Neurocomputing 407 (2020) 121–135

Contents lists available at ScienceDirect

Neurocomputing

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

A systematic review of deep transfer learning for machinery fault diagnosis

Chuan Li[a], Shaohui Zhang[b,*], Yi Qin[c], Edgar Estupinan[d]

[a] National Research Base of Intelligent Manufacturing Service, Chongqing Technology and Business University, Chongqing 400067, China
[b] School of Mechanical and Automotive Engineering, South China University of Technology, Guangzhou 510641, China
[c] State Key Laboratory of Mechanical Transmission, Chongqing University, Chongqing 400044, China
[d] Department of Mechanical Engineering, University of Tarapacá, Arica, Chile

ARTICLE INFO

Article history:
Received 20 December 2019
Revised 8 February 2020
Accepted 6 April 2020
Available online 12 May 2020

Communicated by Dr. Ma Lifeng Ma

Keywords:
Machinery fault diagnosis
Deep transfer learning
Intelligent computing
Prognostics and health management

ABSTRACT

With the popularization of the intelligent manufacturing, much attention has been paid in such intelligent computing methods as deep learning ones for machinery fault diagnosis. Thanks to the development of deep learning models, the interference of the human experience can be greatly reduced, and the fault diagnosis accuracy can also be increased under certain conditions. To improve the generalization ability of the intelligent fault diagnostics, the deep transfer learning consisting of both transfer learning and deep learning components was accordingly developed. This paper reviews the research progress of the deep transfer learning for the machinery fault diagnosis in recently years. It is summarizing, classifying and explaining many publications on this topic with discussing various deep transfer architectures and related theories. On this basis, this review expounds main achievements, challenges and future research of the deep transfer learning. This provides clear directions for the selection, design or implementation of the deep transfer learning architecture in the field of the machinery fault diagnostics.

1. Introduction

Fault diagnosis is of great significance to improve production efficiency and to reduce accident rate for complex mechanical systems. Academic and industrial circles have attached great importance to the machinery fault diagnostics, and have put forward some diagnosis methods for real applications. Those fault diagnosis techniques can generally be divided into two categories: model-based diagnosis method and data-driven one. To reveal the relation between fault mechanism and parameters of the mechanical system, model-based diagnosis researches usually establish dynamic models of the machinery by means of dynamics, finite element and modal analyses. For example, a lumped-parameter model with 20 degrees of freedom was developed for a planetary gear [1]. Finite element models were developed for the meshing stiffness of sun-planet and ring-planet gears. Moreover, advanced signal processing methods have been applied to extract corresponding fault features from collected signals according to the characteristic frequency. Generalized synchrosqueezing transform [2,3], operating parameters identification [4], variable mode decomposition [5], and spectral segmentation [6] methods were reported as different signal processing techniques for the machinery fault diagnosis. Gao et al. [7] investigated the fault diagnosis problem for time-varying systems with multiplicative noises over sensor networks. A modified residual evaluation technique was presented to detect sensor faults [8]. Zhang et al. [9] developed an adaptive dynamics learning for weak fault detection. Based on the signal processing techniques, the fault detection problem has been studied for different systems [10], such as the multi-rate time-varying system [11], uncertain time-varying nonlinear stochastic system [12], nonlinear stochastic systems [13], and nonlinear Markovian jump system [14].

In recent years, some researchers have made a lot of fault diagnosis contributions based on machine learning and pattern recognition, in shallow learning, deep learning, or deep transfer learning frameworks. Some representative works are listed as follows.

(i) In shallow learning. For diagnosing machinery faults, support vector machine (SVM) model [15] global-local margin fisher analysis [16], continuous-scale mathematical morphology [17], fuzzy logic [18–20], and Bayesian approach [21] were reported, respectively. Hu et al. [22] combined intrinsic time scale decomposition, wavelet packet transform with correlation dimension algorithm to diagnose a wind gearbox using non-stationary vibration signals. Li et al. [23] systematically interpreted the primary literature on fuzzy formalisms for bearing fault diagnosis. Javed

* Corresponding author.
E-mail addresses: chuanli@21cn.com (C. Li), zhangsh@dgut.edu.cn (S. Zhang), qy_808@aliyun.com (Y. Qin), eestupin@academicos.uta.cl (E. Estupinan).

https://doi.org/10.1016/j.neucom.2020.04.045
0925-2312/

et al. [24] proposed a novel network, named summation wavelet extreme learning machine, for condition classification. A joint SPA and LKPCA technique was presented for the condition monitoring [25]. Different fuzzy clustering algorithms, such as fuzzy c-means, Gustafson-Kessel, FN-DBSCAN, and FCMFP methods, were compared in [26] for bearing fault diagnosis. Based on ant colony algorithm, Zhang et al. [27] developed a new SVM method for rotating machinery fault diagnosis. To improve forecasting accuracy, Li et al. [28] combined the random forest and ensemble empirical mode decomposition for improving forecasting accuracy. We developed several intelligent diagnosis algorithms, such as semi-supervised self-organizing mapping, neighbor preserving embedding with self-organizing mapping, hidden semi-Markov model, wavelet packet entropy with Gaussian mixture model, for dealing with the machinery condition monitoring [29–33]. All the above representative contributions are based on shallow learning mode, which is affected by the human experience and leads to poor generalization ability of the extracted sensitive features to some extent. To address this problem, Hinton and Salakhutdinov [34] proposed the deep learning model that can reduce the human experience and could effectively reveal the complex inputs.

(ii) In deep learning. Some other intelligent models have been introduced in the machinery fault diagnostics based on deep learning theory. Aghazadeh et al. [35] combined spectral subtraction and convolutional neural network (CNN) to monitor the state of tool grinding process. Deep statistical feature learning was developed for the fault diagnosis of rotating machinery [36]. A local connection network was constructed by normalized sparse autoencoders for planetary gearbox and motor bearing fault diagnosis [37]. A large capacity storage retrieval neural network was reported in [38]. A deep learning method named multimodal deep support vector classification was developed for gearbox fault diagnosis [39]. Autoencoders were employed for the fault diagnosis of rotating machinery [40,41]. Ding and He [42] applied deep CNNs to diagnose ring main bearings. Jiang et al. [43] proposed a multi-scale CNN method for the fault diagnosis of the wind turbine gearbox. Yang et al. [44] applied long short-term memory (LSTM) for the fault detection of electro-mechanical actuators. Sina et al. [45] proposed dynamic neural networks for the fault detection of gas turbine engines. Wu et al. [46] used vanilla LSTM method to obtain good result of remaining useful life estimation for aircraft turbofan engine. Hu et al. [47] employed denoised autoencoders to predict fan rotation speed. Gan et al. [48] constructed a bearing fault diagnosis model based on deep belief networks (DBNs). Mao et al. [49] proposed a deep output kernel learning method for bearing fault diagnosis. Wang et al. [50] applied a dropout method to prevent over-fitting of the deep learning model, which was used for the fault identification of the fan gear transmission system. Wang et al. [51] applied DBNs to detect multiple faults of axial piston pump. Chen et al. [52] presented a data-driven and deep learning-based method to detect early faults. In the fault diagnostics of rolling element bearings, the sparsity and neighborhood preserving deep extreme learning machine were combined to detect the fault from motor current signals [53]. Multi-scale features of the bearing state were extracted by using CNN for residual life prediction [54]. Jia et al. [55] applied deep neural networks for the fault diagnosis of planetary gearbox. Shao et al. [56] proposed a joint DBN and local linear embedding method for the bearing fault detection. Ren et al. [57] proposed an integrated deep learning method for the bearing life prediction. Peng et al. [58] presented an unsupervised health indicator construction method by using DBN and particle filter for remaining useful life prediction of the aircraft engine. Hu et al. [59] combined deep Boltzmann machine (DBM) with multi-granularity cascade forest method for the fault diagnosis. Liu et al. [60] proposed time series CNN for mechanical fault diagnosis. Zhang et al. [61] developed a local-global deep neural network to extract local and global features for detecting motor rolling bearing failure and wind turbine gears failure. Gao et al. [62] developed a Wasserstein generative adversarial network (GAN) with gradient penalty to address the problem of low-data or imbalanced data from the industrial process. We researched deep learning in the machinery fault diagnostics, by using different techniques such as sparse self-coding based on information fusion for bearing condition monitoring [63], deep fuzzy echo state networks [64], Bayesian and time series dimensionality reduction with LSTM model [65], deep hybrid state network with feature reinforcement [66], sparse echo autoencoder network [67], manifold sparse autoencoder [68], and evolved deep echo state network [69].

To achieve better result, training dataset and test dataset should be within the same distribution for classical deep learning algorithms. However, for practical applications, distributions for the training dataset (source domain) and the test dataset (target domain) are usually different, since they are collected from different working conditions or different mechanical devices. In this case, the deep learning model is difficult to achieve good recognition result. To overcome this problem, transfer learning was proposed to improve classification rate. The purpose of the transfer learning is to reduce the distribution difference between the source and target domains. Thus, the deep transfer learning method can effectively extract the common features of the source domain and the target domain.

(iii) In deep transfer learning. A deep transfer network model was proposed for bearing fault diagnosis based on the joint distribution adaptive method [70]. Shao et al. [71] used the deep transfer learning method for the fault diagnosis of motors, gearboxes and shaft bearings. Guo et al. [72] proposed a deep convolutional transfer network to realize the fault diagnosis of bearings in different machines. Sun et al. [73] combined sparse atuoencoders with deep transfer learning method to predict tool residual life. Hasan and Kim [74] combined CNN and transfer learning method for the bearing fault diagnosis. Pan et al. [75] applied the deep transfer learning method for the fault diagnosis of high voltage circuit breakers. Shao et al. [76] proposed an auxiliary classification model based on GANs, which learned from mechanical signals and generated corresponding one-dimensional signals for motor fault diagnosis. Liu et al. [77] proposed classification antagonism autoencoders for the fault diagnosis of rolling bearings. Wang et al. [78] combined GANs with stacked sparse atuoencoders to diagnose planetary gearboxes. Xiang et al. [79] used deep GANs to solve the domain adaptive problem and to diagnose the bearing fault of the train bogie.

Although deep learning and its derivative methods have achieved good diagnosis results, hyper parameters optimization performed by intelligent optimization algorithms [80,81] for the complex deep learning models may consume a lot of computational time. This brings difficulties for the real-time monitoring. In the actual industrial environment, it is extremely difficult to obtain enough system status samples. Experimental data and industrial data are difficult to meet the conditions of independent and same distributions. This makes the optimization model established by the laboratory difficult to work in the industrial scene. However, there are some contradictions, e.g., the one between the adjustment of complex structural parameters (for accurate diagnosis) and the real-time monitoring (for fast diagnosis); the one between the data dependence of the diagnosis model (more data for modeling) and the difficulty of obtaining industrial data (less data in real applications); and, the one between experimental diagnosis model and the industrial diagnosis demand. All those contradictions make diagnosis methods difficult to be popularized in the real environment.

To this end, this paper reviews the progress of the deep transfer learning and discusses future directions in this field. Main contri-

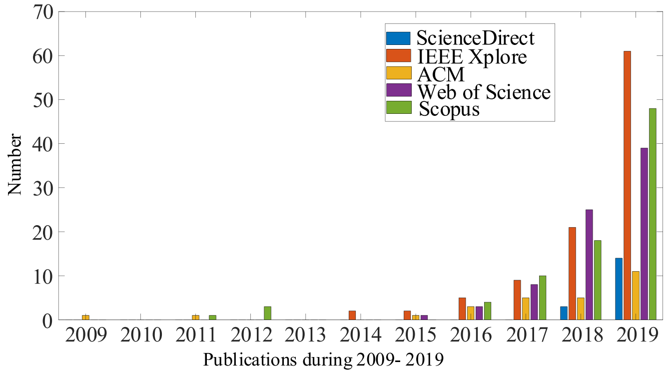

Fig. 1. Number of publications during 2009–2019 in five scientific index engines.

butions of this review are as follows: (i) A systematic review of the main literature on the machinery fault diagnostics with the deep transfer learning is presented. It is summarizing, classifying and explaining many publications on this topic. (ii) In this review, we provide a good background for the state-of-the-art research achievements on this subject, and point out the imperfections in the existing literature. And, (iii) analyze pros and cons of the deep transfer learning, so as to put forward new directions for the machinery fault diagnosis.

The rest of this paper is organized as follows. Literature collection and metrics are introduced in Section 2. Typical deep transfer learning models are described in Section 3. Applications of the deep transfer learning in fault feature extraction and fault classification are introduced in Section 4. Section 5 presents future developments in the related topics. Finally, conclusions are given in Section 6.

2. Literature collection

To attempt a liberal and reproducible review, a systematic literature collection of studies on the deep transfer learning for machinery fault diagnosis has been carried out. The collected works are from 2009 to now. The following scientific index engines and search strategy were employed.

(i) IEEE Xplore: (Full text & metadata: “transfer learning” OR “domain adaptation”) and then retrieval in result used (“fault diagnosis” OR “health detection”).
(ii) Science Direct: (Title, abstract, keywords: “transfer learning” OR “domain adaptation”) AND (Title, abstract, keywords: “fault diagnosis” OR “health detection”) AND (publication date: 2009-2019).
(iii) Web of Knowledge: (Theme: “transfer learning” OR “domain adaptation”) AND (theme: “fault diagnosis” OR “health detection”).
(iv) ACM: (Anywhere: “transfer learning” OR “domain adaptation”) AND (anywhere: “fault diagnosis” OR “health detection”) AND (publication date: 2009-2019). And,
(v) Scopus: (TITLE-ABS-KEY: ((“transfer learning” OR “domain adaptation”) AND (“fault” AND “diagnosis” OR “health detection”)) AND (publication date: 2009-2019) AND (EXCLUDE (DOCTYPE, “cr”)).

With the development of the transfer learning proposed by Pan et al. [82], the application of the transfer learning in the machinery fault diagnosis has become more and more popular. The publication time distribution is shown in Fig. 1. A new upsurge of the research on the transfer learning applied in machinery fault diagnosis is in the making. From 2009 to 2019, there are up to 300 publication records (duplications accounted) in this field. In 2019, for example, Science Direct, IEEE Xplore, ACM, Web of Knowledge and Scopus included 14, 61, 11, 39 and 48 publications on transfer learning for the fault diagnosis, respectively.

Table 1
Reported machinery diagnosed based on transfer learning.

Machinery	Literature
Bearing	[72,98,115–135,144,161,163–165,169,172,175,178,179,181,183–186,188,191,194–199,201]
Gearbox	[110–114,163,171,174,182,184,186,193,195,202]
Common rotating machinery	[88–97,165,166,168,177,190,200]
Motor	[98–109,163,189,192]
Others	[76,128–133,160,162,170,173,180,187]

The philosophy of the deep transfer learning (and some shallow versions) include transfer component analysis (TCA) [83], joint distribution adaptation (JDA) [84], deep adaptation networks (DANs) [85], adversarial domain adaptation (ADA) [86], and so on. By analyzing the retrieval results of various scientific index engines, one can see some influential research teams in this field including: Liang Gao, Ruqiang Yan, Qian Ding, Tao Zhang, Xiang Li, Haidong Shao, Weihua Li, Diego Carbrera, Mohsen Moghaddam, Chuan Li, … In addition to traditional mechanical fault diagnosis, such as common rotating machinery [87–97], motors [98–109], gearboxes [110–114], rolling element bearings [98,115–136], there were some other systems diagnosed such as 3D printers [83], analog circuits [137,138], catenary support component [139], microelectronics attachment [140], pipeline [141], tool [142] and embedded operating systems [143]. The fault diagnosis systems based on transfer learning are detailed in Table 1. It shows that most of transfer learning models were used for gearboxes and bearings. Most of the existing works considered the difference between working conditions for the same type of the machinery. When the data acquired from different components and the dimensions for the source and the target domains are different, it is necessary to further study how to ensure the input consistency for the transfer learning model.

3. Typical deep transfer learning models

3.1. Basics of deep transfer learning in fault diagnostics

As shown in Fig. 2a typical procedure of the intelligent fault diagnosis can be outlined as follows: (i) Collect raw data under different faulty modes from sensors installed on the machinery; (ii) Divide the dataset into training and testing ones; (iii) Develop a diagnosis model for feature extraction by using the training dataset; (iv) Develop a classifier based on the training dataset; and (v) Calculate the accuracy using the testing data by employing the developed intelligent model.

Machinery fault diagnosis using intelligent computing such as artificial neural networks can effectively identify the health condition. However, traditional artificial neural networks require a large number of labeled samples for training. This greatly limits the application of intelligent computing methods for the machinery fault diagnosis. Moreover, the adaptability of different working conditions is poor.

For industrial machinery, collected data are usually not evenly distributed. This means that the datasets from different operating conditions are not within the same distribution. Many researchers have proposed several solutions to overcome this problem. However, most of them tried to extract features of the original dataset to unify them. This is the original philosophy of the transfer learning. After this, the transfer learning has been applied more and more in the intelligent fault diagnosis of the machinery. It solves the problem of related fields (target domain) through known do-

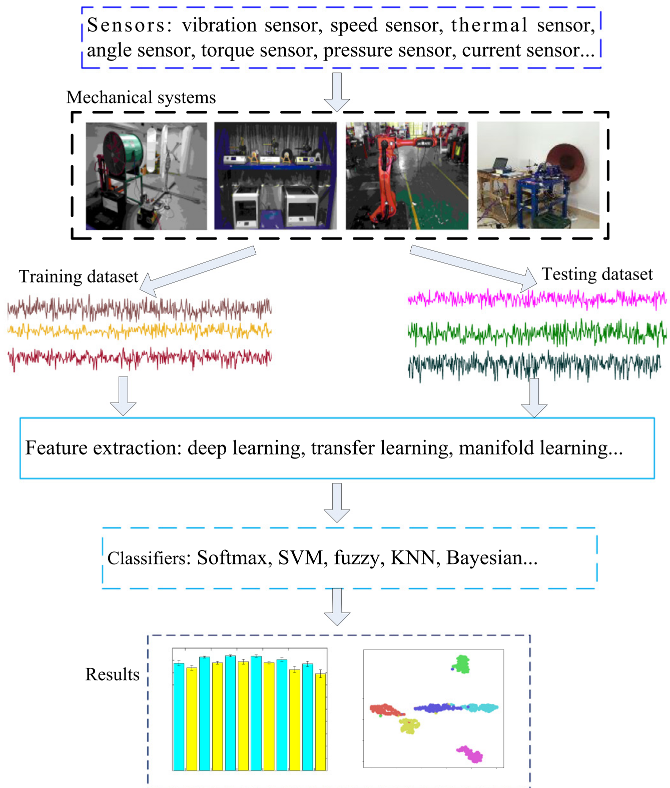

Fig. 2. A typical procedure of the intelligent fault diagnosis approach.

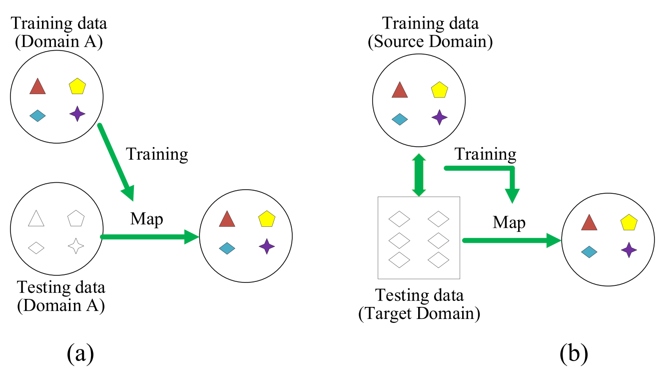

Fig. 3. Framework of the domain adaptation approach: (a) traditional intelligent method; and (b) domain adaption based intelligent method.

main knowledge (source domain). This leads to the development of domain adaptation algorithms such as TCA, JDA, and correlation comparison (CORAL). The domain adaptation framework is shown in Fig. 3 [144].

A machine learning method based on the shallow structure cannot acquire abstract features by minimizing the probability distribution difference among different domains [145]. It needs to extract features based on deep learning and probability distribution adaptation. As one of the most potential technologies for big data, deep learning can extract compact, hierarchical and abstract data representation, and has the ability to transfer between domains [146]. It can be seen that under the unified framework of the deep transfer learning, domain invariant deep representation learning and domain probability distribution difference can be modified simultaneously. Domain separation network [147], zero-shot learning [148], few-shot adversarial domain adaptation [149,150], residual transfer network [151], state frequency matrix [152] have been used in cross domain public opinion analysis, computer vision, spam classification, behavior recognition, stock price prediction, and so on.

3.2. Transfer component analysis (TCA)

TCA was firstly proposed by Pan et al. [82] for learning across domain feature representations under shallow framework. It may also be used for deep transfer learning. TCA assumes that there exists a feature mapping φ, which makes the distribution of data in source and target domains closer after mapping. The distance called maximum mean discrepancy (MMD) is used to measure the difference between source and target domains. Given source labeled data $D_S=\{(x_{S1},y_{S1}),\ldots,(x_{Sn1},y_{Sn1})\}$ and target unlabeled data $D_T=\{x_{T1},\ldots,x_{Tn2}\}$, $P(X_S)$ and $Q(X_T)$ denote the marginal distributions from source and target domains, respectively. The distance between two distributions P and Q can be empirically measured as

$$Dist\left(X_S', X_T'\right) = \left\| \frac{1}{n_1}\sum_{i=1}^{n_1}\phi\left(x_{S_i}\right) - \frac{1}{n_2}\sum_{i=1}^{n_2}\phi\left(x_{T_i}\right)\right\|_{\mathcal{H}}^2 \tag{1}$$

where n denotes the number of data, and H means a universal reproducing kernel Hilbert space. Then introduce a kernel matrix K and a condition matrix L, i.e.,

$$K = \begin{bmatrix} K_{S,S} & K_{S,T} \\ K_{T,S} & K_{T,T} \end{bmatrix} \in R^{(n_1+n_2)*(n_1+n_2)} \tag{2}$$

$$L = \left[L_{i,j}\right] \begin{cases} \dfrac{1}{n_1^2} & x_i, x_j \in X_S \\ \dfrac{1}{n_2^2} & x_i, x_j \in X_T \\ -\dfrac{1}{n_1 n_2} & otherwise \end{cases} \tag{3}$$

The objective function can be then written as

$$\min_{K \geq 0} Dist\left(X_S', X_T'\right) = \min_{K \geq 0} tr(KL) - \lambda tr(K), \tag{4}$$

where $\lambda \geq 0$ is a trade-off parameter, which can be determined empirically. One can accordingly decompose the kernel matrix $\mathbf{K}$ into

$$K = \left(KK^{-\frac{1}{2}}\right)\left(K^{-\frac{1}{2}}K\right) \tag{5}$$

The use of a $(n_1+n_2) \times m$ matrix $\tilde{W} \in R^{(n_1+n_2)\times \mathrm{m}}$ is considered to simplify $\mathbf{K}$ as

$$\tilde{K} = (KK^{-1/2}\tilde{W})(\tilde{W}^T K^{-1/2} K) = KWW^T K \tag{6}$$

where $W = K^{-1/2}\tilde{W} \in \mathbb{R}^{(n_1+n_2)\times \mathrm{m}}$. The distance is then simplified as

$$Dist(X_S', X_T') = tr((KWW^T K)L) = tr(W^T KLKW) \tag{7}$$

The regularization term $tr(W^{\mathrm{T}}W)$ is usually needed to control the complexity of W. To minimize the above equation, the kernel learning is then reduced to

$$\begin{aligned} &min_W \quad tr\left(W^{\mathrm{T}}W\right) + \mu\ tr\left(W^{\mathrm{T}}KLKW\right) \\ &s.t. \quad W^{\mathrm{T}}KHKW = I_m \end{aligned} \tag{8}$$

where the trade-off parameter is μ, the identity matrix is $I_m \in \mathbb{R}^{m\times m}$, and the centering matrix is $\mathbb{H} = I_{n_1+n_2} - \frac{1}{n_1+n_2}11^{\mathrm{T}}$. $I_{n_1+n_2} \in \mathbb{R}^{(n_1+n_2)\times(n_1+n_2)}$ and $1 \in \mathbb{R}^{n_1+n_2}$ are the identity matrix and the column vector with all ones, respectively. In this way, TCA algorithm is transformed into an optimization problem. To solve this optimization efficiently, Lagrangian constraint is introduced for Eq. (8) to transform the optimization problem as

$$\min_W tr\left(\left(W^{\mathrm{T}}KHKW\right)^{\dagger} W^{\mathrm{T}}(I + \mu KLK)W\right) \tag{9}$$

The solution of W is m leading eigenvectors of $(\mu I + KLK)^{-1}KHK$, where $m \leq n_1 + n_2 - 1$. Then, the outputs are $\theta_s = X_S' \cdot W = \{(x_i, y_i)\}, i \in k$ as the training dataset and $\theta_t = X_T' \cdot W = \{x_i\}, i \in k$ as the testing set, respectively.

The advantage of TCA is that it is an unsupervised method so that the implementation is simple. However, it still needs a lot of computation time for large matrices and its nonlinear fitting is poor.

3.3. Joint distribution adaptation (JDA)

JDA was firstly proposed by Long et al. [84] in 2013. The goal of JDA is to reduce the distance of the joint distribution between source and target domains for the domain adaptation. In other words, JDA takes use of two distances to approximate the difference between the source and target domains.

Let a domain D consisting of an m-dimensional feature space X and a marginal probability distribution $P(x)$, $D=\{X, P(x)\}$, where $x \in X$. A task T is composed of a label set Y and a classifier $f(x)$, i.e. $T=\{y, f(x)\}$, where $y \in Y$ and $f(x) = Q(y|x)$ can be represented as the conditional probability distribution. One can define the problem of joint distribution adaptation as below.

For a given labeled source domain $D_s = \{(x_1, y_1), \ldots, (x_{ns}, y_{ns})\}$ and a given unlabeled target domain $D_t = \{x_{ns+1}, \ldots, x_{ns+nt}\}$, they are under the assumptions that $X_s = X_t$, $Y_s = Y_t$, $P_s(x_s) \neq P_s(x_s)$, $Q(y_s|x_s) \neq Q(y_t|x_t)$. The target of the adaptation is to learn a feature representation in which the distribution differences between i) $P_s(x_s)$ and $P_s(x_s)$, and ii) $Q(y_s|x_s)$ and $Q(y_t|x_t)$ are explicitly reduced.

The joint distributions are proposed to adapt by a feature transformation T so that the joint expectations of the feature x and label y are matched between domains:

$$\begin{aligned} &\min_T \left\| E_{P(x_s,y_s)}[T(x_s), y_s] - E_{P(x_t,y_t)}[T(x_t), y_t] \right\|^2 \\ &\approx \left\| E_{P(x_s)}[T(x_s)] - E_{P(x_t)}[T(x_t)] \right\|^2 \\ &\quad + \left\| E_{Q_s(y_s|x_s)}[y_s|T(x_s)] - E_{Q_t(y_t|x_t)}[y_t|T(x_t)] \right\|^2 \end{aligned} \tag{10}$$

It contains two adaptation parts, i.e., marginal distribution adaptation and conditional distribution adaptation. Denoting $\mathbf{X}=[x_1,\ldots, x_n] \in \mathrm{R}^{m\times n}$ as the input data matrix, for adapting the marginal distribution, the distance between the sample means of the source and target data in the k-dimensional embeddings can be written as

$$\left\| \frac{1}{n_s}\sum_{i=1}^{n_s} \mathbf{A}^{\mathbf{T}} x_i - \frac{1}{n_t}\sum_{j=n_s+1}^{n_s+n_t} \mathbf{A}^{\mathbf{T}} x_j \right\|^2 = tr(\mathbf{A}^{\mathbf{T}}\mathbf{X}\mathbf{M_0}\mathbf{X}^{\mathbf{T}}\mathbf{A}) \tag{11}$$

where $\mathbf{A}$ is an orthogonal transformation matrix, and $\mathbf{M_0}$ is the MMD matrix given by

$$(\mathbf{M_0})_{ij} = \begin{cases} \dfrac{1}{n_s n_s} & x_i, x_j \in D_s \\ \dfrac{1}{n_t n_t} & x_i, x_j \in D_t \\ -\dfrac{1}{n_s n_t} & otherwise \end{cases} \tag{12}$$

As for adapting the conditional distribution, the distance between the class-conditional distributions $Q_s(x_s|y_s = c)$ and $Q_t(x_t|y_t = c)$ can be described as

$$\left\| \frac{1}{n_s^{(c)}}\sum_{x_i \in D_s^{(c)}} \mathbf{A}^{\mathbf{T}} x_i - \frac{1}{n_t^{(c)}}\sum_{x_j \in D_t^{(c)}} \mathbf{A}^{\mathbf{T}} x_j \right\|^2 = tr(\mathbf{A}^{\mathbf{T}}\mathbf{X}\mathbf{M}_c\mathbf{X}^{\mathbf{T}}\mathbf{A}) \tag{13}$$

where $c \in \{1,\ldots, C\}$ in the label Y, $D_s^{(c)} = \{x_i : x_i \in D_s \wedge y(x_i) = c\}$ is the set of examples belonging to class c in the source data, $y(x_i)$ is the true label of x_i, and $n_s^{(c)} = |D_s^{(c)}|$. Correspondingly, $D_t^{(c)} = \{x_j : x_j \in D_t \wedge \hat{y}(x_j) = c\}$ is the set of examples belonging to class c in the target data, $\hat{y}(x_j)$ is the pseudo label of x_j, and $n_t^{(c)} = |D_t^{(c)}|$. Thus, the MMD matrices $\mathbf{M_c}$ involving class labels are computed as

$$(\mathrm{M_c})_{ij} = \begin{cases} \dfrac{1}{n_s^{(c)} n_s^{(c)}} & X_i, X_j \in D_s^{(c)} \\ \dfrac{1}{n_t^{(c)} n_t^{(c)}} & X_i, X_j \in D_t^{(c)} \\ \dfrac{-1}{n_s^{(c)} n_t^{(c)}} & \begin{cases} X_i \in D_s^{(c)}, X_j \in D_t^{(c)} \\ X_j \in D_s^{(c)}, X_i \in D_t^{(c)} \end{cases} \\ 0 & otherwise \end{cases} \tag{14}$$

To achieve effective and robust transfer learning, the differences in both the marginal distributions and conditional distributions across domains should be simultaneously minimized. To this end, the JDA optimization problem can be expressed as

$$\min_{\mathbf{A^T K H K^T A = I}} \sum_{c=0}^{C} tr(\mathbf{A^T K M_c K^T A}) + \lambda \|\mathbf{A}\|_F^2 \tag{15}$$

where λ is the regularization parameter to guarantee the optimization problem to be well-defined, and $\mathbf{A} \in \mathrm{R}^{n\times k}$ is the adaptation matrix for the kernel of JDA.

JDA adapts source and target distributions to one optimization goal. Hence it improves the performance of the model by combining a weak classifier. Unlike TCA method, however, JDA is a supervised learning model and therefore needs an iterative process.

3.4. Deep adaptation network (DAN)

Based on the deep domain confusion method, Long et al. [85] developed DAN. It introduced three domain adaptation layers at three layers before a classifier by taking use of multiple kernel variant of MMD (MK-MMD) to measure the distance between source and target domains. The MK-MMD can be presented as

$$K \triangleq \{k = \sum_{u=1}^{m} \beta_u k_u : \beta_u \geq 0, \forall u\} \tag{16}$$

where the constraints on coefficients $\{\beta_u\}$ are imposed to guarantee that the derived multi-kernel k is characteristic. The optimization objective consisting of the loss function and the distance of the distributions can be presented as

$$\min_{\Theta} \frac{1}{n_a} \sum_{i=1}^{n_a} J(\theta(x_i^a), y_i^a) + \lambda \sum_{l=l_1}^{l_2} d_k^2(D_s^l, D_t^l) \tag{17}$$

where Θ denotes all weights and bias parameters of the network, l_1 and l_2 are layer 6 and layer 8 of the Alexnet model [153], respectively, n_a means a collection of all annotated data in both source and target domains, $\theta(x_i{}^a)$ is the conditional probability that the CNN assigns $x_i{}^a$ to label $y_i{}^a$, and $J(\cdot)$ is a loss function using the cross-entropy in the deep networks.

The innovations of DAN are the multi- domain adaptation and multiple kernel variant of maximum mean discrepancies. Since DAN model is based on AlexNet architecture, enough samples with labels information are necessary in the training process.

3.5. Adversarial domain adaptation (ADA)

With the popularity of GAN [154], the concept of ADA employing the adversarial training for domain adaptation was also emerged. The theme of ADA is to make the generator as a feature extractor of the target domain. This keeps learning the feature of data from two domains and the domain discriminator cannot distinguish where domains data come from. The loss function of ADA is given by

$$l = l_c(D_s, y_s) + \lambda l_d(D_s, D_t) \tag{18}$$

where l_c denotes the loss of net and l_d stands for the loss of the domain discriminator.

Ganin et al. [86] was the first researcher introduced adversarial training into domain adaptation procedure. The proposed algorithm was called domain-adversarial neural network (DANN), which aimed to generate features for distinguishing features draw from two domains and to confuse the discriminator as much as possible. The loss function of DANN can be written as

$$l_d = \max \left[-\frac{1}{n} \sum_{i=1}^{n} L_d^i(\mathbf{W}, \mathbf{b}, \mathbf{u}, z) - \frac{1}{n'} \sum_{i=n+1}^{N} L_d^i(\mathbf{W}, \mathbf{b}, \mathbf{u}, z) \right] \tag{19}$$

where (**W, b**) is a matrix pair, (**u**, z) is a vector-scalar pair and L_d can be represented as

$$L_d(G_d(G_f(x_i)), d_i) = d_i \log \frac{1}{G_d(G_f(x_i))} + (1 - d_i) \log \frac{1}{G_d(G_f(x_i))} \tag{20}$$

where G_d is a domain regressor, G_f is a hidden layer, and d_i denotes the binary variable (domain label) for the i-th sample, which indicates whether x_i come from the source distribution (d_i=0) or the target distribution (d_i=1).

Bousmalis et al. [155] expanded DANN into domain separation networks (DSNs). DSN considers that source and target domains consist of public and private sectors, respectively. The public sector is used for learning the common features and the private sector for keeping the independence of features between the two domains. The loss function is accordingly defined as

$$l = l_{task} + \alpha l_{recon} + \beta l_{difference} + \gamma l_{similarity} \tag{21}$$

where l_{task} means the training loss, l_{recon} means the reconstruction loss keeping the private sector useful, $l_{diffrence}$ means the loss of difference between two sectors, and $l_{similarity}$ means the similarity loss of public sector between source and target domains.

Tzeng et al. [156] proposed an adversarial training framework called adversarial discriminative domain adaptation which set various options such as: the loss to be chosen, the weight to be shared, and the base model to be chosen. Inspired by Wasserstein GAN [157], Shen et al. [158] expanded the standard adversarial loss into Wasserstein distance guided loss for ADA. Long et al. [159] proposed conditional domain adversarial networks (CDANs) to tackle two problems that ADA methods are faced. The first one is that ADA fails to capture multimodal structures. The second one is that it is only to align features rather than labels when performing the adaptation. CDANs were applied for multilinear conditioning and entropy conditioning to solve these two problems. The objective function is given by

$$\begin{aligned} &\min_G \mathrm{E}_{(x_i^s, y_i^s) \sim D_s} L(G(x_i^s), y_i^s) \\ &+ \gamma \left(\mathrm{E}_{x_i^s \sim D_s} \log \left[D(T(\mathbf{h}_i^s)) \right] + \mathrm{E}_{x_j^t \sim D_t} \log \left[1 - D(T(\mathbf{h}_j^t)) \right] \right) \\ &\times \max_D \mathrm{E}_{x_i^s \sim D_s} \log \left[D(T(\mathbf{h}_i^s)) \right] + \mathrm{E}_{x_j^t \sim D_t} \log \left[1 - D(T(\mathbf{h}_j^t)) \right] \end{aligned} \tag{22}$$

where γ is a hyper-parameter between source classifier G and conditional domain discriminator D, **f** and **g** mean the domain-specific feature representation and classifier prediction, respectively. So $\mathbf{h} = (\mathbf{f}, \mathbf{g})$ is the joint variable for the adversarial adaptation.

Based on GAN theory, ADA can enhance positive transfer and weaken negative transfer. Moreover, ADA is an unsupervised method. The shortcoming is that ADA cannot acquire better performance in the fault diagnosis under different working conditions.

3.6. Other transfer learning methods

In addition to the abovementioned transfer learning methods, there have been some other reported methods such as CORAL and TriTL, and their combined models. According to the model architecture, one can divide transfer learning into shallow and deep ones. Compared to the shallow transfer learning, the most of researches in transfer learning are based on the deep structure, such as deep GANs [88,94,108,119,129,160–162], CNNs [72,98,101,123,140,163–166], stack sparse auto-encoders (SAEs) [91,92,99,108,114,124,129,134,165–168], deep convolutional transfer networks, as well as the improvement methods for special problems. The basic purpose of these transfer learning methods is to apply the knowledge acquired from one problem to another different but related problem. Transfer learning has achieved good results in image processing, speech recognition, and other fields. In the machinery fault diagnostics, furthermore, transfer learning methods can deal with small-sample problems where the normal dataset is effortless to collect, while it is expensive to obtain the fault mode dataset. The intelligent fault diagnosis methods need enough dataset which is hard to achieve in industrial environment. Thus, transfer learning can be used to overcome those small-sample problems.

Shallow transfer network cannot deal with nonlinear data, whereas deep transfer learning model needs enough source domain samples with enough training process. For one-shot or few-shot conditions in the source domain, the above transfer learning methods is difficult to achieve good results.

4. Applications of deep transfer learning to machinery fault diagnosis

4.1. Deep transfer learning for the fault feature extraction

Though had been developed in 2009, transfer learning was recently used for feature extraction in the machinery fault diagnostics. In 2016, transfer learning was introduced by Shen et al. [169] to improve bearing diagnostic performance under different working conditions. It adjusted selective auxiliary data and auxiliary target data classification weights to enhance feature extraction ability. In this work, negative transfer was avoided through the similarity judgment. Yang et al. [170] proposed a radar emitter identification method based on three-dimensional distribution characteristics and transfer learning. This method firstly describes the intra-pulse modulation information of the radar emitter by cubic feature of time-frequency-energy distribution, and then reconstructs the feature by using transfer learning for obtaining the robust feature against signal-noise-rate (SNR) variation. Wang et al. [171] proposed a transfer learning method, i.e., transfer factor analysis, to reduce regional differences caused by various working conditions for gearbox diagnosis. The features are transferred to low-dimensional potential spaces to minimize regional differences while preserving data attributes to find the main features in different domains under different operating conditions. Deep learning can extract hierarchical representation features of raw data, while transfer learning provides a good way to perform learning tasks on different but related distributed datasets. It should be noted that there are different aspects between deep transfer learning and multi-task learning. Deep transfer learning focuses on reducing the distribution difference between the source and target domains for improving the classification rate. On the other hand, multi-task learning focuses on the multiple characteristics of the data source in the training process, making the indicators more reliable. Wen et al. [172] proposed a deep transfer learning method by introducing deep networks into transfer learning. A three-layer sparse autoencoder was employed to extract original data features. The maximum average error term between the training and the testing data features was applied to reduce the error loss. This method has a good feature extraction effect on the bearing dataset of Case Western Reserve University and the improved deep transfer learning structure is shown in Fig. 4 [172]. Yu et al. [173] built a transfer feature extraction model for the defect identification of wind turbine blades.

In recent years, adaptive neural feature extraction methods based on deep neural networks have attracted much attention. Those methods usually require a large amount of data for training. Many researchers have developed a lot of ways to solve this problem. Cao et al. [174] proposed a transfer learning method based on CNN. Pre-trained deep neural networks were first used to automatically extract features from input gear experimental data. A fully connected stage was then presented to classify features to be trained. This method can not only obtain free adaptive feature extraction of preprocessing, but need relatively less training data as well. To improve model feature extraction and fault classification performance, Tong et al. [175] proposed the domain adaptation using feature transfer learning under variable conditions. The datasets of normal and faulty bearings were obtained by fast Fourier transform on vibration signals collected under different motor speeds and load conditions. Then, by refining pseudo-test labels based on the maximum mean difference and the domain-invariant clustering in the normal space, the marginal and conditional distributions were simultaneously reduced between the

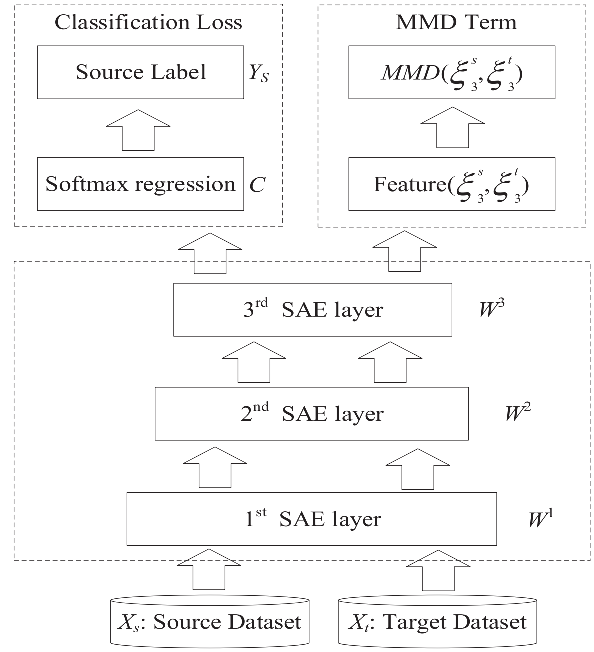

Fig. 4. Improved deep transfer learning.

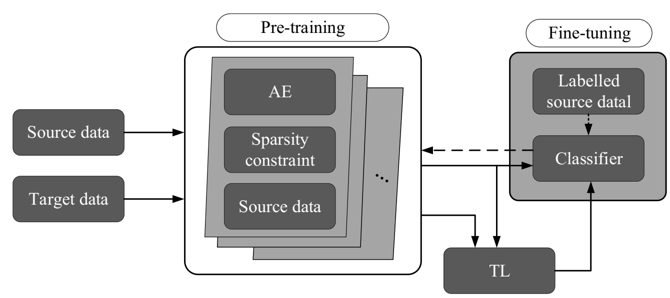

Fig. 5. The framework of the optimized transfer learning algorithm.

training data and the testing data. Finally, a transferable feature representation for training data and testing data was implemented.

In practical applications, the sample imbalance in the machinery fault diagnosis is accompanied by the domain adaptive problem. On the other hand, transfer learning cannot extract domain-invariant features that do not change with the rotational speed. This leads to poor results when the diagnosis model is applied to other working conditions without transferring the rotating speed. Xu et al. [176] developed a convolutional transmission feature discrimination network structure. Scaled exponential linear unit activation function was used in this network. This method adopted a two-branch network architecture of weight sharing, including a deep feature extraction network for fusing feature extraction, a transfer learning network for domain-adaptive problems, and an unbalanced-sample feature discrimination network for feature similarity determination. To achieve better accuracy, Zhang et al. [177] developed a transfer subspace learning framework for sucker rod pumping wells. Sun et al. [178] introduced an optimized transfer learning algorithm to solve the domain adaptation problem. By directly inheriting the features obtained from the pre-training process in the source domain and changing only the fine-tuning process, the complexity of the algorithm was dramatically reduced. The framework is shown in Fig. 5 [178].

When applying the deep learning to initial troubleshooting, the initial number of fault samples is usually very limited. Chen et al. [179] proposed a DNN-based transfer learning model, which

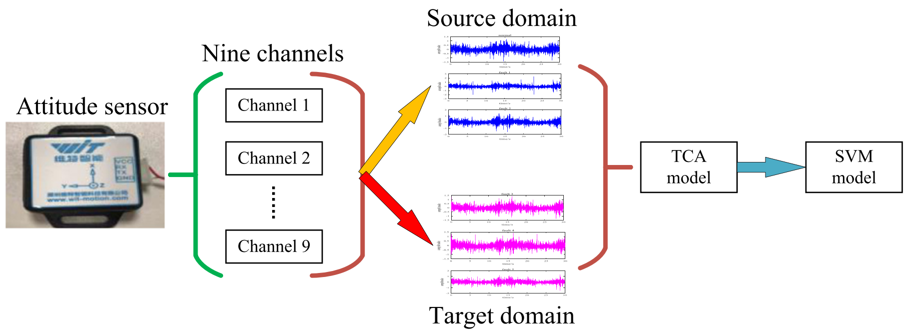

Fig. 6. Overview of the TSVM approach.

can extract fault features involved in a large number of important fault samples with less initial fault samples. Duan et al. [180] presented a transfer learning method to deal with the lack of fault samples for deep learning. Each fault location was defined as a classification label. These labels were classified by using different CNNs to obtain fault location results. Image segmentation was then performed to extract the features of the fault area and to simplify the amount of data. Ma et al. [181] proposed a transfer learning- CNN (TLCNN) based on AlexNet for bearing fault diagnosis. A two-dimensional image representation was first used to convert a vibration signal into a two-dimensional time-frequency image. The TLCNN model was used to extract the features of the two-dimensional time-frequency image for classifying bearing conditions. The training speed is faster and the precision is higher in this method. Wen et al. [164] presented a fault diagnosis transfer learning method based on pre-training VGG-19 (TranVGG-19). Time domain signals were converted into RGB images. The pre-trained VGG-19 was exploited as a feature extractor for the converted images.

4.2. Deep transfer learning for the fault classification

For real applications, there are few failure modes for operating mechanical systems. As mentioned before, it is usually expensive to obtain these specific failure modes. Moreover, the failure is unpredictable in the operation process. This means that the contradiction between high reliable diagnosis and few fault samples highly affects the machinery fault diagnosis. With the help of transfer learning algorithms, fortunately, fault diagnosis models from the laboratory can be transformed into the industrial field. In this way, an intelligent fault classification model can be established for the industrial application. To this end, some publications have introduced the fault classification based on the transfer learning [182,183] for the machinery fault diagnosis under variable speeds and variable load conditions. Those methods employed transfer component analysis [182] and neural networks [183] to assist the training classifier for the diagnosis. Transfer learning and deep learning methods were combined for modeling transferable fault classification [184,185]. To tackle small-sample problems, a transfer support vector machine (TSVM) technique was introduced for the fault classification of delta 3D printers [83]. The overview of the TSVM approach is shown in Fig. 6.

Usually, the speed variation and the load change may trigger the deviation between different datasets. This seriously affects the performance of the machine learning-based fault classification methods. Qian et al. [90] proposed a joint high-order Kullback-Leibler divergence and transfer learning algorithm for the fault classification. To classify fault types for rotating machinery under variable working conditions, Qian et al. [91] constructed a three-level deep fault diagnosis network using adaptive batch normalization. It does not require repeated training on the target dataset. An et al. [144] employed domain adaptation with multiple-kernel method to promote the successful applications for the intelligent fault classification. Qian et al. [186] proposed an improved joint distribution adaptation to align the edge distribution and conditional distribution of datasets. On this basis, a robust fault classification method based on vibration signals was proposed. A data enhancement method was first proposed to generate more useful samples for unbalanced vibration signals with innovatively using noise to improve network performance. Sparse filtering was then used to reduce the input dimension. Finally, shared features and main features were exploited for the fault classification.

To improve the accuracy of the fault classification, a diagnostic model framework based on transfer learning technique was reported in Ref. [187]. It extracted additional diagnostic knowledge by transmitting data information from other femtocells. Domain-adversarial neural networks were introduced to overcome the over-fitting problem of the actual training data [188]. This improves the generalization ability for new machines and new working conditions. Xiao et al. [189] used deep transfer learning for motor fault classification, where the knowledge of mark dataset under constant conditions was transferred to unmarked dataset under continuous changing conditions. CNN was first used as the basic framework to extract multi-level features from original vibration signals. Mean maximum deviation regularization term was then introduced during the training process. CNN parameters were therefore constrained to reduce the mismatch of the source and target domain feature distributions. In dynamically changing production processes, different data do not obey the same distribution. To tackle this problem, a two-stage digital double-aided method based on deep transfer learning was developed to realize the faults classification in the development phase and the maintenance phase [190]. Ultra-high fidelity model was run in the virtual space, followed by the training of the deep neural network-based classification model. By using deep transfer learning, the trained model was migrated from virtual space to physical space for real-time monitoring and predictive maintenance. Moreover, a deep transfer learning method was introduced for bearing fault classification. This method transferred the diagnostic knowledge learned from supervised data of multiple rotating machinery to the target device through domain adversarial training.

Traditional fault diagnosis requires a large amount of samples. Wang et al. [167] developed a heterogeneous transfer learning method for the fault classification. In the present technique, the source dataset and target domains were represented by heterogeneous features of different dimensions. Moreover, the source domain data were used to train a SVM and the classification accuracy was obtained from the target domain data. Zhang et al. [191] proposed a transfer learning neural network to improve the fault classification performance in different working conditions. The

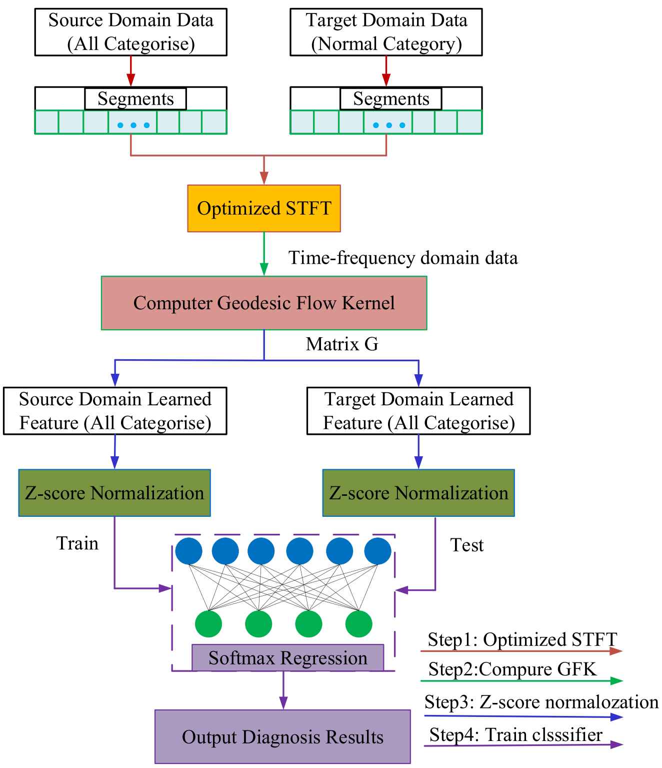

Fig. 7. Domain adaptation model based on geodesic flow kernel.

transfer learning approach extracted characteristics from a large amount of source data and adjusted parameters of the architecture subsequently. Both the source and the target network structures shared the same parameters. The target model was finally trained with a small data under another condition to overcome the small-sample problem. Xiao et al. [192] proposed a small batch target data framework based on transfer learning by using the improved TrAdaBoost algorithm and CNN. A fault classification model was proposed by Chen et al. [193] for dealing with missing data based on the transfer learning. This method can extract additional useful fault features by designing a suitable transfer learning mechanism. Yang et al. [194] proposed a feature-based transfer neural network for the fault classification. The method used a CNN to extract the transferable features of raw vibration data. The regularization term of the multi-layer domain was then adapted. Pseudo-label learning was used to constrain the parameters of CNN to reduce the distribution difference and the inter-class distance of the learned transferable features. Zhang et al. [195] proposed a domain adaptation model based on geodesic flow kernel, which strengthened feature extraction and Z-score normalization for the gear and bearing fault classifications. The procedure of this method is shown in Fig. 7 [195].

In the machinery fault diagnostics, the problem of the generalization ability still exists in an individual deep learning model. To tackle this problem, a negative correlation ensemble transfer learning method was proposed for bearing fault classification with 98.73% accuracy [196]. In this method, a ResNet-50 structure based on transfer learning was proposed with a 50-layer deep learning structure. Cooperative learning was subsequently performed on multiple fully connected layers with negative correlation learning as soft maximum classifiers. Finally, hyper parameters of the present model was determined by cross-validation method. Xie et al. [197] introduced a transfer learning approach with a cycle-consistent GAN model. Based on known conditions, the designed neural network generated new samples under unknown conditions for the pre-trained classifier to perform the fault classification. Hasan and Kim [198] developed a transfer learning classification method based on CNN. The method extracted the knowledge from a large collection of source datasets to identify faults from the target data. The maximum average error term was used to reduce the error loss between the training data and the testing data features. Yang et al. [199] presented a transfer learning method, named convolution adaptive network, for motor bearing condition recognition. In the present method, CNN was employed to extract features, and motor bearings in the laboratory were used for validating the present method. Shao et al. [163] developed a deep learning framework to achieve high-precision fault classification by using transfer learning neural networks. The method converted original sensor data into an image by a wavelet transform. Low-level features were extracted by using the pre-training network. High-level network structure was fine-tuned using the time-frequency image of the marker. Tang et al. [200] proposed a gas path analysis method based on the transfer learning to improve the classification accuracy by 11–20%. This method combined transfer learning with data-driven gas path analysis to learn both failure modes and diagnostic conditions. Zhu et al. [201] introduced a fault diagnosis method to extend CNN to transfer learning. For task-specific features, hierarchical parameters were used to adjust CNN. At the same time, the linear loss of multiple Gaussian kernels was used to calculate the domain loss for improving the adaptive ability. Li et al. [202] developed a domain adaptive diagnosis model to classify early gear pitting faults under multiple working conditions. The improved domain adaptive neural network was optimized by particle swarm optimization algorithm and L2 regularization method. The stability and accuracy of the machinery fault diagnosis are all improved accordingly. Guo et al. [72] proposed a deep convolutional transfer learning network (the structure is shown in Fig. 8 [72]), including two modules: state recognition and domain adaptation. A one-dimensional CNN was used to construct a state recognition module for automatically learning the characteristics and for identifying the health condition of the machine. The domain adaptation module in the one-dimensional CNN was developed for learning domain invariant features by maximizing domain identification errors and minimizing probability distribution distances.

Intelligent diagnosis method has made good developments in the fault recognition, fault feature extraction, dynamic condition monitoring, fault severity evaluation [203], and remaining useful life prediction [204]. However, the existing retrieval results show that deep transfer learning methods were mainly applied in the fault feature extraction and the fault classification. To our knowledge, the applications of the deep transfer learning in healthy condition evaluation, and residual life prediction have not been reported.

5. Future developments

5.1. Challenges of the deep transfer learning in the fault diagnosis

According to the above analysis, there have been in general four different transfer substances in the machinery fault diagnosis. The first one is the transfer between different work conditions, such as different rotation speeds, working loads, based on distance, feature, or parameter criteria [169,171]. The second type is the transfer between mechanical components, such as between gears and bearings. The third one is the transfer from well-established networks to a special task for the deep transfer learning [71]. The last category lies in the transfer between different mechanical environments [72], for example, between a laboratory and an industrial workshop.

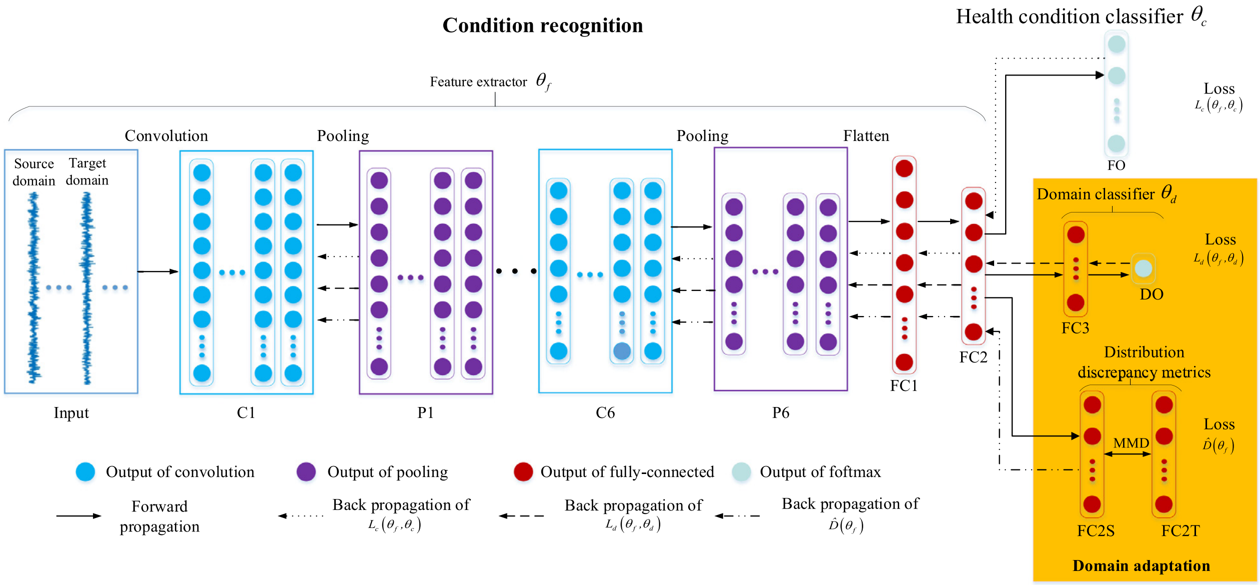

Fig. 8. Structure illustration of the deep convolutional transfer learning network.

Existing deep transfer learning methods have contributed a lot for the machinery fault diagnosis. On the one hand, it overcomes insufficient samples problem for such case as it is expensive to acquire a large amount of the fault pattern samples [83]. On the other hand, it solves the modeling problem without a prior knowledge. Before installing on a real industrial system, the fault diagnosis model can be tested in a laboratory by using the deep transfer learning. Moreover, when employing a trained deep learning network for specific tasks by transfer partial structure, the computation complexity of deep learning models can be reduced dramatically [95].

It should be noted that there remains several problems. Most of deep transfer learning methods depended heavily on the preprocessing such as time-domain feature calculation, frequency-domain analysis and/ or time-frequency domain transform. It requires that the signal source and dimensions between the source domain and target domain are consistent. However, there have been few relevant analyses for the inconsistency in signal source and information dimensions. In addition, for most of transfer cases, high-precision sensors were used. This increased the monitoring cost of the mechanical system. Similarly, there have been few relevant analyses on the transfer effect for low-precision and low-cost sensor cases. In general, there are two main challenges for the deep transfer learning in the machinery fault diagnosis.

On the one hand, it remains challenging for the deep transfer learning to improve the diagnosis accuracy. There are mainly four points to deteriorate the diagnosis result: (i) When transferring between different working conditions, transfer error occurs for the changing features; (ii) Based on the similarity criterion, unique characteristic from different components cannot be well transferred between different components; (iii) Since the laboratory and industrial field are with different scales, scale amplification problem results in transfer errors between mechanical systems; And (iv) the transfer performance is limited by samples and a prior knowledge for the machinery to be diagnosed.

On the other hand, the computational burden is a challenging point in the deep transfer processes. Transfers between source domain and target domain generate additional computational complexity which leads to increased computational burden. Moreover, deep transfer learning methods are based on the deep architecture, which suffers from additional computational burden intrinsically. It has been reported one network transferred to several applications by directly transferring from the trained network to other networks. However, its effectiveness remains to be double checked for different applications.

5.2. Further directions

Having considered the above challenges, one can see that the deep transfer learning is still at the early stage for the machinery fault diagnosis. There are some open problems to be solved by both academia and industry. One should focus on the following directions for further contributing to this community.

(i) Deal with different structures of the condition datasets. It is important to consider nonhomogeneous transfer between different characteristics or dimensions for source domain and target domain, such as between vibration data and acoustic ones, or between one-dimensional signals and two-dimensional images.
(ii) Develop appropriative deep structures for the deep transfer learning. Most of deep transfer learning methods are based on existing deep models. There are not any deep learning networks specially designed for the deep transfer learning. An organic integration of deep transfer learning can be achieved if a special deep model can be developed for the transfer learning.
(iii) Apply to dynamic fault diagnosis of the mechanical system. In the machinery fault diagnosis, there were not any reports on the development of the deep transfer learning for dynamic condition monitoring, such as fault severity assessment and residual useful life estimation. This should be an interesting research direction in the fault diagnosis circle.
(iv) Further improve the diagnosis accuracy. The accuracy of the transfer methods should be improved in working conditions, mechanical components, mechanical systems, and learning networks. It is important to extract useful features and to build more accurate deep transfer models.
(v) Further reduce the computational burden. Efficient transfer algorithms should be developed to reduce computing resources, especially to consider whether there exists a general source domain network.

6. Conclusions

To summarize, plenty of achievements of the deep transfer learning have been reported for the machinery fault diagnosis. Intelligent data-driven diagnosis methods have attracted much attention of both the academia and the industry. Many machine learning algorithms have been used in the fault diagnosis, condition monitoring and life prognosis of the machinery. Based on those achievements, deep transfer learning has become a research hotspot in the machinery fault diagnostics. Transfer learning techniques such as instance-based, feature-based, and shared parameter-based ones have been widely used for the machinery fault diagnosis. Different transfer learning architectures have been developed for various applications. This paper has introduced some requirements and the latest development of the deep transfer learning. Although the effectiveness of these methods has not been completely evaluated, it is reasonable that the deep transfer learning is still the academic frontier of the machinery fault diagnostics.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Chuan Li: Conceptualization, Investigation, Methodology. **Shaohui Zhang:** Data curation, Writing - original draft. **Yi Qin:** Writing - review & editing. **Edgar Estupinan:** Supervision, Writing - review & editing.

Acknowledgments

This work is supported in part by the National Natural Science Foundation of China (51775112, 51975121), the Key Project of the Chongqing Natural Science Foundation (cstc2019jcyj-zdxmX0013), and the CTBU Open Project (KFJJ2018075).

References

[1] S. Xue, I. Howard, Torsional vibration signal analysis as a diagnostic tool for planetary gear fault detection, Mech. Syst. Sig. Process. 100 (2018) (2018) 706–728.
[2] C. Li, M. Liang, Time-frequency signal analysis for gearbox fault diagnosis using a generalized synchrosqueezing transform, Mech. Syst. Sig. Process. 26 (1) (2012) 205–217.
[3] C. Li, V. Sanchez, G. Zurita, M.C. Lozada, D. Cabrera, Rolling element bearing defect detection using the generalized synchrosqueezing transform guided by time-frequency ridge enhancement, ISA Trans. 60 (2016) 274–284.
[4] N. Sawalhi, R.B. Randall, Gear parameter identification in a wind turbine gearbox using vibration signals, Mech. Syst. Sig. Process. 42 (2014) 368–376.
[5] X. Jiang, C. Shen, J. Shi, Z. Zhu, Initial center frequency-guided VMD for fault diagnosis of rotating machines, J. Sound Vib. 435 (2018) 36–55.
[6] C. Li, M. Liang, T. Wang, Criterion fusion for spectral segmentation and its application to optimal demodulation of bearing vibration signals, Mech. Syst. Sig. Process. 64-65 (2015) 132–148.
[7] M. Gao, S. Yang, L. Sheng, D. Zhou, Fault diagnosis for time-varying systems with multiplicative noises over sensor networks subject to Round-Robin protocol, Neurocomputing 346 (2019) 65–72.
[8] C. Xu, S. Zhao, F. Liu, Sensor fault detection and diagnosis in the presence of outliers, Neurocomputing 349 (2019) 156–163.
[9] J. Zhang, C. Yuan, P. Stegagno, W. Zeng, C. Wang, Small fault detection from discrete-time closed-loop control using fault dynamics residuals, Neurocomputing 365 (2019) 239–248.
[10] J. Li, H. Dong, F. Han, N. Hou, X. Li, Filter design, fault estimation and reliable control for networked time-varying systems: a survey, Syst. Sci. Control Eng. 5 (1) (2017) 331–341.
[11] Y. Zhang, Z. Wang, L. Ma, F.E. Alsaadi, Annulus-event-based fault detection, isolation and estimation for multirate time-varying systems: applications to a three-tank system, J. Process Control 75 (2019) 48–58.
[12] J. Hu, Z. Wang, H. Gao, Joint state and fault estimation for time-varying nonlinear systems with randomly occurring faults and sensor saturations, Automatica 97 (2018) 150–160.
[13] L. Ma, Z. Wang, Y. Liu, F.E. Alsaadi, A note on guaranteed cost control for nonlinear stochastic systems with input saturation and mixed time-delays, Int. J. Robust Nonlinear Control 27 (18) (2017) 4443–4456.
[14] L. Ma, Z. Wang, Q. Han, Y. Liu, Dissipative control for nonlinear Markovian jump systems with actuator failures and mixed time-delays, Automatica 98 (2018) 358–362.
[15] S.E. Pandarakone, Y. Mizuno, H. Nakamura, Evaluating the progression and orientation of scratches on outer-raceway bearing using a pattern recognition method, IEEE Trans. Indust. Electron. 66 (2) (2019) 1307–1314.
[16] X. Zhao, M. Jia, Fault diagnosis of rolling bearing based on feature reduction with global-local margin Fisher analysis, Neurocomputing 315 (2018) 447–464.
[17] C. Li, M. Liang, Continuous-scale mathematical morphology-based optimal scale band demodulation of impulsive feature for bearing defect diagnosis, J. Sound Vib. 331 (2012) 5864–5879.
[18] L. Song, H. Wang, P. Chen, Step-by-step fuzzy diagnosis method for equipment based on symptom extraction and trivalent logic fuzzy diagnosis theory, IEEE Trans. Fuzzy Syst. 26 (6) (2018) 553478.
[19] B. Zeng, C. Li, Improved multi-variable grey forecasting model with a dynamic background-value coefficient and its application, Comput. Ind. Eng. 118 (2018) 278–290.
[20] B. Zeng, H. Duan, Y. Zhou, A new multivariable grey prediction model with structure compatibility, Appl. Math. Modell. 75 (2019) 385–397.
[21] C. Li, L. Ledo, M. Delgado, M. Cerrada, F. Pacheco, D. Cabrera, R.V. Sánchez, J. Valente de Oliveira, A Bayesian approach to consequent parameter estimation in probabilistic fuzzy systems and its application to bearing fault classification, Knowl.-Based Syst. 129 (2017) 39–60.
[22] A. Hu, X. Yan, L. Xiang, A new wind turbine fault diagnosis method based on ensemble intrinsic time-scale decomposition and WPT-fractal dimension, Renew. Energy 83 (2015) 767–778.
[23] C. Li, L.J. Valente de Oliveira, M.C. Lozada, D. Cabrera, V. Sanchez, G. Zurita, A systematic review of fuzzy formalisms for bearing fault diagnosis, IEEE Trans. Fuzzy Syst. 27 (7) (2019) 1362–1382.
[24] K. Javed, R. Gouriveau, N. Zerhouni, SW-ELM: a summation wavelet extreme learning machine algorithm with a priori parameter initialization, Neurocomputing 123 (2014) 299–307.
[25] B. Zhou, X. Gu, Multi-block statistics local kernel principal component analysis algorithm and its application in nonlinear process fault detection, Neurocomputing 376 (2020) 222–231.
[26] C. Li, M. Cerrada, D. Cabrera, R.V. Sanchez, F. Pacheco, G. Ulutagay, J. Valente de Oliveira, A comparison of fuzzy clustering algorithms for bearing fault diagnosis, J. Intell. Fuzzy Syst. 34 (6) (2018) 3565–3580.
[27] 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.
[28] C. Li, Y. Tao, W. Ao, S. Yang, Y. Bai, Improving forecasting accuracy of daily enterprise electricity consumption using a random forest based on ensemble empirical mode decomposition, Energy 165 (2018) 1220–1227.
[29] W. Li, S. Zhang, G. He, Semisupervised distance-preserving self-organizing map for machine-defect detection and classification, IEEE Trans. Instrum. Meas. 62 (5) (2013) 869–879.
[30] S. Zhang, W. Li, Bearing condition recognition and degradation assessment under varying running conditions using NPE and SOM, Math. Prob. Eng. 1 (2014) 1–10.
[31] W. Li, J. Li, S. Zhang, Application of continuous hidden semi-Markov model in bearing performance degradation assessment, J. Vib. Eng. 4 (2014) 613–620.
[32] Y. Bai, Z. Sun, B. Zeng, J. Long, L. Li, J.V.D. Oliveira, C. Li, A comparison of dimension reduction techniques for support vector machine modeling of multi-parameter manufacturing quality prediction, J. Intell. Manuf. 30 (2019) 2245–2256.
[33] W. Li, B. Dai, S. Zhang, Bearing performance degradation assessment based on wavelet packet entropy and Gaussian mixture mode, J. Vib. Shock 21 (2013) 35–40 +91, 2013.
[34] G.E. Hinton, R. Salakhutdinov, Reducing the dimensionality of data with neural networks, Science 313 (2006) 504–507.
[35] F. Aghazadeh, A. Tahan, M. Thomas, Tool condition monitoring using spectral subtraction and convolutional neural networks in milling process, Int. J. Adv. Manuf. Technol. 98 (9-12) (2018) 3217–3227.
[36] C. Li, R.V. Sánchez, G. Zurita, M. Cerrada, D. Cabrera, Fault diagnosis for rotating machinery using vibration measurement deep statistical feature learning, Sensors 16 (6) (2016) 895.
[37] F. Jia, Y. Lei, L. Guo, J. Lin, S. Xing, A neural network constructed by deep learning technique and its application to intelligent fault diagnosis of machines, Neurocomputing 272 (2018) 619–628.

[38] M. He, D. He, Deep learning based approach for bearing fault diagnosis, IEEE Trans. Ind. Appl. 53 (3) (2017) 3057–3065.
[39] C. Li, R.V. Sanchez, G. Zurita, M. Cerrada, D. Cabrera, R.E. Vásquez, Multimodal deep support vector classification with homologous features and its application to gearbox fault diagnosis, Neurocomputing 168 (2015) 119–127.
[40] Y. Qi, C. Shen, D. Wang, J. Shi, X. Jiang, Z. Zhu, Stacked sparse autoencoder-based deep network for fault diagnosis of rotating machinery, IEEE Access 5 (2017) 15066–15079.
[41] C. Shen, Y. Qi, J. Wang, G. Cai, Z. Zhu, An automatic and robust features learning method for rotating machinery fault diagnosis based on contractive autoencoder, Eng. Appl. Artif. Intell. 76 (2018) 170–184.
[42] X. Ding, Q. He, Energy-fluctuated multiscale feature learning with deep ConvNet for intelligent spindle bearing fault diagnosis, IEEE Trans Instrum. Meas. 66 (8) (2017) 1926–1935.
[43] G. Jiang, H. He, J. Yan, P. Xie, Multiscale convolutional neural networks for fault diagnosis of wind turbine gearbox, IEEE Trans. Ind. Electron. 66 (4) (2019) 3196–3207.
[44] J. Yang, Y. Guo, W. Zhao, Long short-term memory neural network based fault detection and isolation for electro-mechanical actuators, Neurocomputing 360 (2019) 85–96.
[45] S. Sina, Z.N. Sadough, K. Khorasani, Dynamic neural network-based fault diagnosis of gas turbine engines, Neurocomputing 125 (11) (2014) 153–165.
[46] Y. Wu, M. Yuan, S. Dong, L. Lin, Y. Liu, Remaining useful life estimation of engineered systems using vanilla LSTM neural networks, Neurocomputing 275 (2018) 167–179.
[47] Q. Hu, R. Zhang, Y. Zhou, Transfer learning for short-term wind speed prediction with deep neural networks, Renew. Energy 85 (2016) 83–95.
[48] M. Gan, C. Wang, C. Zhu, Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of rolling element bearings, Mech. Syst. Sig. Process. 72-73 (2016) 92–104.
[49] W. Mao, W. Feng, X. Liang, A novel deep output kernel learning method for bearing fault structural diagnosis, Mech. Syst. Sig. Process. 117 (2019) 293–318.
[50] L. Wang, Z. Zhang, H. Long, J. Xu, R. Liu, Wind turbine gearbox failure identification with deep neural networks, IEEE Trans. Ind. Inf. 13 (3) (2017) 1360–1368.
[51] S. Wang, J. Xiang, Y. Zhong, H. Tang, A data indicator-based deep belief networks to detect multiple faults in axial piston pumps, Mech. Syst. Sig. Process. 112 (2018) 154–170.
[52] H. Chen, B. Jiang, T. Zhang, N. Lu, Data-driven and deep learning-based detection and diagnosis of incipient faults with application to electrical traction systems, Neurocomputing 396 (2020) 429–437.
[53] K. Li, M. Xiong, F. Li, L. Su, J. Wu, A novel fault diagnosis algorithm for rotating machinery based on a sparsity and neighborhood preserving deep extreme learning machine, Neurocomputing 350 (2019) 261–270.
[54] X. Li, W. Zhang, Q. Ding, Deep learning-based remaining useful life estimation of bearings using multi-scale feature extraction, Reliab. Eng. Syst. Saf. 182 (2019) 208–218.
[55] F. Jia, Y. Lei, J. Lin, X. Zhou, N. Lu, Deep neural networks: a promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data, Mech. Syst. Sig. Process. 72-73 (2016) 303–315.
[56] H. Shao, H. Jiang, X. Li, T. Liang, Rolling bearing fault detection using continuous deep belief network with locally linear embedding, Comput. Ind. 96 (2018) 27–39.
[57] L. Ren, J. Cui, Y. Sun, X. Cheng, Multi-bearing remaining useful life collaborative prediction: a deep learning approach, J. Manuf. Syst. 43 (2017) 248–256.
[58] K. Peng, R. Jiao, J. Dong, Y. Pi, A deep belief network based health indicator construction and remaining useful life prediction using improved particle filter, Neurocomputing 361 (2019) 19–28.
[59] G. Hu, H. Li, Y. Xia, L. Luo, A deep Boltzmann machine and multi-grained scanning forest ensemble collaborative method and its application to industrial fault diagnosis, Comput. Ind. 100 (2018) 287–296.
[60] R. Liu, G. Meng, B. Yang, C. Sun, X. Chen, Dislocated time series convolutional neural architecture: An intelligent fault diagnosis approach for electric machine, IEEE Trans. Ind. Inf. 13 (3) (2017) 1310–1320.
[61] X. Zhao, M. Jia, A new local-global deep neural network and its application in rotating machinery fault diagnosis, Neurocomputing 366 (2019) 215–233.
[62] X. Gao, F. Deng, X. Yue, Data augmentation in fault diagnosis based on the Wasserstein generative adversarial network with gradient penalty, Neurocomputing 396 (2020) 487–494.
[63] S. Zhang, Z. Sun, J. Long, C. Li, Y. Bai, Dynamic condition monitoring for 3D printers by using error fusion of multiple sparse auto-encoders, Comput. Ind. 105 (2019) 164–176.
[64] S. Zhang, Z. Sun, M. Wang, J. Long, Y. Bai, C. Li, Deep fuzzy echo state networks for machinery fault diagnosis, IEEE Trans. Fuzzy Syst. (2020) Early Access, doi:10.1109/TFUZZ.2019.2914617.
[65] D. Cabrera, A. Guamán, S. Zhang, M. Cerrada, R.V. Sánchez, J. Cevallos, J. Long, C. Li, Bayesian approach and time series dimensionality reduction to LSTM-based model-building for fault diagnosis of a reciprocating compressor, Neurocomputing 380 (2020) 51–66.
[66] S. Zhang, Z. Sun, C. Li, D. Cabrera, J. Long, Y. Bai, Deep hybrid state network with feature reinforcement for intelligent fault diagnosis of delta 3D printers, IEEE Trans. Ind. Inf. (2020) in press, doi:10.1109/TII.2019.2920661.
[67] J. Long, Z. Sun, C. Li, Y. Hong, Y. Bai, S. Zhang, A novel sparse echo autoencoder network for data-driven fault diagnosis of delta 3-D Printers, IEEE Trans. Instrum. Meas. 69 (3) (2020) 683–692.
[68] S. Zhang, M. Wang, F. Yang, W. Li, Manifold sparse auto-encoder for machine fault diagnosis, IEEE Sens. J. (2020) Early Access, doi:10.1109/JSEN.2019.2925845.
[69] J. Long, S. Zhang, C. Li, Evolving deep echo state networks for intelligent fault diagnosis, IEEE Trans. Ind. Inf. 16 (7) (2020) 4928–4937.
[70] T. Han, C. Liu, W.D. Jiang Yang, Deep Transfer Network with Joint Distribution Adaptation: A New Intelligent Fault Diagnosis Framework for Industry Application, ISA Transactions 97 (2020) 269–281.
[71] S. Shao, S. McAleer, R. Yan, P. Baldi, Highly-accurate machine fault diagnosis using deep transfer learning, IEEE Trans. Ind. Inf. 15 (2018) 1–9.
[72] L. Guo, Y. Lei, S. Xing, T. Yan, N. Li, Deep convolutional transfer learning network: A new method for intelligent fault diagnosis of machines with unlabeled data, IEEE Trans. Indust. Electron. 66 (9) (2018) 7316–7325.
[73] C. Sun, M. Ma, Z. Zhao, S. Tian, R. Yan, X. Chen, Deep transfer learning based on sparse auto-encoder for remaining useful life prediction of tool in manufacturing, IEEE Trans. Ind. Inf. 15 (4) (2018) 2416–2425.
[74] M.J. Hasan, J.M. Kim, Bearing fault diagnosis under variable rotational speeds using stockwell transform-based vibration imaging and transfer learning, Appl. Sci. 8 (12) (2018) 2357.
[75] Y. Pan, F. Mei, H. Miao, J. Zheng, K. Zhu, H. Sha, An approach for HVCB mechanical fault diagnosis based on a deep belief network and a transfer learning strategy, J. Electr. Eng. Technol. 14 (2019) 1–13.
[76] S. Shao, P. Wang, R. Yan, Generative adversarial networks for data augmentation in machine fault diagnosis, Comput. Ind. 106 (2019) 85–93.
[77] H. Liu, J. Zhou, Y. Xu, Y. Zheng, X. Peng, W. Jiang, Unsupervised fault diagnosis of rolling bearings using a deep neural network based on generative adversarial networks, Neurocomputing 315 (2018) 412–424.
[78] Z. Wang, J. Wang, Y. Wang, An intelligent diagnosis scheme based on generative adversarial learning deep neural networks and its application to planetary gearbox fault pattern recognition, Neurocomputing 310 (2018) 213–222.
[79] L. Xiang, W. Zhang, Q. Ding, Cross-domain fault diagnosis of rolling element bearings using deep generative neural networks, IEEE Trans. Ind. Electron. 66 (7) (2019) 5525–5534.
[80] J. Long, Z. Sun, P.M. Pardalos, Y. Hong, S. Zhang, C. Li, A hybrid multi-objective genetic local search algorithm for the prize-collecting vehicle routing problem, Inf. Sci. 478 (2019) 40–61.
[81] J. Long, Z. Zheng, X. Gao, P.M. Pardalos, Scheduling a realistic hybrid flow shop with stage skipping and adjustable processing time in steel plants, Appl. Soft Comput. 64 (2018) 536–549.
[82] S.J. Pan, I.W. Tsang, J.T. Kwok, Q. Yang, Domain adaptation via transfer component analysis, IEEE Trans. Neural Netw. 22 (2) (2010) 199–210.
[83] J. Guo, J. Wu, Z. Sun, J. Long, S. Zhang, Fault diagnosis of delta 3D printers using transfer support vector machine with attitude signals, IEEE Access 7 (2019) 40359–40368.
[84] M. Long, J. Wang, G. Ding, J. Sun, P. Yu, Transfer feature learning with joint distribution adaptation, in: Proceedings of the IEEE International Conference on Computer Vision, 2013, pp. 2200–2207.
[85] M. Long, Y. Cao, J.M. Jordan Wang, Learning Transferable Features with Deep Adaptation Networks. International Conference on Machine Learning 37 (2015) 97–105.
[86] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, V. Lempitsky, Domain-adversarial training of neural networks, J. Mach. Learn. Res. 17 (1) (2016) 1-35.
[87] R. Yan, F. Shen, C. Sun, X. Chen, Knowledge transfer for rotary machine fault diagnosis, IEEE Sens. J. (2020) Early Access, doi:10.1109/JSEN.2019.2949057.
[88] Y. Xie, T. Zhang, A transfer learning strategy for rotation machinery fault diagnosis based on cycle-consistent generative adversarial networks, in: Proceedings of Chinese Automation Congress (CAC), Xi'an, China, 2018, pp. 1309–1313.
[89] C. Zhang, L. Xu, X. Li, H. Wang, A method of fault diagnosis for rotary equipment based on deep learning, in: Proceedings of Prognostics and System Health Management Conference (PHM-Chongqing), Chongqing, China, 2018, pp. 958–962.
[90] W. Qian, S. Li, J. Wang, A new transfer learning method and its application on rotating machine fault diagnosis under variant working conditions, IEEE Access 6 (2018) 69907–69917.
[91] W. Qian, S. Li, J. Wang, Y. Xin, H. Ma, A new deep transfer learning network for fault diagnosis of rotating machine under variable working conditions, in: Proceedings of Prognostics and System Health Management Conference (PHM-Chongqing), Chongqing, China, 2018, pp. 1010–1016.
[92] S. Pang, X. Yang, A cross-domain stacked denoising autoencoders for rotating machinery fault diagnosis under different working conditions, IEEE Access 7 (2019) 77277–77292.
[93] S. Ma, W. Liu, W. Cai, Z. Shang, G. Liu, Lightweight deep residual CNN for fault diagnosis of rotating machinery based on depthwise separable convolutions, IEEE Access 7 (2019) 57023–57036.
[94] X. Li, W. Zhang, Q. Ding, X. Li, Diagnosing rotating machines with weakly supervised data using deep transfer learning, IEEE Trans. Ind. Inf. 16 (3) (2020) 1688–1697.
[95] Z. Chen, K. Gryllias, W. Li, Intelligent fault diagnosis for rotary machinery using transferable convolutional ceural network, IEEE Trans. Ind. Inform. 16 (1) (2020) 339–349.
[96] H. Qiao, T. Wang, P. Wang, L. Zhang, M. Xu, An adaptive weighted multiscale convolutional neural network for rotating machinery fault diagnosis under variable operating conditions, IEEE Access 7 (2019) 118954–118964.

[97] M. Sohaib, J. Kim, Fault diagnosis of rotary machine bearings under inconsistent working conditions, IEEE Trans. Instrum. Meas. 69 (6) (2020) 3334–3347.
[98] D. Xiao, Y. Huang, L. Zhao, C. Qin, H. Shi, C. Liu, Domain adaptive motor fault diagnosis using deep transfer learning, IEEE Access 7 (2019) 80937–80949.
[99] L. Wen, L. Gao, X. Li, A new deep transfer learning based on sparse auto-encoder for fault diagnosis, IEEE Trans. Syst. Man Cybern.: Syst. 49 (1) (2019) 136–144.
[100] B. Yang, Y. Lei, F. Jia, S. Xing, A transfer learning method for intelligent fault diagnosis from laboratory machines to real-case machines, in: Proceedings of International Conference on Sensing, Diagnostics, Prognostics, and Control (SDPC), Xi'an, China, 2018, pp. 35–40.
[101] L. Wen, X. Li, L. Gao, Y. Zhang, A new convolutional neural network-based data-driven fault diagnosis method, IEEE Trans. Ind. Electron. 65 (7) (2018) 5990–5998.
[102] M.A. Khan, Y. Kim, J. Choo, Intelligent fault detection via dilated convolutional neural networks, in: Proceedings of *IEEE International Conference on Big Data and Smart Computing (BigComp)*, Shanghai, China, 2018, pp. 729–731.
[103] W. Song, L. Wen, L. Gao, X. Li, Unsupervised fault diagnosis method based on iterative multi-manifold spectral clustering, IET Collab. Intell. Manuf. 1 (2) (2019) 48–55.
[104] K. Gowthami, L. Kalaivani, Fault classification of induction motor bearing using adaptive neuro fuzzy inference system, in: Proceedings of *Fifth International Conference on Electrical Energy Systems (ICEES)*, Chennai, India, 2019, pp. 1–6.
[105] J. Wang, P. Fu, L. Zhang, R.X. Gao, R. Zhao, Multi-level information fusion for induction motor fault diagnosis, IEEE/ASME Trans. Mechatron. 24 (5) (2020) 2139–2150.
[106] F.B. Abid, M. Sallem, A. Braham, Robust interpretable deep learning for intelligent fault diagnosis of induction motors, IEEE Trans. Instrum. Meas. 69 (6) (2020) 3506–3515 in press, doi:10.1109/TIM.2019.2932162.
[107] M. Sohaib, J. Kim, Fault diagnosis of rotary machine bearings under inconsistent working conditions, IEEE Trans. Instrum. Meas. 69 (6) (2020) 3334–3347, doi:10.1109/TIM.2019.2933342.
[108] B. Wang, C. Shen, C. Yu, Y. Yang, Data fused motor fault identification based on adversarial auto-encoder, in: Proceedings of IEEE 10th International Symposium on Power Electronics for Distributed Generation Systems (PEDG), Xi'an, China, 2019, pp. 299–305.
[109] L.A. García-Escudero, O. Duque-Perez, M. Fernandez-Temprano, D. Morinigo–Sotelo, Robust detection of incipient faults in VSI-fed induction motors using quality control charts, IEEE Trans. Ind. Appl. 53 (3) (2017) 3076–3085.
[110] 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 IEEE International Conference on Prognostics and Health Management (ICPHM), Ottawa, ON, 2016, pp. 1–6.
[111] Z. Liu, B. Lu, H. Wei, X. Li, L. Chen, Fault diagnosis for electromechanical drivetrains using a joint distribution optimal deep domain adaptation approach, IEEE Sens. J. 19 (24) (2020) 12261–12270.
[112] G. Jiang, H. He, P. Xie, Y. Tang, Stacked multilevel-denoising autoencoders: a new representation learning approach for wind turbine gearbox fault diagnosis, IEEE Trans. Instrum. Meas. 66 (9) (2017) 2391–2402.
[113] J. Wang, J. Xie, L. Zhang, L. Duan, A factor analysis based transfer learning method for gearbox diagnosis under various operating conditions, in: Proceedings of International Symposium on Flexible Automation (ISFA), Cleveland, OH, 2016, pp. 81–86.
[114] Z. He, H. Shao, X. Zhang, J. Cheng, Y. Yang, Improved deep transfer auto-encoder for fault diagnosis of gearbox under variable working conditions with small training samples, IEEE Access 7 (2019) 115368–115377.
[115] D. Chen, S. Yang, F. Zhou, Incipient fault diagnosis based on DNN with transfer learning, in: Proceedings of International Conference on Control, Automation and Information Sciences (ICCAIS), Hangzhou, China, 2018, pp. 303–308.
[116] R. Zhang, H. Tao, L. Wu, Y. Guan, Transfer learning with neural networks for bearing fault diagnosis in changing working conditions, IEEE Access 5 (2017) 14347–14357.
[117] Z. Tong, W. Li, B. Zhang, F. Jiang, G. Zhou, Bearing fault diagnosis under variable working conditions based on domain adaptation using feature transfer learning, IEEE Access 6 (2018) 76187–76197.
[118] X. Li, W. Zhang, Q. Ding, Cross-domain fault diagnosis of rolling element bearings using deep generative neural networks, IEEE Trans. Ind. Electron. 66 (7) (2019) 5525–5534.
[119] M. Zhang, D. Wang, W. Lu, J. Yang, Z. Li, B. Liang, A deep transfer model with Wasserstein distance guided multi-adversarial networks for bearing fault diagnosis under different working conditions, IEEE Access 7 (2019) 65303–65318.
[120] F. Shen, C. Chen, R. Yan, R.X. Gao, Bearing fault diagnosis based on SVD feature extraction and transfer learning classification, in: Proceedings of Prognostics and System Health Management Conference (PHM), Beijing, China, 2015, pp. 1–6.
[121] H. Kim, B.D. Youn, A new parameter repurposing method for parameter transfer with small dataset and its application in fault diagnosis of rolling element bearings, IEEE Access 7 (2019) 46917–46930.
[122] A. Zhang, S. Li, Y. Cui, W. Yang, R. Dong, J. Hu, Limited data rolling bearing fault diagnosis with few-shot learning, IEEE Access 7 (2019) 110895–110904.
[123] D. Peng, Z. Liu, H. Wang, Y. Qin, L. Jia, A novel deeper one-dimensional CNN with residual learning for fault diagnosis of wheelset bearings in high-speed trains, IEEE Access 7 (2019) 10278–10293.
[124] X. Li, H. Jiang, K. Zhao, R. Wang, A deep transfer nonnegativity-constraint sparse autoencoder for rolling bearing fault diagnosis with few labeled data, IEEE Access 7 (2019) 91216–91224.
[125] C. Chen, Z. Li, J. Yang, B. Liang, A cross domain feature extraction method based on transfer component analysis for rolling bearing fault diagnosis, in: Proceedings of 29th Chinese Control And Decision Conference (CCDC), Chongqing, China, 2017, pp. 5622–5626.
[126] X. Zhao, Y. Qin, L. Kou, Z. Liu, Understanding real faults of axle box bearings based on vibration data using decision tree, in: Proceedings of *2018 IEEE International Conference on Prognostics and Health Management (ICPHM)*, Seattle, WA, 2018, pp. 1–5.
[127] J. Zhu, N. Chen, C. Shen, A new deep transfer learning method for bearing fault diagnosis under different working conditions, IEEE Sens. J. (2020) Early Access, doi:10.1109/JSEN.2019.2936932.
[128] J. Wang, Z. Mo, H. Zhang, Q. Miao, A deep learning method for bearing fault diagnosis based on time-frequency image, IEEE Access 7 (2019) 42373–42383.
[129] Z. Liu, B. Lu, H. Wei, L. Chen, X. Li, M. Rätsch, Deep adversarial domain adaptation model for bearing fault diagnosis, IEEE Trans. Syst. Man Cybern.: Syst. (2020) Early Access, doi:10.1109/TSMC.2019.2932000.
[130] V. Pandhare, J. Singh, J. Lee, Convolutional neural network based rolling-element bearing fault diagnosis for naturally occurring and progressing defects using time-frequency domain features, in: Proceedings of *Prognostics and System Health Management Conference (PHM-Paris)*, Paris, France, 2019, pp. 320–326.
[131] S. Dong, G. Wen, Z. Zhang, Bearing fault diagnosis under different operating conditions based on cross domain feature projection and domain adaptation, in: Proceedings of *IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, Auckland, New Zealand, 2019, pp. 1–6.
[132] S. Dong, J. Wang, B. Tang, X. Xu, J. Liu, Bearing faults recognition based on the local mode decomposition morphological spectrum and optimized support vector machine, in: Proceedings of *13th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI)*, Xi'an, 2016, pp. 1012–1016.
[133] H. Ahmed, A.K. Nandi, Compressive sampling and feature ranking framework for bearing fault classification with vibration signals, IEEE Access 6 (2018) 44731–44746.
[134] V. Nguyen, T. Dung Hoang, V. Thai, X. Nguyen, Big vibration data diagnosis of bearing fault based on feature representation of autoencoder and optimal LSSVM-CRO classifier model, in: Proceedings of *International Conference on System Science and Engineering (ICSSE)*, Dong Hoi, Vietnam, 2019, pp. 557–563.
[135] R. Yuan, Y. Lv, H. Li, G. Song, Robust fault diagnosis of rolling bearings using multivariate intrinsic multiscale entropy analysis and neural network under varying operating conditions, IEEE Access 7 (2019) 130804–130819.
[136] F. Shen, J. Xu, C. Sun, X. Chen, R. Yan, Transfer between multiple working conditions: a new TCCHC-based exponential semi-deterministic extended Kalman filter for bearing remaining useful life prediction, Measurement 142 (2019) 148–162.
[137] W. He, Y. He, B. Li, C. Zhang, A naive-Bayes-based fault diagnosis approach for analog circuit by using image-oriented feature extraction and selection technique, IEEE Access 8 (2020) 5065–5079.
[138] G. Sheng, Q. Hong, M. Wei, W. Lan, Z. Jie, Research on ELM soft fault diagnosis of analog circuit based on KSLPP feature extraction, IEEE Access 7 (2019) 92517–92527.
[139] Z. Liu, J. Zhong, L. Yang, K. Liu, Y. Han, L. Wang, W. Liu, Location and fault detection of catenary support components based on deep learning, in: Proceedings of IEEE International Instrumentation and Measurement Technology Conference (I2MTC), Houston, TX, 2018, pp. 1–6.
[140] N. Dimitriou, L. Leontaris, T. Vafeiadis, D. Ioannidis, T. Wotherspoon, G. Tinker, D. Tzovaras, Fault diagnosis in microelectronics attachment via deep learning analysis of 3D laser scans, IEEE Trans. Indust. Electron. 67 (7) (2020) 5748–5757.
[141] H. Heng, J.S.T. Shanmugam, M.B. Nair, E.M.A. Gnanamuthu, Acoustic emission source localization on a pipeline using convolutional neural network, in: IEEE Conference on Big Data and Analytics (ICBDA), Langkawi Island, Malaysia, 2018, pp. 93–98.
[142] C. Sun, M. Ma, Z. Zhao, S. Tian, R. Yan, X. Chen, Deep transfer learning based on sparse auto-encoder for remaining useful life prediction of tool in manufacturing, IEEE Trans. Ind. Inf. 4 (15) (2019) 2416–2425.
[143] M. Ezeme, Q. Mahmoud, A. Azim, Hierarchical attention-based anomaly detection model for embedded operating systems, in: Proceedings of IEEE 24th International Conference on Embedded and Real-Time Computing Systems and Applications (RTCSA), Hakodate, 2018, pp. 225–231.
[144] Z. An, S. Li, J. Wang, Y. Xin, K. Xu, Generalization of deep neural network for bearing fault diagnosis under different working conditions using multiple kernel method, Neurocomputing 352 (4) (2019) 42–53.
[145] M. Long, Transfer Learning: Problems and Methods, Tsinghua University, 2014.
[146] Y. Bengio, A. Courville, P. Vincent, Representation learning: a review and new perspectives, IEEE Trans. Pattern Anal. Mach. Intell. 35 (8) (2013) 1798–1828.
[147] K. Bousmalis, G. Trigeorgis, N. Silberman, D. Krishnan, D. Erhan, Domain separation networks, Adv. Neural Inf. Process. Syst. (2016) 343–351.
[148] E. Kodirov, T. Xiang, Z. Fu, S. Gong, Unsupervised domain adaptation for zero-shot learning, in: Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 2015, pp. 2452–2460.

[149] S. Motiian, J. Quinn, I. Seyed, D. Gianfranco, Few-shot adversarial domain adaptation, in: Proceedings of Advances in Neural Information Processing Systems, 2017, pp. 6673–6683.
[150] Z. Luo, Y. Zou, J. Hoffman, F. Li, Label efficient learning of transferable representations across domains and tasks, Adv. Neural Inf. Process. Syst. 12 (2017) 164–176.
[151] M. Long, H. Zhu, J. Wang, M. Jordan, Unsupervised domain adaptation with residual transfer networks, in: Proceedings of Advances in Neural Information Processing Systems, 2016, pp. 136–144.
[152] L. Zhang, C. Aggarwal, G. Qi, Stock price prediction via discovering multi-frequency trading patterns, in: Proceedings of ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD 2017), Halifax, Nova Scotia, Canada, 2017.
[153] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, in: Proceedings of NIPS, 2012.
[154] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, Y. Bengio, Generative adversarial nets, Adv. Neural Inf. Process. Syst. 27 (2014) 2672–2680.
[155] K. Bousmalis, G. Trigeorgis, S.N. ilberman, D. Krishnan, D. Erhan, Domain separation networks, Adv. Neural Inf. Process. Syst. 29 (2016) 343–351.
[156] E. Tzeng, J. Hoffman, K. Saenko, T. Darrell, Adversarial discriminative domain adaptation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 7167–7176.
[157] M. Arjovsky, S. Chintala, L. Bottou, Wasserstein generative adversarial networks, in: Proceedings of International Conference on Machine Learning, 2017, pp. 214–223.
[158] J. Shen, Y. Qu, W.Y. Yu Zhang, Wasserstein Distance Guided Representation Learning for Domain Adaptation. Thirty-Second AAAI Conference on Artificial Intelligence 32 (2018) 1–8.
[159] M. Long, Z. Cao, J. Wang, M. Jordan, Conditional adversarial domain adaptation, Adv. Neural Inf. Process. Syst. 31 (2018) 1640–1650.
[160] S. Lu, T. Sirojan, B. Phung, D. Zhang, E. Ambikairajah, DA-DCGAN: an effective methodology for DC series arc fault diagnosis in photovoltaic systems, IEEE Access 7 (2019) 45831–45840.
[161] Q. Wang, G. Michau, O. Fink, Domain adaptive transfer learning for fault diagnosis, in: Proceedings of 2019 Prognostics and System Health Management Conference (PHM-Paris), Paris, France, 2019, pp. 279–285.
[162] C. Sampedro, J. Rodriguez-Vazquez, A. Rodriguez-Ramos, A. Carrio, P. Campoy, Deep learning-based system for automatic recognition and diagnosis of electrical insulator strings, IEEE Access 7 (2019) 101283–101308.
[163] S. Shao, S. McAleer, R. Yan, P. Baldi, Highly accurate machine fault diagnosis using deep transfer learning, IEEE Trans. Ind. Inf. 15 (4) (2019) 2446–2455.
[164] L. Wen, X. Li, X. Li, L. Gao, A new transfer learning based on VGG-19 network for fault diagnosis, in: Proceedings of IEEE 23rd International Conference on Computer Supported Cooperative Work in Design (CSCWD), Porto, Portugal, 2019, pp. 205–209.
[165] G. Xu, M. Liu, Z. Jiang, W. Shen, C. Huang, Online fault diagnosis method based on transfer convolutional neural networks, IEEE Trans. Instrum. Meas. 69 (2) (2020) 509–520.
[166] J. Li, X. Li, D. He, A directed acyclic graph network combined with CNN and LSTM for remaining useful life prediction, IEEE Access 7 (2019) 75464–75475.
[167] C. Wang, Z. Lv, J. Zhao, W. Wang, Heterogeneous transfer learning based on stack sparse auto-encoders for fault diagnosis, in: Proceedings of Chinese Automation Congress (CAC), Xi'an, China, 2018, pp. 4277–4281.
[168] C. Sun, M. Ma, Z. Zhao, S. Tian, R. Yan, X. Chen, Deep transfer learning based on sparse autoencoder for remaining useful life prediction of tool in manufacturing, IEEE Trans. Ind. Inf. 15 (4) (2019) 2416–2425.
[169] F. Shen, C. Chen, R. Yan, R.X. Gao, Bearing fault diagnosis based on SVD feature extraction and transfer learning classification, in: Proceedings of Prognostics and System Health Management Conference (PHM 2015), 2015, pp. 1–6.
[170] Z. Yang, W. Qiu, H. Sun, A. Nallanathan, Robust radar emitter recognition based on the three-dimensional distribution feature and transfer learning, Sensors 16 (3) (2016) 289.
[171] J. Wang, J. Xie, L. Zhang, L. Duan, A factor analysis based transfer learning method for gearbox diagnosis under various operating conditions, in: Proceedings of International Symposium on Flexible Automation (ISFA), 2016, pp. 81–86.
[172] L. Wen, L. Gao, X. Li, A new deep transfer learning based on sparse auto-encoder for fault diagnosis, IEEE Trans. Syst. Man Cybern.: Syst. 49 (1) (2017) 136–144.
[173] Y. Yu, H. Cao, X. Yan, T. Wang, S. Ge, Defect identification of wind turbine blades based on defect semantic features with transfer feature extractor, Neurocomputing 376 (2020) 1–9.
[174] P. Cao, S. Zhang, J. Tang, Preprocessing-free gear fault diagnosis using small datasets with deep convolutional neural network-based transfer learning, IEEE Access 6 (2018) 26241–26253.
[175] Z. Tong, W. Li, B. Zhang, F. Jiang, G. Zhou, Bearing fault diagnosis under variable working conditions based on domain adaptation using feature transfer learning, IEEE Access 6 (2018) 76187–76197.
[176] K. Xu, S. Li, J. Wang, Z. An, W. Qian, H. Ma, A novel convolutional transfer feature discrimination network for imbalanced fault diagnosis under variable rotational speed, Meas. Sci. Technol. 30 (2019) 105–107.
[177] A. Zhang, X. Gao, Supervised dictionary-based transfer subspace learning and applications for fault diagnosis of sucker rod pumping systems, Neurocomputing 338 (2019) 293–306.
[178] M. Sun, H. Wang, P. Liu, S. Huang, P. Fan, A sparse stacked denoising autoencoder with optimized transfer learning applied to the fault diagnosis of rolling bearings, Measurement 146 (2019) 305–314.
[179] D. Chen, S. Yang, F. Zhou, Incipient fault diagnosis based on DNN with transfer learning, in: Proceedings of International Conference on Control, Automation and Information Sciences (ICCAIS), 2018, pp. 303–308.
[180] J. Duan, Y. He, B. Du, R. Ghandour, W. Wu, H. Zhang, Intelligent localization of transformer internal degradations combining deep convolutional neural networks and image segmentation, IEEE Access 7 (2019) 62705–62720.
[181] P. Ma, H. Zhang, W. Fan, C. Wang, G. Wen, X. Zhang, A novel bearing fault diagnosis method based on 2D image representation and transfer learning–convolutional neural network, Meas. Sci. Technol. 30 (5) (2019) 055402.
[182] 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 IEEE International Conference on Prognostics and Health Management, 2016.
[183] R. Zhang, H. Tao, L. Wu, Y. Guan, Transfer learning with neural networks for bearing fault diagnosis in changing working conditions, IEEE Access 5 (2017) 14347–14357.
[184] W. Lu, B. Liang, Y. Cheng, D. Meng, J. Yang, T. Zhang, Deep model based domain adaptation for fault diagnosis, IEEE Trans. Indust. Electron. 64 (3) (2017) 2296–2305.
[185] L. Wen, L. Gao, X. Li, A new deep transfer learning based on sparse auto-encoder for fault diagnosis, IEEE Trans. Syst. Man Cybern.: Syst. 99 (2017) 1–9.
[186] W. Qian, S. Li, P. Yi, K. Zhang, A novel transfer learning method for robust fault diagnosis of rotating machines under variable working conditions, Measurement 138 (2019) 514–525.
[187] W. Wang, J. Zhang, Q. Zhang, Transfer learning based diagnosis for configuration troubleshooting in self-organizing femtocell networks, in: Proceedings of IEEE Global Telecommunications Conference, GLOBECOM, 2011, 2011, pp. 1–5.
[188] Q. Wang, G.O. Fink Michau, Domain Adaptive Transfer Learning for Fault Diagnosis. Prognostics and System Health Management Conference, 1 (2019) 279–285.
[189] D. Xiao, Y. Huang, L. Zhao, C. Qin, H. Shi, C. Liu, Domain adaptive motor fault diagnosis using deep transfer learning, IEEE Access 7 (2019) 80937–80949.
[190] Y. Xu, Y. Sun, X. Liu, Y. Zheng, A digital-twin-assisted fault diagnosis using deep transfer learning, IEEE Access 7 (2019) 19990–19999.
[191] R. Zhang, H. Tao, L. Wu, Y. Guan, Transfer learning with neural networks for bearing fault diagnosis in changing working conditions, IEEE Access 5 (2017) 14347–14357.
[192] D. Xiao, Y. Huang, C. Qin, Z. Liu, Y. Li, C. Liu, Transfer learning with convolutional neural networks for small sample size problem in machinery fault diagnosis, Proc. Inst. Mech. Eng. Part C: J. Mech. Eng. Sci. 233 (14) (2019) 5131–5143.
[193] D. Chen, S. Yang, F. Zhou, Transfer learning based fault diagnosis with missing data due to multi-rate sampling, Sensors 19 (8) (2019) 1826.
[194] B. Yang, Y. Lei, F. Jia, S. Xing, An intelligent fault diagnosis approach based on transfer learning from laboratory bearings to locomotive bearings, Mech. Syst. Sig. Process. 1229 (2019) 692–706.
[195] Z. Zhang, H. Chen, S. Li, Z. An, J. Wang, A novel geodesic flow kernel based domain adaptation approach for intelligent fault diagnosis under varying working condition, Neurocomputing 376 (2020) 54–64.
[196] L. Wen, L. Gao, Y. Dong, Z. Zhu, A negative correlation ensemble transfer learning method for fault diagnosis based on convolutional neural network, Math. Biosci. Eng. 16 (5) (2019) 3311–3330.
[197] Y. Xie, T. Zhang, A transfer learning strategy for rotation machinery fault diagnosis based on cycle-consistent generative adversarial networks, in: Proceedings of Chinese Automation Congress (CAC), 2018, pp. 1309–1313.
[198] M. Hasan, J. Kim, Bearing fault diagnosis under variable rotational speeds using stockwell transform-based vibration imaging and transfer learning, Appl. Sci. 8 (12) (2018) 2357.
[199] B. Yang, Y. Lei, F. Jia, S. Xing, A transfer learning method for intelligent fault diagnosis from laboratory machines to real-case machines, in: Proceedings of International Conference on Sensing, Diagnostics, Prognostics, and Control (SDPC), 2018, pp. 35–40.
[200] S. Tang, H. Tang, M. Chen, Transfer-learning based gas path analysis method for gas turbines, Appl. Therm. Eng. 155 (2019) 1–13.
[201] J. Zhu, N. Chen, C. Shen, A new deep transfer learning method for bearing fault diagnosis under different working conditions, IEEE Sens. J. (2020) Early Access, doi:10.1109/JSEN.2019.2936932.
[202] J. Li, X. Li, D. He, Y. Qu, A domain adaptation model for early gear pitting fault diagnosis based on deep transfer learning network, Proc. Inst. Mech. Eng. Part O: J. Risk Reliab. 234 (2020) 168–182.
[203] L.L. Cui, Z. Jin, J.F. Huang, H.Q. Wang, Fault severity classification and size estimation for ball bearings based on vibration mechanism, IEEE Access 7 (2019) 56107–56116.
[204] L.L. Cui, X. Wang, H.Q. Wang, J.F. Ma, Research on remaining useful life prediction of rolling element bearings based on time-varying Kalman filter, IEEE Trans. Instrum. Meas. 69 (6) (2020) 2858–2867.

Chuan Li received his Ph.D. degree from the Chongqing University, China, in 2007. He was a Postdoctoral Fellow with the University of Ottawa, Canada; a Research Professor with the Korea University, South Korea; a Senior Research Associate with the City University of Hong Kong, China; a Prometeo Researcher with the Universidad Poliecnica Salesiana, Ecuador; and is a Distinguished Professor with the Chongqing Technology and Business University, China. His research interests include prognostics & health management, and intelligent systems.

Shaohui Zhang received the B.E. degree in Vehicle Engineering from Huaqiao University, Quanzhou, China, in 2009, and the Ph.D. degree in Vehicle Engineering from the South China University of Technology, Guangzhou, China, in 2014. He is currently a teacher with the School of Mechanical Engineering, Dongguan University of Technology, Dongguan, China. His research interests include machine learning, fault diagnosis, and remaining useful life prediction.

Yi Qin received the B.Eng. and Ph.D. degrees in mechanical engineering from Chongqing University, Chongqing, China, in 2004 and 2008 respectively. Since January 2009, he has been with the Chongqing University, Chongqing, China, where he is currently a Professor in the College of Mechanical Engineering. His current research interests include signal processing, fault prognosis, mechanical dynamics and smart structure. Dr. Qin is a Member of IEEE.

Edgar Estupinan received his Ph.D. degree in Mechanical Engineering, from the Technical University of Denmark, Denmark, in 2010. He was a visitor research Fellow at the Institute of Sound and Vibration Research of University of Southampton, UK, in 2006. He is currently Associate Professor with the Department of Mechanical Engineering, University of Tarapaca, Chile. His research interests include predictive maintenance, fault diagnosis, and prognostics of health machinery.