A self-Adaptive CNN with PSO for bearing fault diagnosis

Jungan Chen\textsuperscript{a}, Jean Jiang\textsuperscript{b}, Xinnian Guo\textsuperscript{c} and Lizhe Tan\textsuperscript{d}

\textsuperscript{a}Department of Electronic and Computer Science, Zhejiang Wanli University, Ningbo, People’s Republic of China; \textsuperscript{b}College of Technology, Purdue University Northwest, Hammond, IN, USA; \textsuperscript{c}Department of Electronic Information Engineering, Huaiyin Institute of Technology, Huai’an, People’s Republic of China; \textsuperscript{d}Department of Electrical and Computer Engineering, Purdue University Northwest, Hammond, IN, USA

ABSTRACT

Convolutional neural network (CNN) is now widely applied in bearing fault diagnosis, but the design of network structure or parameter tuning is time-consuming. To solve this problem, a particle swarm optimization (PSO) algorithm is used to optimize the network structure and a self-adaptive CNN is proposed in this paper. In the proposed method, a theoretical method is used to automatically determine the window size of short-time Fourier transform (STFT). To reduce the computation time, PSO is only applied to obtain the optimal key parameters in CNN with a small number of training samples and a small epoch number. To simplify the CNN structure, a fitness function considering the numbers of kernels and neuron nodes is used in PSO. According to the verification experiments for two well-known public datasets, the proposed method can get higher accuracy than other state-of-art methods. Furthermore, the parameters that are required to be input only involve the bearing parameters, so the proposed method can be applied in industry readily.

1. Introduction

Rolling element bearings have been widely applied and are the most critical components in industrial machinery. According to our literature review, 40–55% of broken machines are caused by bearing faults (Boudinar et al., 2016; Hamadache et al., 2015; Harmouche et al., 2015; Rai & Upadhyay, 2016; Sohaib et al., 2017). The degradation of bearings can result in unexpected failures of machines, which lead to long downtimes, large economic losses, and serious human injuries. Therefore, the bearing fault diagnosis is regarded as an important tool for identification, prediction, and prevention of failures in rotating machinery.

Fault diagnosis techniques can be generally categorized into model-based, signal-based, knowledge-based, and hybrid fault diagnosis (Gao et al., 2015).

In signal-based methods, there are many kinds of signals such as vibration and acoustic signals (Liu et al., 2016), electric current (Boudinar et al., 2016), temperature (Yuan et al., 2018), and wear debris. Among these, the vibration signal is the most widely used approach (Lin & Ye, 2019; Sohaib et al., 2017). The signal processing methods generally include Fourier Transform (FT), Short-Time Fourier Transform (STFT), Wavelet Transform (WT), Wavelet Package Transform (WPT), Empirical Wavelet Transform (EWT), Hilbert-Huang Transform (HHT), Empirical Mode Decomposition (EMD), Variation Mode Decomposition (VMD), and Ensemble Empirical Mode Decomposition (EEMD). For more details on the advantages and disadvantages of these methods can be seen in Refs (Lin & Ye, 2019; Ma et al., 2018; Yi et al., 2016).

Nowadays, a signal-based method is always integrated with a knowledge-based method, which can be used to optimize the parameters in the signal-based method. For example, the particle swarm optimization (PSO) and quasi-Newton minimization algorithms are applied to seek optimal parameters of continuous wavelet transform models based on impulse modelling (Huo et al., 2017). Similarly, PSO is applied for the parameter optimization of VMD (Wang et al., 2018; Yi et al., 2016; Yuwono et al., 2016). On the other hand, the signal-based method can be used for handcrafted feature design and feature extraction, and then the knowledge-based method further makes use of these features to learn fault modes through training with historical data. Here, two kinds of methods, i.e. shallow machine learning and deep learning, are used for knowledge learning.

For the shallow machine learning methods, after the right set of features are designed, these features are fed into some shallow machine learning algorithms. For example, in references EMD-SVM (Zhang, Liang et al. 2015), HHT-SVM (Soualhi et al., 2015), DWT-KNN (Jung &
Koh, 2015), WPT-KNN (Wang, 2016), WT-Naive Bayes (NB) (Seshadrinath et al., 2014), HHT-FFT-PCA (Harmouche et al., 2015), and EMD-PSO-SVM (Chen, Tang et al. 2014), signal processing methods including EMD, HHT, DWT, WPT, etc are used to extract features and SVM, KNN, NB, PCA are to perform classifications. PSO is used to determine the optimal parameters of SVM (Chen, Tang et al. 2014).

In the deep learning methods, Ref (Peng et al., 2017) use convolutional neural network (CNN) to diagnose fused signals of two sensors, where FFT is as are preprocessing method. Wavelet scalogram images are fed into CNN to detect faults (Wang et al., 2016). Similarly, three time–frequency analysis methods STFT, WT, and HHT are used to generate image representations of the raw signals which are fed into CNN (Verstraete et al., 2017). Adaptive overlapping CNN (AOCCNN) with a sparse filter is proposed to deal with one-dimensional raw vibration signals directly. The adaptive convolutional layer and the root-mean-square (RMS) pooling layer are proposed to deal with the shift variant problem (Qian et al., 2018). Time-domain and frequency-domain features are extracted from different sensor signals and then fed into multiple two-layer sparse auto encoder (SAE) for feature fusion, which is used to train deep belief network (DBN) for further classification (Chen & Li, 2017). A stacked sparse auto encoder is applied to automatically extract the fault features of sound spectrograms with STFT, and softmax regression is adopted as the method of classifying the fault modes (Liu et al., 2016). In Ref. (Jia et al., 2016), DNN based on a stack auto encoder is applied for roller bearing and planetary gearbox fault diagnosis using frequency spectra. In Ref. (Jiang et al., 2018), deep recurrent neural network (DRNN) is used to extract the features from the input spectrum sequences. In Ref. (Lei et al., 2016), the features are learnt through sparse filtering in an unsupervised way and then fed to softmax regression for classifying health conditions in a supervised manner (Lei et al., 2016).

In these two kinds of knowledge-based methods, deep learning can auto extract highly representative features while shallow machine learning methods require labour-intensive feature engineering, so it is widely researched all over the world. However, the design of network or parameter tuning is one time-consuming procedure, which affects the performance of the neuron network. For example, lots of existing CNN networks have been designed and applied, and different CNNs (Hoang & Kang, 2019; Verstraete et al., 2017; Wang et al., 2016; Wen et al., 2018) consisting of different convolution, pooling, and dense layers are proposed, which result in different performances. On account of this, most current research focuses on the development of the network. In order to design the optimized deep belief network, PSO is used for deciding the optimal network structure, including the optimal number of neurons in each hidden layer, the learning rate, and the momentum (Shao et al., 2015). Similarly, the main parameters of the deep CNN model are also determined with the PSO method (Fuan et al., 2017). In our method, a two-step training idea is introduced, which is different from that in Ref. (Fuan et al., 2017). The first-step is the optimization of the CNN structure, where the fitness function accounts the number of neuron nodes and trains the CNN parameters with a small number of training samples and a small epoch number. The second-step is the final training with the optimized CNN.

Through above literature review, it is obvious that the design of network or tuning parameters is a complex and time-consuming work. To relieve the burden, a Self-Adaptive CNN (SA-CNN) is proposed in this study. Because the PSO has the characteristics of fast convergence speed, small setting parameters and convenient implementation, it is used to optimize the key parameters in CNN (Yuwono et al., 2016).

Although CNN can auto extract features, to accelerate the evolution of PSO, signal processing is taken as a preprocessing method to extract basic time–frequency information. Among signal processing methods reviewed before, WT analysis is restricted by mutually selected wavelet basis function and the number of decomposition levels (Wang et al., 2018); Empirical wavelet transform (EWT) is restricted on choosing the boundaries of Fourier spectrum appropriately (Yi et al., 2016); HHT and EMD exists deficiencies such as model mixing and the end effect (Kumar & Kumar, 2019); Wang et al., 2018; Yi et al., 2016); VMD is not model-adaptive and heavily relies on the penalty parameter selection and the number of components (Wang et al., 2018; Yi et al., 2016); EEMD is computationally expensive and requires selection of noise parameter (Kumar & Kumar, 2019). Most of them are not adaptable and restricted on the selection of parameters. As the simplicity and efficiency of STFT, it is taken as the preprocessing method in this work.

The main contributions of this study are as follows:

1. STFT is a simple and efficient method to transform time domain signals into time–frequency space. Its limitation is that it gives constant time and frequency resolution once the window size is fixed (Kumar & Kumar, 2019). Hence, the setting of the window size is important. In this study, through analysing the characteristics of the bearing fault type, a method of automatically deciding the window size of STFT or its input-size of spectrum is proposed.
PSO is applied to optimize the key parameters of CNN including the kernel number and the number of neuron nodes in CNN. To reduce the computational complexity, a fitness function accounting the number of neuron nodes is provided.

In the implementation of algorithm, to further reduce the time complexity of PSO, a small number of training samples and a small epoch number are used to evaluate the prediction accuracy of each particle. Through a two-step training, a SA-CNN with adaptive input-size for bearing fault diagnosis is proposed.

The rests of this paper are organized as follows. Section II introduces the bearing fault types, the method of automatically deciding parameters in STFT, the optimization with PSO and the implementation of SA-CNN. In Section III, contrastive experiments with two well-known public datasets are conducted and the results are discussed. Finally, Section IV provides some conclusions and suggests topics for future research.

2. Our proposed method

In this study, a self-adaptive CNN method based on STFT (SA-CNN) is proposed for bearing fault diagnosis, as shown in Figure 1. It has the following five steps: (1) calculating the input size \((W, H)\); (2) randomly segmenting the raw signal to many signal segments with \(H \times W\) points; (3) applying STFT to every signal segment, where the rectangle window size is \(W\) and has no overlapping, and obtaining the power spectrum with time–frequency information; (4) in the first-step training, PSO is used to evolve the CNN parameters, i.e. the numbers of kernels and dense units; (5) in the second-step training, the final training with optimized CNN is used to adjust the neuron weight.

2.1. Bearing fault types

Bearing faults are categorized into two types: generalized roughness fault and single-point fault (Eren, 2017; Stack et al., 2004).

The generalized roughness fault is a type of failure where the surface of a bearing has degraded considerably over a large area. It is a kind of distributed and non-cyclic fault caused by improper lubrication, erosion, or pollution. There is no identifiable characteristic frequency associated with this fault type.

The single-point fault is a kind of single and localized fault caused by a small hole, a pit, or a missing material. As we all know, when a fault in one surfaces of a bearing strike another surface, a force impulse is generated, thus exciting resonances in the bearing and the machine. The successive impacts produce a series of impulse responses, the amplitude of which can be modulated (Ho & Randall, 2000). Therefore, the single-point fault creates a periodic impact vibration and produces a characteristic fault frequency.

Since rolling-element bearings are constituted by an inner race, an outer race, a cage and balls, the single-point fault includes four types: cage fault, outer race fault, inner race fault and ball bearing fault, as shown in Figure 2 (Hamadache et al., 2015). Their corresponding characteristic fault frequencies are given by Equations (1–4) separately, where \(N_{\text{ball}}\) is the number of balls in one bearing, \(D_{\text{ball}}\) and \(D_{\text{cage}}\) are the diameters of ball and cage respectively, \(\theta\) is the contact angle, \(f_{r}\) is the shaft speed, and \(f_{s}\) is the sample frequency (Aimer et al., 2019; Eren, 2017).

\[
\begin{align*}
    f_{\text{OR}} &= \frac{N_{\text{ball}}}{2} f_{r} \left(1 - \frac{D_{\text{ball}}}{D_{\text{cage}}} \cos \theta \right) \quad (1) \\
    f_{\text{IR}} &= \frac{N_{\text{ball}}}{2} f_{r} \left(1 + \frac{D_{\text{ball}}}{D_{\text{cage}}} \cos \theta \right) \quad (2) \\
    f_{\text{ball}} &= \frac{D_{\text{cage}}}{2D_{\text{ball}}} f_{r} \left(1 - \left(\frac{D_{\text{ball}}}{D_{\text{cage}}} \right)^2 \cos^2 \theta \right) \quad (3) \\
    f_{\text{cage}} &= \frac{1}{2} f_{r} \left(1 - \frac{D_{\text{ball}}}{D_{\text{cage}}} \cos \theta \right) \quad (4)
\end{align*}
\]

In addition, bearing characteristic fault frequencies can be grouped into four different zones: low frequency zone...
Bearing frequency zones. (shaft speed zone), bearing defect frequency zone, bearing nature resonance frequency zone, and high frequency zone in Figure 3 (El-Thalji & Jantunen, 2015).

Bearing defect frequency zone (Zone II) include two parts. The first part is the low-frequency event which originates from the entry of the ball/roller in the vicinity of defect. This event gives peak in the bearing fault frequency zone. The second part is the high-frequency event, which is the result of the impact of rolling element on the trailing edge of the defect. The defect frequency zone can be derived from Equations (1–4) (Eren, 2017; Kumar & Kumar, 2019). To be notice, this study mainly focuses on the single-point fault or Zone II because the defect frequency will be used to auto decide the parameters related with STFT.

In signal-based methods, the major disadvantage of the time–frequency approach is that the computation time increases with both the number of samples and the number of sought harmonics. Equations (1–4) enable us to know in advance the frequency bands where the fault signature is likely to appear. Therefore, it can reduce the length of the spectrum and consequently the computation time (Aimer et al., 2019; Boudinar et al., 2016) to process the data on these predictable fault frequencies band instead of those on all the spectra.

2.2. STFT with auto-decided parameters

In Figure 1, when a raw signal is input, it will be segmented into many signal segments and every segment is applied with STFT. Therefore, the length of one signal segment and the window size are required to be fixed. As described in Figure 3 in Section 2.1, the bearing defect frequency zone contains lower frequency components and higher frequency components. To prevent any information loss of the lowest frequency component, the time of sampling L points must be longer than the multiple periods of the lowest frequency component, as shown in Equation (7). Thus, W can be given by Equation (8). For Equations (5–8), $\alpha > 2$ is required.

$$T_L = \frac{L}{f_s} > \frac{\alpha}{\min([f_{OR}, f_{IR}, f_{ball}, f_{cage}])}$$  \hspace{1cm} (5)

$$L > \frac{\alpha f_s}{\min([f_{OR}, f_{IR}, f_{ball}, f_{cage}])}$$  \hspace{1cm} (6)

$$T_W = \frac{W}{f_s} > \frac{\alpha}{\max([f_{OR}, f_{IR}, f_{ball}, f_{cage}])}$$  \hspace{1cm} (7)

$$W > \frac{\alpha f_s}{\max([f_{OR}, f_{IR}, f_{ball}, f_{cage}])}$$  \hspace{1cm} (8)

$$H = \frac{L}{W}$$  \hspace{1cm} (9)

After L samples are randomly extracted from the raw signal, they will be split into H frames, every of which has a fixed window size of W with no overlapping. Every frame with W samples is processed through FFT as Equation (10). Equation (12) is the normalization of the power spectrum. Since the power spectrum is symmetrical, only the single-sided data are fed into CNN, i.e. a matrix $[H, W/2]$ is fed into CNN.

$$c_{m,k} = \frac{1}{W} \sum_{n=0}^{W-1} x(n) e^{-j(2\pi kn/W)}$$  \hspace{1cm} (10)

$$p_{m,k} = |c_{m,k}|^2$$  \hspace{1cm} (11)

$$p_{m,k} = \frac{p_{m,k} - \min(p_m)}{\max(p_m) - \min(p_m)}$$  \hspace{1cm} (12)

where $m \in [0,H]$, and $k \in [0,W]$.

2.3. Optimized CNN structure through PSO

2.3.1. Fixed CNN structure

Figure 4 illustrates the proposed CNN structure. Three successive convolution and pooling layers are used to extract high level features. The ReLU is taken as the activation function and the max pooling layer is for preserving the maximum coefficients that contain the main information of fault features. The last three dense layers construct an MLP as a classifier.

According to the universal approximation theorem, three-layer neuron network, with a sufficient number of hidden-layer units, can approximate any continuous
function. But in practice learning, the hidden-layer units are hard to be determined. Feeding the auto extracted features with three-layer CNN into three-layer MLP can make MLP structure as simple as possible and PSO is used to find the optimal network structure, which may be a wise solution.

In CNN, the loss function uses the categorical cross entropy loss (Zhang & Sabuncu, 2018); the optimizer is the Adam method because its hyper-parameters have intuitive interpretations and typically require little tuning (Kingma & Ba, 2015).

2.3.2. Binary PSO
In PSO, each particle has its own position $x$ and velocity $v$ and maintains a record of the position of its previous best performance $p_{\text{best}}$. A global best position $g_{\text{best}}$ is maintained by the algorithm. The $p_{\text{best}}$ represents potential solutions while the global best position is the global optimum solution.

The velocity $v$ and the position $x$ of each particle at the next iteration ($t + 1$) are calculated based on the values at current iteration ($t$). Each iteration comprises the evaluation of each particle, which is discussed in ‘(3) fitness evaluation’; each particle adjusts its own velocity in the direction of $p_{\text{best}}$ and $g_{\text{best}}$. Equation (15) is used for binary encoding.

$$v_{t+1} = w v_t + c_1 r_1 (p_{\text{best}} - x_t) + c_2 r_2 (g_{\text{best}} - x_t) \quad (13)$$

$$x_{t+1} = x_t + v_{t+1} \quad (14)$$

$$x_{t+1} = \begin{cases} 0 & \text{rand()} \leq f(v_{t+1}) \\ 1 & \text{otherwise} \end{cases} \quad (15)$$

$$f(y) = \frac{1}{1 + e^{-y}} \quad (16)$$

where $w$ is called the ‘inertia weight’, which controls the impact of the previous velocity of the particle on its current one; $c_1$ and $c_2$ are positive constants with default values called ‘acceleration coefficients’; $r_1$, $r_2$ and rand() are random numbers uniformly distributed in the interval [0.0, 1.0] (Kennedy & Eberhart, 1997; Ziani et al., 2017).

2.3.3. Fitness evaluation
Let all the kernel size be $F$, all the padding size be $P$, all the stride step be $S$, and all the pooling size be $D$. Then, the output size of one successive convolution and pooling layer can be expressed as:

$$W_i = ((W_{i-1} - F + 2P)/S + 1)/D \quad (17)$$

$$H_i = ((H_{i-1} - F + 2P)/S + 1)/D \quad (18)$$

The encoding information in each particle includes $N_{\text{kernel}}^i$ and $N_{\text{units}}^i \in [0, [N_c, 2^N_d - 1]]$, where $i \in [1, 3], j \in [1, 2]$, $N_c$ is the number of classes and $N_d = 8$ is the number of bits in every dimension. Then, the fitness score can be expressed as Equation (20).

$$\text{score} = \begin{cases} - \text{acc} & \text{acc} < 1 \\ -1 - \text{len} & \text{acc} = 1 \end{cases} \quad (20)$$

where acc is the prediction accuracy of CNN. It is noted that, to reduce the time complexity of PSO, a small number of training samples $N_{ts}$ and a small epoch number $N_{ep}$ are used to evaluate the prediction accuracy of each particle. Equation (19) is to ensure the simplicity of the network structure to reduce the time complexity.

2.4. Implementation of SA-CNN
In Figure 5, steps (1-2) are used to automatically decide the parameters of $W$ and $H$ used in STFT; steps (3.1-3.2) are implemented to preprocess STFT to generate the power spectrum matrix; step 3.3 is to optimize the CNN structure.

3. Experiments
Two public famous fault diagnosis datasets, i.e. the Case Western Reserve University’s (CWRU) bearing dataset (Loparo & Loparo, 2013), and the Machinery Failure Prevention Technology (MFPT) society bearing fault dataset (Bechhoefer, 2016), are used to investigate the effectiveness of the proposed method. In every experiment, let the default batch size be 100 and the epoch number be 50. All results are obtained by running the algorithm for ten times with one GPU. The default proportion of the training data in the whole dataset $Pt$ is 0.8. All prediction accuracies are from the results with the testing dataset.

3.1. CWRU bearing dataset
Two-horsepower reliance electric motors are used for the acquisition of accelerometer data on both the drive end (DE) and fan end (FE) bearings, as shown in Figure 6. In the experiment, artificial single-point faults are seeded in the bearing’s inner raceway (IR), outer raceway (OR), and rolling element (ball) (BF) with an electro discharge machining (EDM) operation. The vibration signal is sampled at 12 kHz through accelerometers attached at the 12 o’clock position to the DE of the motor.

Vibration data are recorded under various engine loads (0–3 hp) running at 1730, 1750, 1772, and 1797 RPM. The depth of the motor shaft bearings with faults ranges
from 0 to 0.028 in. (0, 0.007, 0.014, 0.021, 0.028 in.), and the fault orientation includes the inner race, rolling element, and outer race. As shown in Table 1, all these data have one normal class and 15 fault classes without considering the different RPMs. Notably, in this study, some classes (bold and italic fonts) are merged as one value in the last column of Table 1. Hence, there are totally 12 classes of data to be classified. In this experiment, 10000 data samples are randomly extracted from the raw data file with the same probability.

### 3.1.1. Comparison of the results from different methods

In this experiment, SA-CNN is compared with many deep learning methods including deep recurrent neural network (DRNN) (Jiang et al., 2018), sparse filter (Lei et al., 2016), CNN-Vibration image (Wen et al., 2018), CNN-PSO-WP (Fuan et al., 2017), CNN-Wavelet/STFT (Verstraete et al., 2017), HCNN (Wen et al., 2019b), and an optimized deep belief network with PSO (DBN-PSO) (Shao et al., 2015).

As can be seen from Table 2, our proposed SA-CNN method achieves an accuracy of 99.96% after running for ten times. Figure 7 shows one of the classification confusion matrixes. Furthermore, in our method, the number of classes to be classified is 12, which is more than that in other methods except DRNN (with an accuracy of 96.53%). Therefore, our method is more efficient than the other methods. ‘SA-CNN-E2E’ algorithm, called SA-CNN-End-to-End, means that there is no preprocessor of STFT before feeding into CNN. It get 99.60% accuracy, so the

### Table 1. Classes in CWRU dataset.

| Fault depth | Fault name | Class name used |
|-------------|------------|-----------------|
| 0           | Normal     | Normal          |
| 0.028       | 0.028-Ball | 28-Ball         |
| 0.028       | 0.028-InnerRace | 28-IR        |
| 0.021       | 0.021-Ball | 21-Ball         |
| 0.021       | 0.021-InnerRace | 21-IR        |
| 0.021       | 0.021-OuterRace12 | 21-OR        |
| 0.021       | 0.021-OuterRace3  | 21-OR         |
| 0.007       | 0.007-Ball | 07-Ball         |
| 0.007       | 0.007-InnerRace | 07-IR        |
| 0.007       | 0.007-OuterRace6 | 07-OR        |
| 0.007       | 0.007-OuterRace12 | 07-OR      |

### Table 2. Comparison with CWRU dataset (N_{ap} = 30, N_{ts} = 500, DE\_time, %).

| Algorithm                  | Average accuracy | Class number |
|---------------------------|------------------|--------------|
| SA-CNN                    | 99.96            | 12           |
| SA-CNN-E2E                | 99.60            | 12           |
| DRNN(Jiang et al., 2018)  | 96.53            | 12           |
| HCNN(Wen et al., 2019)    | 99.82            | 10           |
| CNN-Vibration image(Wen et al., 2018) | 99.79         | 10           |
| Sparse filter(Lei et al., 2016) | 99.66         | 10           |
| CNN-PSO-WP(Fuan et al., 2017) | 99.71         | 8            |
| CNN-Wavelet/STFT(Verstraete et al., 2017) | 99.8     | 4            |
| CNN-STFT96×96(Verstraete et al., 2017) | 99.5       | 4            |
| DBN-PSO(Shao et al., 2015) | 87.45           | -            |
preprocessor of STFT is valuable to improve the algorithm performance.

3.1.2. Spectrum image

According to Equations (1–9) and the parameter settings \( f_s = 12000, \alpha = 2, \text{rpm} \approx 1700, D_{ball} = 0.3126, D_{cage} = 1.537, \theta = 0, N_{ball} = 9, \) and \( f_r = 1700/60 \), we can get the characteristic vibration frequencies as follows:

\[
f_{OR} = \frac{9}{2} f_r \left(1 - \frac{0.3126}{1.537} \cos \theta \right) = 3.585 f_r
\]

\[
f_{IR} = \frac{9}{2} f_r \left(1 + \frac{0.3126}{1.537} \cos \theta \right) = 5.415 f_r
\]

\[
f_{ball} = \frac{1.537}{2 \times 0.3126} f_r \left(1 - \left(\frac{0.3126}{1.537} \cos \theta \right)^2 \right) = 2.357 f_r
\]

\[
f_{cage} = \frac{1}{2} f_r \left(1 - \frac{0.3126}{1.537} \cos \theta \right) = 0.3983 f_r
\]

\[
L > 2 \frac{12000}{\text{min}(3.585, 5, 415, 2.357, 0.3983) \times 1700 \over 60} = 2126.69
\]

As Table 3 shows, the spectrum images with STFT according to equations (10-12) are easy to be discriminated by human eyes. Table 4 lists all error results found in the experiment and the corresponding activation maps (Selvaraju et al., 2017). In these activation maps, the maximum coefficients are always mapped to high temperatures in the heat map. In Table 4, 21-OR and 14-IR are the most difficult to be discriminated. Moreover, 14-IR and 14-OR, or 21-Ball and 07-OR are similar in the high temperature area if the location is not considered. The main reason is that: the convolutional units of various layers of CNNs actually behave as object detectors (Zhou et al., 2015), but this ability is lost when fully-connected layers are used for classification (Zhou et al., 2016).

3.1.3. Performance and parameters

As the CNN structure is optimized with PSO, the time complexity needs to be considered carefully. For analysing the performance with different parameters, our method is implemented with different combinations between \( N_{ep} = [30, 50, 70] \) and \( N_{ts} = [500, 1500, 2500] \).

In Tables 5 and 6, the combinations of \( (N_{ep} = 70, N_{ts} = 500) \) and \( (N_{ep} = 30, N_{ts} = 1500) \) can get better prediction accuracies, but their time complexity is not the minimum value. Hence, there is a balance choice between time complexity and prediction accuracy. Of course, the time complexity should be considered only in the beginning of the CNN optimization with PSO, and it will be the same as existing CNN methods once the optimized CNN is fixed.

Furthermore, according to Table 6, \( N_{ts} \) is the main factor that leads to the higher time complexity.

3.1.4. Evolution of CNN with PSO

Figure 8 depicts the curves related with the converge speed, which is proportional to \( N_{ep} \) and \( N_{ts} \). The bottom figure is a partially enlarged figure. The fitness value \(-1\) means that the prediction accuracy is 100% according to

| Signal | Spectrum | Class Name | Signal | Spectrum | Class Name |
|--------|----------|------------|--------|----------|------------|
| 28-IR  |          | 14-OR      | 28-IR  |          | 14-OR      |
| Ball   |          | 14-Ball    | 21-IR  |          | 14-IR      |
|        |          | 21-Ball    | 21-Ball|          | 07-IR      |
| Normal |          | 21-OR      | Normal |          | 07-Ball    |
|        |          |            |        |          | 07-OR      |
Table 4. Class activation mapping.

| Spectrum | Conv 2 | Conv 3 | Dense 1 | Ground Truth | Error |
|----------|--------|--------|---------|--------------|-------|

Table 5. Prediction accuracies with different parameters (DE_time, %).

| N_ep | N_nt | Min   | Max   | Mean  | Std   |
|------|------|-------|-------|-------|-------|
| 30   | 500  | 99.85 | 100.00| 99.96 | 0.05  |
| 50   | 500  | 99.90 | 100.00| 99.98 | 0.04  |
| 70   | 500  | 99.95 | 100.00| 99.99 | 0.02  |
| 30   | 1500 | 99.95 | 100.00| 99.99 | 0.02  |
| 30   | 2500 | 99.90 | 100.00| 99.98 | 0.03  |

Table 6. Time complex with CWRU dataset (DE_time).

| N_nt (N_ep = 30) | Time (s)  | N_nt (N_ep = 500) | Time (s)  |
|-----------------|-----------|-------------------|-----------|
| 500             | 2608.039835 | 30 | 2608.039835 |
| 1500            | 11780.00252  | 50 | 3484.697968 |
| 2500            | 24919.51808  | 70 | 8972.817968 |

Figure 8. Fitness curves (bottom: partially enlarged figure).

Equation (20). After the fitness value reaches to $-1$, the converging speed becomes slowly because of the huge searching space. Furthermore, the optimization of CNN is a multi-optimization problem as the CNN structure from different running results is different (Figure 9).

Figure 9. Network structure.

Figure 10. Fitness curve.

3.1.5. Effect of STFT

As Figure 10 shown, with the preprocessing of STFT, the fitness can be quickly converged to $-1.0$. Therefore, signal processing taken as a preprocessing method to extract basic time–frequency information can accelerate the evolution of PSO and improve the algorithm performance.

3.2. MFPT bearing dataset

This MFPT dataset is recorded from a bearing test rig (normal bearing data, an outer race fault under various loads,
and an inner race fault under various loads), and includes four sets of bearing vibration data: (1) a baseline set under the load of 270 lbs, which is sampled at 97656 Hz for 6 s; (2) an outer race fault set including three outer race faults (ThreeOR) under the load of 270 lbs, which are sampled at 97656 Hz for 6 s; (3) seven additional outer race faults (SevenOR) under various loads of 25, 50, 100, 150, 200, 250, and 300 lbs, which are sampled at 48828 Hz for 3 s; and (4) seven inner race faults (SevenIR) under various loads of 25, 50, 100, 150, 200, 250, and 300 lbs, which are sampled at 48828 Hz for 3 s. In the experiments, 10000 data samples are randomly extracted from the raw data file with the same probability.

### 3.2.1. Spectrum image

According to Equations (1–9) and the parameter settings ($f_s = 97656/6$, $\alpha = 2$, $D_{ball} = 0.235$, $D_{cage} = 1.245$, $\theta = 0$, $N_{ball} = 8$, and $f_r = 25$), we can get the characteristic vibration frequencies as follows:

\[
\begin{align*}
    f_{OR} &= \frac{8}{2} f_r \left(1 - \frac{0.235}{1.245} \cos \theta \right) = 3.245 f_r \\
    f_{IR} &= \frac{8}{2} f_r \left(1 + \frac{0.235}{1.245} \cos \theta \right) = 4.755 f_r \\
    f_{ball} &= 1.245 \frac{2}{2 \times 0.235} f_r \left(1 - \left(\frac{0.235}{1.245}\right)^2 \cos^2 \theta \right) = 2.555 f_r \\
    f_{cage} &= \frac{1}{2} f_r \left(1 - \frac{0.235}{1.245} \cos \theta \right) = 0.406 f_r \\
\end{align*}
\]

If $L > 2 \times \frac{97656}{\min(3.245,4.755,2.555,0.406) \times 25} = 3207.09$

If $W > 2 \times \frac{97656}{\max(3.245,4.755,2.555,0.406) \times 25} = 273.83$

$L = 3208, W = 274, H = L/W = 12$

Table 7 gives the signals and spectrum images used by our method according to Equations (10–12). It seems that the spectrum imaged can be readily discriminated by human eyes.

### Table 7. Signals and spectra.

| Signal   | Spectrum | Class Name | 4-class | 3-class |
|----------|----------|------------|---------|---------|
|          |          | SevenIR    | IR      |         |
|          |          | SevenOR    | OR      |         |
|          |          | ThreeOR    |         |         |
|          |          | Baseline   | Baseline|         |
Figure 11. Comparison of network structure.

Figure 12. Comparison of network structure (Left: prediction accuracy; right: standard deviation).

In Figure 11, the left figures (no any constraint) have more numbers near zero, and the network structure converges to simpler network structure with less neuron nodes because the fitness function considers the simplicity of network structure in Equation (19). Consequently, the simple network structure may lead to worse performance in terms of both prediction accuracy and its standard deviation, as shown in Figure 12.

4. Conclusions

In summary, a self-adaptive CNN method (SA-CNN) with two-step training is proposed on account of the time-consuming problem in the design of a CNN structure or its parameter tuning. To reduce the time complexity, two main mechanisms are used. First, the CNN structure is fixed, and only its key parameters including the kernel number and the number of neuron nodes are optimized by PSO. Second, a small number of training samples and a
small epoch number are used to evaluate the prediction accuracy. Based on the given the bearing parameters, SA-CNN can automatically decide the parameters of W and H used in STFT, and adjust CNN into an optimized structure. Therefore, the proposed method can be easily applied in industry. Through the experiments with two famous public datasets, CWRU and MFPT, it is demonstrated that the SA-CNN method can achieve better results than other state-of-art methods and the preprocessing of STFT can accelerate the evolution speed of PSO. In the future, the location of high temperature area in the rotating machinery will be considered to improve the prediction accuracy; meanwhile, it is necessary to take both fault patterns and fault deterioration into account.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China (61502423), Zhejiang Provincial Natural Science Foundation (LY16G020012), and Zhejiang Province Public Welfare Technology Application Research Project (grant number LGF19F010002, LGF20F010004, LGG21F030014: The Research on Cold-Start Fault Diagnosis Method Based on Immune Algorithm and Deep Learning).

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported in part by the National Natural Science Foundation of China (61502423), Zhejiang Provincial Natural Science Foundation (grant number LY16G020012), and Zhejiang Province Public Welfare Technology Application Research Project (grant numbers LGF19F010002, LGF20F010004, LGG21F030014).

References

Aimer, A. F., Boudinar, A. H., Benouzza, N., et al. (2019). Bearing fault diagnosis of a PWM inverter fed-induction motor using an Improved short time Fourier Transform. Journal of Electrical Engineering & Technology, 1–10. https://doi.org/10.1007/s42835-019-00096-y

Bechhoefer, E. (2016). A quick introduction to bearing envelope analysis. MFPT Data, http://www.mpft.org/FaultData/Fault-Data.htm.Set

Boudinar, A. H., Benouzza, N., Bendiabdellah, A., & Khodja, M.-E.-A. (2016). Induction motor bearing fault analysis using a root-MUSIC method. IEEE Transactions on Industry Applications, 52(5), 3851–3860. https://doi.org/10.1109/TIA.2016.2581143

Chen, Z., & Li, W. (2017). Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network. IEEE Transactions on Instrumentation and Measurement, 66(7), 1693–1702. https://doi.org/10.1109/TIM.2017.2669947

Chen, F., Tang, B., Song, T., & Li, L. (2014). Multi-fault diagnosis study on roller bearing based on multi-kernel support vector machine with chaotic particle swarm optimization. Measurement, 47, 576–590. https://doi.org/10.1016/j.measurement.2013.08.021

El-Thalji, I., & Jantunen, E. (2015). Fault analysis of the wear fault development in rolling bearings. Engineering Failure Analysis, 57, 470–482. https://doi.org/10.1016/j.engfailanal.2015.08.013

Eren, L. (2017). Bearing fault detection by one-dimensional convolutional neural networks. Mathematical Problems in Engineering. https://doi.org/10.1155/2017/8617315

Fuan, W., Hongkai, J., Haidong, S., Wenjing, D., & Shuaipeng, W. (2017). An adaptive deep convolutional neural network for rolling bearing fault diagnosis. Measurement Science and Technology, 28(9), 095005. https://doi.org/10.1088/1361-6501/aa6e22

Gao, Z., Ding, S. X., & Cecati, C. (2015). Real-time fault diagnosis and fault-tolerant control. IEEE Transactions on Industrial Electronics, 62(6), 3752–3756. https://doi.org/10.1109/TIE.2015.2417511

Hamadache, M., Lee, D., & Veluvolu, K. C. (2015). Rotor speed-based bearing fault diagnosis (RSB-BFD) under variable speed and constant load. IEEE Transactions on Industrial Electronics, 62(10), 6486–6495. https://doi.org/10.1109/TIE.2015.2416673

Harmouche, J., Delpha, C., & Diallo, D. (2015). Improved fault diagnosis of ball bearings based on the global spectrum of vibration signals. IEEE Transactions on Energy Conversion, 30(1), 376–383. https://doi.org/10.1109/TEC.2014.2341620

Ho, D., & Randall, R. B. (2000). Optimisation of bearing diagnostic techniques using simulated and actual bearing fault signals. Mechanical Systems and Signal Processing, 14(5), 763–788. https://doi.org/10.1016/mssp.2000.1304

Hoang, D. T., & Kang, H. J. (2019). Rolling element bearing fault diagnosis using convolutional neural network and vibration image. Cognitive Systems Research, 53, 42–50. https://doi.org/10.1016/j.cogsys.2018.03.002

Huo, Z., Zhang, Y., Francq, P., Shu, L., & Huang, J. (2017). Incipient fault diagnosis of roller bearing using optimized wavelet transform based multi-speed vibration signatures. IEEE Access, 5, 19442–19456. https://doi.org/10.1109/ACCESS.2017.2661967

Jia, F., Lei, Y., Lin, J., Zhou, X., & Lu, N. (2016). Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data. Mechanical Systems and Signal Processing, 72-73, 303–315. https://doi.org/10.1016/j.mssp.2015.10.025

Jiang, H., Li, X., Shao, H., & Zhao, K. (2018). Intelligent fault diagnosis of rolling bearings using an improved deep recurrent neural network. Measurement Science and Technology, 29(6), 065107. https://doi.org/10.1088/1361-6501/aab945

Jung, U., & Koh, B.-H. (2015). Wavelet energy-based visualization and classification of high-dimensional signal for bearing fault detection. Knowledge and Information Systems, 44(1), 197–215. https://doi.org/10.1007/s10115-014-0761-z

Kennedy, J., & Eberhart, R. C. (1997). A discrete binary version of the particle swarm optimisation algorithm. In James M. Tien (Ed.), Proceedings of the IEEE International Conference on neural networks (pp. 4104–4108). IEE.
Kumar, A., & Kumar, R. (2019). Role of signal processing, modeling and decision making in the diagnosis of rolling element bearing defect: A review. *Journal of Nondestructive Evaluation*, 38(1). https://doi.org/10.1007/s10921-018-0543-8

Lei, Y., Jia, F., Lin, J., Xing, S., & Ding, S. X. (2016). An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data. *IEEE Transactions on Industrial Electronics*, 63(5), 3137–3147. https://doi.org/10.1109/TIE.2016.2519325

Lin, H. C., & Ye, Y. C. (2019). Reviews of bearing vibration measurement using fast Fourier transform and enhanced fast Fourier transform algorithms. *Advances in Mechanical Engineering*, 11(1). https://doi.org/10.1177/1687814018816751

Liu, H., Li, L., & Ma, J. (2016). Rolling bearing fault diagnosis based on STFT-deep learning and sound signals. *Shock and Vibration*. https://doi.org/10.1155/2016/6127479.

Loparo, K. A., & Loparo, K. A. (2013). Fault diagnosis method using unsupervised feature learning towards mechanical big data. *IEEE Transactions on Industrial Informatics*, 26(11). https://doi.org/10.1109/TII.2013.2271979

Shao, H., Jiang, H., Zhang, X., & Niu, M. (2015). Rolling bearing fault diagnosis based on the MVMD and Teager energy operator. *ISA Transactions*, 80, 297–311. https://doi.org/10.1016/j.isatra.2015.12.037

Soualhi, A., Medjaher, K., & Zerhouni, N. (2015). Bearing health monitoring based on Hilbert–Huang transform, support vector machine, and regression. *IEEE Transactions on Instrumentation and Measurement*, 64(1), 52–62. https://doi.org/10.1109/TIM.2014.2330494

Sovilu, A., Medjaher, K., & Zerhouni, N. (2015). Bearing health monitoring based on Hilbert–Huang transform, support vector machine, and regression. *IEEE Transactions on Instrumentation and Measurement*, 64(1), 52–62. https://doi.org/10.1109/TIM.2014.2330494

Verstraete, D., Ferrada, A., Droguett, E. L., et al. (2017). Deep learning enabled fault diagnosis using time-frequency image analysis of rolling element bearings. *Shock and Vibration*. https://doi.org/10.1155/2017/5067651

Wang, D. (2016). K-nearest neighbors based methods for identification of different gear crack levels under different motor speeds and loads: Revisited. *Mechanical Systems and Signal Processing*, 70-71, 201–208. https://doi.org/10.1016/j.ymssp.2015.10.007

Wang, X. B., Yang, Z. X., & Yan, X. A. (2018). Novel particle swarm optimization-based variational mode decomposition method for the fault diagnosis of complex rotating machinery. *IEEE/ASME Transactions on Mechatronics*, 23(1), 68–79. doi: 10.1109/ASME.2017.7991037

Wang, J., Zhuang, J., Duan, L., & Cheng, W. (2016). A multi-scale convolution neural network for featureless fault diagnosis. *Proceedings of the International Symposium on Flexible Automation*, 65–70. doi: 10.1109/ISFA.2016.7991037

Yuan, Z., Zhang, L., & Duan, L. (2018). A novel fusion diagnosis method for rotor system fault based on deep learning and multi-sourced heterogeneous monitoring data. *Measurement Science and Technology*, 29(11). https://doi.org/10.1088/1361-6501/aad101

Yi, C., Lv, Y., & Dang, Z. (2016). A fault diagnosis scheme for rolling bearing based on particle swarm optimization in variational mode decomposition. *Shock and Vibration*. https://doi.org/10.1155/2016/9372691

Zhang, X., Liang, Y., Zhou, J., & Zang, Y. (2015). A novel bearing fault diagnosis model integrated permutation entropy, ensemble empirical mode decomposition and optimized SVM. *Measurement*, 69, 164–179. https://doi.org/10.1016/j.measurement.2015.03.017

Zhou, B., Khosla, A., Lapedriza, A., et al. (2016). Learning deep representations of different gear crack levels under different motor speeds and loads: Revisited. *Mechanical Systems and Signal Processing*, 70-71, 201–208. https://doi.org/10.1016/j.ymssp.2015.10.007

Zhu, B., Khosla, A., Lapedriza, A., et al. (2016). Learning deep features for discriminative localization. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Zhi, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2015). Object detectors emerge in deep scene cnns. *International Conference on Learning Representations*. Ziani, R., Felkaoui, A., & Zegadi, R. (2017). Bearing fault diagnosis using multiclass support vector machines with binary particle swarm optimization and regularized Fisher’s criterion. *Journal of Intelligent Manufacturing*, 28(2), 405–417. https://doi.org/10.1007/s10845-014-0987-3