DCNN With Explicable Training Guide and Its Application to Fault Diagnosis of the Planetary Gearboxes

PENG LUO, NIAOQING HU, GUOJI SHEN, LUN ZHANG, AND ZHE CHENG
Laboratory of Science and Technology on Integrated Logistics Support, National University of Defense Technology, Changsha 410073, China
College of Intelligence Science and Technology, National University of Defense Technology, Changsha 410073, China
Corresponding author: Niaoqing Hu (hnq@nudt.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFB1702401, and in part by National Natural Science Foundation of China under Grant 51975576 and Grant 51475463.

ABSTRACT The diagnosis performance of Deep Convolutional Neural Network (DCNN) method is closely related to the generalization ability of the training model. An empirical training strategy is to randomly disperse the training samples and train the model with mini-batch training samples. But there are still two problems in the empirical method that need to be solved urgently. Firstly, what is the theoretical basis for random discretization of samples? Secondly, how to scientifically quantify batch division? Aiming at these two problems, the theoretical basis of sample random discretization has been deduced and proved, furthermore, a scientific quantitative batch division method is proposed based on the proved thesis. The fault diagnosis results of the planetary gearbox show that: (1) The model obtained by the training guide proposed in this paper has stronger generalization ability; (2) The DCNN with the training guide can accurately and effectively diagnose the faults of planetary gearbox and obtain ideal diagnosis results.

INDEX TERMS DCNN, generalization ability, explicable training method, fault diagnosis.

I. INTRODUCTION
It is of great importance to monitor the health state of mechanical equipment. Generally, fault diagnosis methods include data acquisition, signal processing, feature extraction and pattern recognition [1]–[4]. Aiming at the four parts aforementioned, many relevant methods, such as EMD, EEMD, VMD and SVM, BPNN et al., have been proposed [5]–[7]. However, these traditional fault diagnosis methods have high requirements for the acquired vibration signals and the signal processing methods, and they also rely on a lot of expert knowledge as well [8], [9].

In recent years, Deep Learning method has been widely used in Computer Vision [10], [11] and Speech Recognition [12]. Due to its powerful data processing and pattern recognition capabilities, some Deep Learning algorithms have been used for the mechanical equipment health monitoring too [13]–[15].

As a typical representative of Deep Learning, Deep Convolutional Neural Network (DCNN) has a strong ability to extract distributed features from the original signal and identify the fault patterns adaptively, which can reduce dependence on expert knowledge [16]–[20]. For example, Heng Li et al. [21] combined the short time Fourier with DCNN to make fault diagnosis of rolling bearings, and the proposed method could avoid the process of feature extraction and classifier design. Zhou et al. [22] carried out fault diagnosis for rotating machinery based on 1D depth convolutional neural network, and compared with the traditional fault methods, their method achieved better performance. Through literature research, it could be found that the diagnosis performance of DCNN is closely related to the generalization ability of the training model. How to fully train the DCNN model with the existed samples and obtain an ideal diagnosis result? Around this issue, many scholars have carried out relevant research work in the fault diagnosis field.

In order to improve the diagnosis performance of DCNN, there are two main strategies at present. One is to perform signal processing first, and then input the processed signal into DCNN. The other is still to input the original signal into the network, but the strategy of adjusting the input data is changed to mini-batch.
DCNN is generally used to process two-dimensional data, such as images. Some scholars transformed one-dimensional vibration data into two-dimensional data, including images, or stacked one-dimensional decomposed vibration data to form a two-dimensional matrix. For example, Wen et al. [23] converted the time-domain vibration signal into images, and input them into the DCNN model for training and diagnosis. Zhao et al. [24] transformed the original time domain signal into the frequency domain signal, and fault diagnosis was carried out based on the spectrum data characteristic diagram. Some other scholars, such as Hu et al. [25] decomposed the signal using EMD method, screened samples in terms of the kurtosis of each decomposition component, and then, input the stacked components into the DCNN model for fault diagnosis. Islam and Kim [26] and Cao et al. [27] combined the DCNN model with the wavelet decomposition method, and applied the method in the bearing fault diagnosis. Han et al. [28] proposed a fault diagnosis method based on the enhanced DCNN, and applied it to the fault diagnosis of planetary gearboxes, the core idea of which was to convert one-dimensional signal into two-dimensional signal and increase the receptive field, so as to extract fault feature information better.

By transforming the dimension or the domain, the above-mentioned methods aim to improve the generalization performance of the training model. However, conducting signal processing will lose the meaning of using DCNN, because it still requires strong professional domain knowledge and expert experience, and the mini-batch strategy has more application and promotion value. Based on this, other researchers devote themselves to taking the original data characteristic diagram. Some other scholars, such as Islam and Kim [26] and Cao et al. [27] combined the DCNN model with the wavelet decomposition method, and applied the method in the bearing fault diagnosis. Han et al. [28] proposed a fault diagnosis method based on the enhanced DCNN, and applied it to the fault diagnosis of planetary gearboxes, the core idea of which was to convert one-dimensional signal into two-dimensional signal and increase the receptive field, so as to extract fault feature information better.

By transforming the dimension or the domain, the above-mentioned methods aim to improve the generalization performance of the training model. However, conducting signal processing will lose the meaning of using DCNN, because it still requires strong professional domain knowledge and expert experience, and the mini-batch strategy has more application and promotion value. Based on this, other researchers devote themselves to taking the original data characteristic diagram. Some other scholars, such as Islam and Kim [26] and Cao et al. [27] combined the DCNN model with the wavelet decomposition method, and applied the method in the bearing fault diagnosis. Han et al. [28] proposed a fault diagnosis method based on the enhanced DCNN, and applied it to the fault diagnosis of planetary gearboxes, the core idea of which was to convert one-dimensional signal into two-dimensional signal and increase the receptive field, so as to extract fault feature information better.

By transforming the dimension or the domain, the above-mentioned methods aim to improve the generalization performance of the training model. However, conducting signal processing will lose the meaning of using DCNN, because it still requires strong professional domain knowledge and expert experience, and the mini-batch strategy has more application and promotion value. Based on this, other researchers devote themselves to taking the original data characteristic diagram. Some other scholars, such as Islam and Kim [26] and Cao et al. [27] combined the DCNN model with the wavelet decomposition method, and applied the method in the bearing fault diagnosis. Han et al. [28] proposed a fault diagnosis method based on the enhanced DCNN, and applied it to the fault diagnosis of planetary gearboxes, the core idea of which was to convert one-dimensional signal into two-dimensional signal and increase the receptive field, so as to extract fault feature information better.

By transforming the dimension or the domain, the above-mentioned methods aim to improve the generalization performance of the training model. However, conducting signal processing will lose the meaning of using DCNN, because it still requires strong professional domain knowledge and expert experience, and the mini-batch strategy has more application and promotion value. Based on this, other researchers devote themselves to taking the original data characteristic diagram. Some other scholars, such as Islam and Kim [26] and Cao et al. [27] combined the DCNN model with the wavelet decomposition method, and applied the method in the bearing fault diagnosis. Han et al. [28] proposed a fault diagnosis method based on the enhanced DCNN, and applied it to the fault diagnosis of planetary gearboxes, the core idea of which was to convert one-dimensional signal into two-dimensional signal and increase the receptive field, so as to extract fault feature information better.

By transforming the dimension or the domain, the above-mentioned methods aim to improve the generalization performance of the training model. However, conducting signal processing will lose the meaning of using DCNN, because it still requires strong professional domain knowledge and expert experience, and the mini-batch strategy has more application and promotion value. Based on this, other researchers devote themselves to taking the original data characteristic diagram. Some other scholars, such as Islam and Kim [26] and Cao et al. [27] combined the DCNN model with the wavelet decomposition method, and applied the method in the bearing fault diagnosis. Han et al. [28] proposed a fault diagnosis method based on the enhanced DCNN, and applied it to the fault diagnosis of planetary gearboxes, the core idea of which was to convert one-dimensional signal into two-dimensional signal and increase the receptive field, so as to extract fault feature information better.
where, \( s \) is the down sampling size, for example, when the mean sampling method is used, \( s \) is 2.

**C. THE FULL CONNECTION LAYER (FC-LAYER)**

In the FC-layer, each value of the input vector is connected to each value of the output vector. If the length of the input and output vectors are \( M \) and \( N \), respectively, the output vector of the \( l \)-th layer can be calculated as follows

\[
x_j^l = f \left( \sum_{i=1}^{M} x_i^{l-1} \cdot w_{ij}^l + b_j^l \right), \quad j = 1, \ldots, N
\]

(5)

where \( w_{ij} \) denotes the weight of the \( j \)-th output value connected to the \( i \)-th input value. The computation for the number of all the parameters of a fully connected layer is described as follows

\[
P = M \times N + 1
\]

(6)

**D. THE OUTPUT LAYER: SOFTMAX CLASSIFIER**

Softmax classifier can be described as

\[
p_{W_0}(y^l = i \mid x^l) = \frac{1}{1 + \exp(-W^l y^l x^l)}
\]

(7)

where, \( p_{W_0}(\cdot)(i \in \{0, 1\}) \) is a sigmoid function with parameters \( W^l \), and the \( x^l \) is the feature learned by the DCNN. The parameter \( W^l \) is learned by a training set. Eq. (7) produces a label between 0 and 1. The predicted class \( \hat{i} \) and prediction score \( \hat{s}(\hat{i}) \) can be described as

\[
\hat{i} = \left\{ i \mid \max p_{W_0}(y^l = i) \right\}, \quad i \in \{0, 1\}
\]

(8)

and

\[
\hat{s}(\hat{i}) = p_{W_0}(y^l = \hat{i}), \quad \hat{i} \in \{0, 1\}
\]

(9)

**III. TRAINING GUIDE METHOD OF DCNN**

**A. BASIS FOR RANDOM DISCRETIZATION OF SAMPLES**

Assuming that \( m \) samples constitute the sample set \( \{x^{(1)}, x^{(1)}, \ldots, x^{(m)}, x^{(m)}\} \), and they are \( n \) categories, respectively, where, \( x^{(i)} \) refers to the input signal vector and \( y^{(i)} \) refers to the target value, namely, the fault-pattern index.

The cost function of the DCNN model can be represented as

\[
R(W, b) = \frac{1}{m} \sum_{i=1}^{m} \frac{1}{2} \left\| h_{W,b}(x^{(i)}) - y^{(i)} \right\|^2
\]

(10)

where, \( W \) is the weight value of each unit and \( b \) is the bias term, \( h_{W,b}(x^{(i)}) \) is the output of the last neural network layer, namely, the fault-pattern index of the sample \( x^{(i)} \). The target of the training network is to find the minimum value of the function \( R(W, b) \) by adjusting the \( W \) and \( b \).

Based on the diagnosis principle of the DCNN, it can obtain the conclusion that the performance of DCNN mainly depends on the parameters of the trained model. This process is mainly achieved through multiple batches of training samples. In other words, when training samples are given, scientific training strategy determines the performance of the DCNN to some extent.

So, the trained DCNN model can be expressed as

\[
\left\{ h \left( \overline{x}_k, \overline{\theta}_k \right), \quad k = 1, 2, \ldots, K \right\}
\]

(11)

where, \( h(\cdot) \) is the training model, \( \overline{x}_k \) is the batch sample set, \( X \) is the total sample set and \( \overline{x}_k \in X, n = K \times n, \) and \( n \) is the batch sample capacity. Each batch samples can train a DCNN, it can be defined as Batch Deep Convolutional Neural Network (BDCNN), \( K \) is the number of batches or the number of BDCNN obtained, \( \overline{\theta}_k \) is the model parameters set obtained from the batch of training.

Based on Eq. (10), the Confidence Function (CF) of DCNN can be defined as

\[
CF(X, Y) = \text{avg} I \left( h \left( \overline{x}_k, \overline{\theta}_k \right) = Y \right) - \max_{J \neq Y} \text{avg} I \left( h \left( \overline{x}_k, \overline{\theta}_k \right) = J \right)
\]

(12)

where, \( X \) and \( Y \) represent the sample set and label set, respectively, \( I(\cdot) \) is the indicator function, \( \text{avg}_k (\cdot) \) is the mean value function, \( Y \) is the correct classified label set, and \( J \) is the misclassified set.

The \( CF(X, Y) \) measures the degree that the number of BDCNN correctly classified exceeds the number of any other
misclassified BDCNN in the process of model training. The larger the value of \( CF(X, Y) \), the stronger the diagnosis ability of the trained DCNN model. In order to measure the performance of DCNN, the Generalization Error (GE) has been introduced and furthermore, the GE of the DCNN model can be defined as

\[
GE = P_{X, Y} (CF (X, Y) < 0)
\]

(13)

where the subscripts \( X, Y \) indicate that the probability \( P_{X, Y} (\cdot) \) is over the \( X, Y \) space.

Next, the factors that are connected to the GE will be found, and the conclusion can be summarized as follows:

**Conclusion:** the Generalization Error of the DCNN model is positively correlated with the correlation between BDCNN and negatively correlated with the classification ability of BDCNN.

Before proving the conclusion, some definitions or properties need to be introduced.

1. The BDCNN correlation is the correlation between the models obtained by batch training, and the detailed mathematical representation will be given in the following parts of the paper.

2. The diagnosis ability of BDCNN is the recognition ability of the model obtained by batch training, and the detailed mathematical forms will be given in the process of subsequent proof in this paper as well.

3. Almost everywhere convergent [30]: suppose that \( \xi \) and \( \{\xi_n\} \) are a sequence of random variables defined on a probabilistic space \( (\Omega, F, P) \), if there is a zero test set \( \Omega_0, \Omega_0 \in F \) and \( P (\Omega_0) = 0 \), \( \forall \omega \in \Omega / \Omega_0 \), if there has \( \xi_n (\omega) \rightarrow \xi (\omega) \), it can be summarized as \( \xi_n \) converge to \( \xi \) almost everywhere, recorded as \( \xi_n \xrightarrow{a.e.} \xi \).

4. Borel strong law of large numbers [31]: suppose that \( \{\xi_n\} \) is an independent sequence of random variables with the same distribution in probabilistic space \( (\Omega, F, P) \), if \( P (\xi_n = 1) = p \) and \( P (\xi_n = 0) = 1 - p \), \( 0 < p < 1 \), \( S_n = \sum_{k=1}^{n} \xi_k \), there is

\[
S_n \xrightarrow{a.s.} p.
\]

5. Chebyshev inequality [32]: for any random variable \( X \), if there are expectations \( E X \) and variances \( D X \), when \( \forall \xi > 0 \), there has \( P \{ |X - E X| \geq \xi \} \leq \frac{D X}{\xi^2} \).

When the training sample sets and number of training batches are given, the batch sample sets \( \{\bar{X}_k\} \) corresponding to \( h (\bar{X}_k, \theta_k) = J \) are limited too, define that

\[
\{\bar{X}_k \mid h (\bar{X}_k, \theta_k) = J\} = S_r.
\]

Suppose that \( K_r \) denote the number of \( S_r \), in other word, \( K_r \) represent the number of misclassified BDCNN, there will be

\[
\frac{1}{K} \sum_{k=1}^{K} I (h (\bar{X}_k, \theta_k) = J) = \frac{1}{K_n} \sum_{k=1}^{K_r} K_r I (\bar{X}_k \in S_r) = \frac{K_r}{K} \sum_{r=1}^{K_r} I (\bar{X}_k \in S_r)
\]

(14)

When there exist \( K \rightarrow \infty \), according to “borel strong law of large numbers” theorem, the formula (15) holds

\[
\frac{K_r}{K} = \frac{1}{K} \sum_{k=1}^{K} I (\bar{X}_k \in S_r) \stackrel{a.s.}{\rightarrow} P_{\theta} (\bar{X}_k \in S_r)
\]

(15)

where the subscripts \( \theta \) indicate that the probability is over the model parameters set \( \theta \) space.

Therefore, based on the “almost everywhere convergent” theorem, for any \( J \), there is a zero test set \( C \) in the value space of \( \{\bar{X}_k\} \), the following expression holds

\[
\frac{1}{K} \sum_{k=1}^{K} I (h (\bar{X}_k, \theta_k) = J) \xrightarrow{K \rightarrow \infty} P_{\theta} (h (\bar{X}_k, \theta_k) = J)
\]

(16)

so, Eq. (17) holds.

\[
\frac{1}{K} \sum_{k=1}^{K} I (h (\bar{X}_k, \theta_k) = J) \xrightarrow{K \rightarrow \infty} P_{\theta} (h (\bar{X}_k, \theta_k) = J)
\]

(17)

so, Eq. (18) holds.

\[
\lim_{K \rightarrow \infty} GE = P_{X, Y} (P_{\theta} (h (\bar{X}_k, \theta_k) = Y) \quad - \max_{J \neq Y} P_{\theta} (h (\bar{X}_k, \theta_k) = J) < 0)
\]

(18)

Note that

\[
J_{\text{max}} = \arg \max_{J \neq Y} P_{\theta} (h (\bar{X}_k, \theta_k) = J)
\]

(19)

Eq. (12) can be re-written as follows

\[
CF (X, Y) = P_{\theta} (h (\bar{X}_k, \theta_k) = Y) - P_{\theta} (h (\bar{X}_k, \theta_k) = J_{\text{max}})
\]

\[
= E_{\theta} (I (h (\bar{X}_k, \theta_k) = Y)) - I (h (\bar{X}_k, \theta_k) = J_{\text{max}})
\]

(20)

Note that

\[
\text{rmg} (\theta, X, Y) = I (h (\bar{X}_k, \theta_k) = Y)
\]

\[
- I (h (\bar{X}_k, \theta_k) = J_{\text{max}})
\]

(21)

Further, there is

\[
CF (X, Y) = E_{\theta} \text{rmg} (\theta, X, Y)
\]

(22)

It has been proved that Eq. (18) holds, the upper bound of generalization error GE can be obtained by analyzing \( P_{X, Y} (CF (X, Y) < 0) \). In order to prove that the diagnosis result of DCNN model is reliable, there is \( E_{X, Y} CF (X, Y) > 0 \), where \( E_{X, Y} CF (X, Y) \) represents the degree of expectation of classification results of each sample by DCNN, and \( E_{X, Y} CF (X, Y) > 0 \) indicates that the
classification result is reliable. According to the “Chebyshev inequality”, there is

\[ P_{X,Y}(CF(X, Y) < 0) = P_{X,Y}(CF(X, Y) - E_{X,Y}CF(X, Y) < E_{X,Y}CF(X, Y)) < P_{X,Y}((CF(X, Y) - E_{X,Y}CF(X, Y)) > E_{X,Y}CF(X, Y)) \]

\[ \leq \frac{\text{var}_{X,Y}(CF(X, Y))}{E_{X,Y}CF(X, Y)^2} \] \hspace{1cm} (23)

so,

\[ GE \leq \frac{\text{var}_{X,Y}(CF(X, Y))}{E_{X,Y}CF(X, Y)^2} \] \hspace{1cm} (24)

The classification ability of BDCNN is defined as \( s \), and the average correlation between BDCNN is \( \overline{\rho} \), and the expressions are as follows

\[ s = E_{X,Y}CF(X, Y) \] \hspace{1cm} (25)

\[ \overline{\rho} = \frac{E_{\theta,\theta^*}(\rho(\theta, \theta^*)sd(\theta)sd(\theta^*))}{E_{\theta,\theta^*}(sd(\theta)sd(\theta^*))} \] \hspace{1cm} (26)

where, \( \rho(\theta, \theta^*) \) represents the correlation between \( \text{rmg}(\theta, X, Y) \) and \( \text{rmg}(\theta^*, X, Y) \), and \( sd(\theta) \) represents the standard deviation of \( \text{rmg}(\theta, X, Y) \).

The upper bound of \( GE \) represented by \( s \) and \( \overline{\rho} \) can be obtained by the following proof process.

For independent identical distribution variables \( \theta \) and \( \theta^* \), if \( (E_{\theta,f}(\theta))^2 = E_{\theta,\theta^*}(f(\theta)f(\theta^*)) \) exists, then \( \text{var}_{X,Y}(CF(X, Y)) \) can be calculated as follows

\[ \text{var}_{X,Y}(CF(X, Y)) = E_{X,Y}(CF(X, Y))^2 - (E_{X,Y}CF(X, Y))^2 = E_{X,Y}E_{\theta,\theta^*}\text{rmg}(\theta, X, Y)\text{rmg}(\theta^*, X, Y) - E_{\theta,\theta^*}(E_{X,Y}\text{rmg}(\theta, X, Y)E_{X,Y}\text{rmg}(\theta^*, X, Y)) = E_{\theta,\theta^*}(\rho(\theta, \theta^*)sd(\theta)sd(\theta^*)) \] \hspace{1cm} (27)

The result can be obtained as follows

\[ \text{var}_{X,Y}(CF(X, Y)) = \overline{\rho}(E_{\theta,\theta}(\theta)sd(\theta))^2 \leq \overline{\rho}E_{\theta,\theta}(\theta)sd(\theta)^2 = \overline{\rho}E_{\theta}(E_{X,Y}(\text{rmg}(\theta, X, Y))^2 - (E_{X,Y}\text{rmg}(\theta, X, Y))^2) \leq \overline{\rho}(E_{\theta}E_{X,Y}(\text{rmg}(\theta, X, Y))^2 - (E_{\theta}E_{X,Y}\text{rmg}(\theta, X, Y))^2) \leq \overline{\rho}(1 - s^2) \] \hspace{1cm} (28)

So, the following result can be obtained

\[ GE \leq \frac{\overline{\rho}(1 - s^2)}{s^2} \] \hspace{1cm} (29)

Known from the foregoing definition, \( \overline{\rho} \) represents the average correlation between BDCNN, and \( s \) represents classification intensity of BDCNN. Here, the key factors that affect the diagnosis ability of DCNN have been found.

Therefore, the Generalization Error of DCNN model is positively correlated with the correlation between BDCNN and negatively correlated with the classification ability of BDCNN.

Based on this conclusion, the generalization ability of the DCNN model can be enhanced and the confidence in diagnosis results can be improved by reducing the correlation between BDCNN and improving the classification ability of BDCNN.

From Eq. (26), it can be seen that the correlation between BDCNN is closely related with the correlation between the training samples. The sample correlation in the same health state is certainly higher than that in different health state. So, if the correlations between samples have been reduced by sample random discretization, and then, the correlation between BDCNN could be reduced. This is the reason that why the generalization ability of the DCNN can be enhanced by dispersing the training samples.

B. OPTIMIZATION OF THE BATCH DIVISION

Based on the conclusion obtained, when we do the training sample batch division, we should consider how to improve the classification ability of BDCNN scientifically.

After the experimental research shown in Fig. 2, it can be found that the \( GE \) of diagnosis model changes along with the different iterations and the batch sample capacity. Proper parameters can effectively improve the classification ability of BDCNN. Therefore, it is very important to find the balance between the iterations and the batch sample capacity.

Suppose that \( X \) is the sample set, and \( \text{size}(X) = K \times n \), \( K \) is the number of batches, and \( n \) is the batch sample capacity. The experience strategy used to adopt multi-batch and mini-capacity strategy, so that it could train the diagnosis model with much more times and obtain better trained model. The experimental results are the same. Fig. 3 shows that the \( GE \) of diagnosis model is getting smaller along with bigger number of iterations, and the classification ability of the model can be improved by increasing training times or iterations.

However, just like the experimental result shown in Fig. 4, when the batch sample capacity is too small, the sample information learned by the training model will also be meager. It is
P. Luo et al.: DCNN With Explicable Training Guide and Its Application to Fault Diagnosis

FIGURE 3. Generalization Error under different iterations and DCNN structures.

FIGURE 4. Generalization Error under different batch sample capacity.

FIGURE 5. Generalization Error under different iterations and batch sample capacity.

not conducive to improving the generalization ability of the diagnosis model.

The method selected in this paper is to traverse and optimize the batch sample capacity, just like Fig. 5. On the premise of ensuring the prediction accuracy, the goal is to search for the best sample batch capacity with less iteration.

The specific application method is the linear interpolation, and the batch division method can be divided into the following steps:

(1) Training samples preprocessing and batch samples inputting.

(2) Set the parameter of DCNN initialization and increase the number of batch sample capacity and iteration by equal steps.

(3) Obtain the GE of DCNN model under different parameter sets and seek to obtain model diagnosis performance inflection points as much as possible.

(4) Use the linear interpolation method to obtain the Landform Map of the GE of DCNN model under different iterations and batch sample capacity.

(5) Find the bottom of the Landform Map on the basis of larger sample batch capacity and less iteration, and it is the optimal parameter set.

C. THE OVERALL FRAMEWORK OF DCNN’S TRAINING GUIDE

Based on the training strategy proposed in this paper, the overall framework of the DCNN’s training guide can be constructed as shown in Fig. 6. The detailed operation process is expressed as follows.

(1) In the process of sample processing, shift sample the original vibration data under different health status with a sampling window of a certain width.

(2) Convert one-dimensional signal into two-dimensional signal matrix using MATLAB’s own dimensional conversion function.

(3) Mix the signal matrix of all health status together and disperse them randomly.

(4) Conduct the operation process expressed in section 3.2.

(5) Enter the test samples and get the diagnosis results. The training guide proposed in this paper has two highlights. One, the key factors that affect the diagnosis ability of the trained DCNN model have been found, and it is more scientifically instructive to the designation of training strategies. The other, based on the highlight one, the batch division quantification could be directed and a new division method could be proposed.

IV. EXPERIMENTAL AND VERIFICATION

A. TEST RIG AND EXPERIMENT SETTING UP

In order to verify the feasibility of the training guiding method, the data were obtained on the planetary gearbox fault experimental platform. The planetary gearbox test rig is shown in Fig. 7.

The experiment was carried out on the planetary gearbox 1. Four kinds of faults were planted artificially on the sun gear, such as worn tooth, eccentric, pitting and chipped tooth (shown in Fig. 8). The acceleration sensor was installed on the planetary gearbox 1. The motor speed is 1200 r/min. The sampling frequency is 5 kHz. The load is 41.2 N·m. The sampling points are 196608 points respectively. The time domain waveform of the collected vibration signals are shown in Fig. 9.

B. TRAINING SAMPLES PREPARING

The sample method used in this paper is shown in Fig. 10. The original vibration signal is shift sampled with a sampling window of a certain width (for example, 1024 points), and finally, n samples are obtained.
FIGURE 6. DCNN’s training guide.
The diagnostic object has five different types of health states, and $5 \times n$ samples could be obtained. Then the dimension deformation of each sample signal could be carried out, and the sample data can be transformed from one-dimensional (1024 points) into two-dimensional ($32 \times 32$). In order to represent the sample form more intuitively, the sample matrix is given in the form of confusion matrix, as shown in Fig. 6. The sample matrix of all health states is combined to form the training sample set, and then, all the samples are randomly dispersed.

**C. CONSTRUCTION OF DCNN MODEL AND OPTIMIZATION OF PARAMETERS**

The parameters related to the DCNN model are shown in Table 1 12C-2S-24C-2S means that DCNN has two...
convolutional layer and two down sampling layer, the number of the kernels are 12 and 24 respectively, the method that down sampling layer used is mean sampling. The iterations and batch sample capacity are determined by the final optimization result.

Input the dispersed sample set into the DCNN model, and the diagnostic results under different iterations and batch sample capacity are shown in Table 2 and Fig. 11 below. The methods used contain the grid search method and the interpolation method, so as to obtain the Landform Map of

| TABLE 1. 1 DCNN related structure and parameters. |
|-----------------------------------------------|
| A     | B     | C     | D     | E     | F       | G     |
| 12C-2S-24C-2S | 5×5   | 1     | 900   | 50    | To be determined | To be determined |

A: DCNN’s structure; B: Convolutional kernel size; C: Learning rate; D: No. of training samples; E: No. of testing samples; F: Iterations; G: Batch sample capacity.

the GE of DCNN model under different iterations and batch sample capacity.
As shown in Table 1, the number of training samples is 900. Considering that the number of training batches in a single cycle should be an integer in the specific algorithm, the batch sample capacity should be divisible to 900. In the optimization process, with the decrease of sample capacity, the number of batches will increase, but the training efficiency will be reduced. So, based on the ideal diagnostic accuracy, the iterations and batch sample capacity should be fewer and larger respectively as possible. The optimum number of iterations and batch sample capacity can be found as 150 and 10, respectively.

### TABLE 2. Diagnostic results for different iterations and batch sample capacity (A represents iterations, B represents misclassification, C represents batch sample capacity).

| A | 140 | 150 | 160 | 170 | 180 |
|---|-----|-----|-----|-----|-----|
| B | 2   | 3   | 5   | 6   | 9   |
| C | 3   | 1   | 0   | 0   | 0   |
|   | 1   | 0   | 0   | 0   | 1   |
|   | 1   | 0   | 0   | 0   | 1   |
|   | 1   | 0   | 0   | 2   | 2   |
|   | 1   | 0   | 0   | 2   | 8   |
|   | 50  | 5   | 6   | 12  | 9   |

**FIGURE 17.** Probabilistic distributed feature \( f_v \) (New DBN).

**FIGURE 18.** Target output before activation (New DBN).

**FIGURE 19.** Target output after activation (New DBN).

**FIGURE 20.** Diagnosis accuracy (%) (New DBN).

**FIGURE 21.** Probabilistic distributed feature \( f_v \) (Original DCNN).

### D. INFLUENCE OF RANDOM DISCRETIZATION OF SAMPLES ON THE DIAGNOSIS ACCURACY

In order to carry out the comparison, the Deep Belief Network (DBN) method and the sample non-random discretization experiment have been adopted. The results of DBN with sample non-random discretization are shown...
FIGURE 22. Target output before activation (Original DCNN).

FIGURE 23. Target output after activation (Original DCNN).

FIGURE 24. Diagnosis accuracy (%) (Original DCNN).

FIGURE 25. Probabilistic distributed feature $f_v$ (New DCNN).

FIGURE 26. Target output before activation (New DCNN).

FIGURE 27. Target output after activation (New DCNN).

The results of DBN with training guide are shown in Fig. 17∼Fig. 20. The results of DCNN with sample non-random discretization are shown in Fig. 21∼Fig. 24. The results of DCNN with training guide are shown in Fig. 25∼Fig. 28. In order to understand the process of the diagnosis intuitively, an example of 4-layers DCNN diagnosis process has been shown in Fig. 12. Furthermore, the model learned feature $f_v$, the $f_v * W + b$, the Sigmoid ($f_v * W + b$) and the final diagnosis result are shown visually as follows too.

The distributed data features $f_v$ are shown in Fig. 13, Fig. 17, Fig. 21 and Fig. 25, which are obtained from a single
(3) It is conducive to improving the interpretability of deep learning algorithms.

(4) An explicable training guide has been proposed for the popularization and application of DCNN in mechanical equipment fault diagnosis. Furthermore, when the Deep Learning algorithm is changed, the training strategy is still useful.

ACKNOWLEDGMENT
The authors are grateful for the helpful comments and constructive suggestions of the journal editors and anonymous reviewers.

REFERENCES

[1] L. Zhang and N. Hu, “Fault diagnosis of sun gear based on continuous vibration separation and minimum entropy deconvolution,” Measurement, vol. 141, pp. 332–344, Jul. 2019.

[2] W. Huang, S. Li, X. Fu, C. Zhang, J. Shi, and Z. Zhu, “Transient extraction based on minmax concave regularized sparse representation for gear fault diagnosis,” Measurement, vol. 151, Feb. 2020, Art. no. 107273, doi: 10.1016/j.measurement.2019.107273.

[3] H. Liu, W. Huang, S. Wang, and Z. Zhu, “Adaptive spectral kurtosis filtering based on Morlet wavelet and its application for signal transients detection,” Signal Process., vol. 96, pp. 118–124, Mar. 2014.

[4] B. Cai, Y. Liu, Y. Liu, X. Kong, H. Wang, H. Xu, and W. Ge, “Remaining useful life estimation of structure systems under the influence of multiple causes: Subsea pipelines as a case study,” IEEE Trans. Ind. Electron., vol. 67, no. 7, pp. 5737–5747, Jul. 2020.

[5] Y. Yang, P. Luo, L. Gan, and J. Cheng, “SADBNN and its application in fault classification and identification of rolling bearings,” Vib. Shock, vol. 38, pp. 11–16, Aug. 2019.

[6] P. Luo, Y. Yang, and J. Cheng, “AFDCNN method and its application in gear quantitative diagnosis,” J. Electron. Meas. Instrum., vol. 32, pp. 33–39, Jan. 2018.

[7] Y. Yang, P. Luo, and J. Cheng, “Generalized variational modal decomposition and its application in gearbox compound fault diagnosis,” China Mech. Eng., vol. 28, pp. 1069–1073, May 2017.

[8] L. Zhang and N. Hu, “Time domain synchronous moving average and its application to gear fault detection,” IEEE Access, vol. 7, pp. 93035–93048, Jun. 2019.

[9] B. Cai, Y. Liu, and M. Xie, “A dynamic-Bayesian-network-based fault diagnosis methodology considering transient and intermittent faults,” IEEE Trans. Autom. Sci. Eng., vol. 14, no. 1, pp. 276–285, Jan. 2017.

[10] R. Socher, B. Hulv, B. P. Bath, C. D. Manning, and A. Y. Ng, “Convolutional-recursive deep learning for 3D object classification,” in Proc. NIPS, 2012, vol. 3, no. 7, p. 8.

[11] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” Nature, vol. 521, no. 7553, pp. 436–444, 2015.

[12] L. Deng, J. Li, J.-T. Huang, K. Yao, D. Yu, F. Seide, M. Seltzer, G. Zweig, X. He, J. Williams, Y. Gong, and A. Acero, “Recent advances in deep learning for speech research at Microsoft,” in Proc. IEEE Int. Conf. Acoust., Speech Signal Process., May 2013, pp. 8604–8608.

[13] G. Hinton, L. Deng, D. Yu, G. Dahl, A.-R. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. Sainath, and B. Kingsbury, “Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups,” IEEE Signal Process. Mag., vol. 29, no. 6, pp. 82–97, Nov. 2012.

[14] C. Lu, Z.-Y. Wang, W.-L. Qin, and J. Ma, “Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification,” Signal Process., vol. 130, pp. 377–388, Jan. 2017.

[15] H. Chen, J. Wang, B. Tang, K. Xiao, and J. Li, “An integrated approach to planetary gearbox fault diagnosis using deep belief networks,” Meas. Sci. Technol., vol. 28, no. 2, Feb. 2017, Art. no. 025010.

[16] J. An, P. Ai, S. Xu, C. Liu, J. Xia, and D. Liu, “An intelligent fault diagnosis method for rotating machinery based on one dimensional convolution neural network,” J. Nanjing Univ. (Natural Sci.), vol. 55, pp. 133–142, Jan. 2019.

[17] C. Wu, P. Jiang, F. Feng, T. Chen, and X. Chen, “Faults diagnosis method for gearboxes based on a 1-D convolutional neural network,” J. Vib. Shock, vol. 37, pp. 51–56, Nov. 2018.

---

**TABLE 3.** Diagnosis results.

| Methods                        | Average diagnosis accuracy (%) |
|--------------------------------|-------------------------------|
| DBN with sample non-random     | 38                            |
| discrete                        |                               |
| DBN with training guide        | 88                            |
| DCNN with sample non-random    | 76                            |
| discrete                        |                               |
| DCNN with training guide       | 100                           |
[18] G. Chen, T. X. Han, Z. He, R. Kays, and T. Forrester, “Deep convolutional neural network based species recognition for wild animal monitoring,” in Proc. IEEE Int. Conf. Image Process. (ICIP), Paris, France, Oct. 2014, pp. 858–862.

[19] M. Gan, C. Wang, and C. Zha, “Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of rolling element bearings,” Mech. Syst. Signal Process., vol. 72–73, pp. 92–104, May 2016.

[20] Q. Li, Z. Jin, C. Wang, and D. D. Zeng, “Mining opinion summarizations using convolutional neural networks in Chinese microblogging systems,” Knowl.-Based Syst., vol. 107, pp. 289–300, Sep. 2016.

[21] H. Li, Q. Zhang, X. Qin, and Y. Sun, “Fault diagnosis method for rolling bearings based on short-time Fourier transform and convolution neural network,” J. Vib. Shock, vol. 37, no. 19, pp. 124–131, Sep. 2018.

[22] Q. Zhou, X. Liu, J. Zhou, H. Shen, and X. Xiong, “Fault diagnosis for rotating machinery based on 1D depth convolutional neural network,” J. Vib. Shock, vol. 37, pp. 31–37, Mar. 2018.

[23] L. Wen, X. Li, L. Gao, and Y. Zhang, “A new convolutional neural network-based data-driven fault diagnosis method,” IEEE Trans. Ind. Electron., vol. 65, no. 7, pp. 5990–5998, Jul. 2018.

[24] D. Zhao, T. Wang, and F. Chu, “Deep convolutional neural network based planet bearing fault classification,” Comput. Ind., vol. 107, pp. 59–66, May 2019.

[25] N. Hu, H. Cheng, Z. Cheng, L. Zhang, and Y. Zhang, “Fault diagnosis for planetary gearbox based on EMD and deep convolutional neural networks,” J. Mech. Eng., vol. 55, no. 7, pp. 9–18, May 2019.

[26] M. M. M. Islam and J.-M. Kim, “Automated bearing fault diagnosis scheme using 2D representation of wavelet packet transform and deep convolutional neural network,” Comput. Ind., vol. 106, pp. 142–153, Apr. 2019.

[27] X.-C. Cao, B.-Q. Chen, B. Yao, and W.-P. He, “Combining translation-invariant wavelet frames and convolutional neural network for intelligent tool wear state identification,” Comput. Ind., vol. 106, pp. 71–84, Apr. 2019.

[28] Y. Han, B. Tang, and L. Deng, “An enhanced convolutional neural network with enlarged receptive fields for fault diagnosis of planetary gearboxes,” Comput. Ind., vol. 107, pp. 50–58, May 2019.

[29] W. Zhang, C. Li, G. Peng, Y. Chen, and Z. Zhang, “A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load,” Mech. Syst. Signal Process., vol. 100, pp. 439–453, Feb. 2018.

[30] S. Dong, Z. Huang, and A. Brief, “Theoretical overview of random forests,” J. Integ. Technol., vol. 2, pp. 1–7, Jan. 2013.

[31] Y. Sheng, G. Shi, and Z. Qin, “A stronger law of large numbers for uncertain random variables,” Soft Comput., vol. 22, no. 17, pp. 5655–5662, Sep. 2018.

[32] Y. H. Huo, “Chebyshev inequality and its applications,” J. Changchun Univ. Technol., vol. 33, pp. 712–714, Jun. 2012.

PENG LUO received the B.S. and M.S. degrees in mechanical engineering from Hunan University, in 2012 and 2018, respectively. He is currently pursuing the Ph.D. degree in mechanical engineering with the National University of Defense Technology. His research interests include dynamic modeling, signal processing, and machinery fault diagnosis.

NAOQING HU received the B.S., M.S., and Ph.D. degrees in mechanical engineering from the National University of Defense Technology, in 1989, 1992, and 2001, respectively. From 1993 to 1998, he was a Lecturer with the Department of Mechanical and Electronic Engineering and Instrumentation. From 1998 to 2003, he was an Associate Professor with the Department of Mechatronics and Automation. Since 2004, he has been a Professor with the National University of Defense Technology. He is the author of 4 books, over 260 articles, and more than 7 inventions. His research interests include condition monitoring, prognosis and health management, signal processing, mechanical dynamics, nonlinear systems, structure health monitoring, and artificial intelligence.

GUOJI SHEN received the B.S., M.S., and Ph.D. degrees in mechanical engineering from the National University of Defense Technology, in 1996, 1999, and 2005, respectively. His research interests include dynamic modeling, signal processing, and machinery fault diagnosis.

LUN ZHANG received the B.S., M.S., and Ph.D. degrees in mechanical engineering from the National University of Defense Technology, in 2013, 2015, and 2019, respectively. He is currently a Lecturer with the College of Intelligence Science and Technology, National University of Defense Technology. His research interests include dynamic modeling, signal processing, and machinery fault diagnosis.

ZHE CHENG received the B.S. degree in civil engineering and the M.S. and Ph.D. degrees in mechanical engineering from the National University of Defense Technology, China, in 2004, 2006, and 2011, respectively. He is currently an Associate Professor with the College of Intelligence Science and Technology, National University of Defense Technology. His research interests include dynamical modeling, mechanics analysis, signal processing, condition monitoring and fault prognosis, and prognostics and health management.

***