IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.Doi Number

A Deep Transfer Nonnegativity-constraint Sparse Autoencoder for Rolling Bearing Fault Diagnosis with Few Labeled Data

Xingqiu Li[1], Hongkai Jiang[1], Ke Zhao[1], and Ruixin Wang[1]
[1]School of Aeronautics, Northwestern Polytechnical University, Xi'an 710072, China

Corresponding author: Hongkai Jiang (jianghk@nwpu.edu.cn).

This work is supported by the major research plan of the National Natural Science Foundation of China (No. 91860124), the National Natural Science Foundation of China(No. 51875459), the Aeronautical Science Foundation of China (No. 20170253003) and the Synergy Innovation Foundation of the University and Enterprise for Graduate Students in Northwestern Polytechnical University (No. XQ201901).

ABSTRACT Rolling bearing fault diagnosis can greatly improve the safety of rotating machinery. In some cases, plenty of labeled data are unavailable, which may lead to low diagnosis accuracy. To deal with this problem, a deep transfer nonnegativity-constraint sparse autoencoder is proposed, which takes advantage of deep learning and transfer learning. Firstly, a novel nonnegativity-constraint sparse autoencoder (NSAE) is adopted to enhance sparsity. Then, a base deep NSAE (DNSAE) is established to automatically capture the latent features from raw vibration signals. Next, parameter transfer learning strategy is used to build a deep transfer NSAE (DTNSAE) to tackle the diagnosis problems with few labeled data. Finally, two datasets from different domains are used to verify the effectiveness of the proposed method. The testing results suggest that the proposed method is able to remove manual feature extraction and is more effective than the existing intelligent methods when only few labeled data are available.

INDEX TERMS Rolling bearing, fault diagnosis, deep transfer nonnegativity-constraint sparse autoencoder, parameter transfer learning, few labeled data.

I. INTRODUCTION

Rolling bearing plays an important role in rotating machinery. Any unexpected faults of rolling bearing may result in downtown and economic losses [1]-[3]. To enhance the reliability and security, rolling bearing fault diagnosis has attracted more and more attention in recent years.

Traditional fault diagnosis methods usually utilize artificial neural network (ANN) [4]-[6] and support vector machine (SVM) [7]-[9], for fault recognition. Several results have pointed out that the shallow models have limited nonlinear feature learning capability in fault diagnosis problems [10]. Thus, manual feature extraction is a critical step to ensure high diagnosis accuracy. However, manual feature extraction with signal processing techniques requires related domain expertise and prior knowledge as well as consumes amounts of time [11]. Furthermore, the handcrafted features usually show poor generalization ability, that is, new sensitive features have to be extracted over again for different problems [12].

To get rid of manual feature extraction, deep learning has been introduced for fault diagnosis in most recent years [13]. Through multilayer nonlinear transformations, deep learning methods automatically learn useful features from raw vibration signals [14]-[16], and many satisfactory diagnosis results have been published so far. For example, Shao et al. used convolutional deep belief network and deep wavelet auto-encode for rolling bearing fault diagnosis [17], [18]. Jiang et al. utilized a deep recurrent neural network for rolling bearing fault recognition [19]. Ma et al. applied a deep coupling autoencoder for fault recognition [20]. Pan et al. proposed a convolutional neural network for rolling bearing fault diagnosis [21].

However, deep learning methods mentioned above generally require a substantial set of labeled training data which have same distribution with the testing data. Some research has suggested that the testing accuracy will greatly decrease when only few labeled data are used to train the deep network [22]. In some cases especially in real-world applications, plenty of high-quality labeled data are often

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2926234, IEEE Access

unavailable. To recollect and label sufficient data are sometimes extremely expensive or even impossible. Moreover, some related existing data are often not fully used. Therefore, it is quite meaningful to develop a suitable method for rolling bearing fault diagnosis with few labeled data.

Transfer learning has emerged as a prospective machine learning framework to tackle this problem. The core idea of transfer learning is to improve the performance of the target task utilizing the knowledge acquired from some different but related tasks (source tasks) when few or even no labeled target domain data are available [23]. At present, transfer learning has been successfully applied for image recognition, text classification as well as fault diagnosis [24]. For instance, Guo et al. proposed a maximum mean discrepancy based deep domain adaption network for rolling bearing fault diagnosis with unlabeled target domain data [25]. Because no labeled target domain data were available, which greatly increased the difficulty of fault diagnosis, the final diagnosis accuracy was still lower than 90%. Shao et al. improved the training efficiency and diagnosis accuracy using parameter transfer learning [26]. In that paper, the authors implemented parameter transfer learning based on two unrelated domains and acquire highly accurate results. However, their proposed method still required amount of labeled data and signal preprocessing. Considered that parameter transfer learning is a useful and easy-to-implement method, it is adopted to deal with the fault problems with few labeled target domain data in this paper.

In this paper, a novel deep transfer nonnegativity-constraint sparse autoencoder (DTNSAE) is proposed for rolling bearing fault diagnosis with few labeled target domain data. The proposed method makes full use of deep learning and parameter transfer learning. Firstly, a novel nonnegativity-constraint sparse auto encoder (NSAE) is constructed to enhance sparsity [27]. Secondly, multiple individual NSAEs are stacked to construct a base deep NSAE (DNSAE) for adaptive feature learning. Finally, parameter transfer learning is applied for constructing a DTNSAE by combining with DNSAE. The weights and biases of the base DNSAE and softmax classifier pre-trained with source domain data are transferred to initialize the target network with a same architecture, and then these parameters are fine-tuned with a small set of labeled target domain data. In the paper, two datasets from different domains are used to verify the effectiveness of the proposed method. The testing results show that the proposed method is not only able to remove manual feature extraction but also is more effective than existing traditional methods and deep learning methods. The main contributions of this paper can be summarized as follows:

1) To remove manual feature extraction, a base DNSAE is constructed to automatically capture useful features from raw vibration signals.

2) To tackle the diagnosis problem with few labeled data, parameter transfer learning is used to improve the performance of target task.

3) To make advantage of deep learning and parameter transfer learning, a deep transfer learning framework called DTNSAE is proposed for rolling bearing fault diagnosis. The testing results show that it is much more effective than the traditional methods and deep learning methods.

The rest of the paper is organized as follows. In Section II, basic theory of sparse autoencoder is introduced. In Section III, the proposed method is described in detail. In Section IV, experiment verification is carried out and the results are analyzed. In Section V, conclusions are summarized.

II. BASIC THEORY OF SPARSE AUTOENCODER

As is shown in Fig.1, conventional sparse autoencoder comprises two parts, one is the encoder and the other is the decoder [28]. The encoder is used to obtain the representative features of input data, and the decoder is able to reconstruct the input data from the representative features.

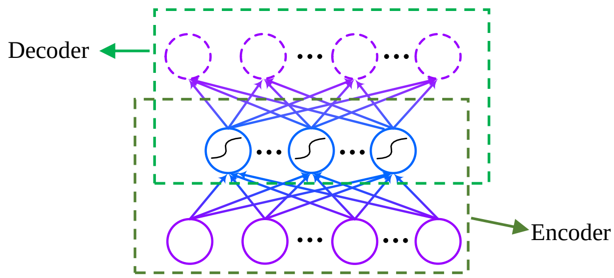

FIGURE 1. Architecture of sparse autoencoder.

Given the input data $\mathbf{X} = \{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_i, \ldots, \mathbf{X}_n\}$, $\mathbf{X}_i = \{x_1, x_2, \ldots, x_j, \ldots, x_m\}$ represents the ith sample, which comprises m data points. The encoder learns the representation by

$$\mathbf{h} = f_s(\mathbf{W}_1\mathbf{X} + \mathbf{b}_1) \tag{1}$$

$$f_s(x) = 1/(1+\exp(-x)) \tag{2}$$

where $\mathbf{W}_1$ denotes the weight matrix between the input layer and the encoding layer, and $\mathbf{b}_1$ is the corresponding bias vector. Then encoding features h are input to the decoder to acquire the reconstruction data by

$$\hat{\mathbf{X}} = f_s(\mathbf{W}_2\mathbf{h} + \mathbf{b}_2) \tag{3}$$

where $\mathbf{W}_2$ represents the weights between the encoding layer and decoding layer, and $\mathbf{b}_2$ is the corresponding bias vector. The learning process is to minimize the cost function

$$\begin{aligned} J_{\mathrm{SAE}}(\boldsymbol{\theta}) &= J_{\mathrm{E}}(\boldsymbol{\theta}) + \beta J_{\mathrm{KL}}(r \,\|\, \hat{\mathbf{r}}) \\ &= \frac{1}{2n}\left(\sum_{i=1}^{n}\left\|\hat{\mathbf{X}}_i - \mathbf{X}_i\right\|^2\right) + \beta\sum_{i=1}^{n_l}\left(r\log\frac{r}{\hat{r}_i} + (1-r)\log\frac{1-r}{1-\hat{r}_i}\right) \end{aligned} \tag{4}$$

where $\boldsymbol{\theta} = \{\mathbf{W}_1, \mathbf{W}_2, \mathbf{b}_1, \mathbf{b}_2\}$ represents the parameter set; $J_{\mathrm{E}}(\boldsymbol{\theta})$ represents the reconstruction error; $J_{\mathrm{KL}}(\mathrm{r}\|\hat{\mathbf{r}})$ is Kullback- Leibler (KL) divergence function, β is the sparsity penalty, n_l is the number of hidden units, r represents the sparsity parameter, and $\hat{r}_i$ is the average activation value of the jth hidden unit.

III. THE PROPOSED METHOD

In this paper, a novel deep transfer learning approach named DTNSAE is proposed for rolling bearing fault diagnosis with few labeled data. The proposed method is dedicated to making advantage of parameter transfer learning and deep learning, and the details are described as follows.

A. MOTIVATION OF TRANSFER LEARNING

Intelligent fault diagnosis methods have achieved great attention in recent years. As is shown Fig. 2, traditional intelligent methods like SVM and ANN usually rely on manual feature extraction to get state-of-the-art results. However, this requires prior knowledge and domain expertise as well as takes amounts of time. Deep learning methods have been demonstrated as effective tools for fault diagnosis, which are able to automatically learn representative features from raw data. Nevertheless, both traditional methods and existing deep learning methods have two common limitations in fault diagnosis problems. One is that they mostly require numerous labeled training data, and the other is that training data have to come from the same distribution with the testing data. In many cases especially in real-world applications, it is very expensive, time-consuming and even impossible to recollect enough labeled data. Thus, it is quite meaningful to develop new methods for rolling bearing fault diagnosis when only few labeled data are available.

At present, transfer learning have been a new focus of fault diagnosis area. Transfer learning aims to help solve a target domain problem by employing the knowledge learned from a different but related domain. As is shown in Fig. 2(c), this paper applies the parameter transfer learning to tackle the problems with few labeled target domain data. Parameter transfer learning is a useful and easy-to-implement method. Suppose that D_s denotes the source domain and X_s is the source domain data; D_t represents the target domain and X_t is the target domain data. It is noted that D_s and D_t are two different but related domains. X_s includes amounts of labeled auxiliary data while X_t only comprises a tiny set of labeled data [29]. The basic process of parameter transfer learning for fault diagnosis can be described as follows. Firstly, construct an intelligent learning system A; secondly, train intelligent learning system A using the source domain training data and validate the well pre-trained learning system using the corresponding testing data; thirdly, construct the other intelligent learning system B, whose architecture is same with system A; fourthly, transfer the knowledge of pre-trained system A to initialize system B; fifthly, fine-tune system B using limited labeled target domain data; finally, test the transfer learning system using the rest unlabeled target domain data.

B. Base DNSAE construction

Recently, a novel sparse autoencoder with nonnegative constraints is proposed for image recognition. Similar to SAE, NSAE is capable of learning robust and sparse representative features from high dimensional data [27]. Thus, NSAE is adopted to construct a deep network for fault diagnosis in this paper.

The most significant difference between NSAE and SAE is that nonnegative constraint is added to the cost function. So the final cost function of NSAE is given as

$$J_{\mathrm{NSAE}}(\boldsymbol{\theta}) = J_{\mathrm{E}}(\boldsymbol{\theta}) + \beta J_{\mathrm{KL}}(r \,\|\, \hat{\mathbf{r}}) + \frac{\mu}{2}\sum_{l=1}^{2}\sum_{i=1}^{n_l}\sum_{j=1}^{n_{l+1}} G(W_{ij}^l) \tag{5}$$

with respect to

$$G(W_{ij}^l) = \begin{cases} (W_{ij}^l)^2 & , W_{ij}^l < 0 \\ 0 & , W_{ij}^l \geq 0 \end{cases} \tag{6}$$

where the third part of the function is nonnegative constraint term, μ is the nonnegative constraint penalty; n_l and n_{l+1} denote the dimensionalities of two adjacent layers; W_{ij}^l is the weight between the ith unit of lth layer and the jth unit of $(l+1)$th layer. Through iteratively minimizing this cost function in the learning process, better sparsity of the encoder and fewer negative weights will be achieved.

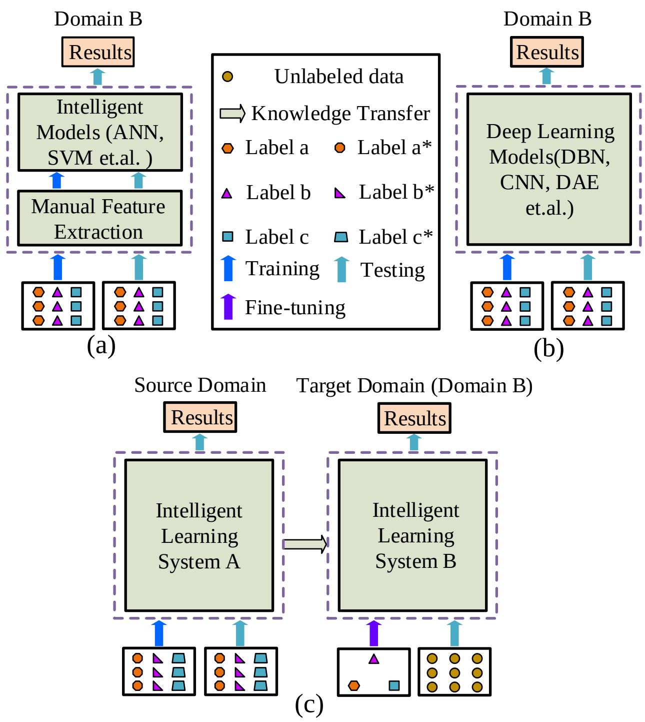

FIGURE 2. **Different fault diagnosis methods: (a) traditional intelligent methods; (b) deep learning methods; (c) Parameter transfer learning methods with few labeled target domain data.**

The parameters of NSAE is updated through the backpropagation algorithm as follows

$$W_{ij}^l = W_{ij}^l - \eta \frac{\partial}{\partial W_{ij}^l} J_{\mathrm{NSAE}}(\boldsymbol{\theta}) \tag{7}$$

$$b_{ij}^l = b_{ij}^l - \eta \frac{\partial}{\partial b_{ij}^l} J_{\mathrm{NSAE}}(\boldsymbol{\theta}) \tag{8}$$

where η is the learning rate, and the corresponding gradients can be calculated by

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2926234, IEEE Access

$$\frac{\partial}{\partial W_{ij}^{l}} J_{\mathrm{NSAE}}(\boldsymbol{\theta}) = \frac{\partial}{\partial W_{ij}^{l}} J_{\mathrm{E}}(\boldsymbol{\theta}) + \beta \frac{\partial}{\partial W_{ij}^{l}} J_{\mathrm{KL}}(r \,\|\, \hat{\mathbf{r}}) + \mu g(W_{ij}^{l}) \tag{9}$$

with respect to

$$g(W_{ij}^{l}) = \begin{cases} W_{ij}^{l} & , \ W_{ij}^{l} < 0 \\ 0 & , \ W_{ij}^{l} \geq 0 \end{cases} \tag{10}$$

$$\frac{\partial}{\partial b_{ij}^{l}} J_{\mathrm{NSAE}}(\boldsymbol{\theta}) = \frac{\partial}{\partial b_{ij}^{l}} J_{\mathrm{E}}(\boldsymbol{\theta}) \tag{11}$$

To get rid of manual feature extraction and adaptively learn sparse representative features from raw vibration signals, a DNSAE is constructed by stacking multiple individual NSAEs. As is shown in Fig. 3, each NSAE is trained by minimizing the cost function (5), and the encoding features of previous autoencoder are referred as the input of next autoencoder.

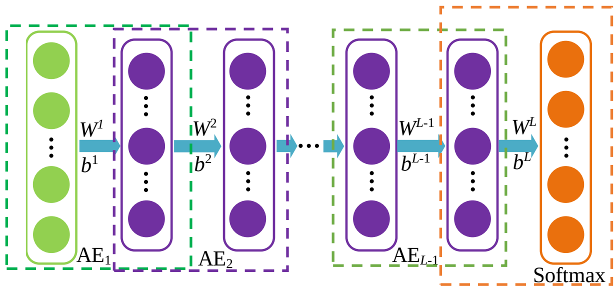

FIGURE 3. Architecture of DNSAE.

Then a nonnegativity-constraint softmax classifier is stacked on top of the last NSAE to recognize the fault modes. The cost function of nonnegativity-constraint softmax classifier is

$$J_{\text{NC-Softmax}}(\mathbf{W}^{L}) = -\frac{1}{n}\sum_{k=1}^{n}(\hat{y}_k \log y_k) + \frac{\mu}{2}\sum_{i=1}^{n_L}\sum_{j=1}^{c} G(W_{ij}^{L}) \tag{12}$$

where $\hat{y}_k$ and y_k are the actual label and predicted label of the kth sample, c is the number of health conditions, n_L represents the hidden unit number of the final autoencoder, and $\mathbf{W}^L$ denotes the weights between the final hidden layer and softmax classifier.

After layer-wise unsupervised feature learning, supervised fine-tuning is a necessary step to ensure great fault diagnosis results. The cost function of DNSAE for fine-tuning is defined as

$$J_{\mathrm{DNSAE}}(\boldsymbol{\theta}_s) = -\frac{1}{n}\sum_{k=1}^{n}(\hat{y}_k \log y_k) + \frac{\mu}{2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{j=1}^{n_{l+1}} G(W_{ij}^{l}) \tag{13}$$

where $\boldsymbol{\theta}_s$ ={ $\mathbf{W}^l$, $\mathbf{b}^l$, l= 1, 2, …, L} comprises all the weights and biases of DNSAE.

C. DTNSAE implementation for fault diagnosis

As is mentioned above, transfer learning is suitable for tackling fault diagnosis problems with few labeled data. Deep learning is capable of adaptively capturing representative features from raw data without manual feature extraction. To take advantage of deep learning and transfer learning, this paper proposes a new method called DTNSAE for rolling bearing fault diagnosis based on parameter transfer learning and DNSAE. As is shown in Fig. 4, the specific procedure of the proposed method is as follows.

- **Step 1:** Organize the source domain data X_s into training samples and testing samples.
- **Step 2:** Construct a base deep network called DNSAE (A) by stacking multiple NSAEs.
- **Step 3:** Train DNSAE (A) using the source domain data and acquire a well-trained deep network with excellent testing performance.
- **Step 4:** Acquire the target domain data X_t and select few available labeled data as training data and unlabeled data as testing data.
- **Step 5:** Construct a target deep network DNSAE (B), which has same architecture with DNSAE (A).
- **Step 6:** Transfer the parameters of well-trained DNSAE (A) and softmax classifier including weights and biases to initialize DNSAE (B) and its classifier as follows

$$\mathbf{U}^{l} = \mathbf{W}^{l}, \mathbf{a}^{l} = \mathbf{b}^{l}, l = 1, 2, \ldots, L \tag{14}$$

where $\mathbf{U}^l$ is the weights between lth layer and (l+1)th layer, $\mathbf{a}^l$ represents the bias vector of (l+1)th layer of DNSAE (B).

- **Step 7:** Fine-tune DNSAE (B) using limited labeled target domain data and optimize these parameters by minimizing the following cost function

$$J_{\mathrm{DTNSAE}}(\boldsymbol{\theta}_t) = -\frac{1}{n}\sum_{k=1}^{n}(\hat{y}_k \log y_k) + \frac{\mu}{2}\sum_{l=1}^{L}\sum_{i=1}^{n_l}\sum_{j=1}^{n_{l+1}} G(U_{ij}^{l}) \tag{15}$$

where $\boldsymbol{\theta}_t$ ={ $\mathbf{U}^l$, $\mathbf{a}^l$, l= 1, 2, …, L} is the parameter set of DNSAE (B).

- **Step 8:** Test the effectiveness of the fine-tuned DTNSAE using unlabeled target domain data.

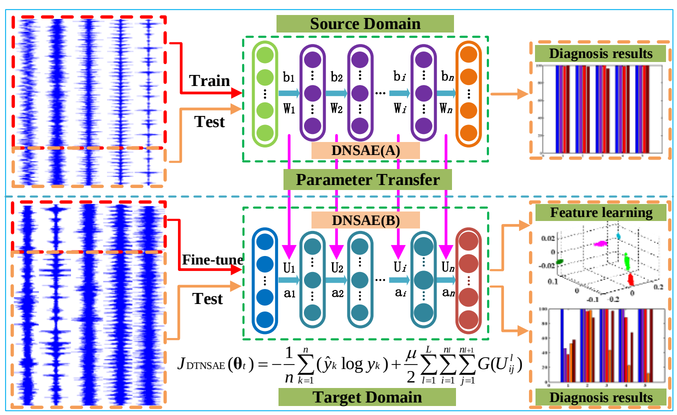

FIGURE 4. Flowchart of DTNSAE.

In this paper, the Limited-memory BFGS (L-BFGS) quasi-Newton method is used to optimize all the cost functions on Matlab 2014a [30].

IV. EXPERIMENT VERIFICATION

A. Data Description

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2926234, IEEE Access

In this paper, a deep transfer learning method called DTNSAE is proposed for rolling bearing fault diagnosis. Two datasets from different domains are required to verify the effectiveness of the proposed method. Here, the motor bearing dataset A from Case Western Reserve University (CWRU) is selected as the source domain dataset, and the other real-world railway locomotive bearing dataset B is used as the target domain dataset. Details about these two datasets are demonstrated as follows.

Source domain data: In this dataset, the vibration signals were acquired with a sampling rate of 12 kHz for ten seconds. Five fault modes are introduced in the dataset, including normal condition, ball fault (BF, 0.007 in.), inner race fault (IRF, 0.007 in.), outer race fault (ORF, 0.007 in.), outer race fault (ORF, 0.021 in.). According to the difference of motor speeds and loads, four data subsets are constructed i.e. A1 (1797 rpm, 0 hp), A2 (1772 rpm, 1 hp), A3 (1750 rpm, 2 hp) and A4 (1730 rpm, 3 hp). Finally, 200 samples are obtained for each fault mode, 150 (75%) of which are selected as the training samples and the others as testing samples. Each sample contains 600 data points. More details about this dataset can be found in [31].

TABLE I
DESCRIPTION OF CWRU BEARING DATASET (SOURCE DOMAIN)

Fault and severity (in.)	Subset A1/A2/A3/A4		No. of training/ testing samples	Label
	Speed	Load		
Normal	1797 rpm /1772 rpm /1750 rpm /1730 rpm	0 hp /1 hp /2 hp /3 hp	150(75%)/50(25%)	1
BF (0.007)			150(75%)/50(25%)	2
IRF (0.007)			150(75%)/50(25%)	3
ORF(0.007)			150(75%)/50(25%)	4
ORF(0.021)			150(75%)/50(25%)	5

Remarks: NO. is the abbreviation of number;

Target domain data: Railway locomotive bearing dataset is used as the target domain data, and the test rig is demonstrated in Fig. 5. The vibration signals are acquired with a sampling rate of 12.8 kHz. Similarly, five fault modes are introduced in this dataset, including normal condition, ball fault (BF), inner race fault (IRF), outer race fault (ORF, slight), outer race fault (ORF, severe). According to TABLE II, different fault modes have different motor speeds and loads. Finally, 200 samples are also obtained for each fault mode, limited samples of which are selected as the training samples and the others as testing samples. Each sample also contains 600 data points. More details about this dataset can be found in [32].

FIGURE 5. The test rig of railway locomotive bearing.

As mentioned above, these two datasets come from different machines with different bearings, speeds, loads and sampling rates yet similar fault components and modes. Therefore, it is very suitable to refer them as the source domain data and target domain data, respectively.

B. DTNSAE validation

As is shown in Section III, a base DNSAE should be constructed to acquire a well-trained deep network first of all. In this paper, four base DNSAEs are obtained, which are trained by training samples from source domain data subsets A1, A2, A3, and A4, respectively. Also, the corresponding testing samples are used to test their effectiveness. Next, the parameters including weights and biases of the four high-performance DNSAEs are individually transferred to initialize the target DNSAE. Finally, few labeled target domain samples are used to fine-tune the target DNSAE, and the final DTNSAE is built for rolling bearing fault diagnosis with few labeled data.

TABLE II
DESCRIPTION OF LOCOMOTIVE BEARING DATASET (TARGET DOMAIN)

Fault modes	Speed	Load	No. of target domain samples	Label
Normal	490 rpm	2002 kg	200(100%)	1
BF	531 rpm	494 kg	200(100%)	2
IRF	498 rpm	501 kg	200(100%)	3
ORF (Slight)	490 rpm	499 kg	200(100%)	4
ORF (Severe)	481 rpm	499 kg	200(100%)	5

(1) Structure selection. Structure selection will greatly influence the performance of machine learning methods. To select a more suitable structure of the proposed method, the testing results of the proposed method with different number of layers and labeled target domain samples are shown in Fig.6. It is obvious that the testing accuracies of the proposed method increase with the increasing of the number of hidden layers. However, the training time will synchronously increase when more hidden layers are constructed. It is noted that the training time represents the average value of four conditions where the number of labeled target domain samples are 1, 2, 4, and 10, respectively. In addition, it can be found that the diagnosis accuracies will reach a very high level and hardly increase too more when three or more hidden layers are selected. Thus, to balance the diagnosis accuracy and computation efficiency, a 3-hidden-layer base DNSAE is finally selected to construct the proposed DTNSAE.

It is noted that the hyper parameters selection is a challenging task in machine learning area. Only number of hidden layers is simply discussed, more related study is not conducted considering that it is not the focus of this paper.

(2) Specific results discussion. The specific hyper parameters of DTNSAE are as follows. Take the condition with 10 (5%) labeled target domain samples for example, the architecture of Base DNSAEs is selected as "600-250-150-50-5", the iteration number of each NSAE is 200, and r,

IEEE Access Multidisciplinary | Rapid Review | Open Access Journal

β, μ are determined as 0.3, 3, and 0.0001 in pre-trained process, respectively. The iteration number of DTNSAE is 35 in fine-tuning process. These parameters are all empirically selected by experiments. Table III shows the testing accuracy comparison among the base DNSAE, deep sparse autoencoders (DSAE), deep denosing autoencoder (DDAE) [33], and deep belief network (DBN) [34], on the source domain data. The testing results shown in Table III are all the average values of 10 trails. The testing accuracy of four Base DNSAEs is 99.85%, 99.72%, 99.80% and 100%, respectively. What's more, the average testing accuracy of the Base DNSAEs is 99.84%, while the testing accuracy of DSAE, DDAE, and DBN is 98.28%, 97.87, and 95.74, respectively. The results suggest that the Base DNSAEs acquire state-of-the-art performance and are more effective than the other methods on the four data subsets. Thus, DNSAE is suitable and effective for rolling bearing fault diagnosis.

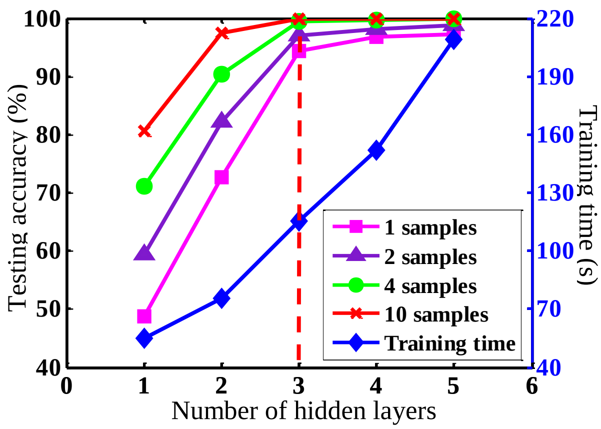

FIGURE 6. The testing results of the proposed method with different number of hidden layers.

TABLE III
RESULTS OF FOUR BASE DNSAES WITH SOURCE DOMAIN DATA

Method	Data subsets				Average accuracy
	A1	A2	A3	A4	
DNSAE (%)	**99.85**	**99.72**	**99.80**	**100**	**99.84**
DSAE (%)	97.48	98.44	98.24	98.96	98.28
DDAE (%)	95.04	99.28	98.56	98.60	97.87
DBN (%)	96.84	97.44	94.00	94.68	95.74

Remarks: The results in this table and the following tables are all the average values of ten-independent-run results.

The hyper parameters of the other three methods are shown as follows. (1) DSAE: the architecture of DSAEs is selected as "600-250-150-50-5", the iteration number of each NSAE is 200, and r, β are selected as 0.3, 3, respectively. (2) DDAE: the architecture of DDAEs is selected as "600-250-150-50-5", the iteration number of each SAE is 200, the weight decay parameter and gauss random noise level are selected as 0.0001 and 0.3, respectively. (2) DBN: the architecture of DBNs is selected as "600-250-150-50-5", the iteration number of each restricted Boltzmann machine is 200, and the weight decay parameter and learning rate are selected as 0.0003 and 0.2, respectively.

Then TABLE IV and Fig.7 show the testing accuracies of the proposed method with different number of labeled target domain samples. Through these results, two points can be basically concluded. One is that the larger the number of the labeled target domain samples is, the higher the testing accuracies of the proposed method are. The other is that when the number of the labeled target domain samples increases to 4(2%) or larger, the average testing accuracy of the proposed method increases to a very high level and doesn't change a lot any more. Therefore, the proposed method is able to accurately recognize the fault modes when only a tiny set of labeled target data are available.

TABLE IV
RESULTS OF THE PROPOSED METHOD WITH DIFFERENT NUMBER OF LABELED TARGET DOMAIN SAMPLES

No. of labeled data	A1-B	A2-B	A3-B	A4-B	Average accuracy
1 (0.5%)	94.82 %	94.69%	94.77%	93.33 %	94.40%
2 (1%)	97.16 %	96.99%	97.09%	96.43 %	96.92%
4 (2%)	99.41 %	99.36%	99.34%	99.52 %	99.41%
10 (5%)	99.68 %	99.85%	99.89%	99.76 %	99.80%
20 (10%)	99.78 %	99.90%	99.88%	100 %	99.89%
40 (20%)	99.94 %	99.99%	99.96%	100 %	99.97%
100(50%)	100 %	100 %	100 %	100 %	100 %
150 (75%)	100 %	100 %	100 %	100 %	100 %

Remarks: A1-B represents that the source domain data is A1and the target dataset is B, others are similar. No. of labeled data denotes the labeled samples number of each fault modes.

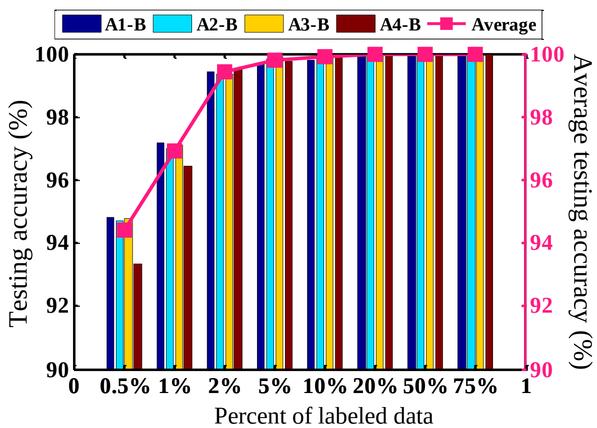

FIGURE 7. The results of the proposed method with different number of labeled target domain samples.

C. Comparison with other methods

In this part, experiment is carried out to compare the proposed method with other existing methods, including DNSAE, DSAE, BP and SVM. DNSAE and DSAE belong to deep learning methods, while BP and SVM belong to traditional intelligent methods. These methods don't involve in transfer learning.

(1) TABLE V and Fig.8 show testing results

comparison of five methods with different number of labeled samples. Based on these results, three points can be intuitively summarized as follows. Firstly, with the number of the labeled samples increasing, the testing accuracies of the five methods all increase synchronously. Secondly, the proposed method performs better than the other methods. Particularly, the testing accuracy of the proposed method is much higher than those of the other methods when the number of the labeled samples is less than 10 (5%). Thirdly, deep learning methods generally have higher testing accuracies than BPNN and SVM. When the number of the labeled samples increases to 150 (75%), DNSAE and DSAE achieve quite high diagnosis accuracies, whereas the shallow methods perform not so well, especially BPNN shows the worst performance.

TABLE V
TESTING RESULTS COMPARISON OF DIFFERENT METHODS

No. of labeled data	Proposed method	DNSAE	DSAE	BPNN	SVM
1 (0.5%)	94.40%	54.12%	52.94%	37.33%	33.84%
2 (1%)	96.92%	57.90%	56.08%	41.45%	41.05%
4 (2%)	99.41%	60.70%	60.57%	42.36%	71.09%
10 (5%)	99.80%	88.26%	84.63%	47.65%	82.36%
20 (10%)	99.89%	96.20%	91.30%	53.62%	87.76%
40 (20%)	99.97%	97.00%	95.63%	57.54%	91.78%
100 (50%)	100 %	98.20%	97.60%	59.28%	94.58%
150 (75%)	100 %	99.20%	98.00%	61.12%	95.20%

(2) In order to conduct more specific comparison, the diagnosis accuracies of each fault modes are shown Fig. 9. The results indicate that the proposed method can recognize each fault modes quite accurately even though only 10(5%) labeled samples are available, whereas the other methods can not accurately recognize the normal condition and the slight outer race faults. This further confirms that the proposed method is more effective than the other methods for fault recognition of rolling bearing with few labeled data.

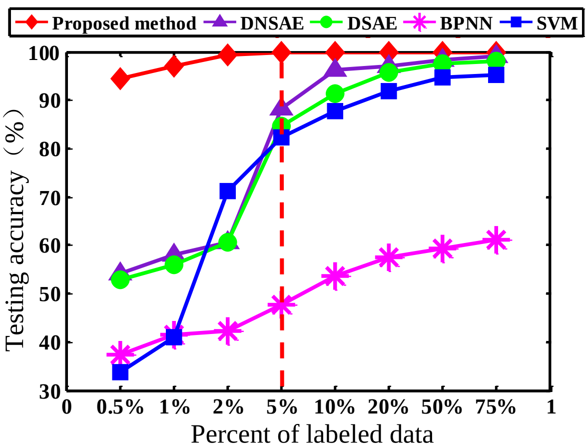

FIGURE 8. Testing results comparison of five methods with different number of labeled samples.

(3) In order to demonstrate the superiority of the proposed method in feature learning, Fig.10 shows three-dimensional visualization of the original data and features learned by different methods. Take the fourth condition for example, where only 10 (5%) labeled samples are available, kernel principle component analysis (KPCA) is adopted to visualize the original data and the third-layer features of the proposed method, DNSAE and DSAE. As is shown in Fig. 11, KPC1 to KPC3 denote the first three principle components, respectively. Marks of different shapes and colors represent different fault modes.

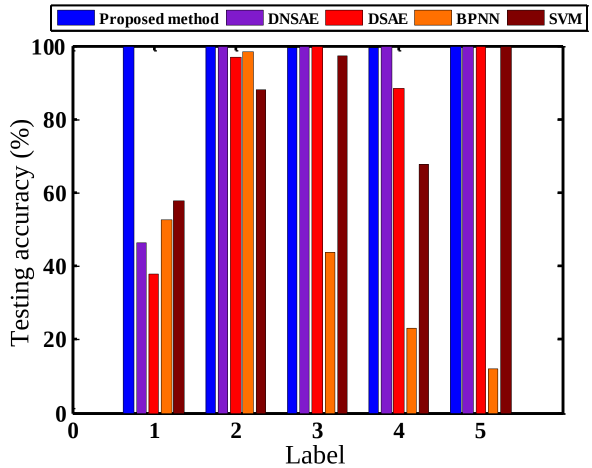

FIGURE 9. Testing accuracies comparison of each fault mode with 10 (5%) labeled samples.

As is shown in Fig. 10 (a), the projection of original data is partly overlapped among different fault modes. The features of each fault modes learned by the proposed method compactly cluster together, which are much more distinguishable and identifiable than those learned by DNSAE and DSAE in the 3-D projection. The main cause is that the proposed method makes advantages of both deep learning and transfer learning. Particularly, transfer learning is able to improve the feature learning of target domain task by applying the knowledge learned from source domain tasks when few labeled target domain data are available.

The evidences above firmly confirm that the proposed method is much more effective for rolling bearing fault diagnosis than DNSAE, DSAE, BPNN and SVM when few labeled samples are available. Even though when plenty of labeled samples are available, the proposed method also have higher diagnosis accuracy.

The hyper parameters of the five methods in the fourth condition (5% labeled samples are available) are as follows. (1) The hyper parameters of the proposed method have been shown in part *B* of Section IV. (2) DNSAE: the architecture is selected as "600-250-150-50-5", r, β, μ are selected as 0.5, 3, and 0.0001, respectively. The iteration number of each NSAE is determined as 200. (2) DSAE: its architecture is also selected as "600-250-150-50-5", r and β are selected as 0.5 and 3, respectively. The iteration number

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2926234, IEEE Access

of each SAE is determined as 200. (3) BP: the architecture is selected as "600-1000-5", the learning rate and momentum are selected as 0.4 and 0.95, respectively. The iteration number is determined as 1000. (4) SVM: the gauss kernel is adopted, the radius of gauss kernel and the penalty factor are selected as 0.625 and 2 by 10-fold cross validation, respectively. It is noted that the above hyper parameters of the five methods are all manually and empirically determined by lots experiments.

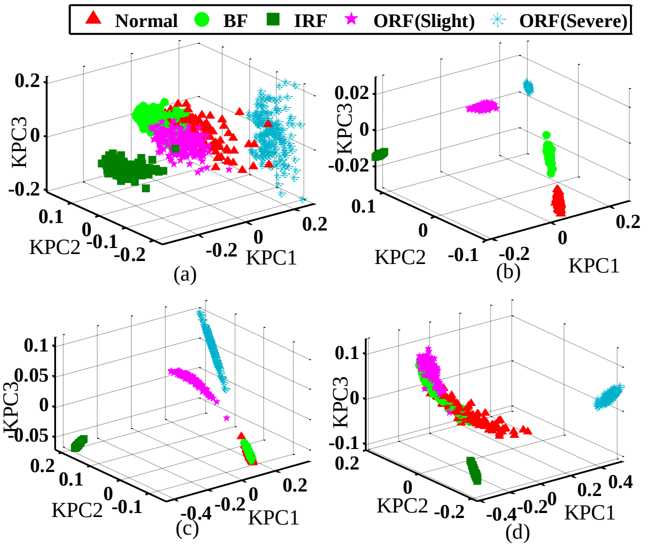

FIGURE 10. 3-dimensional visualization of different features using KPCA: (a) original data; (b) proposed method; (c) DNSAE; (d) DSAE.

D. Comparison between NSAE and SAE

In this paper, to evaluate the weight sparseness of NSAE and conventional SAE, the following sparseness criterion is adopted [35].

$$S(\mathbf{w}) = \frac{\sqrt{m} - \left(\sum_{i=1}^{m} |w_i|\right) \Big/ \sqrt{\sum_{i=1}^{m} w_i^2}}{\sqrt{m} - 1} \tag{16}$$

where $\mathbf{w}$ represents a vector, m and denotes the dimensionality of $\mathbf{w}$. It is noted that the larger $S(\mathbf{w})$ is, the higher the sparseness degree of $\mathbf{w}$ is.

Fig.11 and Fig.12 show the histogram of the sparseness criterion measured on 250 receptive fields and 250 decoding filters, respectively. Generally, the sparseness criterion values of NSAE's receptive fields and decoding filters are both higher than those of SAE. This confirms that the nonnegativity constraints is able to enhance the weights sparseness of encoding layer and decoding layer. Moreover, KL divergence of NSAE is 0.0025 for the testing samples, which is much smaller than 0.0908 of SAE. This suggests that NSAE can learn sparser features.

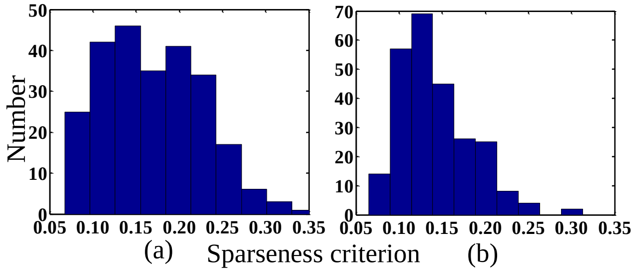

FIGURE 11. Distribution of sparseness criterion measured on 250 receptive fields: (a) NSAE; (b) SAE.

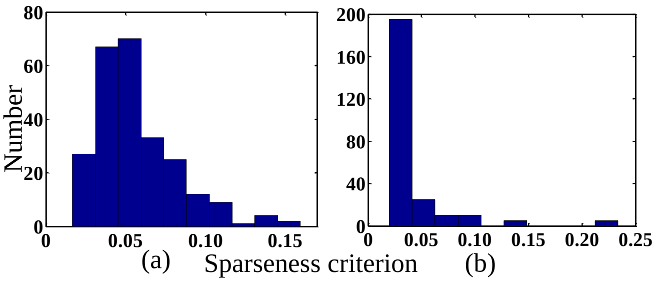

FIGURE 12. Distribution of sparseness criterion measured on 250 decoding filters: (a) NSAE; (b) SAE.

V. CONCLUSION

In this paper, a deep transfer learning method called DTNSAE is proposed to tackle rolling bearing fault diagnosis problem with few labeled data. The proposed method makes full advantage of parameter transfer learning and deep learning. The results suggest that combining parameter transfer learning and deep leaning is promising for tackling the fault diagnosis problems with few labeled data. Compared with traditional methods and pure deep learning methods, the proposed method is able to much more effectively recognize the fault modes of rolling bearing with few labeled data. Moreover, it is more suitable for real-world applications, and it can get rid of manual feature extraction.

ACKNOWLEDGMENT

This work is supported by the major research plan of the National Natural Science Foundation of China (No. 91860124), the National Natural Science Foundation of China (No. 51875459), the Aeronautical Science Foundation of China (No. 20170253003) and the Synergy Innovation Foundation of the University and Enterprise for Graduate Students in Northwestern Polytechnical University (No. XQ201901). Dr. Xingqiu Li conducted most of this work. Besides, professor Hongkai Jiang gave constructive and meaningful instruction to this work, Dr. Ke Zhao checked the source codes and gave useful suggestions, and Dr. Ruixin Wang did the grammar and format checking.

REFERENCES

[1] H.K. Jiang, C.L. Li, and H.X. Li, "An improved EEMD with multiwavelet packet for rotating machinery multi-fault diagnosis," *Mechan. Syst. Signal Process.*, vol. 36 pp. 225-239, 2013.

[2] W.Y. Huang, J.C. Cheng, and Y. Yang, "Rolling Bearing Performance Degradation Assessment Based on Convolutional Sparse Combination Learning," *IEEE Access*, vol. 7, pp. 17834-17846, 2019.

[3] P. wang, H. Wang, and R.Q. Yan, "Bearing Degradation Evaluation Using Improved Cross Recurrence Quantification Analysis and Nonlinear Auto-Regressive Neural Network", *IEEE Access*, vol. 7, pp. 38937- 38946, 2019.

[4] P. Bangalore, and L.B. Tjernberg, "An Artificial Neural Network Approach for Early Fault Detection of Gearbox Bearings," *IEEE Trans. Smart Grid*, vol. 6, no. 2, pp. 980-987, 2015.

[5] R. Dubey, and D. Agrawal, "Bearing fault classification using ANN-based Hilbert footprint analysis," *IET Sci. Meas. Technol.*, vol. 9, no. 8, pp. 1016-1022, 2015.

[6] J.B. Ali, N. Fnaiech, L. Saidi, *et al*., "Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration signals," *Appl. Acoust.*, vol. 89, pp.16-27, 2015.

[7] X.A. Yan, and M.P. Jia, "A novel optimized SVM classification algorithm with multi-domain feature and its application to fault diagnosis of rolling bearing," *Neurocomputing*, vol. 313, pp. 47–64, 2018.

[8] F.B, Abid , S. Zgarni, and A. Braham, "Distinct Bearing Faults Detection in Induction Motor by a Hybrid Optimized SWPT and aiNet-DAG SVM," *IEEE Trans. Energy Convers.*, vol. 33, no. 4, Dec. 2018.

[9] L. Ren, W. Lv, S.W. Jiang, *et al*., "Fault Diagnosis Using a Joint Model Based on Sparse Representation and SVM," *IEEE Trans. Instrum. Meas.*, vol. 65, no. 10, pp. 2313-2320, Oct. 2016.

[10] H.D. Shao, H.K. Jiang, X.Q. Li, *et al.*, "Intelligent fault diagnosis of rolling bearing using deep wavelet auto-encoder with extreme learning machine," *Knowl.-Based Syst.*, vol. 140, pp. 1-14,2018.

[11] H.D. Shao, H.K. Jiang, H.W. Zhao, *et al*., "A novel tracking deep wavelet auto-encoder method for intelligent fault diagnosis of electric locomotive bearings," *Mechan. Syst. Signal Process.*, vol. 110, pp. 193–209, 2018.

[12] F. Jia, Y.G. Lei, N. Lu, *et al*., "Deep normalized convolutional neural network for imbalanced fault classification of machinery and its understanding via visualization," *Mechan. Syst. Signal Process.*, vol. 110, pp. 349–367, 2018.

[13] F. Jia, Y.G. Lei, J. Lin, X. Zhou, and N. Lu, "Deep neural networks: A promising tool for fault characteristics mining and intelligent diagnosis of rotating machinery with massive data," *Mechan. Syst. Signal Process.*, vol. 72-73, pp. 303–315, 2016.

[14] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw*., vol. 61, pp. 85–117, 2015.

[15] J.Q. Xie, G.F. Du, C.Q. Shen, N. Chen, L. Chen, and Z.K. Zhu, "An End-to-End Model Based on Improved Adaptive Deep Belief Network and Its Application to Bearing Fault Diagnosis," *IEEE Access*, vol. 6, pp. 63584- 63596, 2018.

[16] C. Sun, M. Ma, Z.B. Zhao, *et al.*, "Sparse Deep Stacking Network for Fault Diagnosis of Motor," *IEEE Trans. Ind. Informat.*, vol. 14, no. 7, pp. 3261:3269, Jul. 2018.

[17] H.D. Shao, H.K. Jiang, H.Z. Zhang, *et al*., "A novel method for intelligent fault diagnosis of rolling bearings using ensemble deep auto-encoders," *Mechan. Syst. Signal Process.*, vol. 102, pp. 278-297, 2017.

[18] H.D. Shao, H.K. Jiang, H.Z. Zhang, *et al*., "Electric locomotive bearing fault diagnosis using novel convolutional deep belief network," *IEEE Trans. Ind. Electron.*, vol. 65, no. 3, pp. 2727-2736, 2018.

[19] H.K. Jiang, X.Q. Li, H.D. Shao, *et al.*, "Intelligent fault diagnosis of rolling bearings using an improved deep recurrent neural network," *Meas. Sci. Technol.*, vol. 29, no. 6, 2018, Art. no. 065107.

[20] M. Ma, C. Sun, and X.F. Chen, "Deep Coupling Autoencoder for Fault Diagnosis with Multimodal Sensory Data," *IEEE Trans. Ind. Informat.*, vol. 14, no. 3, pp. 1137-1145, 2018.

[21] J. Pan, Y.Y. Zi, J. Chen, *et al.*, "Lifting Net: A Novel Deep Learning Network with Layerwise Feature Learning from Noisy Mechanical Data for Fault Classification," *IEEE Trans. Ind. Electron.*, vol. 65, no. 6, pp. 4973-4982, Jun. 2018.

[22] P. Cao, S. Zhang, and J. Tang. Pre-Processing-Free Gear Fault Diagnosis Using Small Datasets with Deep Convolutional Neural Network-Based Transfer Learning. *IEEE Access*, vol. 6, pp. 26241-26253, 2017.

[23] C. Persello and L. Bruzzone, "Kernel-based domain-invariant feature selection in hyperspectral images for transfer learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 5, pp. 2615-2626, 2016.

[24] S.J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, 2010.

[25] L. Guo, Y.G. Lei, S.B. Xing, et al., "Deep Convolutional Transfer Learning Network: A New Method for Intelligent Fault Diagnosis of Machines With Unlabeled Data," *IEEE Trans. Ind. Electron*., vol. 66, no. 9, pp. 7316-7325, 2019.

[26] S.Y. Shao, S. McAleer, R.Q. Yan, and P. Baldi, "Highly Accurate Machine Fault Diagnosis Using Deep Transfer Learning," *IEEE Trans. Ind. Informat.*, vol. 15, no. 4, pp. 2446-2455, 2019.

[27] H.D. Shao, H.K. Jiang, H.W. Zhao, *et al.*, "An enhancement deep feature fusion method for rotating machinery fault diagnosis," *Knowl.-Based Syst.*, vol. 119, pp. 200-220, 2017.

[28] E. Hosseini-Asl, J. M. Zurada, and O. Nasraoui, "Deep Learning of Part-Based Representation of Data Using Sparse Autoencoders with Nonnegativity Constraints,'' *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 12, pp. 2486 – 2498, Dec. 2016.

[29] D. Tuia, J. Munoz-Mari, L. Gomez-Chova, and J. Malo, "Graph matching for adaptation in remote sensing," *IEEE Trans. Geosci. Remote Sens*., vol. 51, no. 1, pp. 329–341, Jan. 2013.

[30] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, "A limited memory algorithm for bound constrained optimization," *SIAM J. Sci. Comput.*, vol. 16, no. 5, pp. 1190–1208, 1995.

[31] "Case Western Reserve University Bearing Data Center Website" http://csegroups.case.edu/bearingdatacenter/home.

[32] Y.G. Lei, *Intelligent Fault Diagnosis and Remaining Useful Life Prediction of Rotating Machinery*. Butterworth-Heinemann, 2016.

[33] Z. Meng, X.Y. Zhan, J. Li, Z.Z. Pan, "An enhancement denoising autoencoder for rolling bearing fault diagnosis," *Measurement*, vol. 130, pp. 448–454, 2018.

[34] G.E. Hinton, S. Osindero, Y.W. The, "A Fast Learning Algorithm for Deep Belief Nets," *Neural Comput.*, vol. 18, no. 7, pp.1527-1554, 2006.

[35] P.O. Hoyer, "Non-negative matrix factorization with sparseness constraints. Journal of Machine Learning Research," *J. of Mach. Learn. Res.*, vol. 5, pp. 1457-1469, 2004.

Xingqiu Li was born in Anhui Province, China, in 1993. He received the B.S. degree in Electrical Engineering and Automation from Northwestern Polytechnical University in 2017. He is currently pursuing the Ph.D. degree in Means of Transport Applied Engineering from Northwestern Polytechnical University, Xi'an, China.

His current research interests include fault diagnosis, intelligent prognosis, multi-source information fusion and transfer learning.

Hongkai Jiang (M'17) received the Ph.D. degree in Mechanical Engineering School from Xi'an Jiao Tong University, Xi'an, P. R. China, in 2006. He is currently a Full Professor of school of aeronautics, Northwestern Polytechnical University, Xi'an, P. R. China. From January 2011 to January 2012, He was a Visiting Scholar with the University of British Columbia, Vancouver, Canada.

His current research interests include condition monitoring and dynamic signal processing, fault diagnosis and health management, big data analysis and maintenance support.

Ke Zhao was born Shaanxi Province, China, in 1998. He received the B.S. degree in Airworthiness Technology and Management from Northwestern Polytechnical University in 2017. He is currently working toward to the Ph.D. degree in Means of Transport Applied Engineering from Northwestern Polytechnical University, Xi'an, China.

His current research interests include intelligent fault diagnosis, health monitoring and deep learning.

Ruixin Wang was born in Shaanxi Province, China, in 1995. She received the B.S. degree in automation from Northwestern Polytechnical University, Xi'an, China, in 2017, and the M.S. degree in telecommunications from The Hong Kong University of Science and Technology, Hong Kong, China, in 2018. She is currently working toward the Ph.D. degree in means of transport applied engineering at Northwestern Polytechnical University, Xi'an, China.

Her current research interests include fault diagnosis, intelligent prognosis.