

Fault Diagnosis of Rotation Vector Reducer for Industrial Robot Based on a Convolutional Neural Network

Shuai Yang^{1,*} – Xing Luo² – Chuan Li²

¹ Chongqing Technology and Business University, National Research Base of Intelligent Manufacturing Service, China

² Chongqing Technology and Business University, School of Management Science and Engineering, China

As a key component of a mechanical drive system, the failure of the reducer will usually cause huge economic losses and even lead to serious casualties in extreme cases. To solve this problem, a two-dimensional convolutional neural network (2D-CNN) is proposed for the fault diagnosis of the rotation vector (RV) reducer installed on the industrial robot (IR). The proposed method can automatically extract the features from the data and reduce the connections between neurons and the parameters that need to be trained with its local receptive field, weight sharing, and subsampling features. Due to the aforementioned characteristics, the efficiency of network training is significantly improved, and verified by the experimental simulations. Comparative experiments with other mainstream methods are carried out to further validate the fault classification accuracy of the proposed method. The results indicate that the proposed method out-performs all the selected methods.

Keywords: fault diagnosis, convolutional neural network, RV reducer

Highlights

- An improved method based on deep-learning algorithms is proposed in this paper.
- 2D-FFT and 2D-CNN are combined for the fault diagnosis.
- The efficiency of network training is significantly improved.
- Comparative experiments with other mainstream methods are carried out, and the fault classification accuracy of the proposed method is validated.

0 INTRODUCTION

An industrial robot (IR) is defined as a device with multi-joint manipulator or multi-degree of freedom, which has been widely used in various industries, such as electronics, manufacturing, and aerospace [1]. Due to its automaticity and controllability, IR can provide a variety of functions in industrial processing or manufacturing. In the late 1950s, the first-generation IR was designed and manufactured by General Motors for completing simple and repetitive operations. After several decades of evolution, IR has become one of the most common devices in the manufacturing sector.

With the development of manufacturing and computer science, the IR has been increasing in intelligence, flexibility, and interactivity, which can satisfy the diverse needs of various costumers [2].

For most industrial robots, there are three major components and six subsystems. Among them, the driving system (motor) is the key part that provides power to the mechanical system of an IR [3]. Due to the high rotation speed and torque of the servo motor, a motor reducer is required for the driving system, which can match the rotation speed and transmit the torque between driving motor and actuator. The accuracy of the transmission ratio for a reducer directly affects the displacement precision and reduces the service life of the IR [4].

Because of the harsh working conditions, failure of the reducer is one of the most common phenomena in a mechanical system. Wear and leakage are two major types of failure, which can cause the shut-down of a mechanical system. Therefore, fault detection and diagnosis of the reducer are required.

In recent years, the deep-learning (DL) method has become the dominant method in fault diagnosis for extraction, detection, and classification. The mainstream DL methods include deep neural network (DNN), deep belief network (DBN), recursive neural network (RNN) and convolutional neural network (CNN).

Among them, CNN was initially applied to the recognition and classification of images. However, it also shows promising results in fault diagnosis, which can automatically extract features from the original data without manual selection. By training and adjusting the values of the convolution kernel, CNN can achieve the classification and recognition of fault based on the raw data [5] and [6].

Pan et al. [7] propose an improved bearing fault diagnosis method based on CNN and a long-short-term memory (LSTM) recurrent neural network whose input is the raw sampling signal without any pre-processing or traditional feature extraction. Janssens et al. [8] pre-processed the vibration data of bearing shell by fast Fourier transform (FFT) and input it

into CNN for fault detection. Jing et al. [9] focused on the research and development of CNN, which can directly learn features from the frequency data of vibration signals, and test the different performance of feature learning from original data, spectrum, and time-frequency combined data. Azamfar et al. [10] proposed a fault diagnosis method based on the analysis of motor current characteristics. The data obtained from multiple current sensors are fused by a new two-dimensional convolutional neural network structure, which can be directly used for classification without manual feature extraction. DCNN was applied to solve the multi-state fault identification problem of gearboxes. By directly inputting the original signal, a higher diagnosis accuracy is obtained with a lower calculation time cost [11]. The two-dimensional mapping representation method based on cyclic spectrum coherence (CSCoh) and CNN is used to improve the fault identification performance of rolling bearings, which is a new fault diagnosis method based on DL proposed by Chen et al. [12]. A new full closed-loop method based on a deep convolution neural network is proposed to detect and classify the data of power quality disturbances [13]. Xuan and You [14] proposed a hierarchical convolutional neural network (HCNN) based on DL, which was used to detect pancreatic tumours. This method can improve the performance of the classifier and reduce the cost of the medical Internet of Things (IoMT). A framework based on DL was applied to classify the target data, which adopted deep CNN to learn the compact and effective representation of each signal from wavelet coefficients. This method is used to classify the ultrasonic signals of carbon fibre reinforced polymer (CFRP) specimens with voids and delamination [15]. Lu et al. [16] proposed a new method of rice disease recognition based on a deep convolution neural network and achieved good results. In [17], a time-varying reliability method of rotation vector (RV) reducer for industrial robots with multiple failure modes based on the Kriging model is proposed.

Based on the previous research, the application of the CNN model for fault detection and classification has achieved tremendous progress in many fields. However, for RV reducers of IR, the application of the CNN model remains rare. As is well-known, a convolutional neural network has unique advantages in processing massive data and can learn features from massive data. In this paper, a 2D-CNN based on the experimental vibration data after 2D-FFT is proposed for the fault diagnosis of an RV reducer for IR. Comparisons are conducted with other mainstream methods. The results indicate that the

proposed method shows better performance in the fault diagnosis of an RV reducer.

1 METHODS

1.1 Back Propagation (BP) Algorithm

Since Rumelhart et al. [18] and [19] reinvented and described the application of the BP algorithm in neural networks in 1986, this algorithm has become a standard learning method for neural network models. Due to its high efficiency, the BP algorithm is also selected to determine the network parameters of the CNN model in this paper.

With the BP algorithm, the matrix of universal parameters can be optimized by error-oriented motion, which is used for the classified task of the CNN. The residual error between the actual and expected output is sent back to the hidden layer and eventually back to the input layer. Based on the results, the weighted coefficient will be adjusted to minimize the error till the optimal coefficient is obtained. The loss function for a single sample is written as follows:

$$J(W, b; x, y) = \frac{1}{2} h_{W,b}(x) - y^2, \quad (1)$$

where $h_{W,b}(x)$ is the actual output, y is the desired output, W is the weight-matrix, b is a bias vector. Assume there are m samples in each data set, the overall cost function can be defined as:

$$J(W, b) = \left[\frac{1}{m} \sum_1^m \left(\frac{1}{2} h_{W,b}(x^{(i)}) - y^{(i)2} \right) \right] + \Gamma, \quad (2)$$

where Γ is the regularization term or weight decay term, which can reduce the magnitude of the weight to prevent overfitting, m is the number of samples. Γ can be determined by:

$$\begin{cases} \Gamma = \frac{\lambda}{2} \sum_{l=1}^{n_l-1} \sum_{i=1}^{s_j} \sum_{j=1}^{s_{j+1}} (W_{ij}^l)^2 \\ l = n_l - 1, n_l - 2, \dots, 2 \end{cases}, \quad (3)$$

where λ stands for the weight decay parameter; W_{ij}^l is the weight matrix between the i^{th} neuron in layer l and the j^{th} neuron in layer $l+1$, n_l is the output layer. The gradient descent method updates the parameters w and b :

$$W_{ij}^{l_{\text{new}}} = W_{ij}^{l_{\text{old}}} - \alpha \frac{\partial}{\partial W_{ij}^{l_{\text{old}}}} J(W, b), \quad (4)$$

$$b_i^{l_{\text{new}}} = b_i^{l_{\text{old}}} - \alpha \frac{\partial}{\partial b_i^{l_{\text{old}}}} J(W, b), \quad (5)$$

where α represents the learning rate; b_i^l is the bias vector of the i^{th} neuron in layer l . Using the backpropagation algorithm to calculate the partial derivative, the residual calculation formula for each output unit of the output layer is listed as follows:

$$\delta_i^n = \frac{\partial}{\partial z_i^n} \frac{1}{2} (y - h_{W,b}(x))^2 = -(y_i - a_i^n) \cdot f'(z_i^n), \quad (6)$$

where a_i^n is the i^{th} input of output layer; $f'(z_i^n)$ is the derivative of the activation function that takes the partial derivative of a_i^n ; δ_i^n is the residual. The residual calculation formula for each output unit of the other layer can be written as:

$$\delta_i^l = \left(\sum_{j=1}^{s_m} W_{ij}^l \delta_j^l \right) f'(z_i^l). \quad (7)$$

The calculation method of partial derivative can be presented by:

$$\frac{\partial}{\partial W_{ij}^l} J(W, b; x, y) = a_i^l \delta_i^n, \quad (8)$$

$$\frac{\partial}{\partial b_i^l} J(W, b; x, y) = \delta_i^n. \quad (9)$$

Thus, the residual calculation formula for the output layer and other layer are rewritten as:

$$\delta_i^n = -(y - a_i^n) \cdot f'(z_i^n), \quad (10)$$

$$\delta_i^l = \left((W^l)^T \delta^{l+1} \right) \cdot f'(z_i^l). \quad (11)$$

Calculating the final required partial derivative value:

$$\nabla_{W^l} J(W, b) = \delta^{l+1} (a^l)^T, \quad (12)$$

$$\nabla_{b^l} J(W, b) = \delta^{l+1}. \quad (13)$$

The calculation formula for the change of parameters W and b are as follows:

$$\Delta W^l = \Delta W^l + \nabla_{W^l} J(W, b), \quad (14)$$

$$\Delta b^l = \Delta b^l + \nabla_{b^l} J(W, b). \quad (15)$$

The final parameters update formulas are as follows:

$$W^l = W^l - \alpha \left[\left(\frac{1}{m} \Delta W^l \right) + \lambda W^l \right], \quad (16)$$

$$b^l = b^l - \alpha \left[\left(\frac{1}{m} \Delta b^l \right) \right]. \quad (17)$$

1.2 Typical Structure of CNN Method

CNN is a feedforward neural network with a deep structure that includes convolution computation [20]. The initial concept of CNN was proposed by LeCun et al. [21]. However, the actual application of CNN began in 2012, due to the appearance of AlexNet [22]. This method has been widely used in various fields, such as image recognition, speech separation and video classification.

For most CNN models in fault classification, feature extraction, feature selection, and fault classification are three major parts. As shown in Fig. 1, there are five layers in the structure of CNN: input

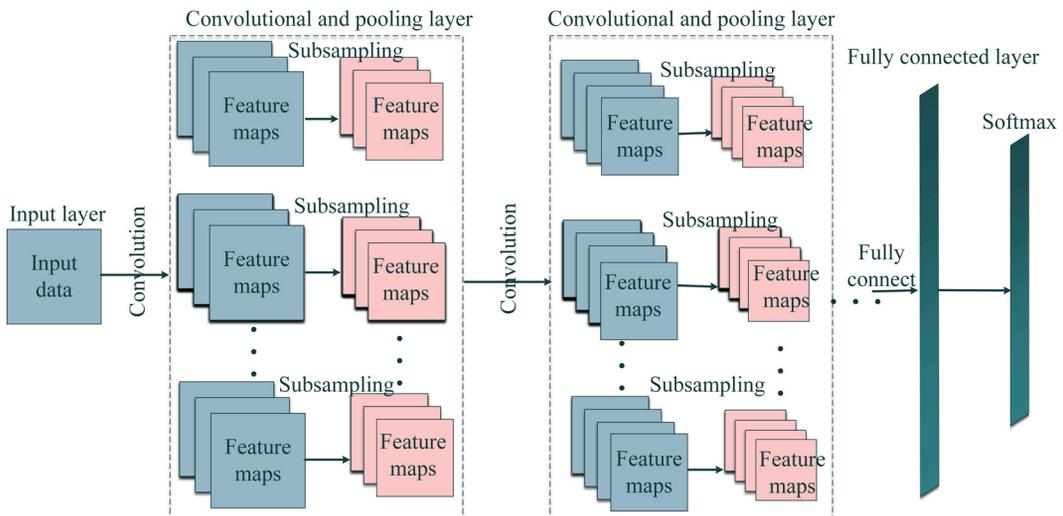


Fig. 1. Typical structure of CNN

layer, convolution layer, pooling layer, fully connected layer, and output layer.

(1) Convolution layer: as shown Fig. 1, there are many feature maps in the convolution layer; for each map, multiple neurons are distributed. The feature map of the former layer is connected to those neurons by the convolution kernel. As a feature matrix, the convolution kernel can extract the partial features of the input data. Due to the weight sharing of the convolution layer, the parameters in this layer can be minimized, which will improve the time of computation and avoid overfitting.

The output area of the convolution layer can be determined by:

$$Omap_{con} = \frac{Imap_{con} - kers + 2 \cdot padding}{stride} + 1. \quad (18)$$

In this equation, $Omap_{con}$ is the output area of the convolution layer, $Imap_{con}$ is the input area of convolution layer, $kers$ is the convolution kernel, $stride$ stands for the sliding length of convolution kernel, $padding$ represents the padding size.

The specific convolution formula involved in the convolution process is as follows:

$$y^{l(i,j)} = f\left(w_i^{l(j)} \cdot x^{l(r^j)} + b_i^l\right) = f\left(\sum_{j'=0}^{k-1} w_i^{l(j')} x^{l(j+j')} + b_i^l\right), \quad (19)$$

where $w_i^{l(j)}$ is the j th weight matrix of the i th convolution kernel in layer l ; $x^{l(r^j)}$ is the j th convoluted local region in layer l ; k is the width of convolution kernel; b_i^l is the bias terms in layer l and a vector; $f(\cdot)$ is the activation function.

There are five frequently-used activation functions: Sigmoid function, hyperbolic tangent (Tanh) function, rectified linear units (ReLU) function, exponential linear unit (ELU) function and MaxOut function. In this research, the ReLU function is selected as the activation function, because it can solve both explosion and disappearance of the gradient. Meanwhile, the convergence rate of ReLU function is much higher than the sigmoid and tanh functions [23]. The curve diagram of the ReLU function is shown in Fig. 2, which can be written as:

$$ReLU(x) = \max(0, x) = \frac{(x + |x|)}{2}, \quad (20)$$

$$ReLU'(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases}. \quad (21)$$

(2) Pooling layer: Similar to the convolutional layer, the pooling layer is composed of multiple

feature maps. Although the number of feature maps is the same as the former layer, the size of each map is smaller. By pooling, the feature vector will be reduced, which leads to the second-extraction of feature data. In this paper, mean pooling is chosen instead of maximum pooling. The number of output feature maps of the pooling layer in CNN can be determined by:

$$Omap_{pool} = \frac{Imap_{pool} - kers}{stride} + 1, \quad (22)$$

where $Omap_{pool}$ is the size of the output feature map of the pooling layer; $Imap_{pool}$ is the size of the feature map of the input pool layer.

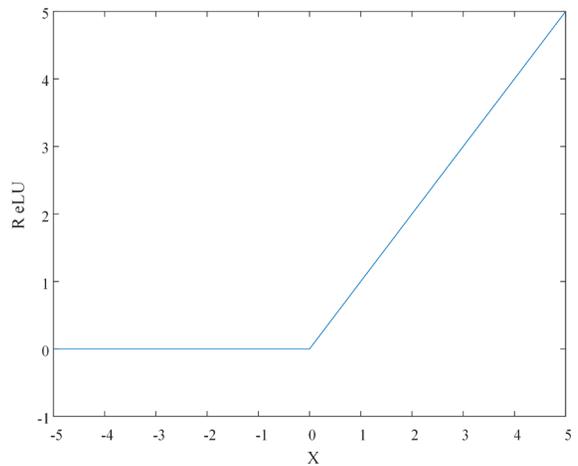


Fig. 2. ReLU graph

(3) Fully connected layer: in this layer, each neuron is fully connected with all neurons in the previous layer. The full connection layer can integrate the local information with class discrimination in the output of the pooling layer. The forward propagation in this layer can be represented as:

$$Z_j^{l+1} = \sum_{i=1}^n W_{ij}^l a_i^l + b_j^l. \quad (23)$$

In the formula, W_{ij}^l is the weight matrix between the i th neuron in layer l and the j th neuron in layer $l+1$; Z_j^{l+1} is the output value of the j th neuron in $l+1$ layer; b_j^l is the bias vector of all neurons in layer l to the j th neuron in layer $l+1$; a_i^l is the output feature maps of l layer.

(4) Output layer: The output layer uses softmax logistic regression to classify the input data. Mathematically speaking, the softmax operation can be described as:

$$p(z)_i = \frac{e^{z_i}}{\sum_{i=1}^{N_c} e^{z_i}}, \quad (24)$$

where $p(z)_i$ represents the probability that z belongs to category i , $i \in N_c$.

1.3 Advantages of CNN for Fault Classification

As mentioned before, CNN can automatically learn the features from the raw data without manual selection. Through training, CNN will optimize the model's parameters and achieve the recognition and classification of data. Because of this characteristic, the original vibration signal can be directly used as input for CNN, which will bring two benefits. 1) By using the raw data, pre-processing of the data can be avoided, which will save time and maximumly ensure the authenticity of data. 2) For the signal processing of a complex system, the selection of feature functions usually requires rich mechanical mathematical knowledge, which means it is difficult for a common technician to extract effective features. However, data features can be extracted directly by using CNN.

The weight sharing of the convolution layer in CNN is the same as sparse connection, which reduces the parameters in the network and lowers the computation time and difficulty. Meanwhile, due to the regularization effect, the stability and generalization ability of network structure are improved, and overfitting is also avoided [24]. As mentioned before, by pooling layers, the number of neurons in the model will be minimized, and the robustness of translation invariance in input space is improved [25]. Moreover, the structure of CNN is highly extensible, and a deeper layer can be designed in experiments. Therefore, CNN can be used for complex classification due to the ability of fault data extraction.

As a symbolic algorithm of deep learning, CNN can extract high-order features from input data. Specifically, the convolutional layer and pooling layer in the CNN can respond to the translation invariance of input features; in other words, they can identify similar features located in different spatial positions. Therefore, the CNN model has become one of the most popular models by researchers for fault identification and classification.

Since the data samples from the RV reducer are huge and complex, the CNN model is chosen for the fault identification and classification.

2 FAULT DIAGNOSIS METHOD BASED ON CNN

According to the different characteristics of the processing object and the dimension of the input data, CNN is divided into one dimension, two dimensions, and three dimensions, which play a role in different fields. The 2D-CNN weight filters take the dot product of the entire input as they move along the spatial dimension and sum them up. Unlike filters of 1D-CNN, which move in the time dimension, filters of 2D-CNN move in the space dimension, with two dimensions of height and width, which can more accurately and comprehensively extract the features of reconstructed data. In this paper, the 2D-CNN model is used to diagnose the fault data of the RV reducer used in an IR, and the vibration signal is directly used as the input to realize end-to-end fault diagnosis.

2.1 The Data Collection

In this paper, the RV reducer of an IR (model No. BRTIRUS1510A) is chosen as the research subject. There are six types faults: normal, broken teeth of planetary gear A, pitting of sun gear B, broken teeth of planetary gear B, crack of planetary gear B and broken teeth of planetary gear B respectively. For each fault, the rotation speed of the IR maintains 600 r/min without any load. Meanwhile, the motion track of IR will repeat 20 times, which is fixed (maximum motion range). For both motion and accelerate sensors, the sampling frequency is 100 kHz, the sampling time is 20 seconds, which will repeat 10 times. The experimental rig is shown in Fig. 3.



Fig. 3. Experimental device for data collection

The experimental design involves six different conditions corresponding to the faults, which includes one healthy mode (NR) and five fault modes (fault 1 to fault 5). Broken teeth are usually divided into fatigue fracture and overload fracture in the reducer, which are caused by fatigue stress and short-term overload or impact load. The pitting of teeth is generated by the teeth with small cracks mesh continuously, which lead to a small piece of metal slides off the tooth surface. The cracks of the teeth in the reducer are mainly triggered by machine overloaded, insufficient design strength, and defective material. Fig. 4 shows an example of a vibration signal acquired in each one of the fault types. The fault categories are listed in Table 1. Fig. 5 shows an example of each one of the five types of faults.

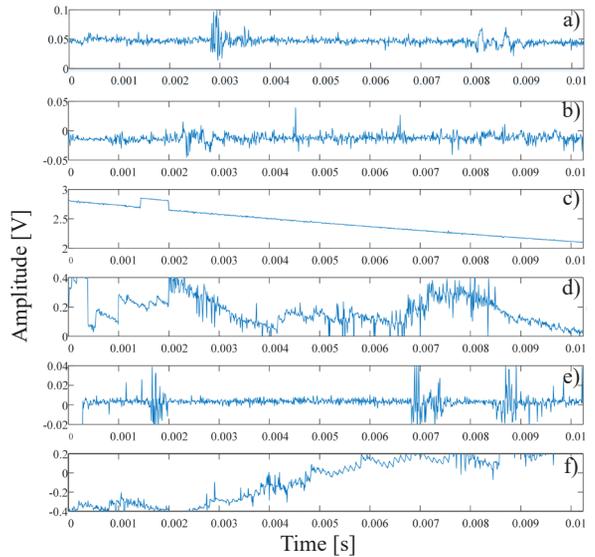


Fig. 4. Vibration signals; a) normal, b) fault 1, c) fault 2, d) fault 3, e) fault 4, and f) fault 5

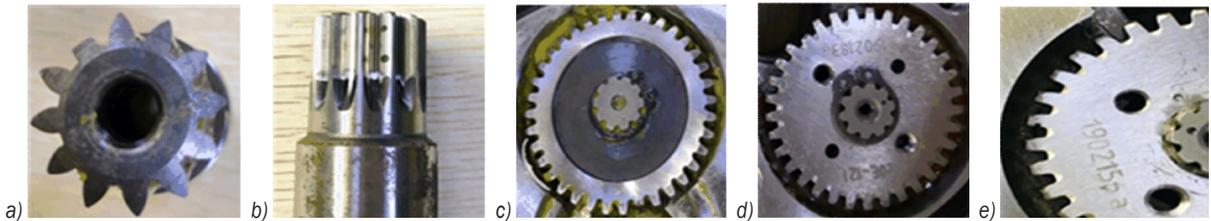


Fig. 5. Five fault conditions; a) broken tooth in sun gear B, b) pitting in sun gear B, c) broken tooth in planetary gear A, d) cracking in planetary gear B, and e) cracking in planetary gear B

Table 1. Fault categories

| | Serial number | Fault type | Fault location | Degree of fault |
|------------------|---------------|--------------|--|---------------------------|
| Normal | NR | None | None | \ |
| Planetary gear A | Fault 1 | Broken teeth | Fault of tooth breakage of the second shaft RV40E-121 planetary gear | Completely broken teeth |
| Sun gear B | Fault 2 | Pitting | Tooth pitting failure of the third shaft RV20E-121 sun gear | Moderate |
| | Fault 3 | Broken teeth | Tooth breakage fault of the third shaft RV20E-121 sun gear | Completely broken teeth |
| Planetary gear B | Fault 4 | Crack | Crack fault of the third shaft RV20E-121 planetary gear teeth | Width 0.5 mm Depth 0.5 mm |
| | Fault 5 | Broken teeth | Fault of tooth breakage of the third shaft RV20E-121 planetary gear | Completely broken teeth |

In this paper, for each fault, there will be 6×1000 samples. Among them, 4200 will be chosen for training, and the rest will be used for verification.

2.2 Experimental Environment

The hardware and software settings of this experimental environment are shown in Table 2.

Table 2. Development environment

| Hardware | Configuration or installation information |
|-------------------------|--|
| Processor | Intel(R) Core(TM) i7-6700 CPU @3.40GHZ 3.41GHZ |
| RAM | 8.0 GB |
| Operating system | Windows 10 Professional 64-bit |
| Compilation environment | MATLAB R2018a |

2.3 Data Processing

There are 6000 samples in this paper, and each sample has 1024 data points (i.e. 1024×1). The model used in this paper is 2D-CNN, so one-dimensional data is changed into two-dimensional data (i.e. $32 \times 32 \times 1$) by pre-processing. Then, after 2D-FFT processing, the time domain diagram is changed into the frequency domain diagram, as shown in Fig. 6. The 2D-FFT is usually calculated by the row and column decomposition algorithm; in other words, by using its separability, the two-dimensional FFT is divided into row-oriented FFT and column-oriented FFT and then calculated in sequence. Generally speaking, the row-column sequence of 1D-FFT algorithm has no great influence on the operation results of 2D-FFT, as long as the row-oriented FFT and column-oriented FFT can be calculated separately. The Fourier transform can transform the signal from time domain to frequency domain, and then study the spectrum structure and variation rule of the signal. Some signals in the time domain have more obvious features in the frequency domain, which enables the CNN model to extract data features more accurately.

2.4 Model Design

The CNN model used in this paper is shown in Fig. 7, which consists of an input layer, convolution layer C1, pooling layer P2, convolution layer C3, pooling layer P4, fully connected layer and output layer. The activation function of the convolution layer is

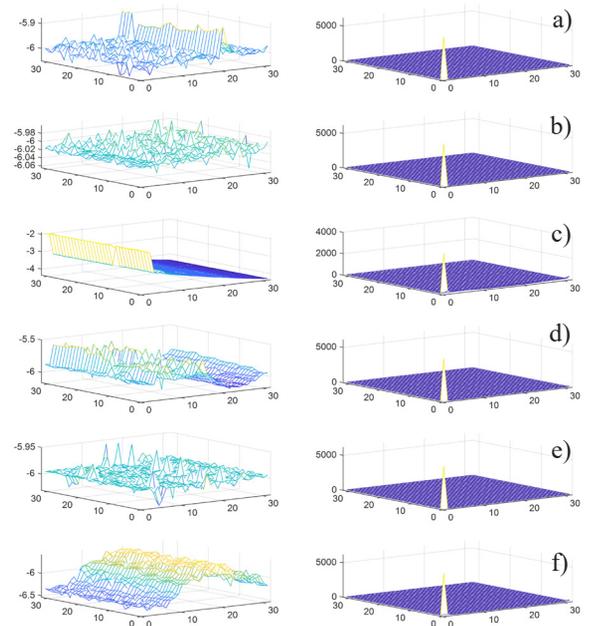


Fig. 6. 2D raw data and frequency domain diagram; a) normal, b) fault 1, c) fault 2, d) fault 3, e) fault 4, and f) fault 5

the ReLU function, the pooling layer adopts mean pooling, and the fully connected layer contains 250 processing units.

In order to automatically adapt to the different characteristics of vibration signals, the key parameters of CNN need to be optimized. The parameters of the network layer structure, convolution kernel size, number of iterations, learning rate, number of

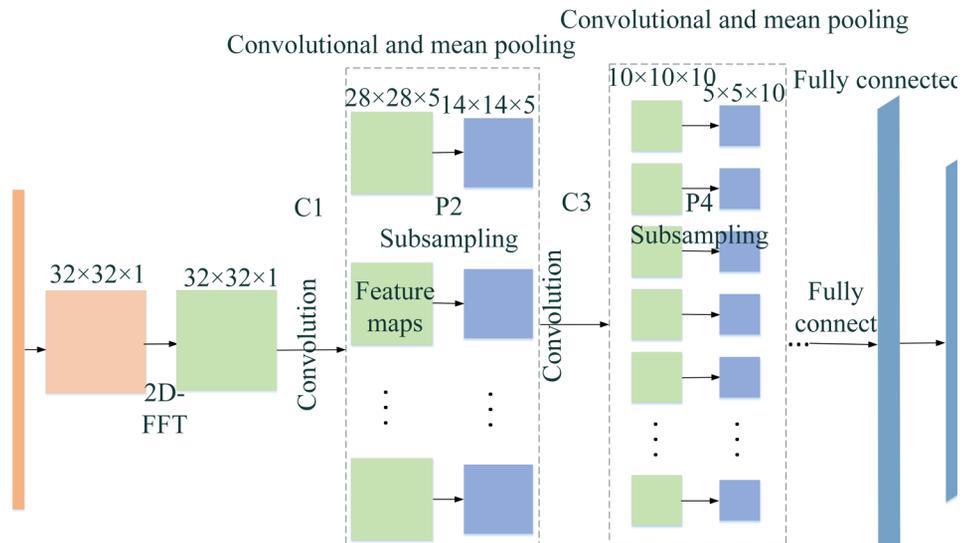


Fig. 7. CNN framework diagram used in this paper

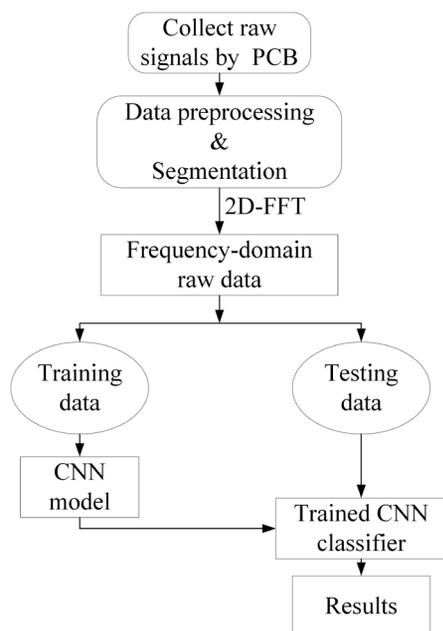
Table 3. Parameters of CNN model

| Layers | Types | Model parameters | Training parameters |
|--------|----------------------|---|--|
| 2 | Convolutional layer | Filter width = 5, Filter height = 5, Filter channel = 1, Filter number = 5, Bias = 5 | Minibatch size = 200 Learning rate = 0.01 Momentum = 0.1 Max epochs = 300 Training sample rate = 70 % Activation='ReLU' |
| 3 | Pooling layer | Sub-sampling rate = 2 | |
| 4 | Convolutional layer | Filter width = 5, Filter height = 5, Filter channel = 5, Filter number = 10, Bias = 10 | |
| 5 | Pooling layer | Sub-sampling rate = 2 | |
| 6 | Full-connected layer | 250 nodes | |
| 7 | Softmax classifier | 6 outputs | |

convolution kernels, activation function, and pooling method all need to be adjusted. The optimal structure parameters are shown in Table 3.

The diagnosis process proposed in this paper is shown in Fig. 8, and the processing steps are listed as follows:

- Step 1: Collecting vibration signals of different faults from the RV reducer through sensors installed on the machine;
- Step 2: Pre-process each kind of fault data and divide it into training samples and test samples;
- Step 3: Carrying out two-dimensional Fourier transform on the data;
- Step 4: Initializing the 2D-CNN model;
- Step 5: Fine-tuning the parameters;
- Step 6: Learning the features of the training samples by 2D-CNN model, and then using the trained model for fault diagnosis of the test samples;
- Step 7: Output the fault diagnosis results.

**Fig. 8.** Fault diagnosis flow chart of CNN

2.5 Experimental Results

The 2D-CNN model designed in this paper is used to classify six kinds of fault data of robots, and the classification results of test sets are shown in Table 4.

Table 4. Fault classification results of test sets

| Fault category | Sample size | Accuracy [%] |
|------------------|-------------|--------------|
| NR | 300 | 99 |
| 1 | 300 | 100 |
| 2 | 300 | 99.33 |
| 3 | 300 | 100 |
| 4 | 300 | 100 |
| 5 | 300 | 99.67 |
| Overall accuracy | | 99.67 |

Based on the above table, the proposed CNN model in this paper can achieve 99.67 % overall accuracy in recognition rate. For Fault 1, Fault 3, and Fault 4, the recognition rate can reach 100 %. The accuracy of normal data is 99 %, the accuracy of Fault 2 is 99.3 %, and the accuracy of Fault 5 is 99.67 %. It can be seen from Table 4 that the classification accuracy rate of Fault 2 is the lowest, and Fault 2 is pitting. As shown in Figure 4, the vibration signal of Fault 2 is quite different from the other four fault signals.

All the results indicate that the CNN model is an effective method for fault classification of a reducer on an industrial robot (IR).

3 SIMULATION COMPARISON

To further verify the efficiency of the proposed method, simulations of comparison are carried out in this paper. Four mainstream methods are chosen as comparative subjects, which are 1D-CNN, DBN, Stacked AutoEncoder (SAE), RNN and Neural Network (NN), respectively.

Similar to 2D-CNN, 1D-CNN can automatically extract feature data through weight sharing to avoid overfitting. Table 5 shows the parameters of the 1D-CNN model used in this paper.

Table 5. Parameters of 1D-CNN model

| Parameter | Value |
|---------------|-------------------------------------|
| Con1 | Omaps = 5; kers = 15; actv = 'sigm' |
| Pool1,2 | Scale = 5; pool = 'mean' |
| Con2 | Omaps = 5; kers = 13; actv = 'sigm' |
| Numepochs | 500 |
| Batchsize | 50 |
| Learning rate | 0.015 |
| Output layer | Softmax |

DBN is a probabilistic generative model, which is formed by stacking several restricted Boltzmann machines (RBM) [26]. Its training process is to use unsupervised greedy layer-wise pre-train to obtain weights. The bottom layer is mainly used to receive the input data and transform the input data into a hidden layer through RBM, that is, the input of higher layer RBM comes from the output of lower layer RBM. Stacked RBM networks form a DBN model to extract the features of fault data, and then the classifier is used for classification. Table 6 shows the parameters of the DBN model used in this paper.

Table 6. Parameters of DBN model

| Parameter | Value |
|---------------------|----------|
| Dbnsizes | 10×10 |
| Learning rate | 0.000003 |
| Numepochs | 500 |
| Batchsize | 100 |
| Momentum | 0.00001 |
| Activation function | tanh |
| Output layer | Softmax |

SAE is a deep neural network model composed of multi-layer AutoEncoder (AE), and AE is a three-layer unsupervised nonlinear neural network for feature extraction, including input layer, hidden layer, and output layer. The key idea of SAE is to let the network learn the features of input data in an unsupervised way [27]. Firstly, the greedy layer-wise training method is used to train the SAE network unsupervised, and the features of the fault data are obtained. Then, the trained SAE is connected with the softmax layer to classify the fault data of the test set. Table 7 shows the parameters of the SAE model used in this paper.

NN is composed of the input layer, hidden layer, and output layer, the latter of which have only one

layer, and the number of hidden layers can be arbitrary. The input layer is introduced from the instance feature vectors in the training set, and is transferred to the next layer according to the weights between the connection points. Table 8 shows the parameters of the NN model used in this paper.

Table 7. Parameters of SAE model

| Parameter | Value |
|-------------------------|---------|
| Hidden layer | 100 |
| Numepochs | 20 |
| Batchsize | 200 |
| InputZeroMaskedFraction | 0.5 |
| Learning rate | 0.0026 |
| Activation function | tanh |
| Output layer | Softmax |

Table 8. Parameters of NN model

| Parameter | Value |
|----------------------|----------|
| Hidden layer | 100 |
| Numepochs | 500 |
| Batchsize | 100 |
| Momentum | 0.5 |
| Learning rate | 0.21 |
| Scaling_learningrate | 1 |
| Activation function | sigm |
| Output layer | Logistic |

Six types of RV reducer data (one normal data and five fault) are collected in our experiments. The diagnosis results are shown in Table 9 and Fig. 9.

The fault diagnosis accuracy of five tests is shown in Fig. 9. It can be seen from the figure that the accuracy of fault diagnosis results of the 2D-CNN model is the highest, with the highest being 100 % and the lowest being 99.33 %, followed by 1D-CNN, NN and SAE. DBN performed the worst in these models.

Table 9 shows the average classification accuracy of the above five methods. In the classification performance, the fault classification method based on 2D-CNN has achieved better classification performance than other deep networks, and its average accuracy can reach 99.67 %. 1D-CNN's classification effect is ranked after 2D-CNN, and its average classification accuracy is 98.46 %. NN's classification effect is inferior to that of 1D-CNN, and its average classification accuracy is 83.47 %. The classification performance of SAE based on AEs is general, and the average accuracy of SAE is 83.05 %. The classification accuracy of DBN based on RBMs is low, and the average classification accuracy of DBN is only 58.08 %. To summarize, the 2D-CNN-based classification method has the highest average

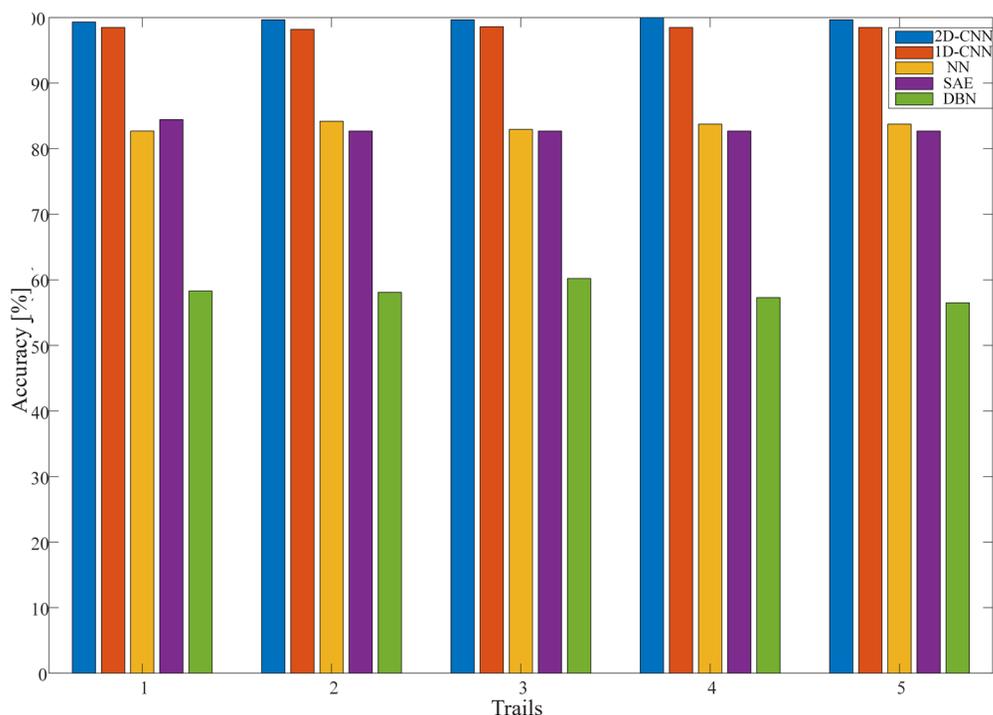


Fig. 9. Comparison of model accuracy

accuracy among the five methods, which is 1.21 % higher than 1D-CNN, 16.2 % higher than NN, 41.59 % higher than DBN and 16.62 % higher than SAE.

Table 9. Mean accuracy of model

| Model | Mean accuracy [%] |
|--------|-------------------|
| 2D-CNN | 99.67 |
| 1D-CNN | 98.46 |
| DBN | 58.08 |
| NN | 83.47 |
| SAE | 83.05 |

4 CONCLUSION

As a key component of industrial robots, the RV reducer often has various faults due to its bad operating environment. The failure of the reducer can bring huge economic losses to a company and even lead to serious casualties in some extreme cases. Therefore, the fault detection and diagnosis of the reducer is very important. In order to solve this problem, a 2D-CNN based on the experimental vibration data after 2D-FFT is proposed for the fault diagnosis of RV gear for IR. In the proposed algorithm, the time-frequency transformation of the original data is firstly performed by the 2D-FFT. Secondly, the convolutional layer and

pooling layer of 2D-CNN are used to extract the fault features contained in the data. Finally, the Softmax layer classifies the extracted features.

2D-CNN can automatically learn the features in the data, and it has three features: 1) The spatial connection of data is local. Each neuron does not need to perform convolution computation on all the data. The global information can be obtained by integrating different local neurons obtained by convolution computation at a higher level, which can reduce the number of connections; 2) The weight sharing between different neurons can reduce the parameters that need to be solved, and multiple feature maps can be obtained by using multiple filters to convolute the input data; 3) Subsampling can reduce the amount of data to be processed while retaining useful information. These features are conducive to improving the efficiency of network training.

The proposed method is applied to the rotation vector reducer for industrial robots. The data used in this paper include one normal data and five fault data, and the diagnosis accuracy reaches 99.67 %. However, the comparison models include 1D-CNN, DBN, NN, SAE. And their accuracy rates are 98.46 %, 58.08 %, 83.47 %, 83.05 %, respectively. Experimental results show that, compared with other peer models, the proposed model has good fault diagnosis accuracy.

5 ACKNOWLEDGMENTS

This research is supported by National Natural Science Foundation of China (51905058), Natural Science Foundation of Chongqing (cstc2020jcyj-msxmX0182, cstc2019jcyj-zdxmX0013), Key Project of Technology Innovation and Application of Chongqing (cstc2019jscx-fxydX0077), Research Start-Up Funds of Chongqing Technology and Business University (1856018), Key Research Platform Project of Chongqing Technology and Business University (ZDPTTD201918). The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

6 REFERENCES

- [1] Eguchi, K., Wu, Q., Liu, Y., Wu, C., Chen, T. (2018). An overview of current situations of robot industry development. *ITM Web of Conferences*, vol. 17, DOI:10.1051/itmconf/20181703019.
- [2] Zhang, R., Zhang, C., Zheng, W. (2018). The status and development of industrial robots. *Materials Science and Engineering*, vol. 423, art. ID 012051, DOI:10.1088/1757-899X/423/1/012051.
- [3] Chen, J. (2018). Industrial robot technology and its typical application analysis. *Journal of Electronic Research and Application*, vol. 2, no. 2, DOI:10.26689/jera.v2i3.379.
- [4] Li, T., Xu, H., Tian, M. (2020). A loaded analysis method for rv cycloidal-pin transmission based on the minimum energy principle. *Strojniški vestnik - Journal of Mechanical Engineering*, vol. 66, no. 11, p. 655-667, DOI:10.5545/sv-jme.2020.6868.
- [5] Girshick, R., Donahue, J., Darrell, T., Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, p. 580-587, DOI:10.1109/CVPR.2014.81.
- [6] Tompson, J., Goroshin, R., Jain, A. (2015). Efficient object localization using convolutional networks. *IEEE Conference on Computer Vision and Pattern Recognition*, p. 648-656, DOI:10.1109/CVPR.2015.7298664.
- [7] Pan, H., He, X., Tang S., Meng, F. (2018). An improved bearing fault diagnosis method using one-dimensional CNN and LSTM. *Strojniški vestnik - Journal of Mechanical Engineering*, vol. 64, no. 7-8, p. 443-452, DOI:10.5545/sv-jme.2018.5249.
- [8] Janssens, O., Slavković, V., Vervisch, B., Stockman, K., Locuffier, M., Verstockt, S., Van de Walle, R., Van Hoesche, S. (2016). Convolutional neural network based fault detection for rotating machinery. *Journal of Sound and Vibration*, vol. 377, p. 331-345, DOI:10.1016/j.jsv.2016.05.027.
- [9] Jing, L., Zhao, M., Li, P., Xu, X. (2017). A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox. *Measurement*, vol. 111, p. 1-10, DOI:10.1016/j.measurement.2017.07.017.
- [10] Azamfar, M., Singh, J., Bravo-Imaz, I., Lee, J. (2020). Multisensor data fusion for gearbox fault diagnosis using 2-D convolutional neural network and motor current signature analysis. *Mechanical Systems and Signal Processing*, vol. 144, art. ID 106861, DOI:10.1016/j.ymssp.2020.106861.
- [11] Qiu, G., Gu, Y., Cai, Q. (2019). A deep convolutional neural networks model for intelligent fault diagnosis of a gearbox under different operational conditions. *Measurement*, vol. 145, p. 94-107, DOI:10.1016/j.measurement.2019.05.057.
- [12] Chen, Z., Mauricio, A., Li, W., Gryllias, K. (2020). A deep learning method for bearing fault diagnosis based on cyclic spectral coherence and convolutional neural networks. *Mechanical Systems and Signal Processing*, vol. 140, p. DOI:10.1016/j.ymssp.2020.106683.
- [13] Wang, S., Chen, H. (2019). A novel deep learning method for the classification of power quality disturbances using deep convolutional neural network. *Applied Energy*, vol. 235, p. 1126-1140, DOI:10.1016/j.apenergy.2018.09.160.
- [14] Xuan, W., You, G. (2020). Detection and diagnosis of pancreatic tumor using deep learning-based hierarchical convolutional neural network on the internet of medical things platform. *Future Generation Computer Systems*, vol. 111, p. 132-142, DOI:10.1016/j.future.2020.04.037.
- [15] Meng, M., Chua, Y. J., Wouterson, E. and Kong, C. P. K. (2017). Ultrasonic signal classification and imaging system for composite materials via deep convolutional neural networks. *Neurocomputing*, vol. 257, p. 128-135, DOI:10.1016/j.neucom.2016.11.066.
- [16] Lu, Y., Yi, S., Zeng, N., Liu, Y., Zhang, Y. (2017). Identification of rice diseases using deep convolutional neural networks. *Neurocomputing*, vol. 267, p. 378-384, DOI:10.1016/j.neucom.2017.06.023.
- [17] Qian, H. M., Li, Y.F., Huang, H.Z. (2020). Time-variant reliability analysis for industrial robot RV reducer under multiple failure modes using Kriging model. *Reliability Engineering & System Safety*, vol. 199, art. ID 106936, DOI:10.1016/j.res.2020.106936.
- [18] Rumelhart, D.E., Hinton, G.E., Williams, R.J. (1987). Learning Internal Representations by Error Propagation. In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations*, vol. 1, p. 318-362.
- [19] Rumelhart, D.E., Hinton, G.E., Williams, R.J. (1986). Learning representations by back-propagating errors. *Nature*, vol. 323, p. 533-536, DOI:10.1038/323533a0.
- [20] Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J., Chen, T. (2018). Recent advances in convolutional neural networks. *Pattern Recognition*, vol. 77, p. 354-377, DOI:10.1016/j.patcog.2017.10.013.
- [21] LeCun, Y., Bottou, L., Bengio, Y., Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, vol. 86, no. 11, p. 2278-2324, DOI:10.1109/5.726791.
- [22] Krizhevsky, A., Sutskever, I., Hinton, G.E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, vol. 60, p. 84-90, DOI:10.1145/3065386.
- [23] Xu, B., Wang, N.Y., Chen, T.Q., Li, M. (2015). Empirical evaluation of rectified activations in convolution network. *Computer Science*.

- [24] O'Shea, K., Nash, R. (2015). An introduction to convolutional neural networks. *Computer Science*.
- [25] Huang, J.T., Li, J.Y., Gong, Y.F. (2015). An analysis of convolutional neural networks for speech recognition. *IEEE International Conference on Acoustics, Speech and Signal Processing*, p. 4989-4993, **DOI:10.1109/ICASSP.2015.7178920**.
- [26] Xu, F., Fang, Z., Tang, R., Li, X., Tsui, K.L. (2020). An unsupervised and enhanced deep belief network for bearing performance degradation assessment. *Measurement*, vol. 162, art. ID 107902, **DOI:10.1016/j.measurement.2020.107902**.
- [27] Vincent, P., Larochele, H., Lajoie, I., Bengio, Y., Manzagol, P.-A. (2010). Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion. *Journal of Machine Learning Research*, vol. 11, p. 3371-3408.