Article

Bearing Fault Diagnosis Based on Improved Convolutional Deep Belief Network

Shuangjie Liu ¹, Jiaqi Xie ¹, Changqing Shen ^{1,*} , Xiaofeng Shang ², Dong Wang ³
and Zhongkui Zhu ¹

¹ School of Rail Transportation, Soochow University, Suzhou 215000, China; sjliu12@stu.suda.edu.cn (S.L.); jqxie@stu.suda.edu.cn (J.X.); zhuzhongkui@suda.edu.cn (Z.Z.)

² Wuxi Metro Group Co, Ltd., Wuxi 214000, China; shangxf@wxmetro.net

³ State Key Laboratory of Mechanical System and Vibration, Shanghai Jiao Tong University, Shanghai 200000, China; dongwang4-c@sjtu.edu.cn

* Correspondence: cqshen@suda.edu.cn

Received: 29 July 2020; Accepted: 9 September 2020; Published: 12 September 2020

Abstract: Mechanical equipment fault detection is critical in industrial applications. Based on vibration signal processing and analysis, the traditional fault diagnosis method relies on rich professional knowledge and artificial experience. Achieving accurate feature extraction and fault diagnosis is difficult using such an approach. To learn the characteristics of features from data automatically, a deep learning method is used. A qualitative and quantitative method for rolling bearing faults diagnosis based on an improved convolutional deep belief network (CDBN) is proposed in this study. First, the original vibration signal is converted to the frequency signal with the fast Fourier transform to improve shallow inputs. Second, the Adam optimizer is introduced to accelerate model training and convergence speed. Finally, the model structure is optimized. A multi-layer feature fusion learning structure is put forward wherein the characterization capabilities of each layer can be fully used to improve the generalization ability of the model. In the experimental verification, a laboratory self-made bearing vibration signal dataset was used. The dataset included healthy bearings, nine single faults of different types and sizes, and three different types of composite fault signals. The results of load 0 kN and 1 kN both indicate that the proposed model has better diagnostic accuracy, with an average of 98.15% and 96.15%, compared with the traditional stacked autoencoder, artificial neural network, deep belief network, and standard CDBN. With improved diagnostic accuracy, the proposed model realizes reliable and effective qualitative and quantitative diagnosis of bearing faults.

Keywords: mechanical fault diagnosis; bearing; feature extraction; convolutional deep belief network

1. Introduction

With the explosive progress of modern science and industry, machinery and equipment in fields such as aerospace, rail, and wind power are becoming faster, more automated, and meticulous than before. However, increasingly complex operating conditions inevitably lead to failures. Thus, monitoring and diagnosing the health of rotating machinery and equipment to ensure operational safety and reliability are immensely necessary and urgent; fortunately, intelligent diagnosis approaches, the combination of classification algorithms, and signal processing techniques have produced promising results [1].

Based on vibration signal processing, the traditional fault diagnosis methods extract fault components from the initial noise signal by using rich professional knowledge. At present, signal processing methods, such as empirical mode decomposition (EMD) [2], wavelet packet transform [3],

and morphological filter [4], are commonly used for time-frequency analysis. For separating bearing fault ingredients in a wind turbine gearbox, Dong et al. [5] applied improved convolutional neural network with anti-interference to rolling bearing performance degradation assessment. Gong et al. [6] proposed a novel deep learning method called improved convolutional neural networks and support vector machines with data fusion for intelligent fault diagnosis. To realize bearing fault detection, Shi et al. [7] formed a quantity matrix of bearing vibration signal features on the basis of EMD and local mean decomposition. In recent years, machine learning methods, such as support vector machines [8] and artificial neural networks (ANNs) [9], have been gradually introduced into mechanical fault diagnosis. Such an introduction has enabled the minimization of human intervention and reliance on professional skills for development toward an intelligent direction in this field.

As an urgent field in machine learning research, deep learning has begun to appear in bearing failure in recent years, including stacked autoencoder (SAE) [10], deep belief networks (DBNs) [11], and convolutional neural networks (CNNs) [12]. Zhao et al. [13] summarized the emerging research on machine health monitoring based on deep learning, then discussed new trends in machine learning monitoring methods of deep learning. In considering ways of enhancing the robust features for rotating machinery, Shen et al. [14] proposed an automatic learning method with robust features based on contractive autoencoder. Chen et al. [12] proposed cyclic spectral coherence and convolutional neural networks for bearing fault diagnosis. Shao et al. [15] compressed the data with autoencoders and constructed a convolutional deep belief network (CDBN) to diagnose bearing faults. In view of compressed sensing and deep learning theory, Wen [16] studied a fault diagnosis method of bearing vibration signals that can extract features automatically.

In summary, traditional methods have several shortcomings, as follows:

1. Traditional vibration signal processing and analysis methods rely on certain professional skills;
2. Existing shallow machine learning methods rely on the accuracy of manual feature extraction;
3. Improper selection of parameters for standard deep learning models can easily result in failure to effectively converge, thus, diagnostic accuracy is difficult to guarantee;
4. Existing research on the quantitative diagnosis of bearing fault is relatively inadequate compared with that on qualitative diagnosis.

It can be concluded that the general deep models have challenges: (1) a suitable signal preprocessing method is needed to enhance features; (2) necessary measures need to be taken during the model training process to make it more stable; (3) the models should determine not only the fault type, but also the fault degree. In response to these problems, a CDBN has the superiority of fast calculations and feature extraction. Based on the standard CDBN model, this study introduces Adam's optimization algorithm and optimizes the structure of the CDBN. An improved CDBN model is proposed to enhance diagnosis accuracy. Unlike image processing, in the case of fault diagnosis, the raw data will probably be noisy. The feature extraction capability of a standard CDBN is insufficient, which is the reason of conducting band-pass filter to pre-process the data. The main constructions of the proposed method are listed as follows:

1. A band pass filter has been introduced to the preprocessing step to filter out noise;
2. Qualitative and quantitative diagnoses of bearing faults can be effectively implemented;
3. Both single fault and compound faults can be effectively identified;
4. A comparative experiment under different operation loads further confirmed the reliability of the model.

The rest of the paper is organized as follows. Section 2 contains an introduction of the theoretical background of a CDBN and the structure of the proposed method. The experimental results are discussed in Section 3. Finally, in Section 4, conclusions are summarized.

2. Theoretical Background and Proposed Method

2.1. Restricted Boltzmann Machine

DBN is composed of several restricted Boltzmann machines (RBMs). The definition of the RBM model is based on the energy function. It is a basic component of DBNs and a key preprocessing unit in deep learning [17]. RBM is a basic single-layer machine learning network and has extensive applications. The formulas involved in the following text refer to the study of Chen and Li [17]. As shown in Figure 1, an RBM structure has two layers, namely, visible and hidden layers. A connection exists between each neuron node in the adjacent layer, and in each layer, the neuron nodes remain in a connectionless state with each other. The connection weight between two layers is W . RBM neurons are Booleans, implying that only two states exist, 0 and 1. State 1 indicates the activation of a neuron, and state 0 indicates suppression. In a way, for a given visible and hidden layer neuron in a binary state, the energy function of the RBM is as follows:

$$E(v, h; \theta) = \sum_{i=1}^n a_i v_i - \sum_{i=1}^n \sum_{j=1}^m w_{ij} v_i h_j - \sum_{j=1}^m b_j h_j \quad (1)$$

where $\theta = \{W_{ij}, a_i, b_j\}$ is the parameter set of RBM, w_{ij} denotes the weight of the i th cell in the visible layer and the j th cell in the hidden layer, a_i denotes the offset of the i th unit of the visible layer, b_j denotes the offset of the j th cell of the hidden layer, v_i is the state of the i th unit in the visible layer, and h_j is the state of the j th unit in the hidden layer.

Figure 1. Structure of restricted Boltzmann machine.

After the parameters are determined, the joint probability distribution of (v, h) is exported by Formula (1), where $Z(\theta)$ represents the normalized constant.

$$P(v, h; \theta) = \frac{1}{Z(\theta)} e^{-E(v, h; \theta)} \quad (2)$$

$$Z(\theta) = \sum_{v, h} e^{-E(v, h; \theta)} \quad (3)$$

From the joint probability distribution, the edge probability and the conditional probability of visible and hidden neurons are driven by Formulas (4) and (5), respectively:

$$P(v; \theta) = \frac{\sum_h e^{-E(v,h;\theta)}}{\sum_v \sum_h e^{-E(v,h;\theta)}} \quad (4)$$

$$P(v|h; \theta) = \frac{e^{-E(v,h;\theta)}}{\sum_v e^{-E(v,h;\theta)}}$$

$$P(h; \theta) = \frac{\sum_v e^{-E(v,h;\theta)}}{\sum_v \sum_h e^{-E(v,h;\theta)}} \quad (5)$$

$$P(h|v; \theta) = \frac{e^{-E(v,h;\theta)}}{\sum_h e^{-E(v,h;\theta)}}$$

Given a visible neuron and a hidden neuron state, the probability of activating the j th hidden neuron and the i th visible neuron are respectively provided as follows, where $\sigma = 1/(1 + e^{-x})$:

$$p(h_j = 1|v; \theta) = \frac{\sum_{h_{kj}} p(h_j=1, h_{kj}, v; \theta)}{\sum_h p(h, v; \theta)} \quad (6)$$

$$= \sigma(\sum_{i=1}^{n_h} w_{ij} v_i + b_j)$$

$$p(v_i = 1|h; \theta) = \frac{\sum_{v_{ki}} p(v_i=1, v_{ki}, h; \theta)}{\sum_v p(v, h; \theta)} \quad (7)$$

$$= \sigma(\sum_{j=1}^m w_{ij} h_j + a_i)$$

The contrast divergence algorithm [18] is used in RBM during model training. The task of training RBM is to find the optimal value of parameter θ so that the edge probability of the visible neurons in the distribution represented by RBM is maximized; that is, the goal is to maximize the log-likelihood function.

$$\theta^* = \operatorname{argmax}_{\theta} \sum_v \ln P(v; \theta) \quad (8)$$

2.2. Convolutional Deep Brief Network and Its Improvement

2.2.1. Convolutional Restricted Boltzmann Machine

Convolutional RBM (CRBM) is an improvement on the basis of the original RBM, and the structure of CRBM is similar to that of RBM. CRBM is an improved model composed of two random variable matrix layers, namely, the visible and hidden layers. An image with salient features of the local receptive field and weight sharing is regarded as the input layer of CRBM. The hidden layers are locally connected with visible layers, and their weights are shared by convolution.

The CRBM model, as shown in Figure 2, comprises the view layer V, the hidden layer H, and the pooling layer P, for a total of three layers. We assume that the size of the input layer matrix is $N_v \times N_v$, the number of groups of matrices in the hidden layer is k , each group is a binary array with the size of $N_H \times N_H$, and $N_H^2 K$ hidden layer units are present. Each group of hidden layers is associated with an $N_w \times N_w$ size filter.

Figure 3 shows the process of obtaining the hidden layer from the visible layer. The size of the convolution kernel is 3×3 , and the hidden layer units are divided into k sub-matrices. $W_1, W_2, \dots, W_K$ connects the visible and hidden layers. Each hidden layer unit represents a specific feature extracted from a neighborhood unit of a visible layer. Moreover, b_k indicates the value of each hidden layer unit bias, and C is the bias globally shared by all visible units. We obtain the energy number of CRBM as Equation (9):

$$E(v, h) = -\sum_{k=1}^k h^k (w^k * v) - \sum_{k=1}^k b_k \sum_{i,j} h_{ij}^k - c \sum_{i,j} v_{ij} \quad (9)$$

where $*$ indicates convolution, v_{ij} denotes the input values of the i th visible layer unit and the j th hidden layer unit, h^k denotes the k th hidden layer, h_{ij}^k denotes the values of the i th visible layer unit and the j th unit of the k th hidden layer, and w^k denotes the convolution kernel of the k th hidden unit.

As with standard RBM, the conditional probability distribution is given as follows:

$$p(h_j^k = 1|v) = \sigma((\tilde{W}^k * v)_j + b_k) \quad (10)$$

$$p(v_i = 1|h) = \sigma((W^k * h^k)_i + c) \quad (11)$$

where $\sigma = 1/(1 + e^{-x})$, $\tilde{W}_j^k \triangleq W_{N_w-j+1}^k$.

A CRBM is a single-layered network. It can be considered as the basis of a CDBN. Stacking multiple CRBMs, with the hidden layer of the previous RBM as the visible layer of the next CRBM, constitutes a CDBN. In each training, the RBM of the lowest layer is trained, one layer at a time, until the top layer.

Figure 2. Structure diagram of a convolutional restricted Boltzmann machine (CRBM).

Figure 3. The process of obtaining hidden layers from the visible layer based on CRBM.

2.2.2. CDBN and Its Improvement

Proposed by Lee in 2009, a CDBN is a network model that consists of a CRBM. Multiple CRBMs are connected to form a CDBN. The outcome of the previous CRBM layer is regarded as the input of its subsequent layer. The model fitting ability is further improved by multi-layer linking.

The objective function of a CDBN is to maximize the log-likelihood function $\theta(\cdot)$ in order to get the optimal parameters, and reconstruction error is used to evaluate the model. The reconstruction error refers to the difference between the original data after Gibbs sampling with the training sample as the initial state and the distribution of the RBM model. A smaller reconstruction error indicates better training. Reconstruction error is given by Equation (12), with $\hat{X}_i$ and X_i indicating the real output and ideal output:

$$Error = \frac{1}{N} \sum_{i=1}^N (\hat{X}_i(W^k, b^k) - X_i)^2 \quad (12)$$

Based on the standard CDBN model, an improved structure is proposed. The standard CDBN uses only single-layer output features but ignores the comprehensive utilization of the features of each layer. As a consequence, the classification results are not representative. An improvement of the model connection mode is proposed in this study. The outputs of a two-layer CRBM are combined into a vector and input to the softmax classifier by multi-level feature fusion to utilize features, further improving classification accuracy.

Model training can be summarized according to the following steps:

1. Forward propagation:
  - (a) Use CD algorithm to pre-train W and b and determine the opening and closing of the corresponding hidden element;
  - (b) Propagate upward layer by layer, calculate the excitation value of each hidden element, and use the sigmoid function to complete the standardization;
2. Backpropagation:
  - (a) Use the minimum mean square error criterion for the backward error propagation algorithm and update the parameters of the network.
  - (b) Update the weight and bias of the network with Adam optimizer.

2.3. Band-Pass Filter for Signal Preprocessing

A band-pass filter is a commonly used data processing method to filter out clutter. A band-pass filter can retain frequency components in a certain frequency range [19] while filtering out those in other ranges. Its bandwidth is determined by parameter λ , following positive correlation.

Assuming that the input image size is $N \times N$, its passband bandwidth is given by Equation (13) [19]:

$$f_0 = \lambda N \quad (13)$$

Therefore, the filter function is given by Equation (14), with x and y denoting the angular frequency of a two-dimensional filter [20]:

$$filter(x, y) = \rho(x, y)e^{-(\rho(x, y)/f_0)^4} \quad (14)$$

where $\rho(x, y) = \sqrt{x^2 + y^2}$.

Figure 4 shows the spectrogram of a band-pass filter. It can effectively remove high-frequency noise and low frequency-connected domain parts in the image, retaining the texture features in the image and increasing the suitability of the samples for CNNs to process.

Figure 4. The spectrogram of a band-pass filter.

2.4. Fault Diagnosis Model Based on Improved CDBN

Based on the CDBN model in 2.2.2 2.2, we further improved the model according to two aspects: changes to the preprocessing stage and the introduction of an optimizer. First, in the preprocessing stage, we introduced a band-pass filter to process the original samples after fast Fourier transform (FFT) and folding. The center frequency of the band-pass filter is set to 50% of the bandwidth, and the passband width is set to 40% of the bandwidth. The spectrogram of the band-pass filter is shown in Figure 4. Second, during the model training phase, the Adam optimizer [20] is introduced, enabling the model to adjust to different learning rates for each parameter. For frequently changing parameters, smaller steps are applied, whereas larger steps are better for sparse parameters. This algorithm is used to further improve the model's convergence speed and to reduce the number of model errors.

Figure 5 is the flowchart of the improved CDBN-based bearing fault diagnosis model proposed in this study. The input vibration signal is at a length of 1024. After FFT, it is normalized to between 0 and signals of different fault types and degrees of failure. Moreover, band-pass filtering is conducted to enhance texture features and reduce the noise effect. The CDBN model is a combination of two CRBM layers. For the first CRBM layer, the convolution kernel size is 7×7 , the number of link weight matrices is 9, each size matches the size of the convolution kernel, the pooling kernel size is 2×2 , and the maximum pooling method is used. The output of the first layer with a size of 13×13 is a feature map, and it is used as the input of the second CRBM layer. The second hidden layer adopts a size of 9×9 with 16 link weight matrices. The size of the second convolution kernel is 5×5 , and the size of the second pooling kernel is 2×2 . The proposed model also uses the maximum pooling method. At the end of the model, softmax is used for classification. The extracted features are mapped to 13 different types, divided according to fault types or degrees. Unlike the standard CDBN, before classification, the outputs of the first layer and second layer use feature fusion to further increase the diagnostic accuracy. Table 1 displays the specific parameters of the network.

Figure 5. Bearing fault diagnosis model based on an improved convolutional deep belief network (CDBN). FFT; fast Fourier transform.

Table 1. Specific parameters of the network.

Learning Rate	Batch Size	Epoch	Number of Layer
0.001	200	20	2
Size of Convolution Kernel in the 1st Layer	Size of Convolution Kernel in the 2nd Layer	Size of Pooling Kernel in the 2nd Layer	Pooling Method
7×7	5×5	2×2	Maximum pooling

3. Experimental Validation

3.1. Dataset Description

To confirm the performance of the proposed model in bearing fault diagnosis, a laboratory-made bearing fault test platform was used to collect multiple faults and vibration signals at various levels.

As shown in Figure 6, the test platform included variable frequency motors, normal bearings, test bearings, loading systems, and acceleration sensors. The test bearing model was 6205-2RS SKF, and specific bearing parameters are listed in Table 2. Single-point or compound faults were set on the bearing surface by wire cutting. Figure 7 shows the real pictures of four bearings with four different health conditions. Types of faults included outer-race faults, ball faults, inner-race faults, inner-race compound ball faults, outer-race compound ball faults, and inner-race compound outer-race faults (IO). In this table, IB indicates that inner-race fault and ball faults exist at the same time on the test bearing, both under a fault degree of 0.2 mm. In the same way, IO indicates that inner-race fault and outer-race fault exist at the same time, and OB indicates that outer-race fault and ball fault exist at the same time. Specific data sets are described in Tables 3 and 4; dataset1 was collected under 0 kN load and dataset2 was collected under 1 kN load. The motor speed was set to 961 rpm, signal sampling frequency was set as 10 kHz, and the entire experiment was under a load of zero condition.

Figure 6. Bearing fault diagnosis model based on an improved CDBN.

Figure 7. The real pictures of four bearings with four different health conditions.

Table 2. Specifications of the test bearings.

Contact Angle	Roller Diameter	Number of Rollers
0°	7.938 mm	9
Outer Diameter	Internal Diameter	Pitch Diameter
52 mm	25 mm	38.5 mm

Table 3. Description of compound faults for dataset1. IB: inner-race fault and ball faults exist at the same time on the test bearing; IO: inner-race fault and outer-race fault exist at the same time; OB: outer-race fault and ball fault exist at the same time.

Fault Type	Load	Fault Degree (mm)	Train Sample	Test Sample	Label
Normal		\	200	100	1
		0.2	200	100	2
Outer race		0.3	200	100	3
		0.6	200	100	4
		0.2	200	100	5
Ball fault	0 kN	0.3	200	100	6
		0.6	200	100	7
		0.2	200	100	8
Inner race		0.3	200	100	9
		0.6	200	100	10
IB		0.2	200	100	11
IO		0.2	200	100	12
OB		0.2	200	100	13

Table 4. Description of compound faults for dataset2.

Fault Type	Load	Fault Degree (mm)	Train Sample	Test Sample	Label
Normal		\	200	100	1
		0.2	200	100	2
Outer race		0.3	200	100	3
		0.6	200	100	4
		0.2	200	100	5
Ball fault	1 kN	0.3	200	100	6
		0.6	200	100	7
		0.2	200	100	8
Inner race		0.3	200	100	9
		0.6	200	100	10
IB		0.2	200	100	11
IO		0.2	200	100	12
OB		0.2	200	100	13

The original vibration signal was intercepted into samples with a length of 1024. From each label, 300 samples were obtained. Two-thirds of the total samples were used as training data, and the remaining third was used as testing data.

The raw data of the bearing vibration signals is shown in Figure 8. It can be observed that the signal is noisy and lacks key information. The first step of the data preprocessing is to fold it into the size of 32×32 by FFT. Figure 9 shows the state of 13 sample types after FFT processing and folding. The second step of data preprocessing is filtering and normalization, for the purpose of filtering out noise. Figure 10 shows the samples after band-pass filtering preprocessing. The figures clearly show that the noise was reduced without weakening the fault signal. By contrast, the texture features of the samples were enhanced, which is more conducive for convolutional neural networks to extract features.

Figure 8. Raw data of bearing vibration signals of different labels.

Figure 9. Visualization of the data after FFT.

Figure 10. Visualization of data after filtering and normalization. The main purpose of this step is to filter out noise.

3.2. Diagnosis Results and Comparative Analysis

The comparison experiment was split into two parts. The proposed model was first compared with SAE, ANN, and DBN. Then, it was compared with standard CDBN.

3.2.1. Comparison with SAE, ANN, and DBN

The test results of the proposed model are shown in Figure 11b,d. The proposed model can achieve 100% accuracy for most types and degrees of faults. By contrast, misclassifications occur in the case of detecting an IO fault type with a small degree, as a result of the features of the IO fault type are less obvious than those of other types of faults. Operating on the collected experimental data set, the proposed improved CDBN model generally achieved an accuracy of 98.15% with 0 kN load and 96.15% with 1 kN load.

Figure 11. (a) Testing result of dataset1 using the standard CDBN. (b) Testing result of dataset1 using the improved CDBN. (c) Testing result of dataset2 using the standard CDBN. (d) Testing result of dataset2 using the improved CDBN.

Tables 5 and 6 illustrate the comparison results of the proposed improved CDBN model with SAE, ANN, and DBN. The length of the input signal was 1024. A single hidden layer with 100 hidden neurons was used in SAE; ANN also had the same structure as SAE; DBN had two hidden layers, whose first and second hidden layers each have 100 neurons. The listed results indicate that the proposed model had an obvious advantage over the other models. When the detection performance of the other models was considerably decreased, the proposed model was still able to maintain a high detection accuracy.

Table 5. Comparison of testing results of dataset1 under load 0 kN by different models.

Label Model	1	2	3	4	5	6	7	8	9	10	11	12	13	Average Accuracy %
SAE	100	83	33	0	20	82	100	8	100	100	100	100	100	71.23
ANN	100	96	76	59	0	61	100	100	100	100	100	0	100	76.31
DBN	100	100	97	98	90	87	96	79	91	100	100	65	95	92.15
Proposed model	100	100	100	100	100	98	96	96	100	100	100	86	100	98.15

Table 6. Comparison of testing results of dataset2 under load 1 kN by different models.

Label Model	1	2	3	4	5	6	7	8	9	10	11	12	13	Average Accuracy %
SAE	100	100	100	100	95	99	40	100	0	100	100	100	100	87.23
ANN	90	74	79	38	0	50	100	98	100	20	100	0	100	65.31
DBN	0	96	95	90	96	89	43	91	71	94	77	85	100	79.00
Proposed model	98	100	100	100	73	99	90	100	100	100	94	86	100	96.15

3.2.2. Comparison with Standard CDBN

The test results of the standard CDBN are illustrated in Figure 11a,c. The figure shows that the improved CDBN was able to achieve higher accuracy than the standard CDBN, especially in outer-race faults and ball faults. However, for IO fault types, namely, label 12, the classification accuracies of both models had no obvious optimization because of the difficulty in catching distinguishable features. In general, the proposed improved CDBN model had better performance compared with the standard CDBN model.

Figure 12 reveals the comparison of the second-layer reconstruction error between the standard and the improved CDBN models. Take dataset1 for example. During training, the reconstruction errors of both models decreased smoothly with the increase of the training batch, whereas the improved model had a much smaller reconstruction error. The same phenomenon occurred in the reconstruction error of the first layer of the two models, and the training accuracy rate increased slightly with the increase of epoch. The difference between final training accuracy and test accuracy was about 2%, which is not much different.

Figure 12. Comparison of second-layer reconstruction error when training (a) standard CDBN and (b) improved CDBN.

A commonly used method for high-dimensional data analysis is t-distribution stochastic neighbor embedding (tSNE). tSNE can reduce any dimension to less than three at will. As shown in Figure 13,

an intuitive feature distribution was obtained by using tSNE to visualize the features of the standard and improved CDBN.

Figure 13. Feature visualization of (a) dataset1 using the standard CDBN, (b) dataset1 using the improved CDBN, (c) dataset2 using the standard CDBN, and (d) dataset2 using the improved CDBN.

The results indicate the obvious optimization effect of the proposed method on feature processing. CDBN integrates convolution operation into origin DBN, which makes it have a stronger feature extraction capability. On the basis of the standard CDBN, a filter was introduced to filter out noise, so that the improved CDBN model has a more stable learning ability. The feature distribution of the original data is relatively loose, and mixed regions exist without obvious class boundaries. These phenomena are not conducive to the subsequent classification. For instance, features of labels 6 and 7 are mostly muddled; thus, misclassification occurred, as shown in Figure 11. By contrast, the feature distribution is clustered tightly with the improved CDBN, and it has easily observed boundaries and a wider distance, which are helpful for further feature extraction and classification.

Reconstruction error is a key indicator for measuring model learning ability. Different optimizers were employed for comparative analysis, and the reconstruction errors by different optimizers are shown in Figure 14. As shown in the figure, during the training, the reconstruction error of the CDBN model with Adam optimizer dropped smoothly and had the smallest reconstruction error.

Figure 14. Comparison of second-layer reconstruction error when training the CDBN with different optimizers: (a) Adam, (b) Gradient Descent, (c) Nesterov, (d) RMSprop.

4. Conclusions

An improved CDBN-based fault diagnosis model was proposed for the effective extraction and learning of quantitative and qualitative features of different bearing fault types and sizes. The proposed model was able to extract deep features of the dataset, effectively implementing the diagnosis of multiple degrees and types of bearing faults. First, a band-pass filter was used to preprocess the original signal to obtain optimized features. Simultaneously, the Adam optimizer was introduced to speed up training and improve the model convergence speed. Finally, in combination with the softmax two-layer connection, multi-layer feature fusion was used to fully enable the feature representation capabilities of each layer of the model. Compared with the standard CDBN, the improved CDBN model displayed higher accuracy. The improved model also ensured the characterization of the learned features and ultimately enabled qualitative and quantitative diagnosis of bearing faults. The experimental results show that the proposed method had reduced training error, a smoother decline, and a higher diagnostic accuracy than SAE, ANN, and DBN. The proposed model had a positive impact on the classification of single and compound fault types of the datasets. Future research will focus on dealing with unbalanced data or small data, since these cases are more meaningful for practical applications. In future investigations, comparative experiments of different optimizers, other networks (such as recurrent neural network and generative adversarial networks), and attention-based mechanisms will be conducted.

Author Contributions: Conceptualization, S.L. and J.X.; Methodology, S.J., C.S. and X.S.; Validation, S.L., J.X. and D.W.; Formal Analysis, S.L. and C.S.; Data Curation, X.S., D.W. and Z.Z.; Writing – Original Draft Preparation, S.L., C.S. and D.W.; Writing – Review & Editing, X.S. and Z.Z.; Supervision, C.S. and Z.Z.; Funding Acquisition, C.S., D.W. and Z.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Foundation of China (No. 51875376, 51875375), as well as the Suzhou Science Foundation (No. SYG201802).

Conflicts of Interest: The authors have no conflicts of interest to declare.

References

1. Jiang, X.; Shen, C.; Shi, J.; Zhu, Z. Initial center frequency-guided VMD for fault diagnosis of rotating machines. *J. Sound Vib.* **2018**, *435*, 36–55. [\[CrossRef\]](#)
2. Kumar, P.S.; Kumaraswamidhas, L.A.; Laha, S.K. Selecting effective intrinsic mode functions of empirical mode decomposition and variational mode decomposition using dynamic time warping algorithm for rolling element bearing fault diagnosis. *Trans. Inst. Meas. Control* **2019**, *41*, 1923–1932. [\[CrossRef\]](#)
3. Huang, W.T.; Kong, F.Z.; Zhao, X.Z. Spur bevel gearbox fault diagnosis using wavelet packet transform and rough set theory. *J. Intell. Manuf.* **2018**, *29*, 1257–1271. [\[CrossRef\]](#)
4. Kaplan, K.; Kaya, Y.; Kuncan, M.; Minaz, M.R.; Ertunc, H. An improved feature extraction method using texture analysis with LBP for bearing fault diagnosis. *Appl. Soft Comput.* **2020**, *87*, 106019. [\[CrossRef\]](#)
5. Dong, S.J.; Wu, W.L.; He, K.; Mou, X.Y. Rolling bearing performance degradation assessment based on improved convolutional neural network with anti-interference. *Measurement* **2020**, *151*, 107219. [\[CrossRef\]](#)
6. Gong, W.; Chen, H.; Zhang, Z.; Zhang, M.; Wang, R.; Guan, C.; Wang, Q. A Novel Deep Learning Method for Intelligent Fault Diagnosis of Rotating Machinery Based on Improved CNN-SVM and Multichannel Data Fusion. *Sensors* **2019**, *19*, 1693. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Shi, Q.J.; Guo, X.Z.H.; Liu, D.S.H. Bearing fault diagnosis based on feature fusion and support vector machine. *J. Electron. Meas. Instrum.* **2019**, *33*, 104–111.
8. Hongfeng, T.; Chaochao, Z.; Liu, Y.; Yajun, T. Decision Tree SVM Fault Diagnosis Method of Photovoltaic Diode-Clamped Three-Level Inverter. U.S. Patent Application 10/234,495, 19 March 2019.
9. Moosavian, A.; Jafari, S.M.; Khazaei, M.; Ahmadi, H. A Comparison Between ANN, SVM and Least Squares SVM: Application in Multi-Fault Diagnosis of Rolling Element Bearing. *Int. J. Acoust. Vib.* **2018**, *23*, 432–440.
10. Meng, Z.; Zhan, X.; Li, J.; Pan, Z. An enhancement denoising autoencoder for rolling bearing fault diagnosis. *Measurement* **2018**, *130*, 448–454. [\[CrossRef\]](#)
11. Liang, T.; Wu, S.; Duan, W.; Zhang, R. Bearing fault diagnosis based on improved ensemble learning and deep belief network. *J. Phys. Conf.* **2018**, *1074*, 012154. [\[CrossRef\]](#)
12. Chen, Z.; Mauricio, A.; Li, W.; Gryllias, K. A Deep Learning method for bearing fault diagnosis based on Cyclic Spectral Coherence and Convolutional Neural Networks. *Mech. Syst. Signal Process.* **2020**, *140*, 106683. [\[CrossRef\]](#)
13. Zhao, R.; Yan, R.; Chen, Z.; Mao, K.; Wang, P.; Gao, R.X. Deep learning and its applications to machine health monitoring. *Mech. Syst. Signal Process.* **2019**, *115*, 213–237. [\[CrossRef\]](#)
14. Shen, C.; Qi, Y.; Wang, J.; Cai, G.; Zhu, Z. An automatic and robust features learning method for rotating machinery fault diagnosis based on contractive autoencoder. *Eng. Appl. Artif. Intell.* **2018**, *76*, 170–184. [\[CrossRef\]](#)
15. Shao, H.; Jiang, H.; Zhang, H.; Liang, T. Electric locomotive bearing fault diagnosis using a novel convolutional deep belief network. *IEEE Trans. Ind. Electron.* **2018**, *65*, 2727–2736. [\[CrossRef\]](#)
16. Wen, J.T.; Yan, C.H.; Sun, J.D.; Qiao, Y. Bearing fault diagnosis method based on compressed acquisition and deep learning. *Chin. J. Sci. Instrum.* **2018**, *39*, 171–179.
17. Chen, Z.; Li, W. Multi sensor Feature Fusion for Bearing Fault Diagnosis Using Sparse Autoencoder and Deep Belief Network. *IEEE Trans. Ind. Electron.* **2017**, *66*, 1693–1702.
18. Hinton, G.E. Training Products of Experts by Minimizing Contrastive Divergence. *Neural Comput.* **2014**, *14*, 1771–1800. [\[CrossRef\]](#) [\[PubMed\]](#)

19. Cui, L.; Wang, X.; Wang, H.; Ma, J. Research on Remaining Useful Life Prediction of Rolling Element Bearings Based on Time-Varying Kalman Filter. *IEEE Trans. Ind. Electron.* **2020**, *69*, 2858–2867. [[CrossRef](#)]
20. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).