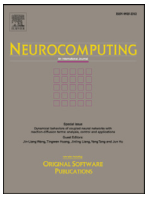
Neurocomputing 310 (2018) 77–95

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom



A robust intelligent fault diagnosis method for rolling element bearings based on deep distance metric learning

Xiang Li[a,*], Wei Zhang[b], Qian Ding[c]

[a] College of Sciences, Northeastern University, Shenyang 110819, China
[b] School of Aerospace Engineering, Shenyang Aerospace University, Shenyang 110136, China
[c] Department of Mechanics, Tianjin University, Tianjin 300072, China

ARTICLE INFO

Article history:
Received 16 October 2017
Revised 22 February 2018
Accepted 4 May 2018
Available online 23 May 2018

Communicated by Shen Yin

Keywords:
Fault diagnosis
Rolling bearing
Deep learning
Deep metric learning
Environmental noise
Domain shift

ABSTRACT

Intelligent data-driven fault diagnosis methods for rolling element bearings have been widely developed in the recent years. In real industries, the collected machinery signals are usually exposed to environmental noises, and the bearing operating condition changes in different working scenarios. That leads to distribution discrepancy between the labeled training data and the unlabeled testing data, and consequently the diagnosis performance deteriorates. This paper proposes a novel deep distance metric learning method for rolling bearing fault diagnosis based on deep learning. A deep convolutional neural network is used as the main architecture. Based on the learned representations through multiple hidden layers, a representation clustering algorithm is proposed to minimize the distance of intra-class variations and maximize the distance of inter-class variations simultaneously. A domain adaptation method is adopted to minimize the maximum mean discrepancy between training and testing data. In this way, the robustness of the fault diagnosis method can be significantly improved against noise and variation of working condition. Extensive experiments on a popular rolling bearing dataset are carried out to validate the effectiveness of the proposed method, and the diagnosis performance is widely evaluated in different scenarios. Comparisons with other approaches and the related works on the same dataset demonstrate the superiority of the proposed method. The experimental results of this study suggest the proposed deep distance metric learning method offers a new and promising tool for intelligent fault diagnosis of rolling bearings.

1. Introduction

Rolling element bearings are one of the most critical components in rotating machines. Since the unexpected failures of rolling bearings usually result in serious loss of safety, production delays and large costs of maintenance in modern industries [1], accurate and timely fault diagnosis of them has always been highly demanded, and received increasing research attention in the past decades [2–8]. While the traditional signal processing methods such as wavelet analysis [2,3], stochastic resonance techniques [4,9,10] etc. have achieved satisfactory diagnosis results based on machinery vibration data [5–8], intelligent fault diagnosis methods are becoming more and more popular nowadays since they do not require prior expertise and can efficiently provide reliable diagnosis results [11–16].

In the past years, a large number of intelligent fault diagnosis methods have been proposed based on machine learning and statistical inference techniques, such as artificial neural networks (ANN) [11,12,17], support vector machines (SVM) [15,16], random forest (RF) [18], fuzzy inference and other improved algorithms [13,14]. Generally, neural networks are one of the most popular data-driven methods to identify faulty and healthy machine conditions, where fault diagnosis is treated as a classification problem through feature extraction. Especially, deep learning has recently emerged as a highly effective network architecture for pattern recognition, that holds the potential to overcome the obstacles in the current intelligent fault diagnosis. Deep learning is characterized by the deep network structure where multiple layers are stacked in the network to fully explore the collected signal information [19]. High level abstract data representations can be efficiently learned through multiple linear and non-linear transformations for machine health condition classification. In general, better diagnosis results have been obtained comparing with shallow architectures [20–26].

[*] Corresponding author.
E-mail address: xiangli@mail.neu.edu.cn (X. Li).

https://doi.org/10.1016/j.neucom.2018.05.021
0925-2312/© 2018 Elsevier B.V. All rights reserved.

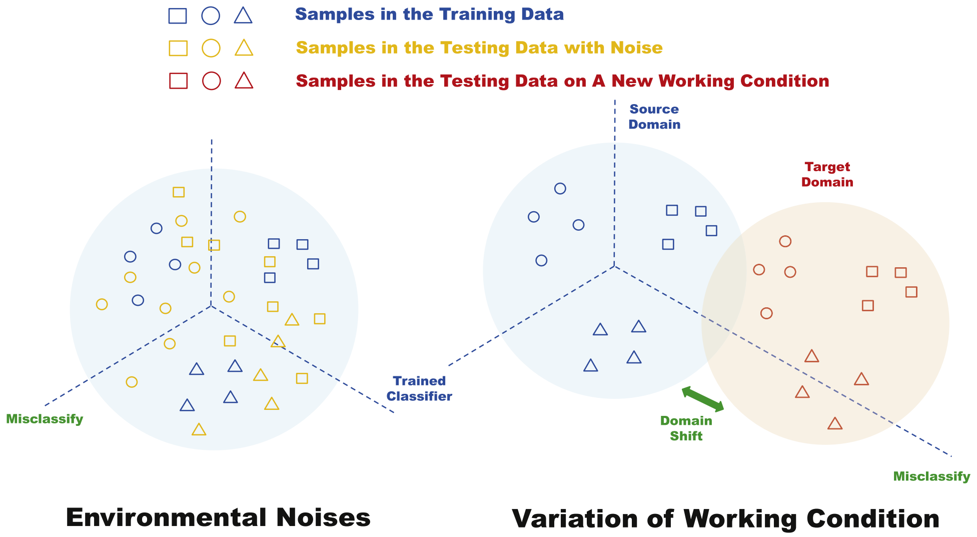

Samples in the Training Data
Samples in the Testing Data with Noise
Samples in the Testing Data on A New Working Condition
Source Domain
Target Domain
Misclassify
Trained Classifier
Domain Shift
Misclassify
Environmental Noises
Variation of Working Condition

Fig. 1. Illustration for bearing fault diagnosis problem with environmental noises and variation of working condition.

Most data-driven methods for fault diagnosis are currently implemented under the assumption that training and testing data are subject to the same distribution. However, in real industries, environmental noise and variation of machine operating condition inevitably make the distributions of training and testing data different from each other and consequently, it is difficult for the well-trained neural network to generalize the learned pattern knowledge from the labeled training data, denoted as *source domain*, to the new unlabeled testing data, denoted as *target domain*. This challenge of pattern learning validity is known as the domain shift problem [27] and therefore, one of the main challenges for intelligent fault diagnosis is the robustness of the algorithm [28] against domain shift.

Fig. 1 presents an illustration of the domain shift problem. In general, the deep architecture can be effectively trained and performs excellent machine health condition classifications on the source domain. However, due to interference of noise and variation of working situation, domain shift phenomenon may significantly deteriorate the network classification performance on the target domain. In real-world industries, such problem is very common since rotating machines are subject to random environmental noises all the time during operations [28]. Moreover, the labeled vibration data for model training are usually collected under one working load of rolling bearings, but testing data are possibly from different working loads and thus subject to new distributions. The distribution discrepancy poses an obstacle in adapting the well-trained models across domains. The signal characteristics may change remarkably for different working scenarios even with the same bearing fault.

Generally, applying the learned fault patterns on new data distributions requires specific customization to accommodate the new domain information. One straight-forward solution is by the means of collecting a certain number of labeled data in the target domain for training. However, that is very expensive and almost impossible in many cases. As an alternative way, the available labeled source domain data and unlabeled target domain data can be further explored to calibrate the established model, that is relatively easy to be carried out in real-world applications.

In order to address the aforementioned domain shift problem and enhance the algorithm robustness, this paper proposes a novel deep distance metric learning method. Recent machine learning-based studies show that learning a distance metric from the available data has large potential to achieve promising results, compared with the use of hand-crafted distance metrics [29,30]. In the past years, a number of metric learning methods have been successfully developed and applied in many research tasks such as person re-identification [31], human activity recognition [32], image classification [29,30] and so forth.

In this study, as Fig. 1 shows, higher diagnosis accuracies are expected to be obtained if the distribution discrepancy between training and testing data is minimized and data samples belonging to the same fault type cluster better. Therefore, two techniques of distance metric learning are proposed in this paper to improve the model generalization ability, i.e. representation clustering [33–35] with respect to different fault types and domain adaptation, which is a particular case of transfer learning that leverages labeled data in the source domain to learn a classifier for unlabeled data in the target domain [27]. In the recent years, advanced representation clustering methods have been successfully developed [36–38], and domain adaptation has also attracted much research attention [39,40]. The scheme of the two techniques are illustrated in Fig. 2.

While promising results have been achieved by deep distance metric learning, limited researches can be found with respect to its application on machinery fault diagnosis. An adaptive batch normalization method was proposed by Zhang and colleagues [39] to improve the cross-domain fault diagnosis performance of neural network. Lu et al. [40] proposed a deep neural network-based domain adaptation method for diagnosis, where the feature maximum mean discrepancy (MMD) is minimized, and a weight regularization term is used to strengthen the representative features. Xie et al. [41] addressed the cross-domain feature extraction and fusion from time and frequency-domain with spectrum envelop pre-processing and time domain synchronization average principle using transfer component analysis (TCA).

This paper proposes a novel data-driven fault diagnosis method for rolling bearings based on deep convolutional neural network. Industrial domain shift problem due to environmental noise and variation of working condition is addressed using deep distance metric learning algorithm. Labeled source domain data for training and unlabeled target domain data for testing are assumed to be available. Different from existing researches, an integrated optimization objective function is used to enhance the generalization ability of the proposed method to new data distribution, which consists of classification error, domain discrepancy, and inter-class and intra-class representation distance optimization. Experiments

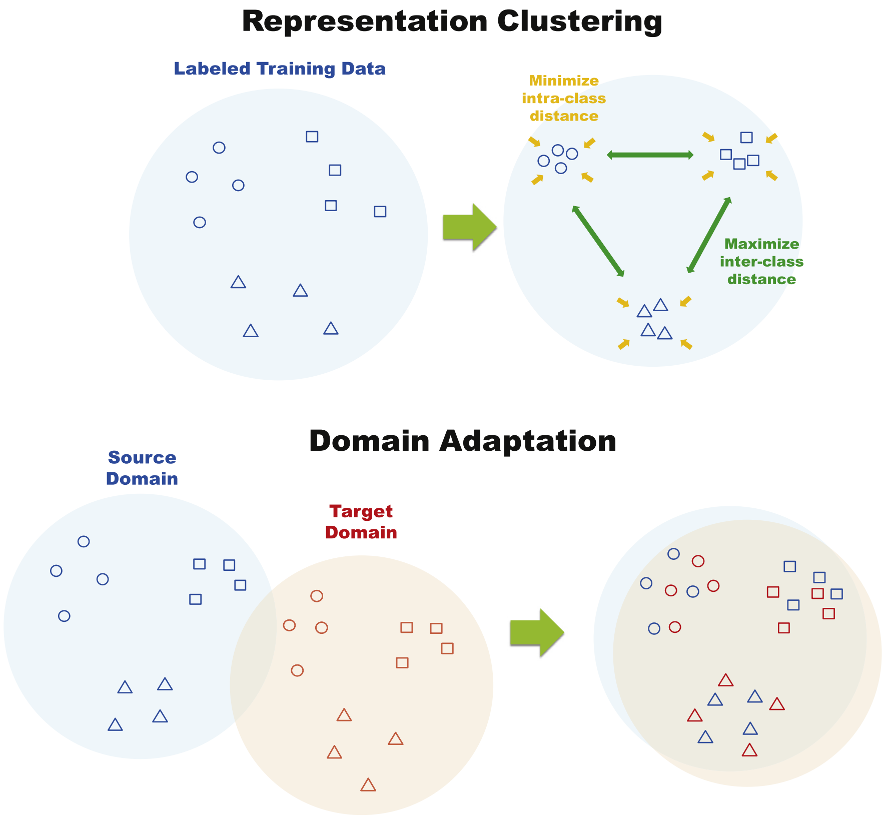

Representation Clustering
Labeled Training Data
Minimize intra-class distance
Maximize inter-class distance
Domain Adaptation
Source Domain
Target Domain

Fig. 2. The two proposed algorithms for improving fault diagnosis robustness, i.e. representation clustering and domain adaptation.

on a popular rolling bearing dataset are conducted to validate the effectiveness of the proposed method. The results show the proposed method has strong robustness and adaptation ability with respect to environmental noise and variation of working condition. By comparing with other approaches and related works in the literature, the superiority of the proposed method is demonstrated. The diagnosis performance of the proposed method is extensively evaluated in different situations.

The remainder of this paper starts with the theoretical background in Section 2. The domain shift problem, convolutional neural network, deep distance metric learning and softmax classifier are introduced. The proposed fault diagnosis method is presented in Section 3, and experimentally validated using a popular rolling bearing dataset in Section 4. We close the paper with conclusions in Section 5.

2. Theoretical background

In this section, the domain shift problem for fault diagnosis is presented, and the preliminaries for convolutional neural network, deep distance metric learning and softmax regression are introduced.

2.1. Domain shift problem

Traditionally, machinery fault diagnosis aims to identify fault location, severity *etc*. based on a prior known set of faults. It is assumed that the source and target domain distributions are the same, and the learned fault patterns from the labeled training samples can be directly applied on the unlabeled testing samples. However, discrepancy between the source and target domains inevitably exists in practical tasks, which makes the model generalization ability deteriorate across domains. Generally, this study is carried out under the assumptions:

1. The fault diagnosis task remains the same for different scenarios, i.e. the class labels are identical.
2. The source and target domains are related to each other, but have different distributions.
3. Labeled samples from the source domain are available for training.
4. Unlabeled samples from the target domain are available for training and testing.

In domain adaptation problems with respect to fault diagnosis, let X denote the input space and $Y = \{1, 2, \ldots, N_c\}$ represents the set of N_c possible machine health conditions. We are given a source domain $\mathcal{D}_s = \left\{\left(\mathbf{x}_i^s, \mathbf{y}_i^s\right)\right\}_{i=1}^{n_s}$ of n_s labeled samples and a target domain $\mathcal{D}_t = \left\{\left(\mathbf{x}_i^t\right)\right\}_{i=1}^{n_t}$ of n_t unlabeled samples. $\mathcal{D}_s$ and $\mathcal{D}_t$ are sampled from joint distributions $P(X, Y)$ and $Q(X, Y)$ respectively, and $P \neq Q$. The aim of this paper is to construct a deep neural network $\mathbf{y} = f(\mathbf{x})$ that is able to reduce the cross-domain shifts in joint distributions and learn domain-invariant features and classifiers, in order to minimize the target risk $R_t(f) = \Pr_{(\mathbf{x},\mathbf{y})\sim Q}[f(\mathbf{x}) \neq \mathbf{y}]$ with source supervision.

2.2. Convolutional neural network

Convolutional neural networks (CNNs), that are specifically designed for variable and complex signals, are utilized in this study. In the past few years, a large number of researches [23–26,42] have benefited from CNN's characteristics of local receptive fields, shared weights and spatial sub-sampling. CNN's ability to maintain data information regardless of scale, shift and distortion invariance has been shown [43].

The convolutional layers convolve multiple filters with raw input data and generate features, and the following pooling layers extract the most significant local features afterwards. The input data are usually 2-dimensional (2D) data for CNNs to learn abstract spatial features by alternating and stacking convolutional kernels and pooling operation. Since the input data in this research is a sequence of machinery vibration signal, the 1-dimensional (1D) CNN is briefly introduced in the following.

The input sequential data is assumed to be $\mathbf{x} = [x_1, x_2, \ldots, x_N]$ where N denotes the length of the sequence. The convolution operation in the convolutional layer can be defined as a multiply operation between a filter kernel $\mathbf{w}$, $\mathbf{w} \in R^{F_L}$, and a concatenation vector representation $\mathbf{x}_{i:i+F_L-1}$, which can be expressed as,

$$\mathbf{x}_{i:i+F_L-1} = x_i \oplus x_{i+1} \oplus \cdots \oplus \ x_{i+F_L-1}, \tag{1}$$

where $x_{i:i+F_L-1}$ represents a window of F_L length sequential signal starting from the ith point, and $\oplus$ concatenates the data samples into a longer embedding. The final convolution operation is defined as,

$$z_i = \varphi(\mathbf{w}^T \mathbf{x}_{i:i+F_L-1} + b), \tag{2}$$

where $*^T$ denotes the transpose of a matrix $*$, and b and φ represent the bias term and non-linear activation function, respectively. The output z_i can be considered as the learned feature of the filter kernel $\mathbf{w}$ on the corresponding subsequence $\mathbf{x}_{i:i+F_L-1}$. By sliding the filter window from the first point to the last point in the sample data, the feature map of the jth filter can be obtained, which is denoted as,

$$\mathbf{z}_j = [z_j^1, z_j^2, \ldots, z_j^{N-F_L+1}]. \tag{3}$$

In CNNs, multiple filter kernels can be applied in the convolution layer with different filter length F_L.

Usually, a pooling layer is applied to the feature maps generated by the convolutional layer. On the one hand, the pooling is able to extract the most significant local information in each feature map. On the other hand, the feature dimensionality, i.e. the number of model parameters, can be remarkably reduced by this operation. In general, average-pooling and max-pooling are widely used. In this paper, the max-pooling function is applied in the network.

The max-pooling operation is carried out in the feature maps with a pooling length of g. The extracted feature corresponding to the filter kernel can be obtained as,

$$\mathbf{p}_j = [p_j^1, p_j^2, \ldots, p_j^s], \tag{4}$$

$$p_j^k = \max\left(z_j^{(k-1)g+1}, z_j^{(k-1)g+2}, \ldots, z_j^{kg}\right), \tag{5}$$

where $\mathbf{p}_j$ is the output of the pooling layer applied to the jth feature map and has s dimensions. By alternating the convolutional and max-pooling layers, fully-connected layer and softmax regression are usually added as the top layers to make classification. The framework for 1D CNN is displayed in Fig. 3.

2.3. Deep distance metric learning

In pattern recognition problems such as fault diagnosis, the latest researches show that learning a distance metric from the available data is able to achieve promising results, comparing with the use of hand-crafted metrics [29,30].

Currently, most existing metric learning studies aim to learn a linear distance to transform samples into a linear feature space, where the inter-class distance is maximized for better classification [29]. However, since machine vibration signal inherently possesses highly nonlinear characteristics, the signal information can not be fully explored by the shallow linear projections. In some researches, all pairwise distances between samples are required for metric learning [44]. While promising results can be obtained, that usually results in expensive computations, and respecting all the distances simultaneously may break the optimization balance between classes in the training process. Moreover, many studies are carried out under the assumption that the training and testing data are subject to the same distribution. In this way, performance deterioration is likely to occur with metric learning in many practical tasks with noise disturbance and variation of working condition.

In order to address the aforementioned issues, a novel deep distance metric learning method for fault diagnosis is proposed in this paper. First, feature extraction using deep learning is applied on the raw input of machinery vibration signal, and the distance metric learning is implemented on the extracted high level features. Specifically, the data representations in the top layer are processed for metric learning, and a classifier is attached at last for the final fault classification. The multiple layers of linear and nonlinear transformations contribute to better representations of the original data samples. In the proposed method, two algorithms are used to enhance the model generalization ability, i.e. representation clustering and domain adaptation, which will be presented in the following. Fig. 2 illustrates the scheme of the two algorithms.

2.3.1. Representation clustering

In deep learning tasks, the higher layers especially the top layer in the network, are directly responsible for the final classification. Therefore, better clustering of each class and separability between classes of the high level representations are expected to result in higher diagnosis accuracy. Accordingly, one of the basic ideas of the proposed method is to minimize the distance of intra-class variations and maximize the distance of inter-class variations simultaneously. The left sub-figure in Fig. 2 illustrates the clustering scheme of the proposed representation clustering method.

Let $\mathbf{x}^{(k)}$ denote the raw input samples belonging to the kth class, and the number of the classes is represented by N_{class}. $f^{(m)}(\mathbf{x}^{(k)})$ denotes the output at the mth layer in the network of the input $\mathbf{x}^{(k)}$, and $f^{(m)} : \mathbf{R}^{N_{input}} \to \mathbf{R}^{N_{(m)}}$ is the mapping function determined by the network configuration, where N_{input} represents the input dimension and $N_{(m)}$ denotes the output dimension at the mth layer. The expectation and variance of $f^{(m)}(\mathbf{x}^{(k)})$ can be easily obtained and denoted as $E[f^{(m)}(\mathbf{x}^{(k)})]$ and $Var[f^{(m)}(\mathbf{x}^{(k)})]$, respectively. While $Var[f^{(m)}(\mathbf{x}^{(k)})]$ directly indicates the representation compactness of the kth class, the Euclidean distance between the expectations of two classes subtracted by the two corresponding standard deviations can be used to measure the inter-class separability.

$$\begin{aligned} d_{f^{(m)}}(i,j) = & \left\| E[f^{(m)}(\mathbf{x}^{(i)})] - E[f^{(m)}(\mathbf{x}^{(j)})] \right\|_2 \\ & - \sqrt{Var[f^{(m)}(\mathbf{x}^{(i)})]} - \sqrt{Var[f^{(m)}(\mathbf{x}^{(j)})]}, \end{aligned} \tag{6}$$

where $d_{f^{(m)}}(i,j)$ denotes the inter-class distance between the ith and jth classes, $1 \le i < j \le N_{class}$.

While all the pairwise distances are suggested to be maximized simultaneously for metric learning by some existing studies in image processing field, the distance optimization balance for multiple classes has large opportunity to be broken in the fault diagnosis problems. In that situation, one or more distances between classes can be extended, while the other pairwise distances may be shortened on the contrary, that will significantly deteriorate the testing performance. Therefore, rather than optimizing the summation of all the pairwise distances, this paper proposes to only depart the two classes with the shortest distance at each training step, after evaluation of all the pairwise class distances. A supervised distance metric learning objective $J_{cluster}$ for representation clustering can be formulated as,

$$\min_{f^{(m)}} J_{cluster} = -D_{inter} + \eta D_{intra}, \tag{7}$$

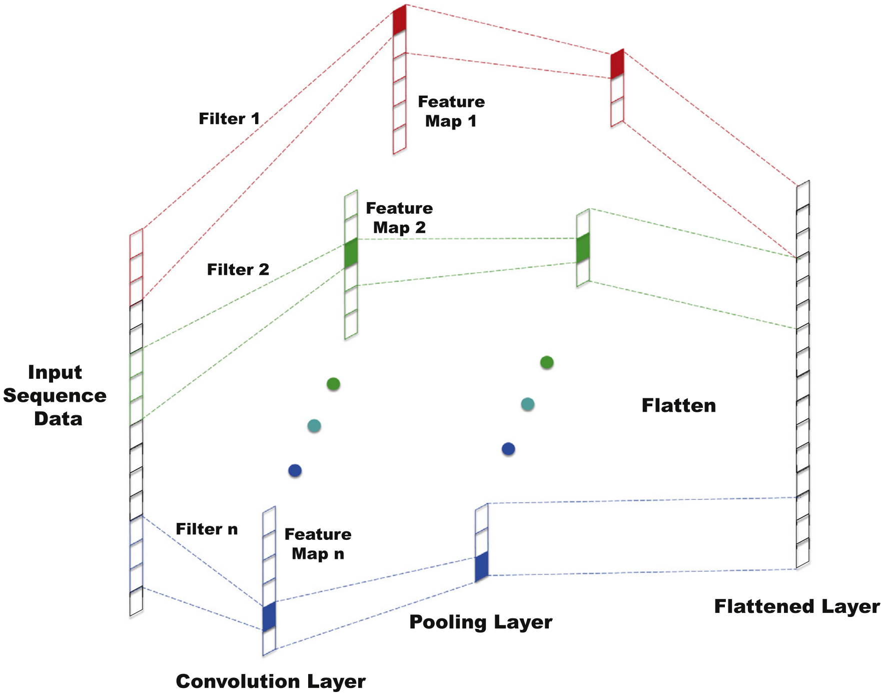


Fig. 3. Illustration for 1D CNN operation.

where

$$\begin{aligned} D_{inter} &= d_{f^{(m)}}(p,q), \\ D_{intra} &= \sum_{i=1}^{N_{class}} \sqrt{Var[\mathbf{x}^{(i)}]}, \\ p,q &= \underset{p,q}{\operatorname{argmin}}\, d_{f^{(m)}}(p,q), \\ & 1 \le p < q \le N_{class}. \end{aligned} \tag{8}$$

D_{inter} and D_{intra} measure the inter-class separability and intra-class compactness respectively, and η is a scaling coefficient. By optimizing $J_{cluster}$, the deep learning algorithm is expected to offer more distinct regions for different classes in the high-level feature space, that facilitates the following final classification of the data samples. In this way, the proposed fault diagnosis method can be more robust against environmental noises.

2.3.2. Distribution discrepancy

In rolling bearing health condition monitoring, variation of machine working condition such as rotating speed has significant influence on the collected vibration signal characteristics, and that results in serious domain shift problem as Fig. 1 shows. When the data distributions of the source and target domains have remarkable difference, the well-trained classifier usually fails to provide accurate diagnosis for the testing data. Therefore, a domain adaptation algorithm is proposed in this paper to minimize the distribution discrepancy.

Domain adaptation establishes knowledge transfer from the source domain to the target domain by exploring domain-invariant structures that bridge the distribution discrepancy [45]. In the recent years, it has been adopted by some researchers in different fields and achieved promising results in many tasks such as sentiment analysis [46], object recognition [47,48], facial recognition [49], speech recognition [50], video recognition [51] etc. In [52–54], shallow domain-invariant features are learned by minimizing the discrepancy between domains. Furthermore, latest researches have revealed that deep learning architectures for domain adaptation are able to learn more transferable features and thus are more promising [55,56].

In this study, in order to measure and further minimize the distribution discrepancy, the maximum mean discrepancy (MMD) is adopted in this paper [57]. MMD is defined as the squared distance between the kernel embeddings of marginal distributions in the reproducing kernel Hilbert space (RKHS).

$$\mathrm{MMD}_k(P,Q) \triangleq \left\| \mathbf{E}_P[\phi(\mathbf{x}^s)] - \mathbf{E}_Q\left[\phi\left(\mathbf{x}^t\right)\right] \right\|^2_{\mathcal{H}_k}, \tag{9}$$

where $\mathcal{H}_k$ denotes the RKHS endowed with a characteristic kernel k. The most important property is $\mathrm{MMD}_k(P,Q) = 0$ iff $P = Q$.

As stated in [58], kernel choice is critical to ensure the testing power and low testing error of MMD, since different kernels may embed probability distributions in different RKHSs where different orders of sufficient statistics can be emphasized. Therefore in this paper, multiple kernels of MMD are used to leverage different kernels and formulate a principled approach for optimal kernel selection. Specifically, a mixture of N_k RBF kernels are utilized,

$$k(\mathbf{x}^s, \mathbf{x}^t) = \sum_{i=1}^{N_k} k_{\sigma_i}(\mathbf{x}^s, \mathbf{x}^t), \tag{10}$$

where k_{σ_i} represents a Gaussian kernel with bandwidth parameter σ_i. In the experiments, it is noticed that using simple values of the bandwidth parameters and a mixture of 5 kernels is able to obtain good results [59]. So the default bandwidth parameters are selected as 1, 2, 4, 8 and 16 in this paper, and their weights are kept equal for simplicity.

The distribution discrepancy optimization objective J_{mmd} is defined as,

$$\min_{f^{(m)}} J_{mmd} = \mathrm{MMD}_k(P^{f^{(m)}}, Q^{f^{(m)}}), \tag{11}$$

where $P^{f^{(m)}}$ and $Q^{f^{(m)}}$ denote the m-th layer representations for the source and target samples respectively, and $\mathrm{MMD}_k(P^{f^{(m)}}, Q^{f^{(m)}})$ represents the multi-kernel MMD between the source and target domains evaluated on the mth layer representations.

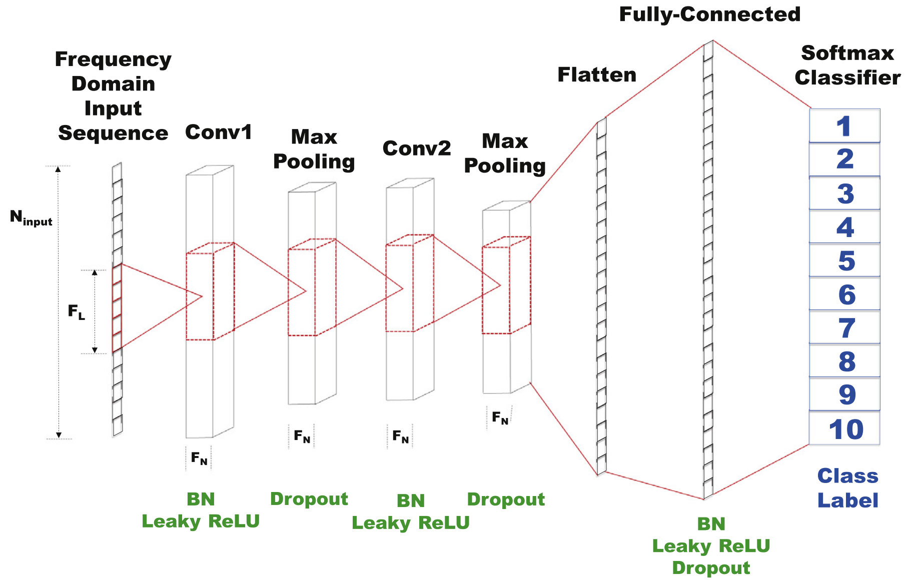


Fig. 4. Proposed deep learning architecture for bearing fault diagnosis. Conv1 and Conv2 denote the 2 convolutional layers, which are followed by batch normalizations (BN) and Leaky ReLU activations.

2.4. Softmax classifier

The softmax regression is adopted in this study for the final fault classification, and it is attached on the top layer in the deep network [60]. The extracted high-level feature representations are used as inputs of the supervised classifier followed by global back-propagation optimizations.

Recall the training samples are denoted as $\mathbf{x}_i$ and the corresponding label set is y_i where $i = 1, 2, \ldots, N_{tr}$ and N_{tr} is the number of the training samples. $\mathbf{x}_i \in R^{N \times 1}$ and $y_i \in \{1, 2, \ldots, N_{class}\}$. For an input sample $\mathbf{x}_i$, the softmax regression is able to estimate the probability $p(y_i = j \mid \mathbf{x}_i)$ for each label j $(j = 1, 2, \ldots, N_{class})$. The estimated probabilities of the input data $\mathbf{x}_i$ belonging to each label can be obtained according to the hypothesis function,

$$J_\theta(\mathbf{x}_i) = \begin{bmatrix} p(y_i = 1 \mid \mathbf{x}_i; \theta) \\ p(y_i = 2 \mid \mathbf{x}_i; \theta) \\ \vdots \\ p(y_i = N_{class} \mid \mathbf{x}_i; \theta) \end{bmatrix} = \frac{1}{\sum_{k=1}^{N_{class}} e^{\theta_k^T \mathbf{x}_i}} \begin{bmatrix} e^{\theta_1^T \mathbf{x}_i} \\ e^{\theta_2^T \mathbf{x}_i} \\ \vdots \\ e^{\theta_{N_{class}}^T \mathbf{x}_i} \end{bmatrix}, \tag{12}$$

where $\theta = [\theta_1, \theta_2, \ldots, \theta_{N_{class}}]^T$ denotes the softmax model parameters. This classifier makes sure the outputs are positive and sum to 1, allowing us to interpret the outputs of the network as the probabilities for each class.

3. Proposed fault diagnosis method

3.1. Network structure

In this paper, in order to highlight the domain shift problem and the effectiveness of the proposed method, a conventional deep convolutional neural network architecture is used for simplicity. The network structure is presented in Fig. 4.

Generally, two convolutional layers are first used for feature extraction, which are supposed to share the same configuration for simplicity including the filter size F_L, filter number F_N etc. Zeros-padding operation is adopted to keep the feature map dimension from changing [61]. Max-pooling layer is placed after each of the convolutional layers to reduce the data dimension while keeping the significant spatial information. Next, the extracted high-level feature representations are flattened and connected to a fully-connected (FC) layer. Finally, a softmax regression is attached on the top layer for fault classification.

Table 1
Default parameters of the proposed method and the experimental setting.

Parameter	Value	Parameter	Value
Epoch number	4000	N_{train}	10/2
Learning rate	0.0005	N_{test}	50
α	1	F_N	5
β	0.5	F_L	3
γ	100	Sample dimension ($2 \times N_{input}$)	1024
η	1		

In the past years, batch normalization (BN), which is able to reduce the internal-covariate-shift by normalizing the input distributions of the hidden layers to the desired Gaussian distribution, has been widely used in neural networks to accelerate the training process especially for deep network [62]. Good performance in different tasks has been achieved by batch normalization. In this study, BN is adopted after the convolutional and fully-connected layers in the proposed architecture for faster network training.

To avoid overfitting the training data, dropout technique which is an effective regularization method, is used in multiple layers in the network with rate of 0.5 [21]. In addition, while the rectified linear units (ReLU) activation functions [63] have achieved good results on most deep learning tasks, its disadvantage lies in the zero gradient whenever the unit is not active. That may make the units that are not initially active never be activated as the gradient-based optimization can not change their weights. Therefore, the leaky ReLU which was first introduced in [64] to address the above issue, is generally adopted in the network.

The default network configuration is presented in Fig. 4, and the associated parameters are listed in Table 1.

3.2. Objective function

In domain shift problems, the fault diagnosis task is the same for different domains as stated in Section 2.1. That indicates the fault classes are identical in the training and testing datasets. Therefore, before generalizing to the testing data, minimizing the classification error of the training samples is the primary optimization objective, and the cross-entropy function J_{ce} is used [65]. Corresponding with the softmax classifier in Section 2.4, J_{ce} is defined as,

$$J_{ce} = -\frac{1}{N_{tr}}\left[\sum_{i=1}^{N_{tr}}\sum_{k=1}^{N_{class}} 1\{y_i = k\}\log\frac{e^{\theta_k^T \mathbf{x}_i}}{\sum_{j=1}^{N_{class}} e^{\theta_j^T \mathbf{x}_i}}\right]. \quad (13)$$

Furthermore, as the main contribution of this study, in order to enhance the robustness of the fault diagnosis algorithm against environmental noises and variation of working condition, Eqs. (7) and (11) are integrated in the objective function. In this way, more distinct class regions are expected to be learned, the distribution discrepancy can be minimized and the model generalization ability can be improved. The final optimization problem can be formulated as,

$$\min J = \alpha J_{cluster}^{f^{(FC)}} + \beta J_{mmd}^{f^{(FC)}} + \gamma J_{ce}, \quad (14)$$

where α, β and γ are the regularization parameters, and the superscript $f^{(FC)}$ indicates the distance metric learning is implemented on the fully-connected layer.

It should be pointed out that while some researches suggest domain adaptation on multiple layers may achieve better performance [45,66], the distribution of the data representations on the fully-connected layer is focused on in this study to highlight the effectiveness of the proposed distance metric learning method. The proposed approach can be also easily extended to cover multiple layers.

3.3. Fault diagnosis procedure

The flow chart of the proposed fault diagnosis method is presented in Fig. 5. First, the raw machinery vibration signals are collected by sensors. The labeled source domain data and unlabeled target domain data are supposed to be available. It should be noted that the testing data can be exposed to additional environmental noise, and collected under a different bearing working condition from the training data. The training and testing data samples are prepared with each sample containing $2 \times N_{input}$ points. Fast Fourier transformation (FFT) is then applied on the raw samples to obtain the frequency-domain information. In this way, half of the transformed data points for each sample, i.e. N_{input} points, are used as the proposed model inputs and will be fed into the network.

Next, the network configuration is determined based on the specific fault diagnosis problem and the dataset information. In this paper, in order to highlight the effectiveness of the proposed distance metric learning approach, a conventional deep CNN structure including two convolutional layers and one fully-connected layer is adopted as presented in Section 3.1. The *Xavier* normal initializer is employed for the initializations of the network weights and biases [67].

Afterwards, the model training process starts, and both the labeled and unlabeled data are used for training. The softmax regression classifies the rolling bearing health conditions with the learned features from the deep network. By default, 2000 epochs are first run with the optimization objective in Eq. (13) in order to initialize the parameters. Then the proposed objective function in Eq. (14) is used for another 2000 training epochs. In this way, the extracted features are expected to be domain-invariant and robust against noises, that facilitates the final fault classification.

The back-propagation (BP) algorithm [68] is applied for the updates of all the parameters in the network, and the Adam optimization method [69] is used to minimize the objective (Eq. (14)) with whole batch. After training for 4000 epochs in total, the loss of the proposed network converges in general. When the training process is finished, the testing samples are fed into the proposed model and the testing results can be obtained.

Table 2
The rolling bearing dataset information.

Class label	Fault location	Fault size (mil)	Load (hp)	Sample length
1	N/A (H)	0	0,1,2,3	$2 \times N_{input}$
2	IF	7	0,1,2,3	$2 \times N_{input}$
3	IF	14	0,1,2,3	$2 \times N_{input}$
4	IF	21	0,1,2,3	$2 \times N_{input}$
5	BF	7	0,1,2,3	$2 \times N_{input}$
6	BF	14	0,1,2,3	$2 \times N_{input}$
7	BF	21	0,1,2,3	$2 \times N_{input}$
8	OF	7	0,1,2,3	$2 \times N_{input}$
9	OF	14	0,1,2,3	$2 \times N_{input}$
10	OF	21	0,1,2,3	$2 \times N_{input}$

4. Experimental study

4.1. Experimental setup

The rolling bearing dataset used in this study is provided by the Bearing Data Center of Case Western Reserve University [70]. The dataset is composed of multi-variate vibration signals generated by a bearing test-rig, as presented in Fig. 6. The main components of the experimental apparatus are a 2-horsepower (hp) motor (left side of figure), a torque transducer/encoder (center of figure) and a dynamometer (right side of figure). The motor shaft is supported by 6205-2RS JEM SKF bearings. These bearing data are collected by acceleration transducers under four load conditions (0, 1, 2 and 3 hp) with sampling rates of 12 kHz. The motor rotational speed varies between 1730 and 1797 rpm depending on the load.

The vibration signals used in this study were collected from the drive end of the motor in the test rig on four different health conditions: (1) normal condition (H); (2) outer race fault (OF); (3) inner race fault (IF); and (4) ball fault (BF). All the three kinds of faults are generated by electro-discharge machining with fault diameters of 7 mils, 14 mils and 21 mils (1 mil = 0.001 inches), respectively. Therefore, this dataset contains 10 bearing health conditions under the four loads, where the same health condition under different loads is treated as 1 class. For the convenience of classification, the 10 health conditions with different fault location and fault size are artificially set as class label 1 to 10, respectively. The detailed information of the dataset is presented in Table 2.

In the experiments of this study, N_{train} and N_{test} samples for each health condition under one load are supposed to be available as the labeled source domain and unlabeled target domain data, respectively. For each raw collected signal sequence in the dataset that represents one working condition, the first $120,000$ points are used for preparing samples. The raw data sequence is equally divided into N_{train} or N_{test} sub-signals based on the specific task and each sub-signal contains $2 \times N_{input}$ sequential points. While data overlapping can possibly occur in the sampling process, it is generally avoided in this study.

The default parameters in the proposed method are listed in Table 1.

4.2. Compared approaches

In the case studies of this paper, different neural network-based methods are implemented as comparisons to provide a comprehensive examination of the proposed method. The latest related researches on the same dataset are also presented to show the effectiveness and superiority of the proposed method. Specifically, the following approaches are studied.

1. Baseline
   We implement a basic deep learning method in the case studies as a baseline approach for comparison, where the network architecture is the same with the proposed method. However,

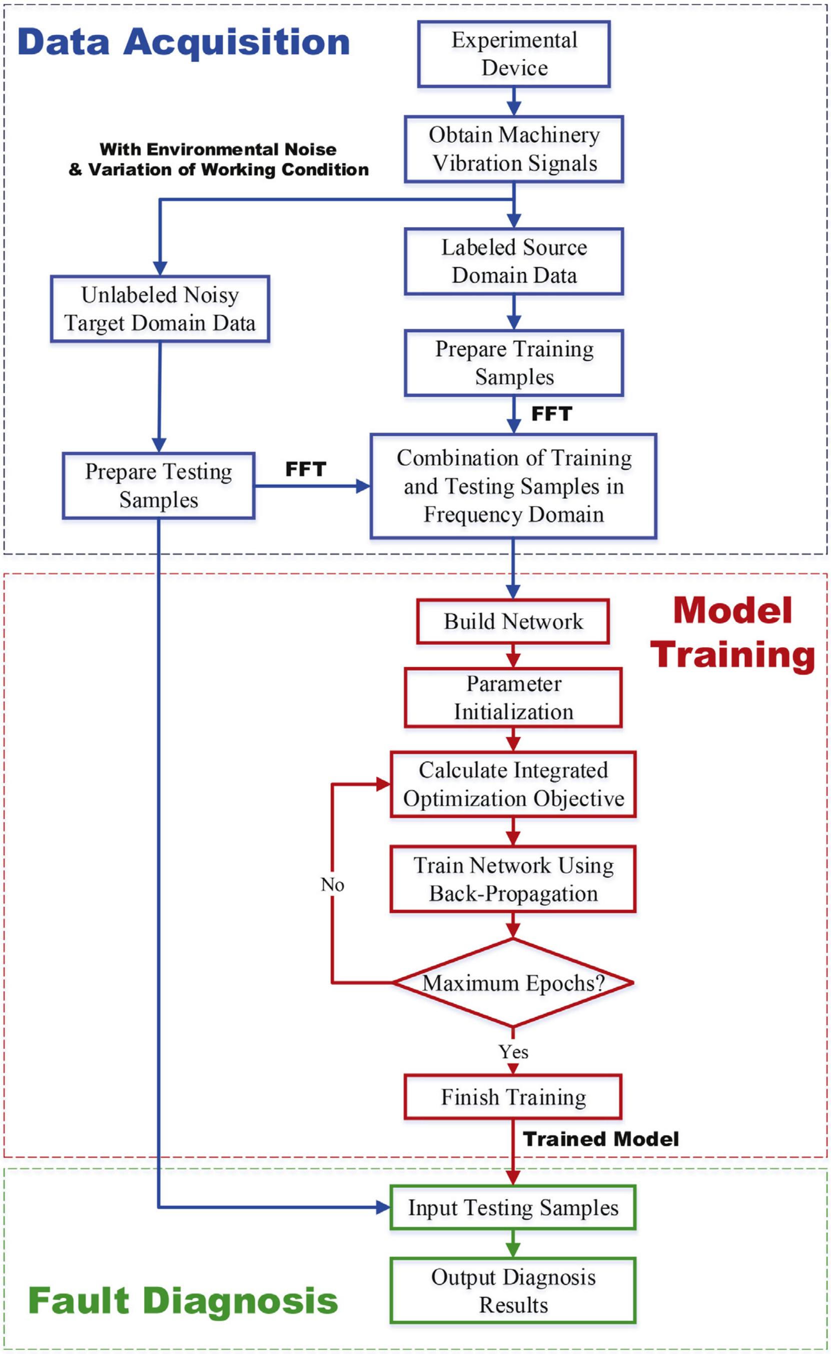

Data Acquisition
Experimental Device
Obtain Machinery Vibration Signals
With Environmental Noise & Variation of Working Condition
Labeled Source Domain Data
Unlabeled Noisy Target Domain Data
Prepare Training Samples
FFT
Prepare Testing Samples
FFT
Combination of Training and Testing Samples in Frequency Domain
Model Training
Build Network
Parameter Initialization
Calculate Integrated Optimization Objective
Train Network Using Back-Propagation
No
Maximum Epochs?
Yes
Finish Training
Trained Model
Input Testing Samples
Output Diagnosis Results
Fault Diagnosis

Fig. 5. Flow chart of the proposed bearing fault diagnosis method.

only the cross-entropy loss in Eq. (13) is used in the objective, and no distance metric learning algorithm is applied. The optimization function is,

$$\min J_1 = J_{ce}. \tag{15}$$

2. W/O clustering
   In order to show the improvements by the proposed representation clustering approach, the W/O Clustering method is implemented, which discards Eq. (7) in the distance metric learning objective. The optimization function of W/O Clustering is thus,

$$\min J_2 = \beta J_{mmd}^{f^{(FC)}} + \gamma J_{ce}, \tag{16}$$

   where the coefficients β and γ are the same with the proposed method in the experiments.
3. W/O MMD
   Similar with the W/O Clustering method, the W/O MMD approach is carried out to show the necessity of domain adaptation. Correspondingly, Eq. (11) is removed from the integrated objective of the proposed method, and the resulting optimization function becomes,

$$\min J_3 = \alpha J_{cluster}^{f^{(FC)}} + \gamma J_{ce}. \tag{17}$$

The above methods provide general evaluations of the proposed approach. All the methods implemented in this study share the same network and computing configurations to provide a fair basis for comparison. The default experimental setting is used for all the

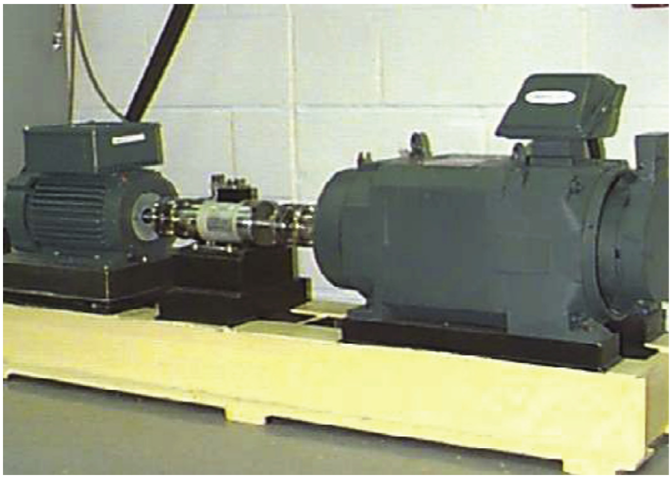


Fig. 6. The bearing test rig used in the experiments [70].

approaches with no special instruction. In the experiments, 2000 training epochs are first run with objective in Eq. (13) for all the compared methods for initialization, and another 2000 epochs are then conducted with the specific objective for different method.

It should be noted that other existing methods are also popular in deep learning tasks such as multi-layer perceptron, recurrent neural network etc. This study focuses on the deep distance metric learning algorithm, and a conventional network structure is thus used for simplicity. However, for further performance improvement, the proposed method can be easily extended to more advanced network architecture and utilize more enhancement techniques such as data augmentation etc.

In this paper, the proposed fault diagnosis method is validated on two tasks generally. First, the robustness of the algorithm is tested against additional environmental noises. Afterwards, the generalization ability of the proposed method with respect to variation of bearing working condition is examined. Furthermore, the testing scenarios with both additional noise and variation of working condition are also investigated. In all the case studies, the experimental results of different compared methods are presented, as well as those of related researches in the literature. The reported experimental results are averaged by 10 trials to reduce the effect of randomness, and the mean values and standard deviations are provided. All the experiments are carried out on a PC with Intel Core i7 CPU, 8-GB RAM and GEFORCE GTX 950M GPU. *Tensorflow* platform is used for the programming, and GPU parallel computing is employed to accelerate the computing.

4.3. Robustness against environmental noises

In real industry, environmental noises are very common, and it is extremely difficult and even impossible to collected the labeled vibration data with respect to all the possible environmental noises. Therefore in this study, additional Gaussian noises are added to the original testing signals in order to examine the robustness of the proposed method against environmental noises. Specifically, noisy data are generated for different signal-to-noise ratio (SNR), which is defined as,

$$\text{SNR (dB)} = 10\log_{10}(P_{signal}/P_{noise}), \tag{18}$$

where P_{signal} and P_{noise} denote the powers of the original signal and the additional Gaussian noise, respectively. In this study, the noisy signal ranging from −8 to 8 dB is used to evaluate the proposed method, as well as the raw collected signal in the dataset. Fig. 7 shows the vibration signals of the ball fault under different level of noises and motor loads, as well as the corresponding frequency spectrum. It can be observed that when the noise interference is strong such as −8 dB SNR, the raw signal is much corrupted which makes it difficult to diagnose the bearing fault. Pattern difference also occurs due to the domain shift phenomenon across different working conditions.

Fig. 8 shows the experimental fault diagnosis results using different methods under different level of environmental noises. All the working conditions under 4 motor loads are considered. $N_{train} = 10$ and $N_{test} = 50$ are used in this case study, that indicates the applied neural networks are trained with $10 \times 4 \times 10 = 400$ labeled samples and tested with $50 \times 4 \times 10 = 2000$ unlabeled samples. It can be observed that the proposed method generally achieves the best diagnosis results in different scenarios.

When the environmental noise is weak, such as SNR ≥ 4, all the compared methods are able to provide accurate diagnosis on the testing samples, and the classification accuracies are close to 100%. That indicates the feature extraction and fault classification based on the frequency-domain information are very effective in bearing fault diagnosis. While the proposed method and W/O Clustering approach achieve nearly 100% diagnosis accuracy for different SNR, the W/O MMD method and the baseline approach perform slightly worse.

Moreover, when the environmental noise is strong, such as SNR ≤ 0, the fault diagnosis task becomes more difficult and the accuracy deteriorates in general. The proposed method and W/O Clustering approach significantly outperform the other two methods in these scenarios, that demonstrates the effectiveness of the use of domain adaptation technique. Specifically, it is observed that the accuracy of the proposed method is higher than that of the W/O Clustering approach, that shows class clustering is able to noticeably improve the network performance especially with strong environmental noises.

Furthermore, in order to show the superiority of the proposed method with small training dataset, experiments with $N_{train} = 2$ are carried out, and the results are presented in Fig. 9. The experimental setting is similar with that presented earlier with $N_{train} = 10$. First, it is noted that training with less labeled source domain data generally leads to lower testing accuracy for all the methods. That is consistent with the related studies on deep learning that sufficient training samples are usually required for good network performance. However, the performance deterioration is not significant based on the experiments, and the proposed method is still able to achieve good diagnosis results under noisy environment. For instance, as high as 94.46% testing accuracy is obtained by the proposed method with additional noise of SNR = −4 dB, and only 80 labeled training samples.

In summary, the experiments in this section show the effectiveness and robustness of the proposed fault diagnosis method against noises. The proposed deep distance metric learning method using representation clustering and domain adaptation techniques are promising for bearing fault diagnosis especially under noisy environment.

4.4. Robustness against variation of working condition

In this section, the robustness of the proposed method is examined against variation of bearing working condition. Two transfer tasks with additional noises of different strength are carried out with different methods. The detailed information of the tasks is presented in Table 3, and the experimental setting is similar with that in Section 4.3. Specifically, transfer tasks T_1 and T_2 across working conditions with motor loads 0 and 3 are investigated. It should be pointed out that while many other transfer tasks with different motor loads can be used for evaluation, the selected two tasks can be generally considered the most difficult ones since the

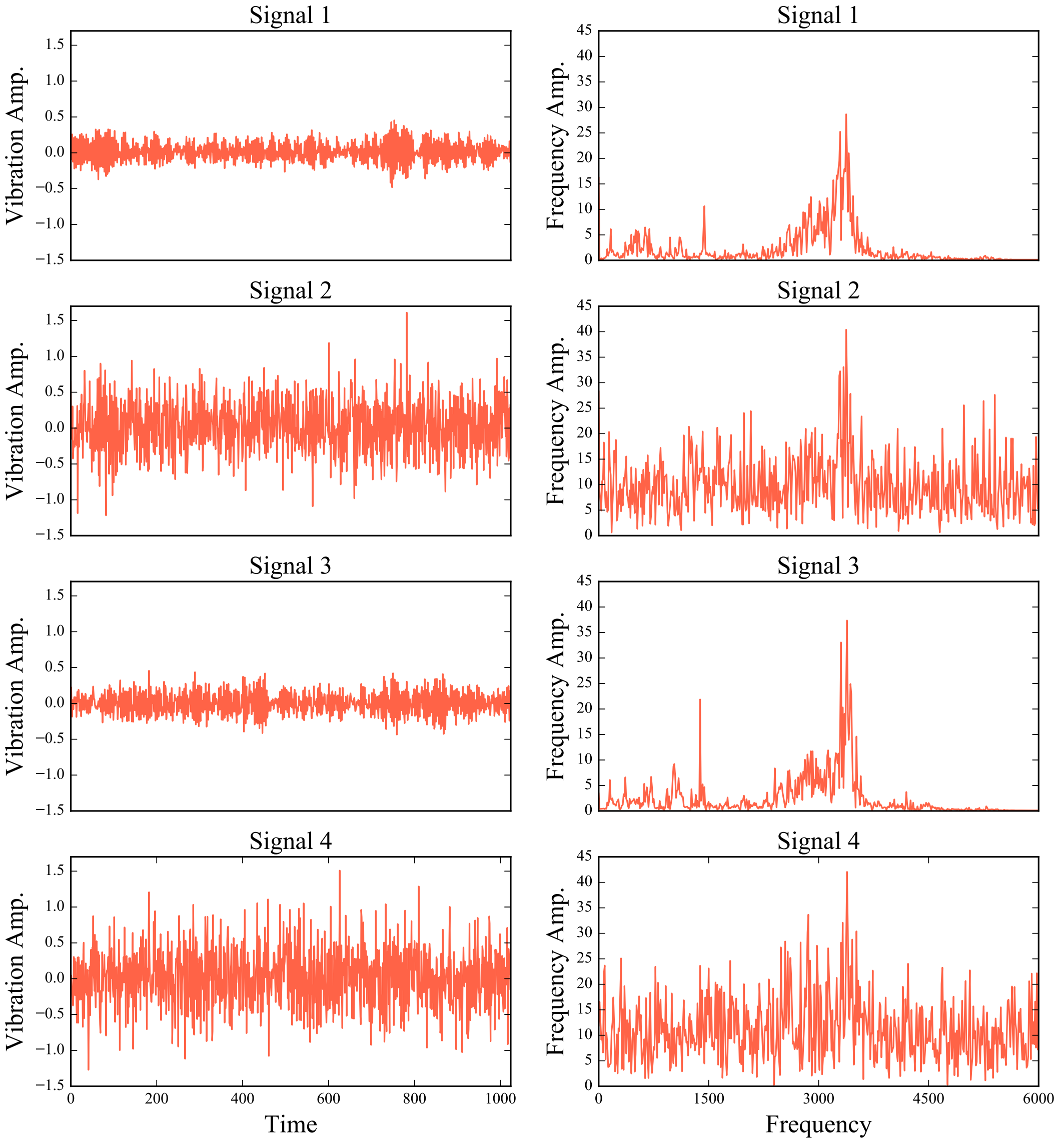


Fig. 7. Illustrations of the bearing vibration signals of ball fault, and the corresponding frequency spectrum of amplitudes. Signal 1: The raw signal collected under motor load 0; Signal 2: Signal 1 with additional environmental noise of SNR = −8 dB; Signal 3: The signal of bearing ball fault, which is collected under motor load 3; Signal 4: Signal 3 with additional environmental noise of SNR = −8 dB.

Table 3
The information of the transfer tasks in this paper.

Transfer Task	Motor load in training data	Motor load in testing data	Additional noise (SNR)	Number of labeled samples	Number of unlabeled samples
T_1	0 hp	3 hp	$-8, -4, 0, 4, 8, +\infty$	$10 \times N_{train}$	$50 \times N_{test}$
T_2	3 hp	0 hp	$-8, -4, 0, 4, 8, +\infty$	$10 \times N_{train}$	$50 \times N_{test}$

difference in the training and testing working conditions is the largest.

Figs. 10 and 11 show the experimental results on the two transfer tasks with additional noise by different methods, where $N_{train} = 10$ is used. Basically, the display patterns observed in Figs. 10 and 11 are similar with each other, and they are both similar with the relationships in Fig. 8. Compared with the experiments under the same working condition as presented in Fig. 8, the transfer tasks are more difficult for fault diagnosis, and lower accuracies are generally obtained with the same level of environmental noises. For the fault diagnosis with variation of working condition, the proposed method achieves the best diagnosis results in most of the scenarios, and it is followed by the W/O Clustering approach. The testing accuracies obtained by the W/O MMD method and the Baseline approach are not competitive, while the former one is generally slightly better than the latter one.

Variation of bearing working condition and additional environmental noise pose significant obstacles for fault diagnosis. For instance, with respect to the Baseline method which achieves 99.92% accuracy in the ordinary task as Fig. 8 shows, its testing accuracy drops to 93.72% in T_1 without noise, and it further goes down to 51.76% with SNR = −4 dB. The proposed method shows good robustness in this situation, obtaining 99.43% accuracy in T_1 without noise, and when SNR = −4 dB is further applied, as high as 84.82% accuracy can be still achieved.

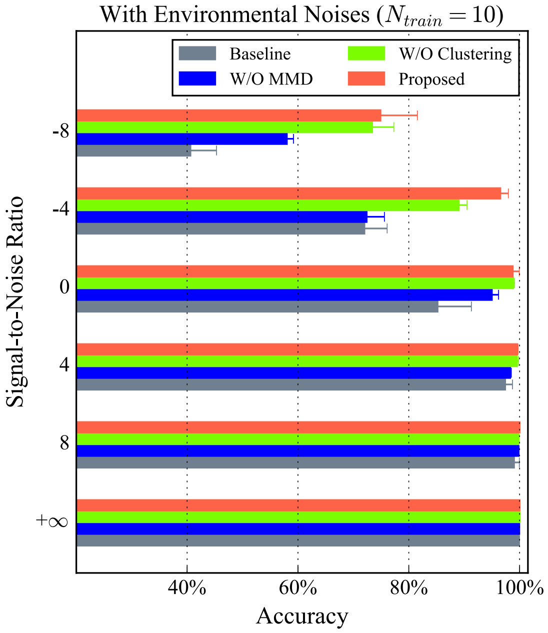


Fig. 8. Testing diagnosis results with additional environmental noise of different level by different methods. $N_{train} = 10$ is used.

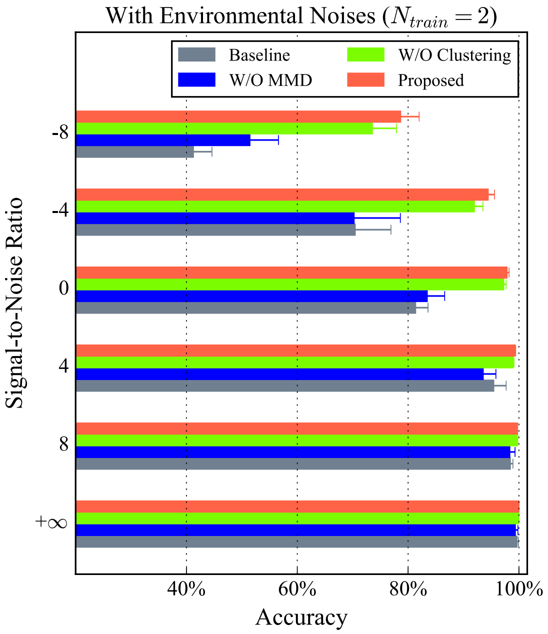


Fig. 9. Testing diagnosis results with additional environmental noise of different level by different methods. $N_{train} = 2$ is used.

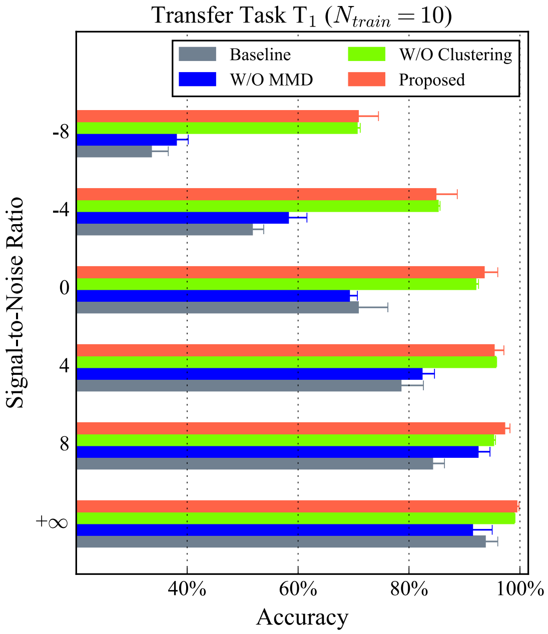


Fig. 10. Testing diagnosis results of transfer task T_1 with environmental noises by different methods. $N_{train} = 10$ is used.

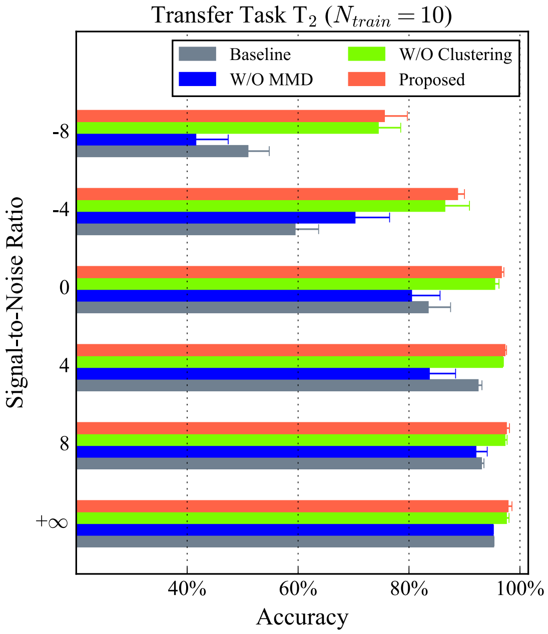


Fig. 11. Testing diagnosis results of transfer task T_2 with environmental noises by different methods. $N_{train} = 10$ is used.

Furthermore, in order to evaluate the effectiveness of the proposed method with limited training data, experiments are carried out with $N_{train} = 2$, that means only 20 labeled samples are used for training in the transfer tasks. The results for T_1 and T_2 are presented in Figs. 12 and 13, respectively. Compared with the numerical results with larger training dataset presented in the corresponding Figs. 10 and 11, no significant deterioration in testing accuracy is observed. That indicates the proposed method is able to provide reliable diagnosis in the cross-domain tasks even with very limited labeled training samples.

The experiments in this section show that the proposed deep distance metric learning method can significantly improve the diagnostic performance with variation of bearing working condition and additional environmental noise. Therefore, it is very promising for industrial applications.

4.5. Visualization of learned representation

In this section, the effectiveness of the proposed fault diagnosis method is illustrated qualitatively based on the visualizations of the learned data representations. Based on the fact that the last fully-connected layer in the network is directly responsible for the final classification performance, the visualizations of the fully-connected layer in different scenarios are presented for comparison.

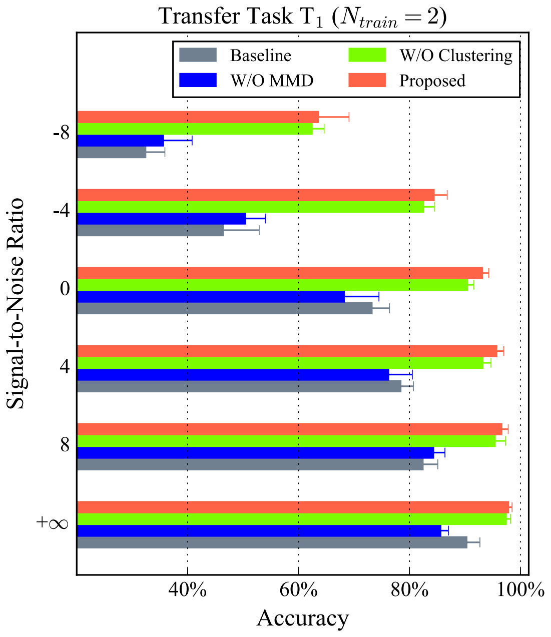


Fig. 12. Testing diagnosis results of transfer task T_1 with environmental noises by different methods. $N_{train} = 2$ is used.

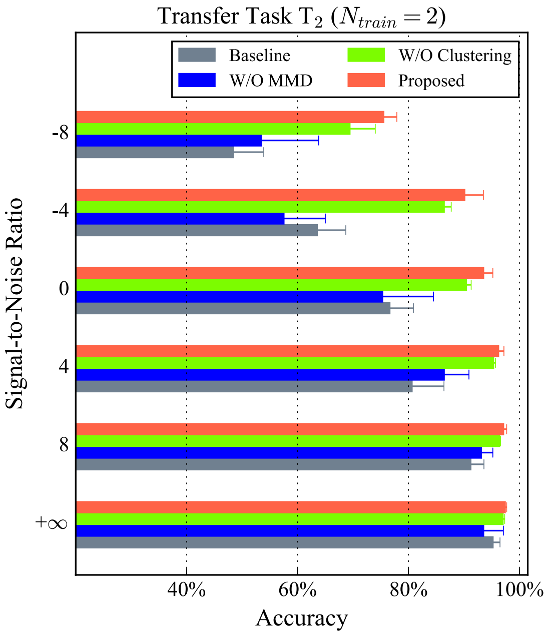


Fig. 13. Testing diagnosis results of transfer task T_2 with environmental noises by different methods. $N_{train} = 2$ is used.

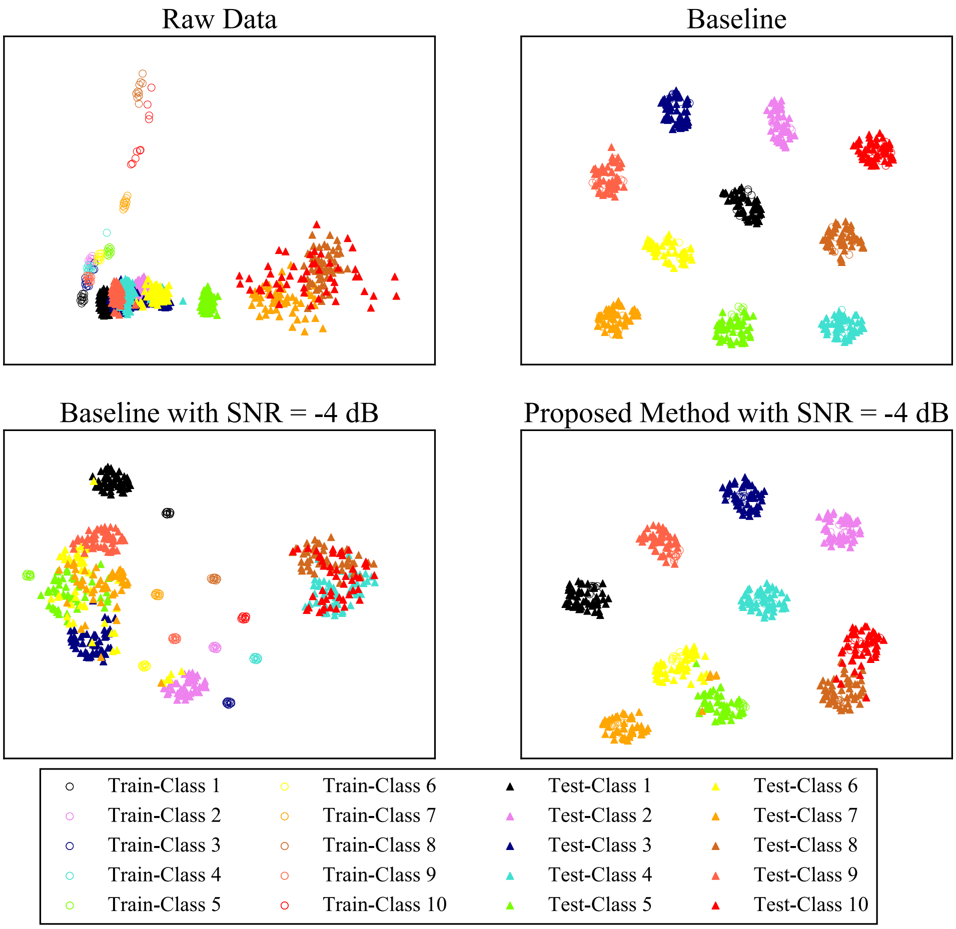


Fig. 14. Visualization of the learned representations in the fully-connected layer by the Baseline and the proposed method. Different levels of additional environmental noises are applied.

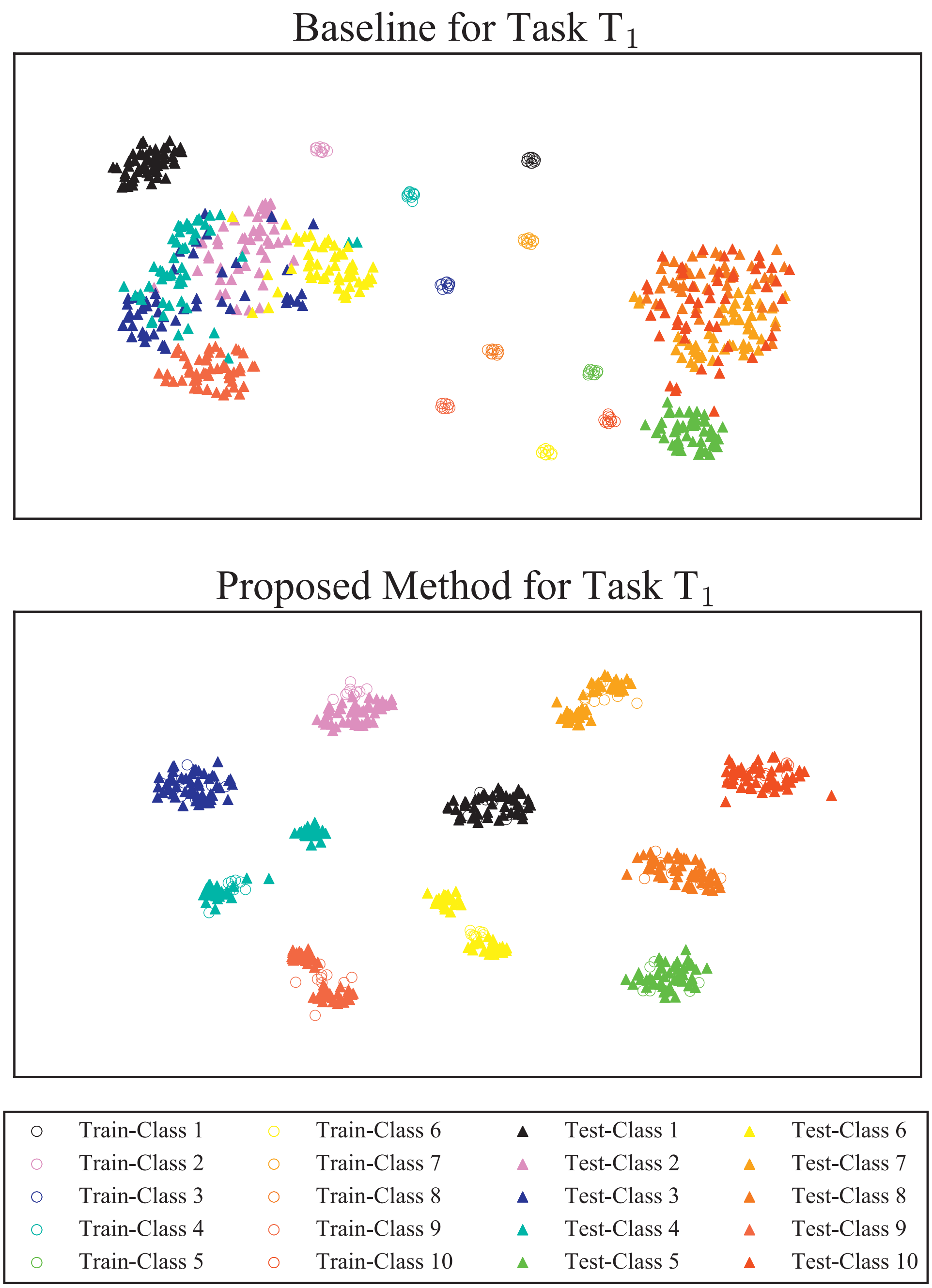


Fig. 15. Visualization of the learned representations in the fully-connected layer by the Baseline and the proposed method in the transfer task T_1.

An effective technique “t-SNE” is used to visualize the high-dimensional data representation by mapping the samples from the original feature space into a 2-dimensional space map [71]. The principal component analysis (PCA) is first adopted to reduce the dimensionality of the feature data to 50 and suppress signal noise. Afterwards, the technique “t-SNE” is used to convert the 50-dimensional learned representation to a 2-dimensional map.

4.5.1. With environmental noises

Fig. 14 shows the raw data distribution, and the maps of the learned representations for both the training and testing data by the Baseline and the proposed method. Different levels of additional environmental noise are applied, all the working conditions are considered, and $N_{train} = 10$ and $N_{test} = 50$ are used. The experimental setting is similar with that in Section 4.3. It can be observed that serious data overlappings of different fault classes exist in the raw data visualization, that indicates direct fault classification based on raw data can be extremely difficult and feature extraction is necessary. According to the numerical results presented in Fig. 8, the diagnosis task without additional noise is relatively simple and fairly high diagnosis accuracy close to 100% can be obtained by different methods. Excellent clustering phenomenon can be observed with the Baseline approach.

However, when strong additional environmental noise of SNR $= -4$ dB is applied, significant distribution discrepancy is observed between the training and testing data by the Baseline approach, and overlappings of different classes are remarkable. In this way, the trained classifier loses effectiveness in the testing dataset.

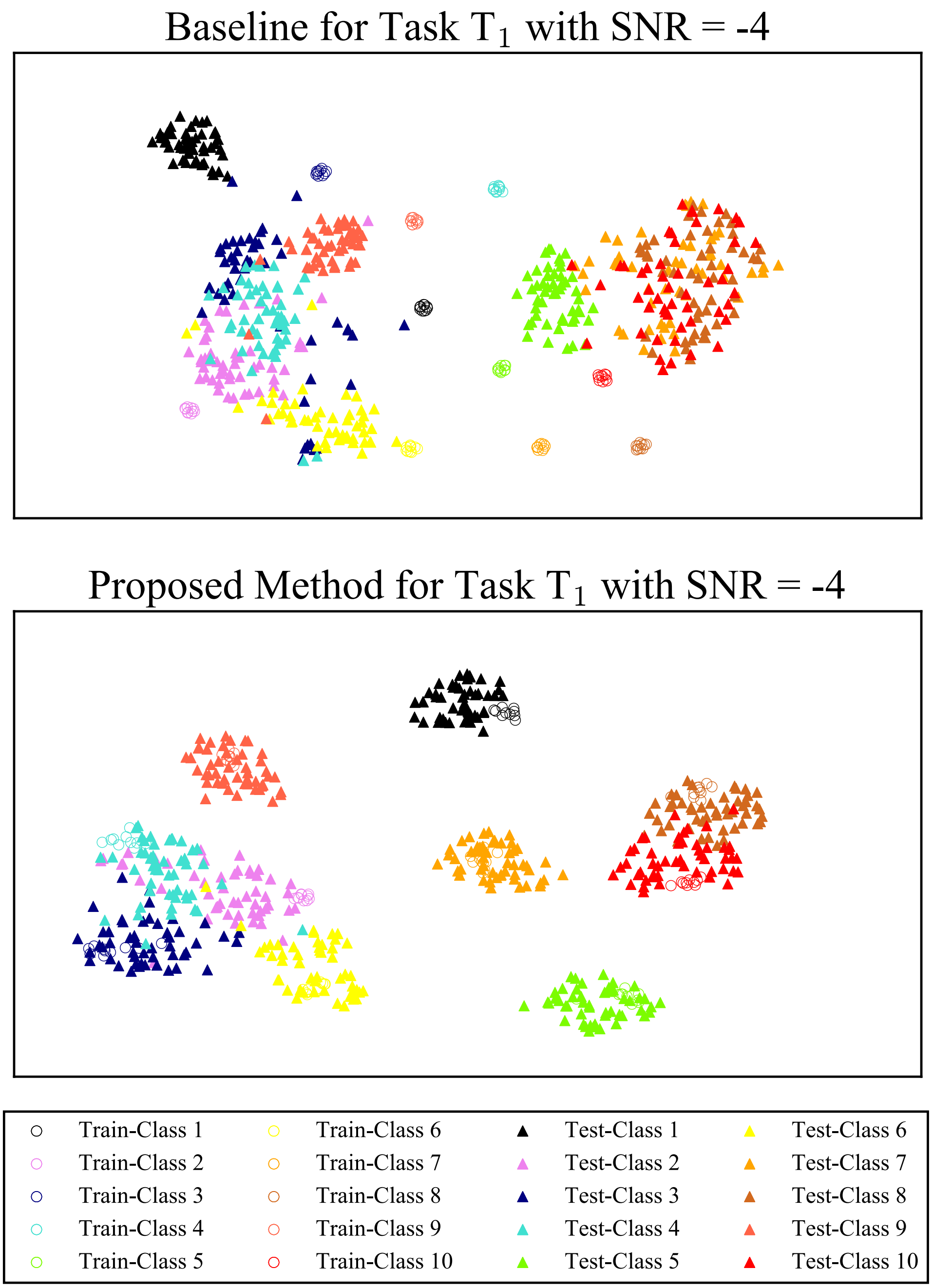


Fig. 16. Visualization of the learned representations in the fully-connected layer by the Baseline and the proposed method in the transfer task T_1. Additional environmental noise SNR $= -4$ dB is applied.

Therefore, the Baseline method is vulnerable to potential noise disturbance. On the other hand, the proposed method is able to perform good clustering with strong noise. The data samples of different fault types are separated well, and no noticeable distribution discrepancy between training and testing data is observed. Hence, the proposed method has shown strong robustness against environmental noises.

4.5.2. With variation of working condition

Next, the experiments with variation of working condition are visualized. Fig. 15 shows the resulting maps of the learned representations by different methods for the transfer task T_1. No additional noise is applied and $N_{train} = 10$ and $N_{test} = 50$ are used. The corresponding diagnosis results are presented in Fig. 10. The Baseline approach shows poor performance in clustering of different classes. While the training data can be separated well, a large number of data overlappings of different fault types in the testing data are observed, due to the serious domain shift problem. The Baseline method thus fails to precisely diagnose bearing faults with variation of working condition, and only 93.72% testing accuracy is obtained.

In the lower sub-figure in Fig. 15, the proposed method shows good clustering phenomenon of different classes, and the training and testing samples are mostly subject to the same distribution. That is the basis of accurate fault diagnosis, and consequently as high as 99.43% cross-domain testing accuracy is obtained.

Furthermore, Fig. 16 shows the visualizations of the learned representations by the Baseline and the proposed method for the

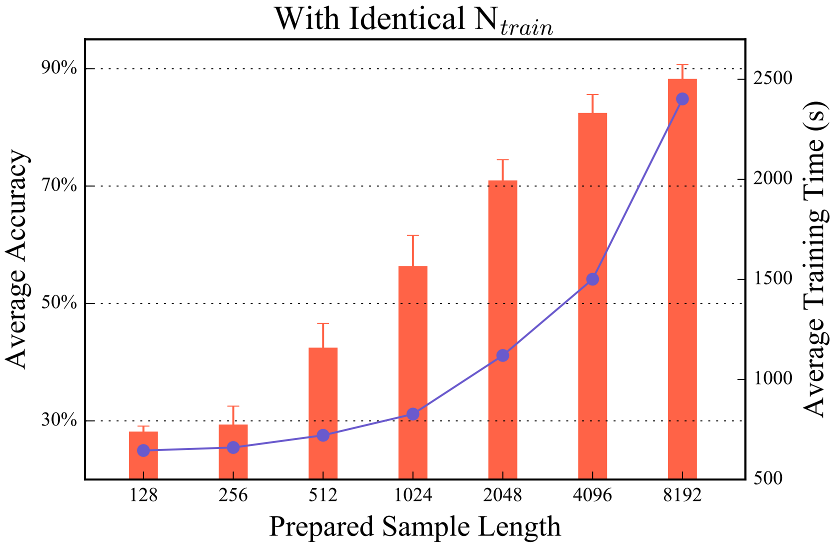


Fig. 17. Effects of the prepared sample length on the diagnosis performance. $N_{train} = 10$ is used. The red bars denote the average testing accuracies, and the blue lines represent the average training time (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

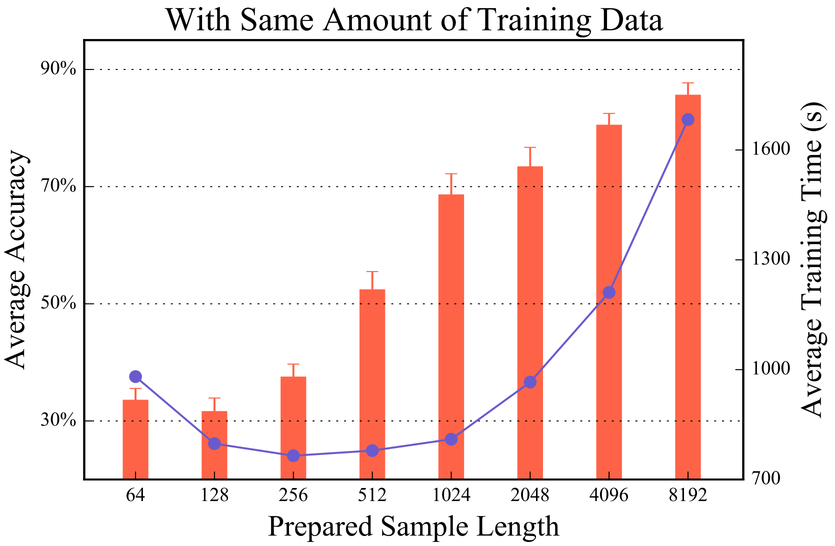


Fig. 18. Effects of the prepared sample length on the diagnosis performance with identical amount of training data. The red bars denote the average testing accuracies, and the blue lines represent the average training time (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

transfer task T_1 under strong environmental noise SNR = −4 dB. Significant distribution discrepancy between the training and testing data is observed by the Baseline approach, and the testing data of different classes merge into the same regions. Environmental noise brings strong disturbance to the collected signal, and only 51.76% testing accuracy is obtained by the Baseline approach in this noisy transfer task.

On the other hand, the proposed method still shows good distributions of the training and testing data in the difficult fault diagnosis task, and up to 84.82% testing result is achieved. Therefore, the proposed deep distance metric learning algorithm is able to significantly improve the fault diagnosis performance with variation of bearing working condition and strong environmental noise.

In summary, the visualizations of the learned representations by different methods are presented in this section. Since the features in the fully-connected layer are directly responsible for the final fault classification, the corresponding data distributions provide a general insight into the learning capacities of the neural network-based methods. It should be pointed out that the final classification is carried out in a high-dimensional space nonlinearly. Therefore, acceptable point overlappings for different health conditions in visualization agree with the high numerical classification accuracies.

4.6. Effects of parameters

In this section, the influence of the key parameters in the proposed method on the diagnosis performance is investigated, and the experimental setting is similar with those in previous sections. In order to better illustrate the effects, the relatively more difficult fault diagnosis transfer task T_1 as studied in Section 4.4 is focused on, additional environmental noise SNR = −8 dB is applied on the testing data, and $N_{train} = 10$ is adopted.

Fig. 17 shows the effects of the prepared data sample length on the diagnosis performance by the proposed method. It can be observed that longer sample length significantly leads to higher testing performance. As high as 93.52% testing accuracy can be achieved with the sample length of 8192 under the extreme sit-

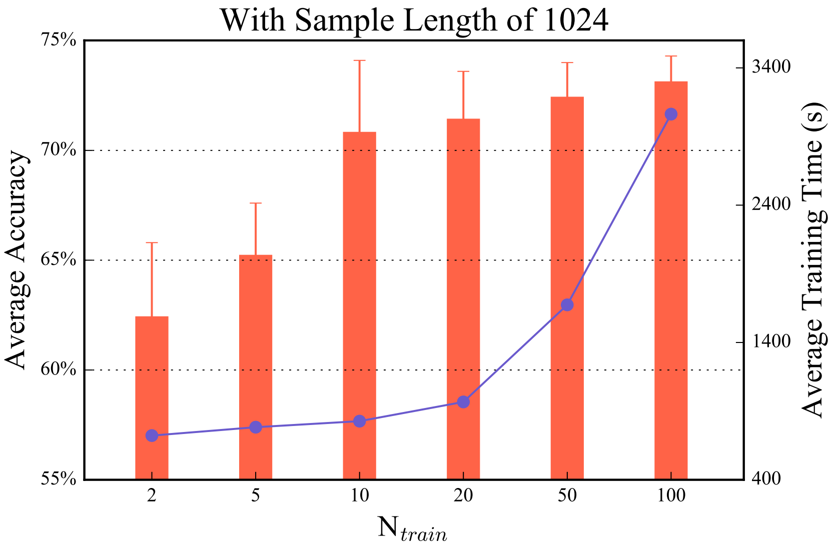


Fig. 19. Effects of the number of the training samples on the diagnosis performance. The prepared sample length is 1024. The red bars denote the average testing accuracies, and the blue lines represent the average training time (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

uation where both the variation of working condition and strong environmental noise are considered. When the sample length is short such as 128, the testing performance is poor and less than 30% accuracy is obtained. Correspondingly, longer sample length generally results in higher computational burden. A tradeoff has to be made between the diagnosis accuracy and off-line training time in applications.

Moreover, it is noted that one drawback in the above experiments lies in that, when identical number of samples are used for training with different sample length, the total amount of the training data changes. Based on the understanding of deep learning in the literature, more training data can generally improve the network performance. In order to eliminate this effect, two additional experiments are carried out.

First, the network performance is evaluated with the same amount of training data and varying sample length. The experiments with shorter sample length are enhanced with larger number of training samples, and vice versa. Specifically, the corresponding training data, whose size equals that of 1024 sample length and $N_{train} = 8,$ are applied for different sample length, and the results are presented in Fig. 18. The display pattern is observed to be similar with that in Fig. 17.

Next, using identical sample length of 1024 points, the effects of the number of training samples on the diagnosis performance are investigated and the results are presented in Fig. 19. While larger N_{train} generally leads to higher testing accuracy, the influence is not as significant as that using larger sample length shown in Fig. 17.

The above experiments indicate that with respect to the proposed fault diagnosis method based on frequency-domain feature extraction, extending the prepared sample length is able to significantly improve the testing diagnosis performance, which can not be achieved by increasing the number of training samples alone.

The effects of other hyper-parameters mostly follow the same observed patterns in the literature. For instance, larger convolutional filter size and more filters generally lead to better results with heavier computing load, appropriately deeper network also has the potential to further improve the diagnosis performance and so forth. Additionally, it is observed in the experiments that the weights α, β and γ in the objective function in Eq. (14) do not have significant influence on the diagnosis results, on the condition that γ is sufficiently larger than the rest two coefficients to maintain the numerical balance among different sub-objectives.

4.7. Comparing with related works

The rolling bearing dataset used in this paper is very popular in bearing fault diagnosis studies, and many state-of-the-art classification results have been reported in the past years. However, limited work can be found on the cross-domain problem with additional environmental noises, and most studies focus on diagnosing bearing health condition using the training and testing data from the same domain.

With respect to the general fault diagnosis problem without variation of working condition and additional noise, 95% and higher testing accuracies were achieved in [72–74] where 4 bearing health conditions or fewer were diagnosed. Regarding 10 or more bearing conditions, 88.9, 92.5 and 97.9% testing accuracies were obtained in [75–77], respectively. A two-stage machine learning method was proposed in [60] based on unsupervised feature learning and sparse filtering. Fairly high diagnosis accuracy of 99.66% was achieved.

In [40], the fault diagnosis with variation of working condition was studied, where the labeled data under 0 hp load were used as the training data and the unlabeled data under 3 hp load were tested. The case study is similar with the T_1 transfer task in this paper. Specifically, 4 health conditions were considered in [40], and 1000 samples of $N_{input} = 1200$ sample length were selected from both the two conditions for training. Up to 94.73% cross-domain fault classification accuracy was achieved. In [28], similar cross-domain study was carried out, and for the transfer task from 1 hp motor load to 3 hp, as high as 91.1% testing accuracy was obtained. Based on the results presented in previous sections, when the default experimental setting is adopted where the prepared sample length is 1024 and $N_{train} = 10,$ up to 99.43% testing accuracy can be achieved using the proposed method with respect to the transfer task T_1, which is higher than those in related studies. Especially, if the enhanced experimental setting is used where the prepared sample length is 8192 and $N_{train} = 10,$ excellent testing results of 99.84% accuracy can be achieved.

Regarding the additional environmental noise, 82.05% testing accuracy was obtained in [28] where the signal-to-noise ratio of −4 was applied. Using the proposed method with the default parameters, the testing accuracy is 96.53% under similar condition with SNR = −4. The accuracy increases to 99.98% if the enhanced setting is adopted. Moreover, when even stronger noise is consid-

Table 4
Comparisons of testing diagnosis accuracy of related researches on the same rolling bearing dataset. The proposed method uses the default experimental setting with sample length of 1024 and $N_{train} = 10$. In the enhanced proposed method, the sample length is 8192, and $N_{train} = 10$.

Method	Problem description	Number of fault classes	Testing accuracy (%)	Motor loads (Training → Testing)	Singal-to-noise ratio (dB)
[72]	Training and testing data are subject to the same distribution	4	95.8	0, 1, 2, 3 → 0, 1, 2, 3	$+\infty$
[75]		10	88.9	0, 1, 2, 3 → 0, 1, 2, 3	$+\infty$
[76]		10	92.5	0, 1, 2, 3 → 0, 1, 2, 3	$+\infty$
[77]		11	97.91	0, 1, 2, 3 → 0, 1, 2, 3	$+\infty$
[60]		10	99.66	0, 1, 2, 3 → 0, 1, 2, 3	$+\infty$
Proposed		10	**100**	0, 1, 2, 3 → 0, 1, 2, 3	$+\infty$
[40]	With variation of working condition only	4	94.73	0 → 3	$+\infty$
[28]		10	91.1	1 → 3	$+\infty$
[28]		10	90.2	3 → 1	$+\infty$
Proposed		10	**99.43**	0 → 3	$+\infty$
Proposed (Enhanced)		10	**99.84**	0 → 3	$+\infty$
Proposed		10	**97.82**	3 → 0	$+\infty$
Proposed (Enhanced)		10	**99.31**	3 → 0	$+\infty$
[28]	With additional environmental noise only	10	82.05	1, 2, 3 → 1, 2, 3	−4
Proposed		10	**96.53**	0, 1, 2, 3 → 0, 1, 2, 3	−4
Proposed (Enhanced)		10	**99.98**	0, 1, 2, 3 → 0, 1, 2, 3	−4
Proposed		10	**74.90**	0, 1, 2, 3 → 0, 1, 2, 3	−8
Proposed (Enhanced)		10	**98.79**	0, 1, 2, 3 → 0, 1, 2, 3	−8
Proposed	With additional noise and variation of working condition	10	**84.82**	0 → 3	−4
Proposed (Enhanced)		10	**99.34**	0 → 3	−4
Proposed		10	**84.45**	3 → 0	−4
Proposed (Enhanced)		10	**98.87**	3 → 0	−4

ered with $\mathrm{SNR} = -8$, good diagnosis results can be still achieved with the average testing accuracy of 74.90% by the default setting, and up to 98.79% accuracy can be obtained with the enhanced parameters.

Furthermore, the proposed method is able to perform good fault diagnosis with both additional environmental noise and variation of working condition. On the transfer task T_1 with noise of $\mathrm{SNR} = -4$, as high as 84.82% testing accuracy can be obtained with the default setting. Even higher testing accuracy of 99.34% can be achieved in this difficult fault diagnosis task if the prepared sample length is 8192.

The detailed comparison results with the related works on the same rolling bearing dataset are presented in Table 4. It is seen that the proposed method outperforms the other approaches in different scenarios.

It should be noted that all the comparison results by the proposed method are obtained with a small number of training samples $N_{train} = 10$ even in the enhanced experimental setting. Based on the understanding of the neural network, higher testing accuracy is expected to be achieved if more labeled data are used for training. Therefore, the proposed method is very competitive with the related researches and promising for bearing fault diagnosis applications.

5. Conclusions

This paper proposes a novel fault diagnosis method based on deep learning for rolling element bearings. With respect to the extracted high-level features by the deep neural network, a distance metric learning algorithm is proposed to address the practical industrial problems, i.e. domain shift and environmental noise effect. A representation clustering approach is adopted to minimize the distance of intra-class variations and maximize the distance of inter-class variations simultaneously. A domain adaptation method is used in order to minimize the distribution discrepancy between the training and testing data. In this way, the robustness of the proposed method can be largely improved.

The effectiveness of the proposed method is validated by extensive experiments on a popular rolling bearing dataset. The robustness of the proposed method against environmental noises and variation of working conditions is demonstrated. The influence of the key parameters in the proposed framework is investigated, and longer prepared sample length is observed to lead to higher diagnosis accuracy with increased computational burden. Comparisons with other approaches and related researches on the same dataset are provided to verify the superiority of the proposed approach. Based on the experimental results in this study, the proposed method is very promising for bearing fault diagnosis in real industries.

Especially, fairly high testing diagnosis accuracy is achieved by the proposed method with the sample length of 8192, and the average training time is around 40 min. The computing burden is considered acceptable since the neural network is trained off-line. On the other hand, reasonably fewer epochs can be used to reduce the training time which could also lead to converged results.

It should be noted that this paper focuses on the proposed deep distance metric learning method, and a relatively basic neural network architecture is thus used for illustration. Optimization on the network structure can be further carried out to fully explore the potential of the proposed method.

Acknowledgements

The first author would like to thank Dr. Wen-Qi Ren for the valuable advices on this paper. The material in this paper is based on work supported by grants (N170503012, N170308028) from the Fundamental Research Funds for the Central Universities.

References

[1] H. Sun, Z. He, Y. Zi, J. Yuan, X. Wang, J. Chen, S. He, Multiwavelet transform and its applications in mechanical fault diagnosis - a review, Mech. Syst. Signal Process. 43 (1–2) (2014) 1–24.

[2] P.W. Tse, Y.H. Peng, R. Yam, Wavelet analysis and envelope detection for rolling element bearing fault diagnosis - their effectiveness and flexibilities, J. Vib. Acoust. 123 (3) (2001) 303–310.

[3] Z. Ren, S. Zhou, C.E. M. Gong, B. Li, B. Wen, Crack fault diagnosis of rotor systems using wavelet transforms, Comput. Electr. Eng. 45 (2015) 33–41.

[4] X.H. Chen, G. Cheng, X.L. Shan, X. Hu, Q. Guo, H.G. Liu, Research of weak fault feature information extraction of planetary gear based on ensemble empirical mode decomposition and adaptive stochastic resonance, Measurement 73 (2015) 55–67.

[5] P. Zhou, S. Lu, F. Liu, Y. Liu, G. Li, J. Zhao, Novel synthetic index-based adaptive stochastic resonance method and its application in bearing fault diagnosis, J. Sound Vib. 391 (2017) 194–210.

[6] G. He, K. Ding, H. Lin, Fault feature extraction of rolling element bearings using sparse representation, J. Sound Vib. 366 (2016) 514–527.

[7] M. Žvokelj, S. Zupan, I. Prebil, EEMD-based multiscale ICA method for slewing bearing fault detection and diagnosis, J. Sound Vib. 370 (2016) 394–423.

[8] S. Lu, X. Wang, Q. He, F. Liu, Y. Liu, Fault diagnosis of motor bearing with speed fluctuation via angular resampling of transient sound signals, J. Sound Vib. 385 (2016) 16–32.

[9] H.L. He, T.Y. Wang, Y.G. Leng, Y. Zhang, Q. Li, Study on non-linear filter characteristic and engineering application of cascaded bistable stochastic resonance system, Mech. Syst. Signal Process. 21 (7) (2007) 2740–2749.

[10] Q. Li, T. Wang, Y. Leng, W. Wang, G. Wang, Engineering signal processing based on adaptive step-changed stochastic resonance, Mech. Syst. Signal Process. 21 (5) (2007) 2267–2279.

[11] G.F. Bin, J.J. Gao, X.J. Li, B.S. Dhillon, Early fault diagnosis of rotating machinery based on wavelet packets - Empirical mode decomposition feature extraction and neural network, Mech. Syst. Signal Process. 27 (2012) 696–711.

[12] V.T. Tran, B.S. Yang, F. Gu, A. Ball, Thermal image enhancement using bi-dimensional empirical mode decomposition in combination with relevance vector machine for rotating machinery fault diagnosis, Mech. Syst. Signal Process. 38 (2) (2013) 601–614.

[13] Z. Li, H. Fang, M. Huang, Diversified learning for continuous hidden Markov models with application to fault diagnosis, Expert Syst. Appl. 42 (23) (2015) 9165–9173.

[14] A. Youssef, C. Delpha, D. Diallo, An optimal fault detection threshold for early detection using Kullback–Leibler divergence for unknown distribution data, Signal Process. 120 (2016) 266–279.

[15] X. Zhang, W. Chen, B. Wang, X. Chen, Intelligent fault diagnosis of rotating machinery using support vector machine with ant colony algorithm for synchronous feature selection and parameter optimization, Neurocomputing 167 (2015) 260–279.

[16] R. Jegadeeshwaran, V. Sugumaran, Fault diagnosis of automobile hydraulic brake system using statistical features and support vector machines, Mech. Syst. Signal Process. 52–53 (2015) 436–446.

[17] B. Samanta, C. Nataraj, Use of particle swarm optimization for machinery fault detection, Eng. Appl. Artif. Intell. 22 (2) (2009) 308–316.

[18] B.S. Yang, X. Di, T. Han, Random forests classifier for machine fault diagnosis, J. Mech. Sci. Technol. 22 (9) (2008) 1716–1725.

[19] G.E. Hinton, R.R. Salakhutdinov, Reducing the dimensionality of data with neural networks, Science 313 (5786) (2006) 504.

[20] C. Lu, Z.Y. Wang, W.L. Qin, J. Ma, Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification, Signal Process. 130 (2017) 377–388.

[21] W. Sun, S. Shao, R. Zhao, R. Yan, X. Zhang, X. Chen, A sparse auto-encoder-based deep neural network approach for induction motor faults classification, Measurement 89 (2016) 171–178.

[22] W.T. Mao, J.L. He, Y. Li, Y.J. Yan, Bearing fault diagnosis with auto-encoder extreme learning machine: a comparative study, Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci. (2016). 0954406216675896.

[23] X. Guo, L. Chen, C. Shen, Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis, Measurement 93 (2016) 490–502.

[24] W. Sun, R. Zhao, R. Yan, S. Shao, X. Chen, Convolutional discriminative feature learning for induction motor fault diagnosis, IEEE Trans. Ind. Inf. 13 (3) (2017) 1350–1359.

[25] O. Abdeljaber, O. Avci, S. Kiranyaz, M. Gabbouj, D.J. Inman, Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks, J. Sound Vib. 388 (2017) 154–170.

[26] T. Ince, S. Kiranyaz, L. Eren, M. Askar, M. Gabbouj, Real-time motor fault detection by 1-D convolutional neural networks, IEEE Trans. Ind. Electron. 63 (11) (2016) 7067–7075.

[27] G. Csurka, Domain adaptation for visual applications: a comprehensive survey (2017), arXiv:1702.05374.

[28] W. Zhang, C. Li, G. Peng, Y. Chen, 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. 100 (2018) 439–453.

[29] K. Weinberger, L. Saul, Distance metric learning for large margin nearest neighbor classification, J. Mach. Learn. Res. 10 (2009) 207–244.

[30] J.V. Davis, B. Kulis, P. Jain, S. Sra, I.S. Dhillon, Information-theoretic metric learning, in: Proceedings of the Twenty-Fourth International Conference on Machine Learning, 2007, pp. 209–216.

[31] M. Köstinger, M. Hirzer, P. Wohlhart, P.M. Roth, H. Bischof, Large scale metric learning from equivalence constraints, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 2288–2295.

[32] D. Tran, A. Sorokin, Human activity recognition with metric learning, in: D. Forsyth, P. Torr, A. Zisserman (Eds.), Computer Vision - ECCV, Springer Berlin Heidelberg, Berlin, Heidelberg, 2008, pp. 548–561.

[33] X. Peng, L. Zhang, Z. Yi, Scalable sparse subspace clustering, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 430–437.

[34] Z. Li, J. Liu, J. Tang, H. Lu, Robust structured subspace learning for data representation, IEEE Trans. Pattern Anal. Mach. Intell. 37 (10) (2015) 2085–2098.

[35] X. Peng, Z. Yu, Z. Yi, H. Tang, Constructing the L2-graph for robust subspace learning and subspace clustering, IEEE Trans. Cybern. 47 (4) (2017) 1053–1066.

[36] X. Peng, C. Lu, Z. Yi, H. Tang, Connections between nuclear-norm and frobenius-norm-based representations, IEEE Trans. Neural Netw. Learn. Syst. 29 (1) (2018) 218–224.

[37] X. Peng, H. Tang, L. Zhang, Z. Yi, S. Xiao, A unified framework for representation-based subspace clustering of out-of-sample and large-scale data, IEEE Trans. Neural Netw. Learn. Syst. 27 (12) (2016) 2499–2512.

[38] X. Peng, J. Lu, Z. Yi, R. Yan, Automatic subspace learning via principal coefficients embedding, IEEE Trans. Cybern. 47 (11) (2017) 3583–3596.

[39] W. Zhang, G. Peng, C. Li, Y. Chen, Z. Zhang, A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals, Sensors 17 (2) (2017) 425.

[40] W. Lu, B. Liang, Y. Cheng, D. Meng, J. Yang, T. Zhang, Deep model based domain adaptation for fault diagnosis, IEEE Trans. Ind. Electron. 64 (3) (2017) 2296–2305.

[41] J. Xie, L. Zhang, L. Duan, J. Wang, On cross-domain feature fusion in gearbox fault diagnosis under various operating conditions based on transfer component analysis, in: Proceedings of the IEEE International Conference on Prognostics and Health Management, 2016, pp. 1–6.

[42] X. Li, Q. Ding, J.Q. Sun, Remaining useful life estimation in prognostics using deep convolution neural networks, Reliab. Eng. Syst. Saf. 172 (2018) 1–11.

[43] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, in: Proceedings of the Twenty-Sixth Annual Conference on Neural Information Processing Systems, 2, 2012, pp. 1097–1105.

[44] A. Globerson, S. Roweis, Metric learning by collapsing classes, in: Proceedings of the Eighteenth International Conference on Neural Information Processing Systems, Vancouver, British Columbia, Canada, 2005, pp. 451–458.

[45] M. Long, Y. Cao, J. Wang, M. Jordan, Learning transferable features with deep adaptation networks, in: Proceedings of the Thirty-Second International Conference on Machine Learning, 37, 2015, pp. 97–105.

[46] X. Glorot, A. Bordes, Y. Bengio, Domain adaptation for large-scale sentiment classification: a deep learning approach, in: Proceedings of the Twenty-Eighth International Conference on Machine Learning, 2011, pp. 513–520.

[47] B. Fernando, A. Habrard, M. Sebban, T. Tuytelaars, Unsupervised visual domain adaptation using subspace alignment, in: Proceedings of the IEEE International Conference on Computer Vision, 2013, pp. 2960–2967.

[48] B. Gong, Y. Shi, F. Sha, K. Grauman, Geodesic flow kernel for unsupervised domain adaptation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 2066–2073.

[49] A. Sharma, D.W. Jacobs, Bypassing synthesis: PLS for face recognition with pose, low-resolution and sketch, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2011, pp. 593–600.

[50] L. Samarakoon, K.C. Sim, On combining i-vectors and discriminative adaptation methods for unsupervised speaker normalization in DNN acoustic models, in: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, 2016, pp. 5275–5279.

[51] L. Duan, D. Xu, S.F. Chang, Exploiting web images for event recognition in consumer videos: a multiple source domain adaptation approach, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 1338–1345.

[52] X. Wang, J. Schneider, Flexible transfer learning under support and model shift, in: Proceedings of the IEEE Twenty-Seventh International Conference on Neural Information Processing Systems, 2014, pp. 1898–1906.

[53] B. Gong, K. Grauman, F. Sha, Connecting the dots with landmarks: discriminatively learning domain-invariant features for unsupervised domain adaptation, in: Proceedings of the Thirtieth International Conference on Machine Learning, 28, 2013, pp. 222–230.

[54] S.J. Pan, I.W. Tsang, J.T. Kwok, Q. Yang, Domain adaptation via transfer component analysis, IEEE Trans. Neural Netw. 22 (2) (2011) 199–210.

[55] J. Yosinski, J. Clune, Y. Bengio, H. Lipson, How transferable are features in deep neural networks? in: Proceedings of the Twenty-Seventh International Conference on Neural Information Processing Systems, 2014, pp. 3320–3328.

[56] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, T. Darrell, DeCAF: a deep convolutional activation feature for generic visual recognition, in: Proceedings of the Thirty-First International Conference on Machine Learning, 32, 2014, pp. 647–655.

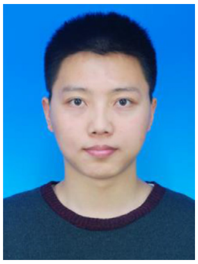
[57] A. Gretton, K. Borgwardt, M. Rasch, B. Schölkopf, A. Smola, A kernel two-sample test, J. Mach. Learn. Res. 13 (2012) 723–773.

[58] A. Gretton, D. Sejdinovic, H. Strathmann, S. Balakrishnan, M. Pontil, K. Fukumizu, B.K. Sriperumbudur, Optimal Kernel Choice for Large-Scale Two-Sample Tests, Curran Associates, Inc., 2012.

[59] Y. Li, K. Swersky, R. Zemel, Generative moment matching networks, in: Proceedings of Thirty-Second International Conference on Machine Learning, 2015, pp. 1718–1727.

[60] Y. Lei, F. Jia, J. Lin, S. Xing, S.X. Ding, An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data, IEEE Trans. Ind. Electron. 63 (5) (2016) 3137–3147.

[61] B. Liu, J. Liu, X. Bai, H. Lu, Regularized hierarchical feature learning with non-negative sparsity and selectivity for image classification, in: Proceedings of Twenty-Second International Conference on Pattern Recognition, 2014, pp. 4293–4298.
[62] S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, in: Proceedings of Thiety-Second International Conference on Machine Learning, 1, Lile, France, 2015, pp. 448–456.
[63] G.E. Dahl, T.N. Sainath, G.E. Hinton, Improving deep neural networks for LVCSR using rectified linear units and dropout, in: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, 2013, pp. 8609–8613.
[64] A.L. Maas, A.Y. Hannun, A.Y. Ng, Rectifier nonlinearities improve neural network acoustic models, in: Proceedings of Thirtieth International Conference on Machine Learning, 28, 2013.
[65] G. Hinton, O. Vinyals, J. Dean, Distilling the knowledge in a neural network (2015), arXiv:1503.02531.
[66] R. Aljundi, T. Tuytelaars, Lightweight unsupervised domain adaptation by convolutional filter reconstruction, in: Proceedings of the European Conference on Computer Vision ECCV Workshops, Springer International Publishing, Cham, 2016, pp. 508–515.
[67] X. Glorot, Y. Bengio, Understanding the difficulty of training deep feedforward neural networks, J. Mach. Learn. Res. 9 (2010) 249–256.
[68] D.E. Rumelhart, G.E. Hinton, R.J. Williams, Learning representations by back--propagating errors, Nature 323 (6088) (1986) 533–536.
[69] D. Kingma, J. Ba, Adam: a method for stochastic optimization (2014), arXiv:1412.6980.
[70] W.A. Smith, R.B. Randall, Rolling element bearing diagnostics using the Case Western Reserve University data: a benchmark study, Mech. Syst. Signal Process. 64–65 (2015) 100–131.
[71] L. Maaten, G. Hinton, Visualizing data using t-SNE, J. Mach. Learn. Res. 9 (2008) 2579–2625.
[72] W. Li, S. Zhang, G. He, Semisupervised distance-preserving self-organizing map for machine-defect detection and classification, IEEE Trans. Inst. Meas. 62 (5) (2013) 869–879.
[73] B.J. van Wyk, M.A. van Wyk, G. Qi, Difference histograms: a new tool for time series analysis applied to bearing fault diagnosis, Pattern Recognit. Lett. 30 (6) (2009) 595–599.
[74] B. Muruganatham, M.A. Sanjith, B. Krishnakumar, S.A.V.S. Murty, Roller element bearing fault diagnosis using singular spectrum analysis, Mech. Syst. Signal Process. 35 (1–2) (2013) 150–166.
[75] W. Du, J. Tao, Y. Li, C. Liu, Wavelet leaders multifractal features based fault diagnosis of rotating mechanism, Mech. Syst. Signal Process. 43 (1–2) (2014) 57–75.
[76] X. Jin, M. Zhao, T.W.S. Chow, M. Pecht, Motor bearing fault diagnosis using trace ratio linear discriminant analysis, IEEE Trans. Ind. Electron. 61 (5) (2014) 2441–2451.
[77] X. Zhang, Y. Liang, J. Zhou, Y. Zang, A novel bearing fault diagnosis model integrated permutation entropy, ensemble empirical mode decomposition and optimized SVM, Measurement 69 (2015) 164–179.



Xiang Li obtained his Ph.D. in Mechanics from Tianjin University in 2017, and received the double Bachelor degrees in Engineering Mechanics and Engineering Management from Tianjin University in 2012. He is currently a lecturer in College of Sciences at Northeastern University, China. His research interests include deep learning, fault diagnosis, machinery health management and multi-objective optimization algorithm.

Wei Zhang obtained her Ph.D. in Mechanics from Tianjin University in 2017. She is currently a lecturer in School of Aerospace Engineering, Shenyang Aerospace University, China. Her research interests include rotor dynamics, multi-objective optimization algorithm and squeeze film damper.

Qian Ding obtained his Ph.D. in Mechanics from Tianjin University in 1997. He is currently a professor in Department of Mechanics at Tianjin University, China. His research interests include rotor dynamics, machinery fault diagnosis, deep learning and control systems.