plays an increasingly important role in actual production, and by introducing fault diagnosis technology in machinery, it can find the potential problems of machinery in time, and also can be solved in a timely manner, greatly reducing the failure rate of equipment.

Traditional fault diagnosis is achieved by extracting bearing fault features in the original signal through signal processing techniques and mathematical transformation, and then identifying the information that can reflect the fault features [7]. However, with the increased productivity and the expanding scale of equipment, the bearing vibration signals collected by the equipment have the features of massive, multi-source and high-dimensional, which leads to inefficiency and low accuracy of the traditional fault diagnosis methods [8,9]. While intelligent fault diagnosis is based on data-driven methods, they can quickly and accurately handle a large amount of complex data, so it has become one of the current research hotspots for condition monitoring and fault diagnosis of mechanical equipment.

The mechanical structure of the printing press is complex, and the printing press generates strong noises. So, the vibration signals of the printing press bearings collect the features with strong noise. It is hard to extract the fault features accurately in the vibration signal with strong noise, and it's easy to leads to the gradient disappearance and overfitting problem of the network model during training of traditional machine learning, such as Support Vector Machine (SVM) and Logistic Regression. As a new network architecture, the ResNet can effectively solve the gradient disappearance or gradient explosion caused by the data or network depth. In addition, the ResNet has a great ability to extract features, which can solve the problem of overfitting when training other network models, and reduce the reliance on fault diagnosis experience of expert and signal processing techniques [10–12]. Therefore, the ResNet is selected as the base network in this proposed method. Lin [13] proposed an improved one-dimensional convolution ResNet fault diagnosis. The extraction and compression of data fault features are first completed by convolution pooling. Then, an improved ResNet is added to avoid network degradation and uneven data distribution in the training model. Zagoruyko and

Komodakis [14] studied the use of a wide residual block, and demonstrated it, experimentally, at a reasonable depth. Konovalenko [15] studied a deep ResNet-based model for rolling steel surface defect recognition and classification, and constructed a classifier to detect three types of defects on flat metal surfaces. Yan [16] proposed the deep order-wavelet convolutional variational autoencoder (DOWCVAE), which improves the feature learning capability of convolutional variational autoencoder ordinary convolutional variational encoder (CVAE). Peng [17] combined ResNet and data fusion to achieve higher recognition accuracy. Yan [18] proposed a deep regularized variational autoencoder (DRVAE) intelligent fault diagnosis model for rotor-bearing systems, which solves the overfitting problem of the original variational autoencoder (VAE) and enhances the feature learning capability of the network model. Wang [19] proposed a deep separable ResNet model, which uses a reduced number of parameters through deep separable convolution, and uses residual connections to transfer features, which can effectively predict the remaining life of rolling bearings. Liang [20] proposed a frequency domain analysis method based on wavelet transform, and improved ResNet with global singular value decomposition (SVD) adaptive. Zhang [1] proposed a blockchain-based distributed joint transfer learning method to improve the accuracy of mechanical fault diagnosis, and applied it to collaborative mechanical fault diagnosis. Yan [22] proposed multiscale cascading deep belief network (MCDBN) for rotating machinery fault diagnosis, which improves fault identification accuracy by learning a wider range of feature representations. Lin [23] proposed a photovoltaic array fault diagnosis scheme using a multi-scale SE-ResNet, and designed a multi-scale perceptual field fusion module to improve the diagnosis performance of the model. Wan [24] used an improved deep ResNet as a feature extractor to extract metastable features from the original vibration signal. The classifier used the extracted domain-invariant features to complete cross-domain fault identification. Yan [25] proposed a wind speed prediction model based on long short-term memory, deep belief network and grasshopper optimization algorithm (GOA) to improve the accuracy and efficiency of wind speed prediction. Yang [26] classified real-time data collected from dissolved oxygen sensors to diagnosis online faults and experimentally proved that ResNet has great performance. Hao [27] proposed a new network structure by replacing the fully connected layer part of traditional ResNet with the global average pooling technique. The problem of too many parameters of the traditional ResNet model was effectively solved. Zhang [28] proposed a fault diagnosis method based on adaptive loss-weighted meta-ResNet (ALWM-ResNet), which uses a weighted network and a meta-network cloned from the original ResNet to establish a mapping of weighting functions, and adaptively learn weights from data containing clean labels. However, the convolution kernel in the ResNet structure has a fixed shape, poor adaptability to changes in unknown images, and poor generalization ability [29,30]. Thus, this proposed method introduces a deformable convolution layer based on the ResNet model, so that the convolution layer can adaptively change the shape of the convolution kernel according to the different input samples, so the convolution kernel can precisely locate and track the target for small fault features, and use the deformable convolution to adaptively identify the fault feature points, thus learning more detailed features and improving the accuracy of feature extraction.

However, the signals collected under actual working conditions often contain complex noises, leading to a decline in accuracy if they are inputting direct into the model. Therefore, this proposed method introduces the signal preprocessing method to extract fault features before input to improve the accuracy of the model. The Short Time Fourier Transform (STFT), Wigner-Ville Distribution (WVD), Empirical Mode Decomposition (EMD), and Wavelet Analysis are widely used preprocessing methods in the field of fault diagnosis. Hartono [31] proposed a joint time-frequency analysis method

for gear fault diagnosis using the combined autoregressive model-based filtering and redistribution with the smoothing pseudo-Wigner Ville distribution (RSPWVD) method. Nezamivand [32] performed empirical modal decomposition and wavelet packet decomposition of vibration signals, and achieved good results for fault state identification of rolling bearings by SVM. Surti [33] proposed a new technique for early bearing fault detection and diagnosis based on discrete wavelet transform (DWT) and K-nearest neighbor (KNN), which showed good accuracy and the ability to distinguish between healthy and unhealthy bearing conditions. Li [34] proposed a deep learning-based remaining useful life (RUL) prediction method to solve the sensor failure problem, and introduced adversarial learning to extract invariant features in generalized sensors. Dubey [1] proposed a Hilbert transform footprint analysis and neural network for ball bearing fault analysis, and the method achieved high fault classification accuracy. Tian [36] proposed a method to detect bearing faults and monitor bearing degradation in electric motors using Programmable Counter Array and semi-supervised KNN distance to combine these features to form health indicators for detection. Amar [37] proposed a novel vibration spectrum imaging (VSI) feature enhancement system for low signal-to-noise ratio (SNR) conditions that enhances and provides a visual representation of feature vibrational spectral features in the form of images. Zhang [38] preliminarily constructed the wavelet-overlapping group sparse (WOGS) optimization model based on the overlapping features of Morlet wavelet transform coefficients, and constructed the weight coefficients in the model by analyzing the pulse features of the signal. Mao [39] used a classification method combining multiscale alignment entropy and support vector machine to achieve the fault-type classification of bearings. Zhao [40] proposed a new convolution neural network scheme based on attention-enhanced convolution blocks (AECB) to achieve higher training accuracy for control moment gyroscope (CMG) fault diagnosis data sets with different sliding window parameters.

The above method used the data collected under the standard data set, and used simple preprocessing into the neural network to achieve bearing fault diagnosis. However, the above-preprocessing methods need to select the appropriate window function or wavelet basis function, as traditional preprocessing methods lead to modal blending and endpoint effects, resulting in low fault diagnosis accuracy, such as STFT, Wavelet Transform and WVD. This is due to the faulty bearing signals being collected within strong noise vibrations in the actual working conditions. The FSWT combines the advantages of STFT and Wavelet Transform, which not only reduces the dependence of wavelets and wavelet packets on wavelet basis functions in reconstructing signals, but also realizes the reconstruction of signals in arbitrary frequency bands and the accurate description of local features, which can flexibly realize the filtering and segmentation of signals [41]. Furthermore, the TFD after FSWT processing contains feature information of the vibration signal in both time domain and frequency domain, which is beneficial to neural networks for feature extraction and improves the efficiency of model identification effectively.

In summary, this method proposes a diagnosis method that integrates FSWT, deformable convolution layer, and ResNet to improve diagnosis efficiency under strong background noise conditions while ensuring diagnosis accuracy. The experimental results show that the proposed method has higher recognition accuracy than other diagnoses. In addition, the diagnosis accuracy is improved while the model training efficiency is enhanced. The main contributions of this proposed method are summarized as follows.

(1) In order to reduce the influence of noise on diagnosis accuracy, this method proposes a method preprocessing bearing vibration signals using FSWT. Compared with other methods, this method can

improve the efficiency and accuracy of subsequent intelligent fault diagnosis significantly.

(2) In this proposed method, the ResNet is selected to solve the problem of gradient disappearance and overfitting during the training of traditional network models. At the same time, in order to enhance the extraction ability of the model for subtle features in the TFD, the deformable convolution layer is introduced into the ResNet, which improves the adaptiveness of the convolution layer shape so that the model can effectively capture the subtle features drowned in noise.

(3) This proposed method is tested with the Case Western Reserve University (CWRU) data set with an accuracy of 99.77%. In the application of actual working conditions, the model is also superior to other methods, and the diagnostic accuracy is 93.90%.

The rest of the paper is organized as follows: Section 2 presents the frequency slice wavelet transform. Section 3 introduces the ResNet model used in this proposed method. Section 4 introduces the structure of the deformable convolution layer and the structure of the DC-ResNet. Section 5 introduces the fault diagnosis process, parameters of the DC-ResNet and model hyperparameters proposed. The experimental validation of the method and the analysis of the experimental results are presented in Section 6. Finally, the conclusions are shown in Section 7.

2. Data preprocessing

The common method of rolling bearing fault diagnosis is based on vibration signals. In traditional bearing fault diagnosis, the one-dimensional time-domain vibration signal is usually spliced into the two-dimensional signal, and the splicing will lead the vibration signal addition of two-dimensional information, which is not available in the original time-domain vibration signal. As the input of the ResNet, the two-dimensional information created artificially will affect the feature extraction of the signal by the ResNet, thus affecting the accuracy of the fault diagnosis. The FSWT have advantages over STFT and wavelet transform, can realize signal filtering and segmentation flexibly [42], and the original time-domain signal can be converted to a time-frequency signal. Due to the rolling bearing fault signals of the printing press, the collection is a signal with strong noise, and the time-frequency domain features of the signal can be better observed by FSWT. Therefore, in this proposed method, FSWT is used to preprocess the bearing vibration signals of the printing press.

2.1. Frequency slice wavelet transform

Let the signal $f(t) \in L^2(R)$, if $\hat{p}(\omega)$ is the Fourier transform of $p(t)$, then its FSWT is defined in the frequency domain as [43].

$$W_f(t, \omega, \lambda, \sigma) = \frac{1}{2\pi} \lambda \int_{-\infty}^{+\infty} f(u) \hat{p} * \left(\frac{u-\omega}{\sigma} \right) e^{iut} du, \quad (2.1)$$

where λ is the energy factor ($\lambda \neq 0$), scale σ is either an invariant or a function of ω , t and u , “*” is the conjugate of the function, ω and t are observation frequency and observation time, u is the frequency of assessment and $\hat{p}(\omega)$ is the FSF. By the Parseval equation, the above equation can be transformed to the time domain as follows:

$$W_f(t, \omega, \lambda, \sigma) = \sigma \lambda \varepsilon^{i\omega t} \int_{-\infty}^{\infty} f(\tau) e^{-i\omega\tau} p * (\sigma(\tau - t)) d\tau \quad (2.2)$$

2.2. Determination of scale factor

In general, taking $\lambda = 1$, let $\sigma = \omega/\kappa$, $\kappa > 0$, then the Eq (2.1) can be rewritten as:

$$W(t, \omega, \kappa) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{f}(u) \hat{p}\left(\kappa \frac{u-\omega}{\omega}\right) e^{iut} du \quad (2.3)$$

Among them, κ is unrelated to ω , and u is mainly used to adjust the sensitivity of the FSWT to frequency or time, and it is called the *time*-frequency resolution factor. From the Heisenberg uncertainty principle, it is not possible to obtain a high resolution in both the frequency and time domains, so we then used σ and ω to estimate the time-frequency resolution factor, while introducing two evaluation coefficients: the frequency resolution ratio η and the amplitude expectation response ratio v ($0 < v \leq 1$), where v is usually taken as $\sqrt{2}/2, 0.5, 0.25$, etc.

2.3. Frequency slicing wavelet inversion

The inversion of frequency slice wavelets can reconstruct the original signal in many different ways, and the commonly used forms of inverse transform are as follows:

$$f(t) = \frac{1}{2\pi\lambda} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} W(\tau, \omega, \lambda, \sigma) e^{i\omega(t-\tau)} d\tau d\sigma \quad (2.4)$$

The Eq (2.4) shows that the inverse transform is independent on $p(t)$, $p(\omega)$ and σ , and the fast Fourier transform algorithm can directly obtain the reconstructed signal. The FSWT realizes time-frequency analysis of the signal, which can filter and segment the signal components in any frequency.

3. Residual neural network

ResNet was proposed by He of Microsoft Research in 2015 [44]. The ResNet solves the training difficulties caused by increasing the depth of the network, and its network performance is far better than the traditional network models.

Suppose a regular convolution neural network has L layers, where the input of layer i ($i \in 1, 2, \dots, L$) is x^i and its corresponding parameter is w^i , and the output of this layer is $y^i = x^i + 1$. For the sake of simplicity of presentation, ignoring the number of layers and bias, the relationship between them can be expressed as Eq (2.5):

$$y = F(x, w_f), \quad (2.5)$$

where F is the nonlinear activation function and w_f is the convolution operation. The depth of ResNet can be expressed by Eq (2.6):

$$y = F(x, w_f) + x \quad (2.6)$$

A simple deformation of Eq (2.6) yields Eq (2.7):

$$F(x, w_f) = y - x \quad (2.7)$$

The function F that the network needs to learn is actually the residual term $y - x$ at the right end of the formula, called the residual function.

As is shown in Figure 1, there are two branches in the residual learning module: The residual function $F(x)$ and the input constant mapping x . These two branches are integrated by adding the

corresponding elements through the nonlinear transformation activation function ReLU, then forms the whole residual learning module, and the structure formed by stacking multiple residual modules is called “ResNet”. This deep learning framework adds residual connectivity, which is more convenient than the original mapping without residual connectivity. It uses a stacked nonlinear combination to fit a constant mapping, making its residuals more likely to converge to zero. Optimizing the residual function $y = F(x, w_f) + x$ is easier than optimizing a complex nonlinear mapping $F(x, w_f) = y - x$, when the network layers are deep enough.

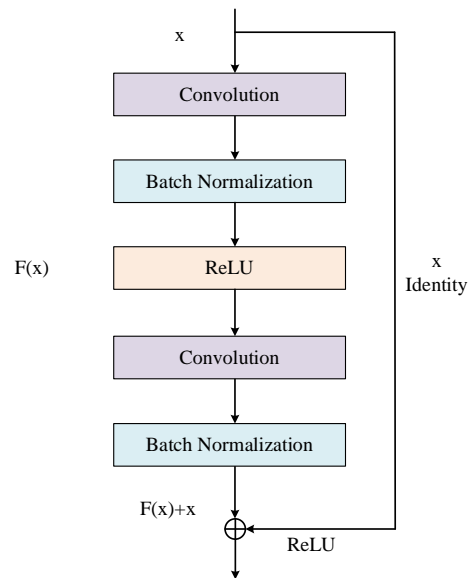


Figure 1. Residual learning modules.

4. Deformable convolution based on ResNet

In deformable convolution [45], an offset parameter $\{\Delta P_n | n = 1, 2, \dots, N\}$ is added to the rectangular region R where the regular convolution neural network acts, as is shown in Figure 2.

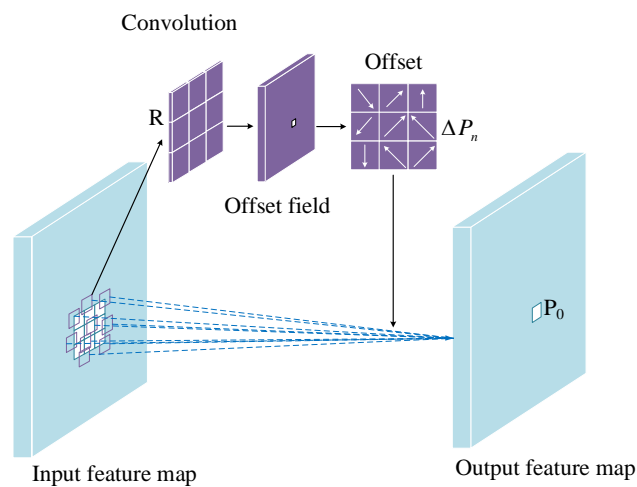


Figure 2. Schematic diagram of deformable convolution.

Where $N = |R|$, the output p_0 in the feature diagram position is calculated as Eq (2.8):

$$y(p_0) = \sum_{r \in R} w(r) * x(p_0 + r + \Delta r) \quad (2.8)$$

Since Δr is a floating-point number, bilinear interpolation is used to sample x . Make $p = p_0 + r + \Delta r$ and use q to represent all possible integer space locations within the feature diagram x , the bilinear interpolation for x are as follows:

$$x(p) = \sum_q G(q, p)x(q), \quad (2.9)$$

where G denotes a two-dimensional bilinear interpolation kernel that can be decomposed into two one-dimensional kernels.

$$G(q, p) = g(q_x, p_x)g(q_y, p_y), \quad (2.10)$$

where $g(a, b) = \max(0, 1 - |a - b|)$. For the vast majority of position q , $G = 0$ and G is the transformation kernel to be learned. The gradient with respect to the offset parameter Δr is calculated as follows:

$$\frac{\partial y(p_0)}{\partial \Delta r} = \sum_{r \in R} w(r) \frac{\partial x(p_0 + r + \Delta r)}{\partial \Delta r} = \sum_{r \in R} \left[w(r) \sum_q \frac{\partial G(q, p_0 + r + \Delta r)}{\partial \Delta r} x(q) \right] \quad (2.11)$$

From the feature diagram of input, the deformable convolution learns the offset parameter through the convolution network with two directions, indicating both x and y directions. With the feature diagram input and the offsets carried out, the computation of the deformable convolution can be performed.

For bearing fault diagnosis, the TFD obtained by FSWT embodies more complex fault features. Since the standard convolution kernel lacks the geometric transformation mechanism inside the diagram, its fix geometry can only be sampled at the fixed position, but the fixed position may lose fault features. Therefore, the network needs deformable convolution kernels to adapt to the feature positions of the diagram. These deformable convolution kernels can perform more detailed feature mining on the TFD to improve the accuracy of bearing fault diagnosis. In this proposed method, the convolution layer in the residual block structure is replaced by a deformable convolution layer, as is shown in Figure 3.

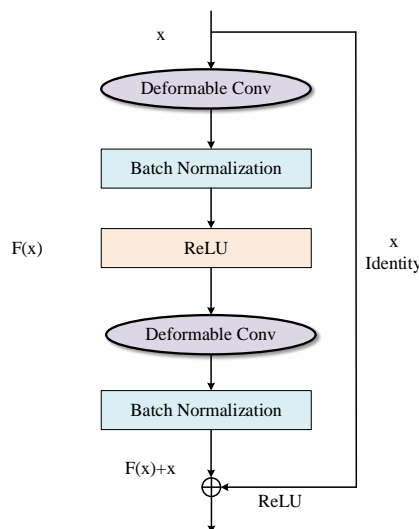


Figure 3. Deformable convolution residual learning module.

5. Fault diagnosis model based on DC-ResNet

The basic ResNet structure is ResNet18. Deformable convolution layers are introduced to the basic network structure, replacing all the traditional 3×3 convolution layers with deformable convolution layers in the original ResNet18. Then, more detailed features are learned by using the adaptive nature of the shape of the convolution kernel and high-precision fault diagnosis of printing press bearings is achieved. The network structure is shown in Figure 4 and Table 1.

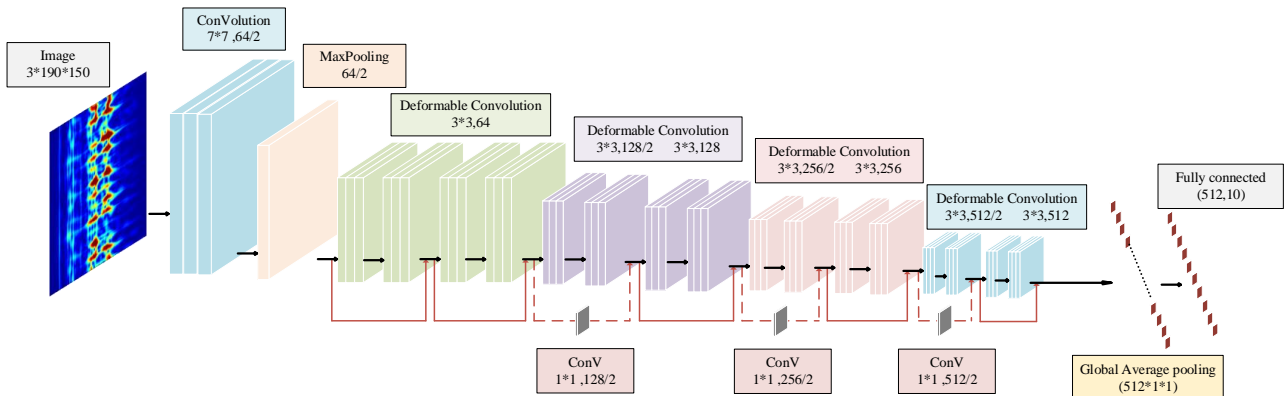


Figure 4. Structure of DC-ResNet.

Table 1. Structural parameters in this proposed method.

DC-ResNet Structure	Output Dimension
Input Layer	(190,150,3)
Convolution Layer 1	(184,144,64)
Max Pooling	(92,72,64)
Residual Block 1	(92,72,64)
Residual Block 2	(92,72,64)
Residual Block 3	(92,72,128)
Residual Block 4	(92,72,128)
Residual Block 5	(92,72,256)
Residual Block 6	(92,72,256)
Residual Block 7	(92,72,512)
Residual Block 8	(92,72,512)
Global Average Pooling	512
Output Layer	10

The flowchart of the bearing fault diagnosis method of the printing press combines the time-frequency processing of FSWT with DC-ResNet, as is shown in Figure 5, and its detailed steps are as follows:

- (1) Use sensor acquisition vibration signals of bearing faults during the operation of the printing

press.

(2) Preprocess the bearing vibration signal collected above using FSWT, and obtain the corresponding TFD of the bearing vibration signal.

(3) Initialize the ResNet18 network model parameters, set its learning rate to 0.001 and the number of iterations to 50, and migrate the model parameters and data to the GPU to accelerate the computational process.

(4) Input the TFD of the vibration signal obtained in step (2) into the network model of ResNet18 and normalize it. Introduce deformable convolution layers to replace all the 3×3 convolution layers in the original network, optimize the gradient descent using the Adagrad optimizer to obtain the optimal model parameters, output the DC-ResNet model parameters, and test them with an unknown category data set.

(5) Get the test results of bearing fault diagnosis, and then get the printing press bearing health status.

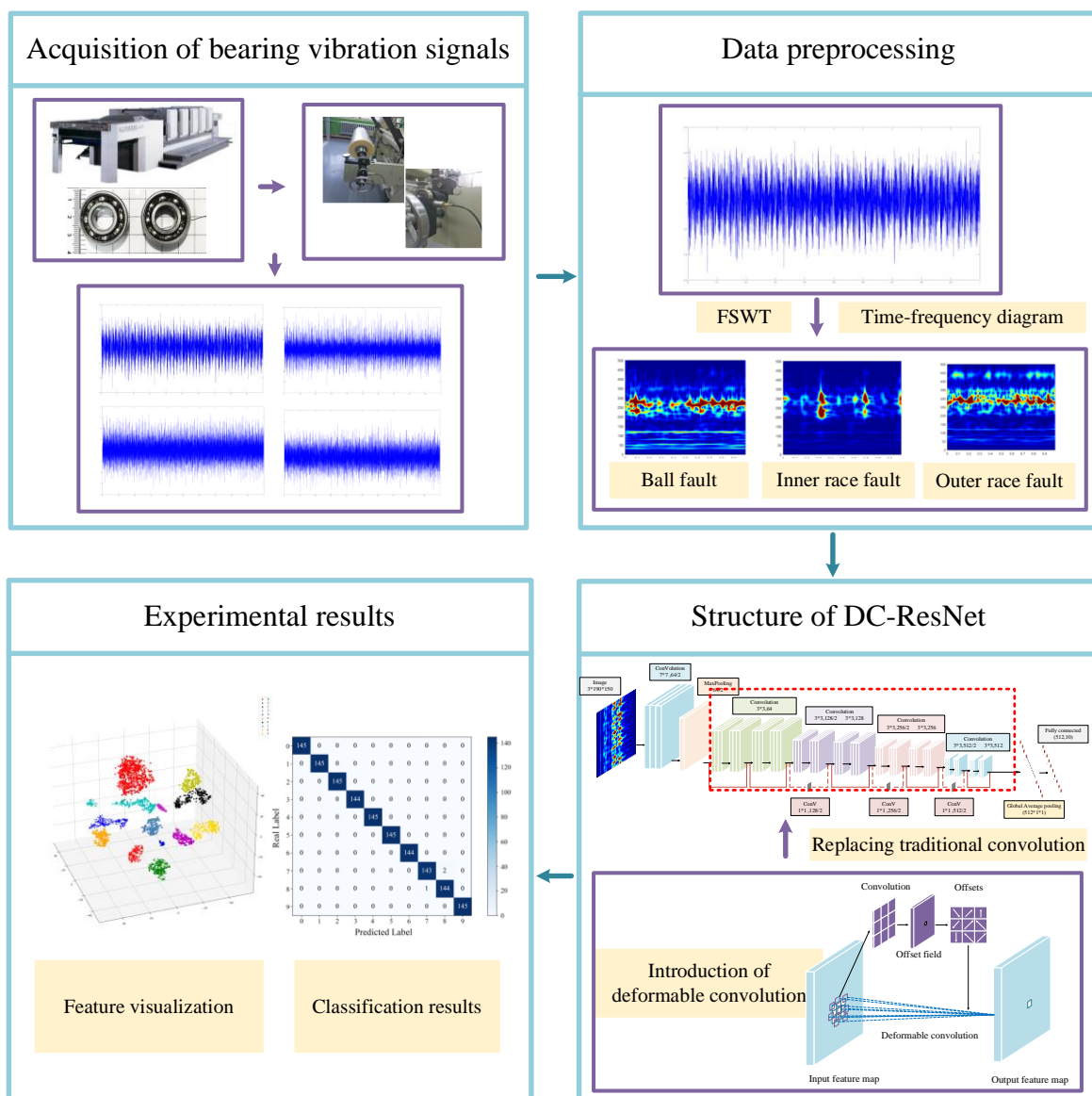


Figure 5. Flowchart of DC-ResNet based bearing fault diagnosis method for printing press.

6. Experiment analysis

6.1. Case 1: bearing data set of case western reserve university

6.1.1. Data description

To verify the effectiveness of the proposed method, the vibration acceleration data of rolling bearings from CWRU are used for experimental analysis [46]. This fault diagnosis experimental setup consists of a torque meter, a power meter and a three-phase asynchronous motor. The sample data contains signals of three different fault types at a sampling frequency of 12 kHz, as well as one normal vibration signal: normal condition, inner race fault, outer race fault and ball fault. Within each fault type, there are three different types of fault depths, 0.18 mm, 0.36 mm and 0.54 mm respectively. The data set contains a total of 9 types of faults and 1 type of normal data.

Figure 6 shows the TFD of the bearing vibration signals of time domain waveforms after Morlet wavelet transform processing and the TFD after FSWT processing, which collected under three typical fault types: inner race fault, outer race fault and ball fault. Figure 6 shows comparison of the TFD obtained after FSWT with the TFD obtained after the Morlet wavelet transform, where the former has better time-frequency focus, more concentrated energy distribution and more obvious fault features, which are more conducive to the identification and classification of bearing faults.

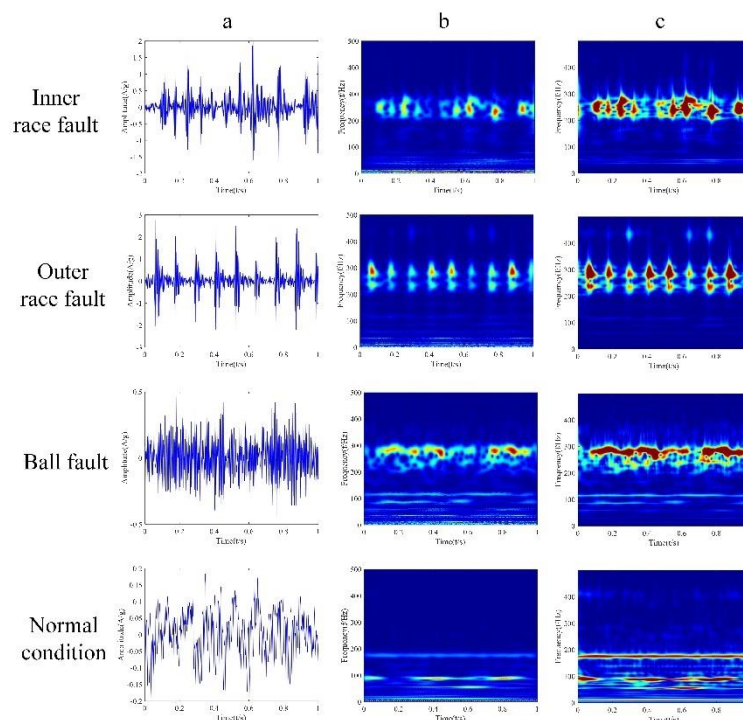


Figure 6. Time domain signal of CWRU bearing vibration and its corresponding TFD. (a) Time domain vibration signals. (b) Morlet wavelet TFD. (c) TFD of FSWT.

Figure 7 shows the control experimental data set. It uses grayscale diagrams formed by simply splicing the one-dimensional time-domain vibration signals into two-dimensional signals.

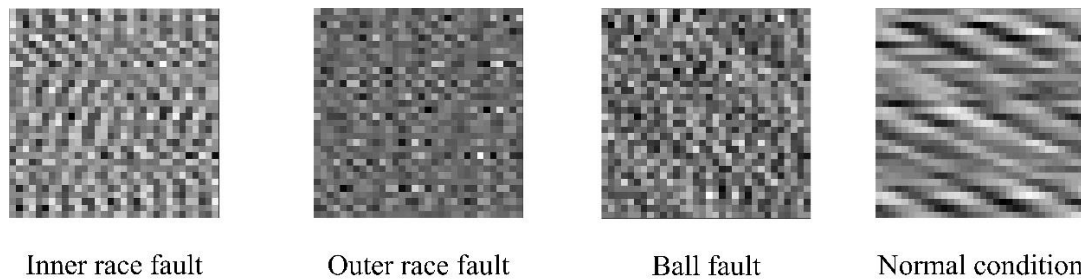


Figure 7. Control experimental data set.

6.1.2. Results and analysis

In order to verify the effectiveness of applying the FSWT to the ResNet in this proposed method, a comparison experiments was conducted. The TFD of FSWT are used as the input of the ResNet18, and the grayscale diagram and Morlet wavelet TFD are used for comparison experiments. Each set of experiments is repeated five times, and the final average diagnosis results obtain by these three preprocessing methods are shown in Table 2. As can be seen from Table 2, compared with the grayscale diagram and Morlet wavelet TFD, the data set obtain by FSWT has a significant improvement in terms of accuracy, and the average accuracy can reach 99.77%, which is higher than the diagnosis results of Morlet wavelet TFD (88.72%) and grayscale diagram (74.92%), proving that the FSWT to the ResNet has more advantages and effectiveness of this proposed method.

Table 2. Average diagnosis results of the three methods.

Preprocessing Methods	Number of Training Samples/Test Samples	Average Accuracy (%)
FSWT	3000/1000	99.77
Morlet Wavelet Transform		88.72
Grayscale		74.92

In order to compare the effects of three different data preprocessing methods on the model diagnosis performance, confusion matrix is introduced [47]. As is shown in Figure 8, the predicted labels of 1000 test samples are compared with the true labels and obtain the confusion matrix of the three data sets based on the ResNet18 model. The color band on the right side of the figure indicates the degree of being accurately classified. The color block at the diagonal position is the number of samples correctly classified in each category, with darker colors representing more correctly classified samples. The color block at the remaining positions is the number of samples misclassified in each category, with lighter colors representing fewer samples misclassified. Among them, Figure 8a shows the confusion matrix heatmap of the grayscale diagram data set, the distribution of each color block is scattered. Each fault is misclassified, and the diagnosis accuracy is low. Figure 8b shows the confusion matrix heatmap of the Morlet wavelet transform TFD data set, from which it is obvious that the color block distribution is significantly better, indicating that, after the Morlet wavelet processing, the ResNet18 model has a certain improvement in the classification effect of bearing faults. Figure 8c shows the confusion matrix heatmap of the TFD data set when FSWT is applied, from which it can be seen that the color block distribution is obviously concentrated, and only individual diagrams are classified wrongly. The above conclusions indicating that the classification accuracy of the ResNet18

model is further improved after the FSWT processing, and has a good classification effect.

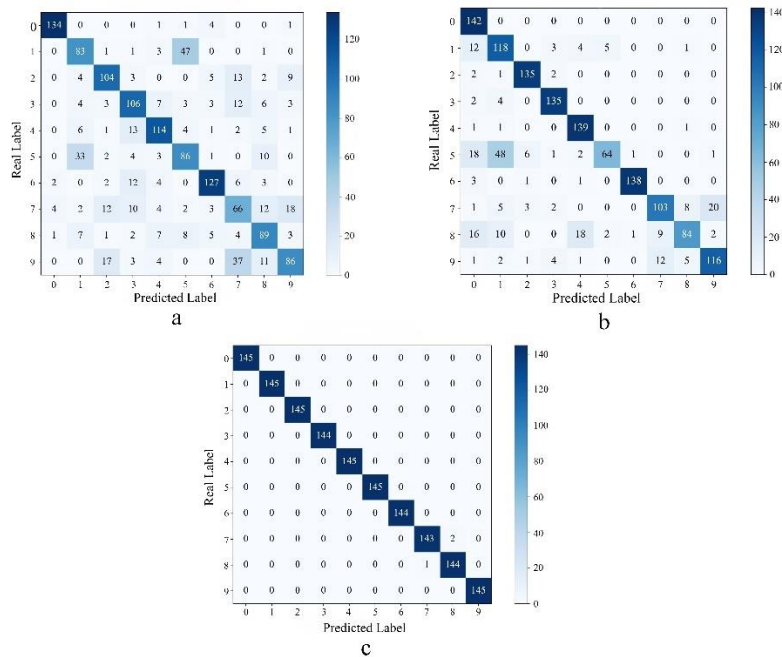


Figure 8. Confusion matrix for the three data set. (a) Time domain grayscale diagrams data set. (b) Morlet wavelet TFD data set. (c) TFD of FSWT data set.

In order to ensure the well training effect of the fault diagnosis algorithm proposed in this paper, three types of hyperparameters in DC-ResNet are tested, which are learning rate, batch size and epoch. A total of 12 different groups of learning rate, batch size and epoch were selected for the experiments, and each group of experiments was repeated five times, and the final accuracy was the average of the five experiments, as shown in Table 3.

Table 3. The accuracy of model validation with different combinations of hyperparameters.

Learning Rate	Batch Size	Epoch	Average Accuracy (%)
0.1	16	30	25.26
0.1	32	50	28.25
0.1	64	70	24.59
0.01	16	30	94.99
0.01	32	50	96.25
0.01	64	70	92.68
0.001	16	30	93.56
0.001	32	50	99.83
0.001	64	70	99.72
0.0001	16	30	85.33
0.0001	32	50	91.69
0.0001	64	70	87.16

From Table 3, when the learning rate = 0.1, the network cannot converge due to the large value of the learning rate, resulting in the phenomenon of gradient explosion; when the learning rate = 0.01, the learning rate is still large, and the network cannot converge precisely to the specified accuracy range; when the accuracy rate = 0.001, the accuracy rates are all at higher values, and when the batch size = 32, epoch = 50, the model accuracy reaches the peak; when the accuracy rate = 0.0001, the learning rate is so small, which leads to the slow convergence of the network and reduces the efficiency. In addition, it can be seen that, when the learning rate is kept constant and the batch size = 32 and epoch = 50, the model accuracy is higher than that of the model with other parameters in the same group. Therefore, the hyperparameters chosen in this paper are learning rate = 0.001, batch size = 32 and epoch = 50. In order to demonstrate the effectiveness of the DC-ResNet, comparative experiments are conducted on two different network models: ResNet50 and ResNet101. Then, using the same data set, deformable convolution layers are introduced in both ResNet50 and ResNet101, named DC-ResNet50 and DC-ResNet101. Each set of experiments are repeated five times, and the accuracy and the computing time of the program are recorded for each experiment. The final average accuracy and the average computing time of the program are obtained, as is shown in Table 4. From Table 4, compared with the ResNet50 and ResNet101, the average accuracy of the ResNet18 can reach 99.77%, and the accuracy is close to that of ResNet50 (99.82%) and ResNet101 (99.70%). However, the ResNet50 and ResNet101 are much larger than the proposed method in terms of program computing time. So, this proposed method has more advantages in combining both program computing time and accuracy, which proves that the DC-ResNet has certain advantages when performing bearing fault diagnosis and proves the effectiveness in this proposed method.

Table 4. Average diagnosis results of different network models.

Experimental Methods	Input Samples	Average Accuracy (%)	Computing time(s)
DC-ResNet	TFD of FSWT	99.83	961.2
ResNet18		99.77	837.6
DC-ResNet50		99.83	9279.6
ResNet50		99.82	1300.2
DC-ResNet101		99.76	18199.2
ResNet101		99.70	1832.0

T-Distributed Stochastic Neighbor Embedding (T-SNE) is a technique that can reduce high-dimensional data to two-dimensional or three-dimensional, and visualize the spatial distribution of the data [48]. Figure 9 shows a schematic diagram of feature visualization after partial training samples are dimensionally reduced by T-SNE. Figure 9a shows the result of partial original training samples after being down scaled, from which it can be seen that the overlap between samples of different categories is very serious. Figure 9b shows the results of the dimensional reduction of the samples in the fully connected layer after DC-ResNet processing. Although there are still a small number of misclassifications, most of the categories can be separated. It is obvious that after DC-ResNet layer-by-layer feature extraction, all kinds of faults tend to separate from each other.

In addition, it can be seen from Table 4 that the diagnosis accuracy of the ResNet18 network model has reached 99.77%, and DC-ResNet has reached 99.83%. The experimental results obtained from DC-ResNet50 and DC-ResNet101 are much greater than the method used in terms of program computing time, with a small increase in both accuracies. It proves that this proposed method has

higher diagnosis accuracy and higher efficiency with less computing time, and it is more suitable for bearing fault diagnosis.

In order to demonstrate, intuitively, the effectiveness of the DC-ResNet model proposed for classification detection on TFD of FSWT data set, the relationship between feature diagram and weight distribution diagram are analyzed by using the Gradient Weighted Class Activation Mapping (Grad-CAM) visual analysis technique [49], it can be judged whether the weight distribution of the model pays more attention to the feature area.

Figure 10, shows the visualization results of Grad CAM by using ResNet18 and DC-ResNet. It is shown that, not only the convolution kernel weights focus more on the region where the fault features are located, but also it has good recognition effect on the tiny fault features. It illustrates that the features extracted by the DC-ResNet are the regions where the fault features are located, proving that the model does not produce overfitting and has certain generalization ability.

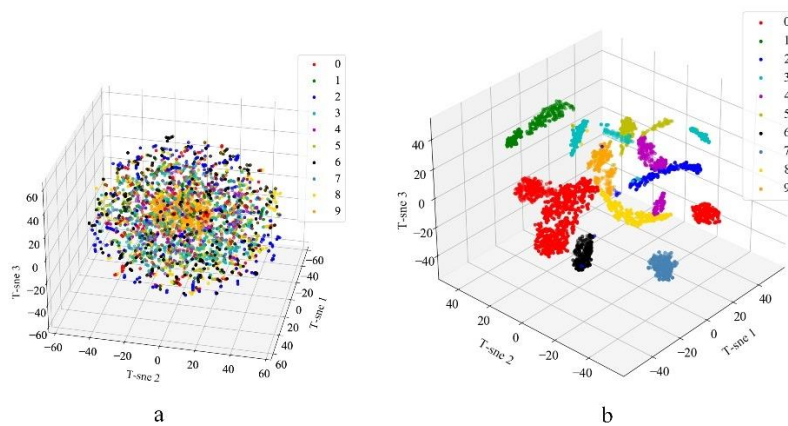


Figure 9. Visualization of T-SNE 3D features. (a) Original training samples. (b) DC-ResNet processing.

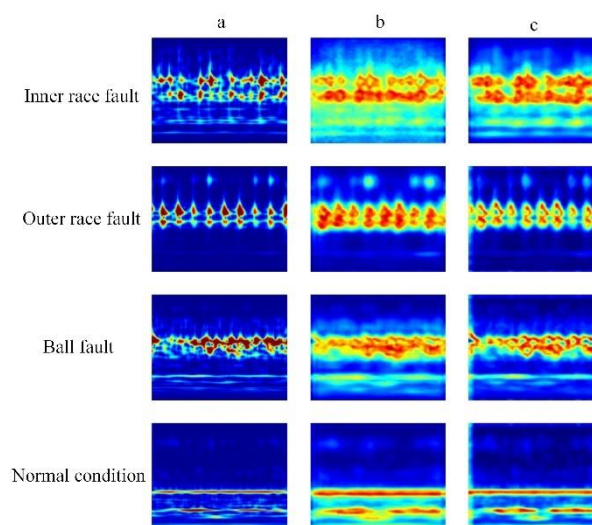


Figure 10. The Grad-CAM visualization of the proposed method. (a) Original TFD. (b) Weight distribution of ResNet18. (c) Weight distribution of DC-ResNet.

6.2. Case 2: printing press bearing data set

6.2.1. Data description

The experimental platform used in this proposed method is the working conditions of the printing press-bearing fault diagnosis experimental platform, as is shown in Figure 11. The sensor used to measure the acceleration signal in the process of signal acquisition is a piezoelectric sensor, which is mounted on the rolling bearing seat for the acquisition of acceleration signal of the bearing, and the acquired signal is saved in the computer for bearing fault diagnosis, subsequently. The bearing model is JYB6004, and its main parameters are shown in Table 5. The bearing fault processing method adopts electron discharge machining, and the fault depths are 0.2 mm, 0.4 mm and 0.6 mm, respectively. Meanwhile, three kinds of faults are introduced into the bearing in three positions: inner race fault, outer race fault and cage fault. After completing the construction of the experimental platform of the rolling bearing of the printing press, the vibration acceleration data of the bearing fault under the actual working condition were collected. In this paper, the speed of the paper delivery shaft of the printing press is set to 800 RPM and the sampling frequency is 12 kHz, and the vibration data of each type of printing press bearing under normal condition is collected in this state, where the bearings with these fault types and the bearings in normal condition constitute the rolling bearing experimental data set used in this experiment. The data set contains a total of nine types of faults and one normal data.

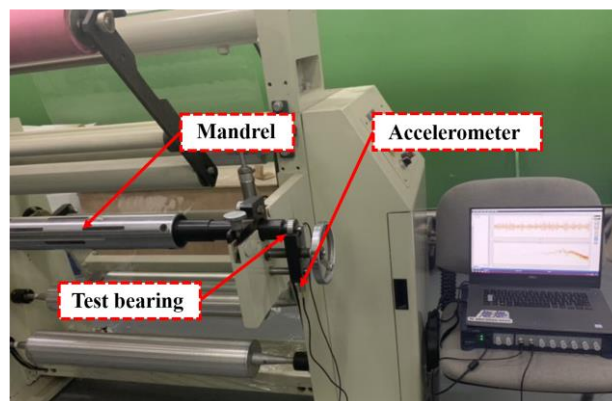


Figure 11. Printing press bearing fault diagnosis experiment platform.

Table 5. Bearing parameters of bearing fault diagnosis platform of the printing press.

Bearing type	Inner diameter /mm	Outer diameter /mm	Width /mm	Weight /g
JYB6004	20	42	12	69

Figure 12 shows the time domain waveforms, the TFD after Morlet wavelet transform processing, and the TFD after FSWT processing of the bearing vibration signals collected under three types of faults: inner race fault, outer race fault and cage fault. From Figure 12, for the bearing vibration signals collected under the actual working conditions with strong noise, they are easily disturbed by noise in the process of time-frequency transform, which leads to the lack of obvious fault features extracted from the TFD after Morlet wavelet transform processing. At the same time, the TFD of FSWT has a

good time-frequency focus, and the energy distribution is more concentrated. The TFD of FSWT is divided into a training set and a test set in the ratio of 7:3, which is the training set and the test set in the training process of the network model of the proposed method.

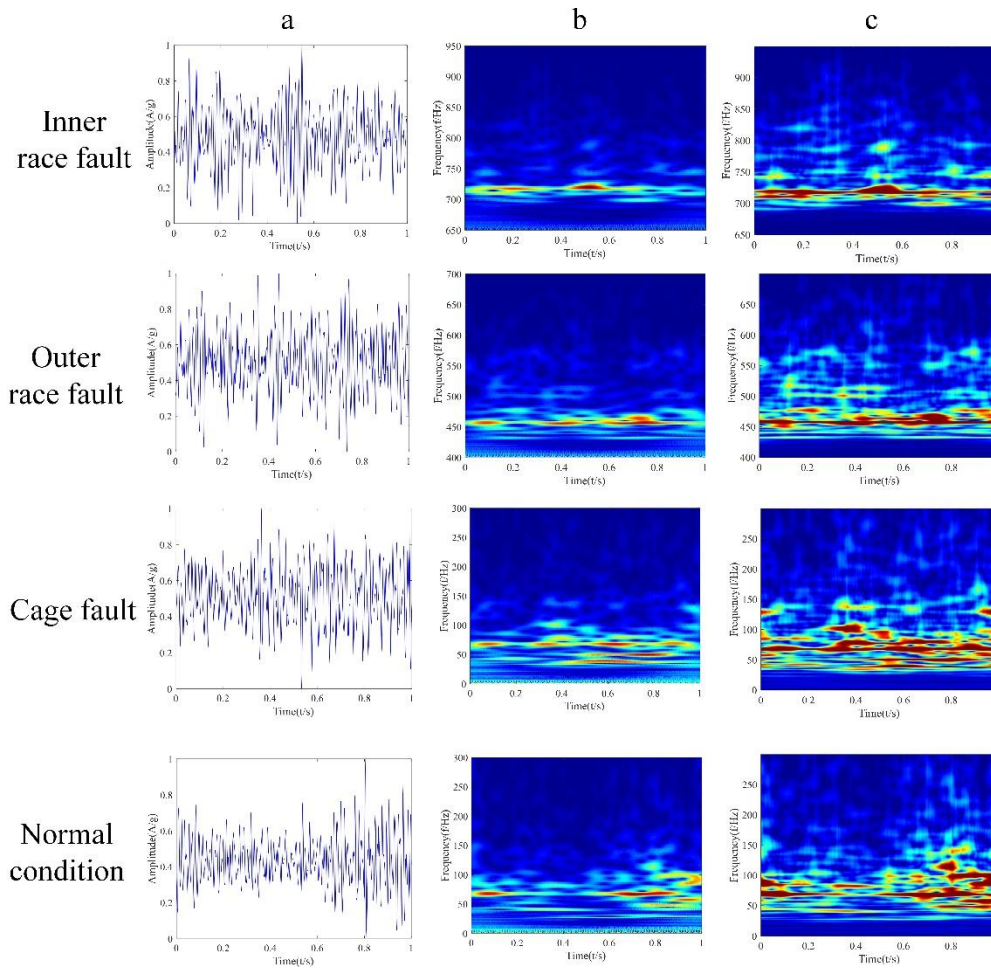


Figure 12. Time domain signal of printing press bearing vibration and its corresponding TFD. (a) Time domain vibration signals. (b) Morlet wavelet TFD. (c) TFD of FSWT.

Figure 13 shows the control experimental data set. It uses grayscale diagrams formed by simply stitching the one-dimensional time-domain vibration signals into two-dimensional signals.

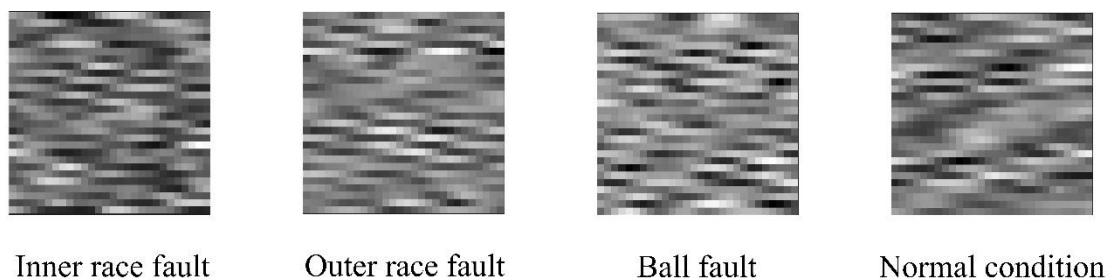


Figure 13. Control experimental data set.

6.2.2. Results and analysis

The accuracy and loss curve in one experimental iteration obtained by inputting the preprocessing TFD of the printing press bearings vibration signal into the DC-ResNet model is shown in Figure 14. It can be seen that the loss value gradually decreases with the increase in the number of iterations. At the same time, the accuracy rate shows a trend of gradual increase, and finally stabilizes at more than 90%. The analysis results show that the fault diagnosis method based on the DC-ResNet can extract the fault features from the TFD, effectively.

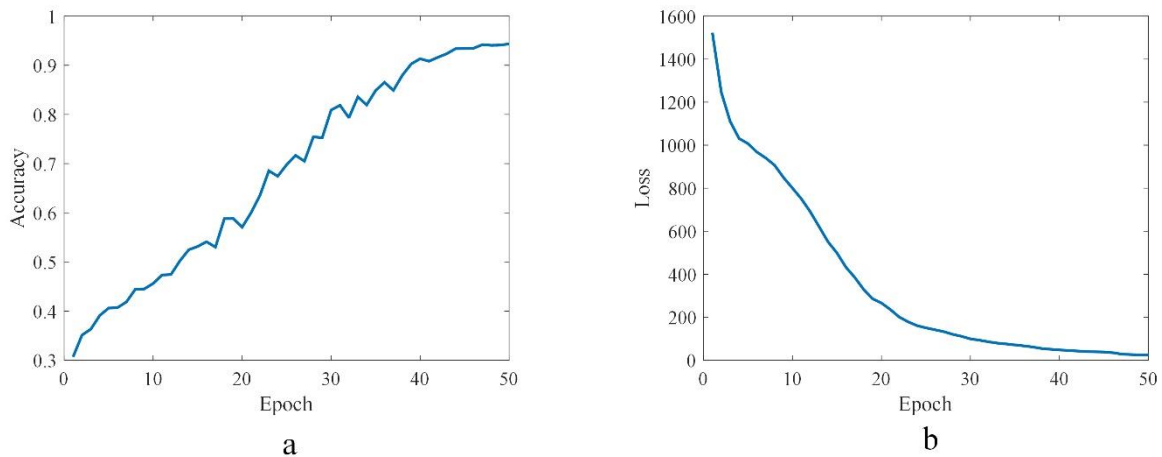


Figure 14. The variation curves of test accuracy and loss.

The DC-ResNet is compared with the AlexNet network and LSTM, and the TFD is used as the input of the neural network. The number of train samples is 10000, and the number of test samples is 3000, and each group of experiments is repeated five times, and the final results obtained by the three models are shown in Figure 15. Compared with the traditional AlexNet and LSTM network models, the DC-ResNet network structure improves fault diagnosis accuracy significantly. The average diagnosis accuracy can reach 93.90%, which proves that the network structure has certain advantages when performing bearing fault diagnosis in actual printing conditions with strong noise. From Figure 15, it can be seen that the accuracy and stability of the experimental results obtained by the three neural networks are significantly improved after using the FSWT preprocessing TFD as the input. When the TFD of FSWT is used as the input of shallow model with weak feature extraction ability, the shallow model diagnosis accuracy is improved. Therefore, the TFD of FSWT has superior performance in fault feature extraction.

In addition, it can also be seen from Figure 15 that the proposed method has high accuracy and well stability when FSWT is used for signal preprocessing. After five averages, the experimental results obtained by the three methods are shown in Table 6. It shows that the average diagnosis accuracy can reach 93.90%, which is significantly better than that of the compared methods, AlexNet (83.27%) and LSTM (83.36%).

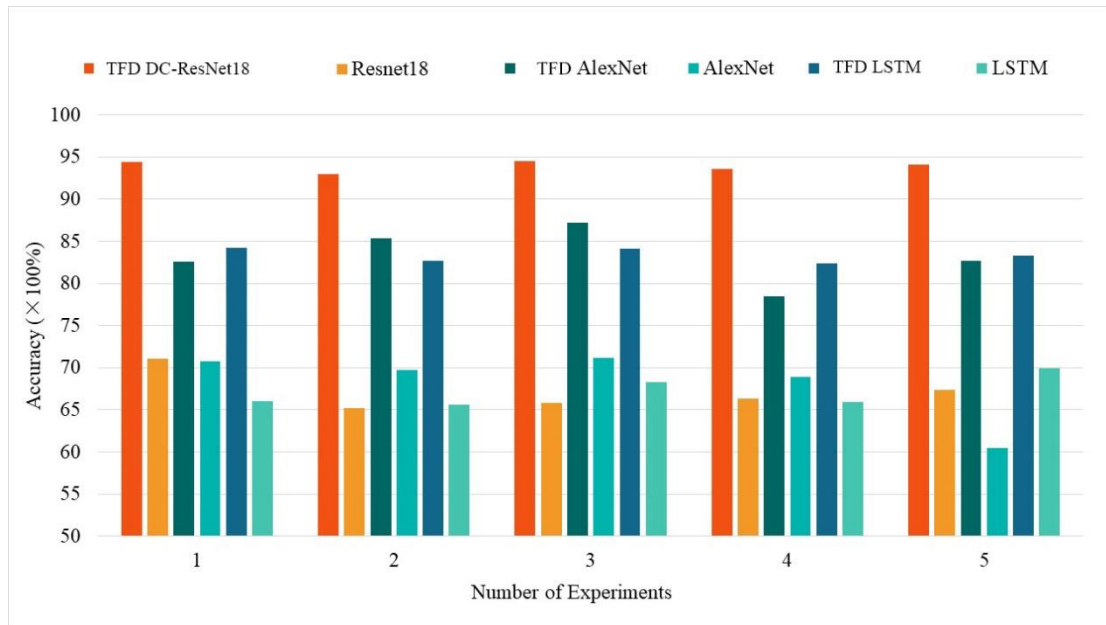


Figure 15. Comparison chart of the analysis results of the different diagnosis methods.

Table 6. Average diagnosis results of the different methods.

Method	Preprocessing Method	Average Accuracy (%)
DC-ResNet	FSWT	93.90%
	Grayscale	77.52%
ResNet18	FSWT	89.98%
	Grayscale	67.14%
AlexNet	FSWT	83.27%
	Grayscale	68.20%
LSTM	FSWT	83.36%
	Grayscale	67.15%

Table 7 shows the average diagnosis results with the introduction of deformable convolution layers. It shows that the DC-ResNet structure has better feature extraction ability than the ordinary ResNet18, and the average diagnosis accuracy can reach 93.90%, which is higher than ResNet18 (89.98%). When the diagnosis accuracy is guaranteed, it can be seen from the mean square error that the method proposed is more stable than other two methods. In summary, the DC-ResNet proposed has higher accuracy and stability in bearing fault diagnosis, and improves the performance of printing press bearing fault diagnosis.

Table 7. Average diagnosis results with the introduction of deformable convolution layers.

Method	Preprocessing Method	Average Accuracy (%)	MSE
DC-ResNet	FSWT	93.90%	0.0038
ResNet18		89.98%	0.0106

The confusion matrix of the DC-ResNet on the printing press-bearing data set is shown in Figure 16. From the figure, it can be seen that the color block distribution is concentrated on the diagonal, where only individual images are misclassified, and the overall classification is good, which indicates the DC-ResNet achieves good classification results on the printing press-bearing data set.

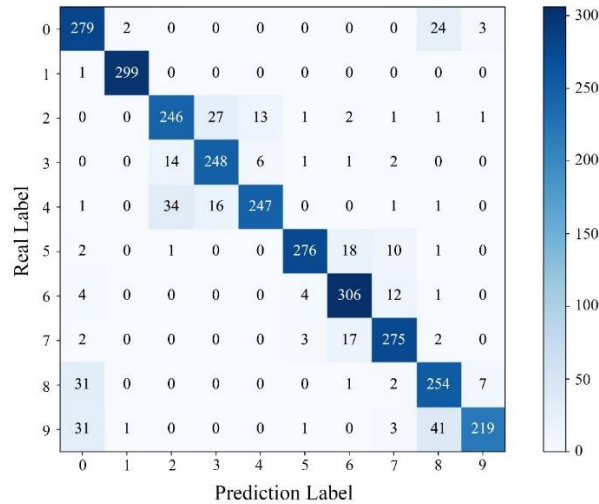


Figure 16. Confusion matrix for DC-ResNet.

Figure 17 shows the feature visualization diagram of some training samples of the printing press-bearing data set after T-SNE dimensional reduction. Figure 17a shows the results of the original training samples after T-SNE dimensional reduction, from which it can be seen that the overlap between samples of different categories is very serious. Figure 17b shows the results of the dimensional reduction of samples in the fully connected layer after DC-ResNet processing. There is still a small number of misclassifications, but most of the categories can be separated. It is obvious that, after the layer-by-layer feature extraction of the network model, the various types of faults tend to separate from each other.

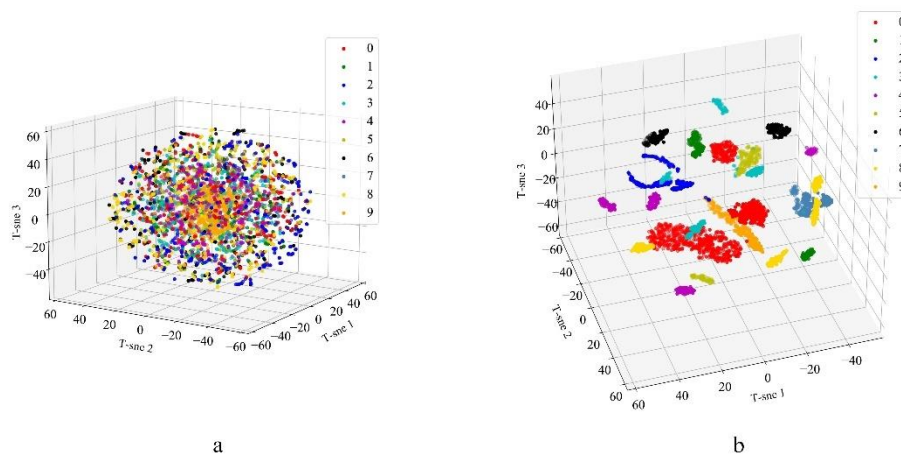


Figure 17. Visualization of T-SNE 3D features. (a) Original training samples. (b) DC-ResNet processing.

Figure 18 shows the visualization results of Grad-CAM in ResNet18 and DC-ResNet. It is obvious from the figure that, for the printing press bearing data set, the ResNet18 is not effective in recognizing the fault features. However, the DC-ResNet pays more attention to the tiny fault features, and the recognition effect is good. It illustrates that the classification features extracted by the network model are the regions where the fault features are located, proving that the model does not produce overfitting and has certain generalization ability.

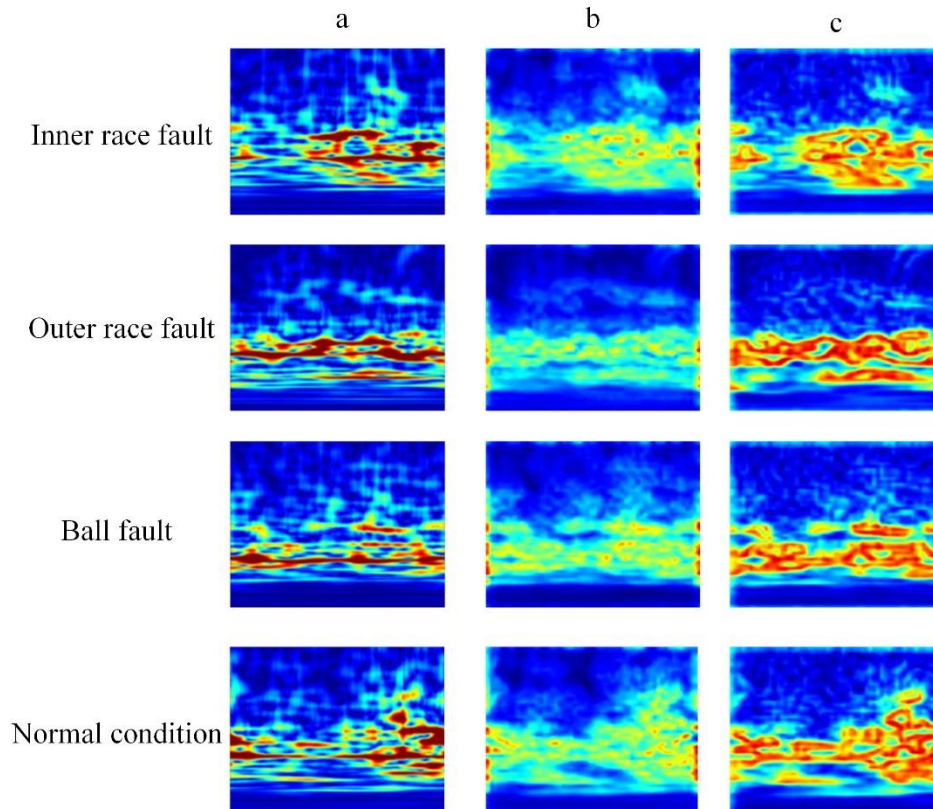


Figure 18. The Grad-CAM visualization of proposed method. (a) Original TFD. (b) Weight distribution of ResNet18. (c) Weight distribution of DC-ResNet.

7. Conclusions

An intelligent bearing fault diagnosis method combining FSWT and DC-ResNet is proposed. Firstly, the vibration signal is preprocessed by FSWT to obtain the TFD. Secondly, the obtained TFD is input to ResNet to extract the fault features for fault identification. In this process, the deformable convolution layer is introduced for the strong noise vibration signal under the actual working condition of the printing press to improve the self-adaptability of the DC-ResNet. In the experimental part, the effectiveness of the proposed method is verified using the data set collected under experimental conditions and the data set of printing press bearings collected under actual working conditions. Thus, the analysis results can be concluded as follows:

(1) The FSWT is used to preprocess the original vibration signal, reducing the dependence on wavelet basis functions, while reconstructing the signal by wavelets and wavelet packets, and realizing the reconstruction of the signal in arbitrary frequency bands and the accurate description of local

features. This preprocessing method can extract the information of bearing fault features effectively, and it can also improve the accuracy of the subsequent intelligent fault diagnosis model.

(2) In this proposed method, the traditional ResNet18 is improved for the complex and tiny fault features of strong noise vibration signals under the actual working conditions of printing presses. The ResNet convolution layer is reconstructed by using the deformable convolution structure. The DC-ResNet can learn more detailed features by using the adaptive nature of the shape of the deformable convolution layer, which improves the accuracy of intelligent fault diagnosis of printing press bearings.

(3) The effectiveness of the method was verified using experiments. Among them, the accuracy reaches 99.77% under experimental conditions and 93.90% for the data set of printing press bearings under actual working conditions. The experimental results show that the DC-ResNet can classify different rolling bearings faults under strong background noise, and the accuracy and stability of fault diagnosis are greatly improved.

Acknowledgments

This research is supported by the Natural Science Foundation of Shanxi Province (Grant No.2022JZ-30), and Scientific Research Program of Shanxi Provincial Department of Education (Grant No.20JY054).

Conflict of interest

We declare that there are no conflicts of interest.

References

1. J. Jiao, M. Zhao, J. Lin, C. Ding, Deep coupled dense convolution network with complementary data for intelligent fault diagnosis, *IEEE Trans. Instrum. Meas.*, **66** (2019), 9858–9867. <https://doi.org/10.1109/tie.2019.2902817>
2. W. Deng, H. Liu, J. Xu, H. Zhao, Y. Song, An improved quantum-inspired differential evolution algorithm for deep belief network, *IEEE Trans. Instrum. Meas.*, **69** (2020), 7319–7327. <https://doi.org/10.1109/tim.2020.2983233>
3. J. Liu, A dynamic modelling method of a rotor-roller bearing-housing system with a localized fault including the additional excitation zone, *J. Sound Vib.*, **469** (2020), 115144. <https://doi.org/10.1016/j.jsv.2019.115144>
4. R. Yan, F. Shen, C. Sun, X. Chen, Knowledge transfer for rotary machine fault diagnosis, *IEEE Sens. J.*, **20** (2019), 8374–8393. <https://doi.org/10.1109/jsen.2019.2949057>
5. Y. Lei, F. Jia, J. Lin, S. Xing, S. Ding, An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data, *IEEE Trans. Ind. Electron.*, **63** (2016), 3137–3147. <https://doi.org/10.1109/tie.2016.2519325>
6. W. Wang, F. Golnaraghi, F. Ismail, Condition monitoring of multistage printing presses, *J. Sound Vib.*, **270** (2004), 755–766. [https://doi.org/10.1016/s0022-460x\(03\)00209-8](https://doi.org/10.1016/s0022-460x(03)00209-8)
7. A. Rai, S. Upadhyay, A review on signal processing techniques utilized in the fault diagnosis of rolling element bearings, *Tribol. Int.*, **96** (2016), 289–306. <https://doi.org/10.1016/j.triboint.2015.12.037>

8. J. Tang, J. Wu, B. Hu, J. Liu, Towards a fault diagnosis method for rolling bearing with Bi-directional deep belief network, *Appl. Acoust.*, **192** (2022), 1–11. <https://doi.org/10.1016/j.apacoust.2022.108727>
9. M. Hmimida, R. Kanawati, Community detection in multiplex networks: A seed-centric approach, *Netw. Heterogen. Media*, **10** (2015), 71–85. <https://doi.org/10.3934/nhm.2015.10.71>
10. M. Zhao, S. Zhong, X. Fu, B. Tang, P. Michael, Deep residual shrinkage networks for fault diagnosis, *IEEE Trans. Industr. Inform.*, **16** (2019), 4681–4690. <https://doi.org/10.1109/tii.2019.2943898>
11. X. Jin, Y. Xu, Research on facial expression recognition based on deep learning, *IEEE EITCE*, 2019, 1144–1147. <https://doi.org/10.1109/eitce47263.2019.9095140>
12. R. Wightman, H. Touvron, H. Jégou, Resnet strikes back: An improved training procedure in timm, *ArXiv*, 2021. <https://doi.org/10.48550/arXiv.2110.00476>
13. Z. Lin, P. Wang, Y. Chen, C. Sun, Fault diagnosis of rolling bearing based on improved convolution neural network, *IEEE DDCLS*, 2022, 643–647. <https://doi.org/10.1109/ddcls55054.2022.9858533>
14. S. Zagoruyko, N. Komodakis, Wide residual networks, *ArXiv*, 2016. <https://doi.org/10.48550/arXiv.1605.07146>
15. I. Konovalenko, P. Maruschak, V. Brevus, Steel surface defect detection using an ensemble of deep Residual neural networks, *J. Comput. Inf. Sci. Eng.*, **22** (2022), 1–7. <https://doi.org/10.1115/1.4051435>
16. X. Yan, D. She, Y. Xu, Deep order-wavelet convolutional variational autoencoder for fault identification of rolling bearing under fluctuating speed conditions, *Expert Syst. Appl.*, **216** (2022), 119479. <https://doi.org/10.1016/j.eswa.2022.119479>
17. B. Peng, H. Xia, X. Lv, M. Annor-Nyarko, S. Zhu, Y. Liu, et al., An intelligent fault diagnosis method for rotating machinery based on data fusion and deep Residual neural network, *Appl. Intell.*, **52** (2022), 3051–3065. <https://doi.org/10.1007/s10489-021-02555-4>
18. X. Yan, D. She, Y. Xu, J. Ping, Deep regularized variational autoencoder for intelligent fault diagnosis of rotor–bearing system within entire life-cycle process, *Knowledge-Based Syst.*, **226** (2021), 107142. <https://doi.org/10.1016/j.knosys.2021.107142>
19. B. Wang, Y. Lei, N. Li, T. Yan, Deep separable convolution network for remaining useful life prediction of machinery, *Mech. Syst. Signal Process.*, **134** (2019), 106330. <https://doi.org/10.1016/j.ymsp.2019.106330>
20. P. Liang, W. Wang, X. Yuan, S. Liu, L. Zhang, Y. Cheng, Intelligent fault diagnosis of rolling bearing based on wavelet transform and improved Residual neural network under noisy labels and environment, *Eng. Appl. Artif. Intell.*, **115** (2022), 105269. <https://doi.org/10.1016/j.engappai.2022.105269>
21. W. Zhang, Z. Wang, X. Li, Blockchain-based decentralized federated transfer learning methodology for collaborative machinery fault diagnosis, *Reliab. Eng. Syst. Safe.*, **229** (2023), 108885. <https://doi.org/10.1016/j.res.2022.108885>
22. X. Yan, Y. Liu, M. Jia, Multiscale cascading deep belief network for fault identification of rotating machinery under various working conditions, *Knowledge-Based Syst.*, **193** (2020), 105484. <https://doi.org/10.1016/j.knosys.2020.105484>

23. P. Lin, Z. Qian, X. Lu, Y. Lin, Y. Lai, S. Cheng, Compound fault diagnosis model for Photovoltaic array using multi-scale SE-Residual neural network, *Sustain. Energy Technol.*, **50** (2022), 101785. <https://doi.org/10.1016/j.seta.2021.101785>
24. L. Wan, Y. Li, K. Chen, K. Gong, C. Li, A novel deep convolution multi-adversarial domain adaptation model for rolling bearing fault diagnosis, *Measurement*, **191** (2022), 110752. <https://doi.org/10.1016/j.measurement.2022.110752>
25. X. Yan, Y. Liu, Y. Xu, M. Jia, Multistep forecasting for diurnal wind speed based on hybrid deep learning model with improved singular spectrum decomposition, *Energy Convers. Manage.*, **225** (2020), 113456. <https://doi.org/10.1016/j.enconman.2020.113456>
26. P. Yang, Q. Liu, B. Wang, W. Li, Z. Li, M. Sun, An empirical study of fault diagnosis methods of a dissolved oxygen sensor based on ResNet-50, *Int. J. Sens. Netw.*, **39** (2022), 205–214. <https://doi.org/10.1504/ijnsnet.2022.124566>
27. X. Hao, Y. Zheng, L. Lu, H. Pan, Research on intelligent fault diagnosis of rolling bearing based on improved deep residual neural network, *Appl. Sci.*, **11** (2021), 10889. <https://doi.org/10.3390/app112210889>
28. K. Zhang, B. Tang, L. Deng, X. Liu, A hybrid attention improved Residual neural network based fault diagnosis method of wind turbines gearbox, *Measurement*, **179** (2021), 109491. <https://doi.org/10.1016/j.measurement.2021.109491>
29. Y. Wang, J. Liang, X. Gu, D. Ling, H. Yu, Multi-scale attention mechanism Residual neural network for fault diagnosis of rolling bearings, *Proc. Inst. Mech. Eng., Part C: J. Mech. Eng. Sci.*, **236** (2022), 10615–10629. <https://doi.org/10.1177/09544062221104598>
30. Z. Lai, X. Zhu, J. Wu, Generative focused feedback Residual neural networks for image steganalysis and hidden information reconstruction, *Appl. Soft. Comput.*, **129** (2022), 109550. <https://doi.org/10.1016/j.asoc.2022.109550>
31. D. Hartono, D. Halim, G. W. Roberts, Gear fault diagnosis using an improved reassigned smoothed Pseudo Wigner-Ville distribution, *Cogent Eng.*, **5** (2018), 1436928. <https://doi.org/10.1080/23311916.2018.1436928>
32. S. Chegini, A. Bagheri, F. Najafi, A new intelligent fault diagnosis method for bearing in different speeds based on the FDAF-score algorithm, binary particle swarm optimization, and support vector machine, *Soft. Comput.*, **24** (2020), 10005–10023. <https://doi.org/10.1007/s00500-019-04516-z>
33. K. V. Surti, C. A. Naik, Bearing condition monitoring of induction motor based on discrete wavelet transform & k-nearest neighbor, *IEEE I2CT*, 2018, 1–5. <https://doi.org/10.1109/i2ct.2018.8529687>
34. X. Li, Y. Xu, N. Li, B. Yang, Y. Lei, Remaining useful life prediction with partial sensor malfunctions using deep adversarial networks, *IEEE/CAA J. Autom. Sinica.*, **10** (2022), 121–134. <https://doi.org/10.1109/jas.2022.105935>
35. R. Dubey, D. Agrawal, Bearing fault classification using ANN-based Hilbert footprint analysis, *Iet. Scl. Meas. Technol.*, **9** (2015), 1016–1022. <https://doi.org/10.1049/iet-smt.2015.0026>
36. J. Tian, C. Morillo, M. H. Azarian, M. Pecht, Motor bearing fault detection using spectral kurtosis-based feature extraction coupled with K-nearest neighbor distance analysis, *Ieee. T. Ind. Electron.*, **63** (2015), 1793–1803. <https://doi.org/10.1109/tie.2015.2509913>

37. M. Amar, I. Gondal, C. Wilson, Vibration spectrum imaging: A novel bearing fault classification approach, *IEEE Trans. Ind. Electron.*, **62** (2014), 494–502. <https://doi.org/10.1109/tie.2014.2327555>
38. W. Zhang, Y. Ding, X. Yan, M. Jia, Weak multiple fault detection based on weighted morlet wavelet-overlapping group sparse for rolling bearing fault diagnosis, *Appl. Sci.*, **10** (2020), 2057. <https://doi.org/10.3390/app10062057>
39. W. Mao, L. He, Y. Yan, J. Wang, Online sequential prediction of bearings imbalanced fault diagnosis by extreme learning machine, *Mech. Syst. Signal. Pr.*, **83** (2017), 450–473. <https://doi.org/10.1016/j.ymsp.2016.06.024>
40. H. Zhao, M. Liu, Y. Sun, Z. Chen, G. Duan, X. Cao, Fault diagnosis of control moment gyroscope based on a new CNN scheme using attention-enhanced convolution block, *Sci. China Technol. Sci.*, **65** (2022), 2605–2616. <https://doi.org/10.1007/S11431-022-2141-9>
41. C. Duan, Q. Gao, X. Xu, Generator unit fault diagnosis using the frequency slice wavelet transform time-frequency analysis method, *Proc. CSEE*, **33** (2013), 96–103. <https://doi.org/10.13334/j.0258-8013.pcsee.2013.32.014>
42. Z. Sheng, Y. Xu, K. Zhang, Applications in bearing fault diagnosis of an improved Kurtogram algorithm based on flexible frequency slice wavelet transform filter bank, *Measurement*, **174** (2021), 108975. <https://doi.org/10.1016/j.measurement.2021.108975>
43. X. Zhong, Research on time-frequency analysis methods and its applications to rotating machinery fault diagnosis, PhD dissertation, Wuhan University of Science and Technology, Wuhan, 2014.
44. K. Hirasawa, M Ohbayashi, M Koga, M Harada, Forward propagation universal learning network, *IEEE ICNN'96*, **1** (1996), 353–358. <https://doi.org/10.1109/icnn.1996.548917>
45. F. Chen, F. Wu, J. Xu, G. Gao, Q Ge, X. Jing, Adaptive Deformable Convolutional Network, *Neurocomputing*, **453** (2021), 853–864. <https://doi.org/10.1016/j.neucom.2020.06.128>
46. W. Smith, R. Randall, Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study, *Mech. Syst. Signal. Pr.*, **64** (2015), 100–131. <https://doi.org/10.1016/j.ymsp.2015.04.021>
47. O. Caelen, A Bayesian interpretation of the confusion matrix, *Ann. Math. Artif. Intel.*, **81** (2017), 429–450. <https://doi.org/10.1007/s10472-017-9564-8>
48. W. Lu, X. Yan, Visualizing high-dimensional industrial process based on deep reinforced discriminant features and a stacked supervised t-distributed stochastic neighbor embedding network, *Expert Syst. Appl.*, **186** (2021), 115389. <https://doi.org/10.1016/j.eswa.2021.115389>
49. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra Grad-cam: Visual explanations from deep networks via gradient-based localization, *Proc. IEEE Int. Conf. Comput. Vis.*, **128** (2020), 336–359. <https://doi.org/10.1007/s11263-019-01228-7>



AIMS Press

©2023 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)