Domain Adaptation with Multilayer Adversarial Learning for Fault Diagnosis of Gearbox under Multiple Operating Conditions

Ming Zhang
Graduate School at Shenzhen
Tsinghua University
Shenzhen, China
zhangming_0706@163.com

Weining Lu*
School of Aerospace Engineering
Tsinghua University
Beijing, China
luweining_thu@163.com

Jun Yang*
Department of Automation
Tsinghua University
Beijing, China
yangjun603@tsinghua.edu.cn

Duo Wang and Liang Bin
Department of Automation
Tsinghua University
Beijing, China

***Abstract*—Deep learning has been widely developed to solve fault diagnosis issues, and it is becoming a crucial technology in the modern manufacturing industry. As an important transmission device of mechanical equipment, gearbox often runs at different speeds and loads, which may lead to changes in data distribution for the actual application. The cross-domain problem caused by the different data distribution may decline the performance of the fault diagnosis model based on deep learning. To overcome this challenge, a new domain adaptation method, named MAAN: Multilayer Adversarial Adaptation Networks, for fault diagnosis of gearbox running at multiple operating conditions. The basic framework of our MAAD is a deep convolutional neural network (CNN) and then an adversarial adaptation learning procedure is used for optimizing the basic CNN to adapt cross different domain. The results of the experiment demonstrate that MAAN has outstanding fault diagnosis and domain adaptation capacity, and it could obtain high accuracies for fault diagnosis of the gearbox with changing mode. For investigating the adaptability in this method, we use t-SNE to reduce the high dimension feature for better visualization.**

Keywords-fault diagnosis; domain adaptation; convolutional neural network; multilayer adversarial learning

I. INTRODUCTION

Fault diagnosis is a critical technology in the modern industrial system, which is aiming at recognizing the cause of failure, preventing the unplanned outage, and reducing the loss of benefits. For fault identification and operating mode prediction, many diagnosis models based on deep learning have been studied [1-3]. However, we found that these deep models can be used well only when the data distribution between the training and testing process keep unchanged. We can consider that the data samples in the training process from the source distribution and the domain distribution of target provide the data sample to test the model. When the distribution changes, the capacity of the model may decline dramatically. Unfortunately, the gearbox as the critical transmission device often changes speed, load, or both of them in the real world applications, which will significantly affect the domain distribution. This phenomenon is considered to a cross-domain issue.

To overcome this problem, some domain adaptation methods have been studied for adjusting the model to fit different domains. In early research, the instance adaptation method [4] has been discussed. Learning the shared feature of the deep model has made great progress in these years, it can be obtained by minimizing the distribution distance of different domains, the key of these methods is designing a suitable distance discriminate function [5-7]. Over the last few years, some adversarial methods have been proven to be very effective for decreasing the domain distribution distance [8]. Some domain adaptation methods with deep learning by minimizing distribution discrepancy have been studied for bearing fault diagnosis when the operating conditions are different [9-11]. Our goal is to develop a deep adaptation model with a multilayer adversarial learning strategy, which can be adapt efficiently for different adaptation missions of fault diagnosis of gearbox running at multiple operating conditions.

In this work, a domain adaptation model, named MAAN: Multilayer Adversarial Adaptation Networks, has been presented. Our model is motivated by Wasserstein generative adversarial nets (WGAN) and takes advantage of CNN to construct the fault diagnosis task. In Section II, the background of related knowledge is described, and the next Section tells about the details of the theoretical framework. Detailed experiments and analyses are carried out in Section IV. At last, we have made the conclusions in Section V.

J. Yang and W. Lu are the corresponding authors.

978-1-7281-0860-5/19/$31.00 ©2019 IEEE

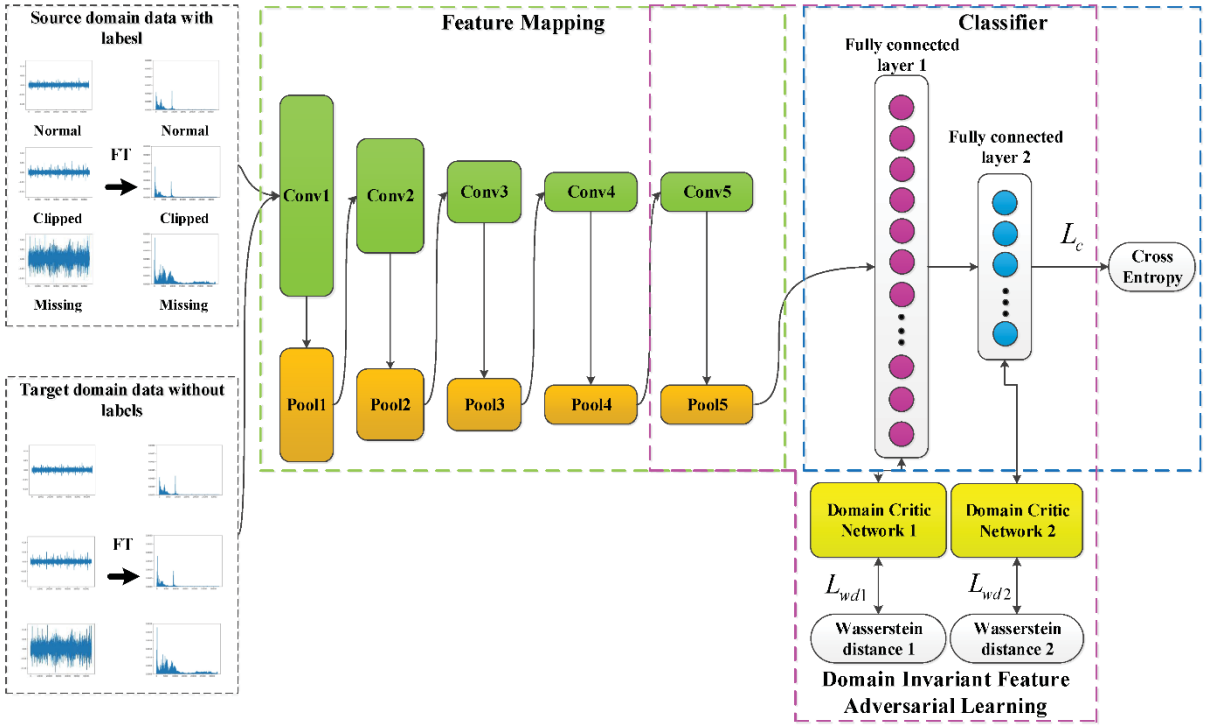


Figure 1. The architecture of Multilayer Adversarial Adaptation Networks (MAAN) method.

II. BACKGROUND

A. Problem Definition

The domain adaptation problem that our main concern is defined as follows. We assume that there are enough labeled samples $X_s = \{(x_i^s, y_i^s)\}_{i=1}^{n^s}$ come from source domain dataset D_s, while in the target domain dataset D_t, there are only samples $X_t = \{x_j^t\}_{j=1}^{n^t}$ without labels. This problem is aiming at training a model $\mathrm{M}: x^s \to y^s$ to adapt to both the source and target domain. It means that the trained model will have high accuracy $\mathrm{Acc}_{D_t} = Pr(\mathrm{M}(x^t) = y^t)$ for the samples from domain of target distribution.

B. Wasserstein Generative Adversarial Nets

The original GAN in [12] has become very popular in recent years since it is a powerful generative model with a minimax game between two networks. The essential of GAN is minimizing the distribution discrepancy between the real and generated image. However, the Jensen-Shannon divergence that GAN minimize will incur gradient vanishing in the critic network, due to the divergence may be uncontinuous [13]. The function of the Wasserstein distance is continuous and differentiable. Therefore, some researchers displace the Jensen-Shannon divergence by Wasserstein distance. Based on the Kantorovich-Rubinstein duality, the objective of WGAN is defined as follows:

$$\min_G \max_D \mathop{\mathbb{E}}_{x\sim P_r}[D(x)] - \mathop{\mathbb{E}}_{\tilde{x}\sim P_g}[D(\tilde{x})] \tag{1}$$

where D denotes the discriminator and G denotes the generator; and $\tilde{x}$ denotes the generative samples from G, which is defined by $\tilde{x} = G(z)$. There is a better way to optimize the WGAN, which can avoid gradient vanishing by increasing the gradient penalty term [14]. The final objective can be obtained:

$$\min_G \max_D \underbrace{\mathop{\mathbb{E}}_{x\sim P_r}[D(x)] - \mathop{\mathbb{E}}_{\tilde{x}\sim P_g}[D(\tilde{x})]}_{WD} - \underbrace{\lambda \mathop{\mathbb{E}}_{\hat{x}\sim P_{\hat{x}}}[(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2]}_{GP} \tag{2}$$

λ is the penalty coefficient.

III. THEORETICAL FRAMEWORK

A. Model Architecture

Fig. 1 displays the architecture of our multilayer adversarial adaptation network (MAAN). Five convolution layers are used for building deep feature representation M, and we make use of two layers of full connection to construct the classifier C. Finally, the domain critic networks with parameters θ_{d_i} are responsible for learning the domain invariant feature, which is the essential goal of the multilayer adversarial training strategy.

B. Classification Training

In the beginning, the deep convolutional neural network architecture with source samples has been trained. The parameters θ_M and θ_c, in feature representation M and

classifier C, is optimized by the loss of classification task. The general objective loss of classification L_C is presented as follows:

$$L_c = -\frac{1}{n^s}\sum_{i=1}^{n^s}\sum_{k=1}^{K} l(y_i^s = k)\cdot logC(M(x_i^s))_k \quad (3)$$

where $C(M(x_i^s))_k$ is the predicted probabilistic distribution; $l(y_i^s = k)$ is the indicator function.

C. Multilayer Adversarial Domain Adaptation Training

CNN is applied not only for implementing the classification model but also for building the domain discriminator in adversarial domain adaptation strategy. Intuitively, if we find the domain invariant feature, the cross-domain problem concerted in this work will be worked out. In this work, we will follow this thought to design the adversarial domain adaptation process. In the research [13], the domain discriminator has been proposed to evaluate the Wasserstein discrepancy of different data distribution, and it will be used for critic distribution distance in the MAAN model. We can estimate the Wasserstein distance by maximizing the loss function L_{wd} expressed as follows:

$$L_{wd} = \frac{1}{n^s}\sum_{x^s\in X^s} D(F(x^s)) - \frac{1}{n^t}\sum_{x^t\in X^t} D(F(x^t)) \quad (4)$$

where x^s are samples come from X^s, and x^t are samples come from X^t. A penalty item on the parameter θ_d of domain critic D need to be added for contenting the 1-Lipschitz restraint condition. The gradient penalty L_{gp} is defined as follows:

$$L_{gp} = (\| \nabla_{\hat{x}\in \mathrm{P}_{\hat{x}}} D(\hat{x}) \|_2 - 1)^2 \quad (5)$$

where $\hat{x}$ denotes the mixed data combine with x^s and x^t, which belongs to $\mathrm{P}_{\hat{x}}$. Eventually, we can acquire the domain invariant feature by the following equation:

$$\min_{\theta_f}\max_{\theta_d}\{L_{wd} - \lambda L_{gp}\} \quad (6)$$

where λ is responsible for the balance of the L_{wd} and L_{gp}. As we can see in equation (6), the learning procedure must be in an adversarial approach.

In our method, the domain invariant feature parameters in θ_{inv}, including θ_{conv_5}, θ_{fc_1}, and θ_{fc_2}, can make the MAAN model to adapt to different data distribution become possible. The domain critic networks D_j are used for estimating Wasserstein distance of different samples domain. The strategy is embodied in the training procedure, which is maximizing the adaptation loss L_{adv_j} to tune the parameter θ_{d_j} of domain critic networks D_j in the first place, and then learning the invariant feature parameters θ_{inv} through minimizing the adaptation loss L_{adv_j}. The adversarial adaptation loss L_{adv_j} is expressed in the following equation:

$$L_{adv_j} = \frac{1}{n^s}\sum_{i=1}^{n^s} D(F_j(M(x_i^s))) - \frac{1}{n^t}\sum_{i=1}^{n^t} D(F_j(M(x_i^t))) - \lambda(\| \nabla_{\hat{x}} D(\hat{x})\| - 1)^2 \quad (7)$$

where,

$$F_j = \begin{cases} Fc_1 & j = 1 \\ Fc_j(F_{j-1}) & j > 1 \end{cases} \quad (8)$$

; λ denotes the balance ratio.

D. The Workflow of Adversarial Training Strategy

The workflow of training our MAAN model is displayed in this part. The domain invariant feature will be achieved when the learning process of adversarial strategy ends. The workflow of our method is summarized as follows:

Step1: Initialization parameters of classification model θ_M and θ_c.

Step2: Pre-training model with classification iterations n_c:

a) Sample $\{x_i^s, y_i^s\}_{i=1}^m$ from source dataset X^s;

b) Update parameters θ_M and θ_c by minimizing the loss function L_c.

Step3: Initialization parameters of domain critic θ_{d_j}.

Step4: The multilayer adversarial training strategy with adversarial iterations n_t:

a) Sample $\{x_i^s, y_i^s\}_{i=1}^m$ and $\{x_i^t\}_{i=1}^m$ from X^s and X^t, respectively;

b) The first adversarial learning: update parameters θ_{d_1} by maximizing the adaptation loss L_{adv_1}, and update parameters θ_{conv_5} and θ_{fc_1} by minimizing the mixed loss consists of L_{adv_1} and L_c;

c) The second adversarial learning: update parameters θ_{d_2} by maximizing the adaptation loss L_{adv_2}, and update parameters θ_{conv_5}, θ_{fc_1},

and θ_{fc_2} by minimizing the mixed loss consists of L_{adv_2} and L_c.

Step5: Testing the trained model with target dataset X^t.

IV. EXPERIMENTS

A. Dataset Description

The dataset was acquired from a two-stage helical gearbox testbed with different faults under different conditions [15]. There are two accelerometers locate on the gearbox for collecting the vibration samples, and a tachometer for testing the rotating speed. The sampling frequency is 66666.67 Hz, and the data were collected when the gearbox running at multiple conditions. There are three states, including the normal, chipped tooth, and missing tooth in this dataset. The length of each sample is 6600. The details can be found in Table I.

B. Compared Methods and Implementation Details

Comparing with SVM [16], TCA [5], TCA-SVM [15], DAFD [9], and TICNN [17], the performance of our proposed model can be verified. For all methods, the samples have been processed by Fast Fourier Transform (FFT) technology before they are input the models. The transformation of the FFT can be referred to as [18]. The basic framework of classification is a convolutional neural network. Table II shows the details of the base CNN architecture, and Table III presents the details of domain critic networks. The hyper-parameters selected for the experiment are $m = 64$, $\alpha = 0.0001$, $n_c = 2000$ and $n_t = 20000$. Based on the evaluation criteria in [11], all compared models are evaluated under optimal conditions, and the best results are reported.

TABLE I. DESCRIPTION OF GEARBOX DATASET

Fault location	Normal	Clipped tooth	Missing tooth	Speed	Load
Category Labels	0	1	2		
Dataset 30L no.	500	500	500	30 Hz	Low
Dataset 30H no.	500	500	500	30 Hz	High
Dataset 35L no.	500	500	500	35 Hz	Low
Dataset 35H no.	500	500	500	35 Hz	High
Dataset 40L no.	500	500	500	40 Hz	Low
Dataset 40H no.	500	500	500	40 Hz	High
Dataset 45L no.	500	500	500	45 Hz	Low
Dataset 45H no.	500	500	500	45 Hz	High
Dataset 50L no.	500	500	500	50 Hz	Low
Dataset 50H no.	500	500	500	50 Hz	High

TABLE II. DETAIL OF FEATURE MAPPING AND CLASSIFIER

	Layer type	Kernel	Stride	Channel	Padding	Activation function
Feature Mapping	Conv 1	64×1	2×1	8	Yes	Relu
	Pool 1	2×1	2×1	8	No	
	Conv 2	32×1	2×1	16	Yes	Relu
	Pool 2	2×1	2×1	16	No	
	Conv 3	16×1	2×1	32	Yes	Relu
	Pool 3	2×1	2×1	32	No	
	Conv 4	8×1	2×1	64	Yes	Relu
	Pool 4	2×1	2×1	64	No	
	Conv 5	8×1	2×1	64	Yes	Relu
	Pool 5	2×1	2×1	64	No	
Classifier	FC 1	512		1		Tanh
	FC 2	3		1		Softmax

TABLE III. DETAIL OF THE DOMAIN CRITIC NETWORKS

	Layer type	Kernel	Stride	Channel	Padding	Activation function
Domain critic	Conv 1	5×1	2×1	64	Yes	LeakyRelu
	FC 1	512		1		LeakyRelu
	FC 2	1		1		Linear

C. Results and Analysis

In the part, the accuracies of all methods have been displayed for the gearbox dataset of multiple operating modes. We follow the experiments in [15] and [9] to compare the typical tradition model and deep model for the domain adaptation problem of gearbox fault diagnosis. In Table IV, the TCA with a frequency spectrum as input performs well at the varying speed of low load, and it is a little worse than the TCA-SVM [15] based on the average accuracy. The TICNN and the base CNN of our method display similar results, they perform well when the distribution discrepancy of different domains is not very large. However, the TICNN presents good efficacy when the source domain samples come from the high-speed state, the result is shown in Table V. This phenomenon may be confirmed that compared with the data samples from the low-speed state, the data samples from the high-speed state will contain more useful information. DAFD [9] has a good result, and it better than the traditional models. Obviously, our MAAN model performs best from all the results, and since the better domain critic can be built by using the multilayer adversarial training strategy, our method owns very good robustness for all these adaptation tasks.

D. Feature Visualization

The high-dimension features have been reduced by t-SNE [19] technology, and the MAAN's domain adaptability has been explained based on the visualization result of two-dimension. Fig.2 displays the two-dimension maps of

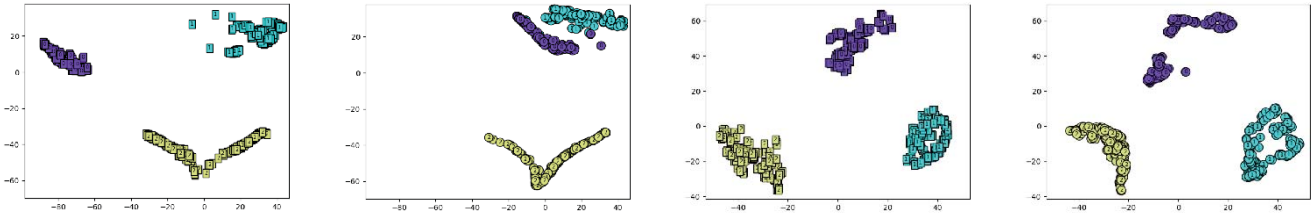
$30L \rightarrow 40H$ adaptation task and the maps of $45L \rightarrow 50H$ adaptation task have been shown in Fig.3. In TICNN model, we can easily find that the features of each category in the source domain have been spread in different locations with a large discrepancy, while it may be very hard to make a distinction between category 0 and 1 in the target domain. Category 0 and 1 of the target domain in the CNN model blend together when the features of each category are easy to identify in the source domain. These results prove that the TICNN and CNN model trained on source data samples cannot complete the diagnosis task for target samples very well. The visualization results of MAAN model demonstrate that the features of each category have been placed in different locations of the map and the distribution shows no difference in both the source and target domain, which means the diagnosis model is valid for the source domain must be effective for the target domain. All in all, we can confirm that our MAAN method has very powerful domain adaptability for making the deep model to overcome the cross-domain problem in fault diagnosis of the gearbox when it running at multiple operating conditions.

TABLE IV. ACCURACY (%) FOR EXPERIMENT SCENARIO 1

	40L→45L	40L→50L	40L→45H	30L→35H	30L→40H	30L→50L	AVG
SVM	40.60%	39.73%	37.00%	50.60%	44.67%	35.53%	41.36%
TCA	97.11%	93.33%	78.89%	63.11%	54.67%	93.55%	80.11%
TCA-SVM [23]	77.50%	83.75%	93.75%	73.75%	76.25%	81.25%	81.04%
TICNN	100.00%	96.22%	89.56%	64.22%	64.60%	66.67%	80.21%
CNN	100.00%	95.11%	95.55%	94.44%	52.89%	66.67%	84.11%
MAAN	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%

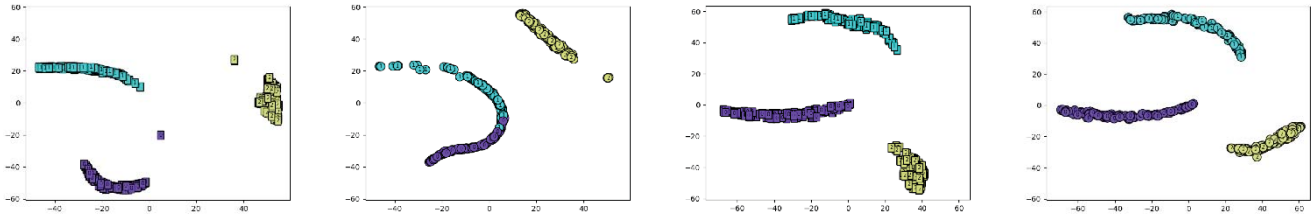
TABLE V. ACCYRACY (%) FOR EXPERIMENT SCENARIO 2

	45L→30H	45L→35H	45L→40H	45L→45H	45L→50H	AVG
SVM	28.73%	34.80%	38.33%	35.20%	39.47%	35.31%
TCA	64.60%	69.11%	58.89%	52.00%	57.78%	60.48%
DAFD [15]	67.60%	73.30%	79.30%	88.90%	87.90%	79.40%
TICNN	66.67%	89.33%	90.89%	96.60%	95.33%	87.76%
CNN	35.56%	61.11%	74.44%	84.44%	80.22%	67.15%
MAAN	95.11%	99.11%	100.00%	100.00%	100.00%	98.84%



(a) source domain of TICNN (b) target domain of TICNN (c) source domain of MAAN (d) target domain of MAAN

Figure 2. Two-dimensional visualization results of full-connection layer for adaptation task 30L→40H



(a) source domain of CNN (b) target domain of CNN (c) source domain of MAAN (d) target domain of MAAN

Figure 3. Two-dimensional visualization results of full-connection layer for adaptation task 45L→45H

V. CONCLUSION

A novel deep adaptation model, named MAAN, to address the cross-domain issue of fault diagnosis for gearbox running at a variety of environments through using multilayer adversarial adaptation strategy. Our model combines the technologies of CNN and WGAN to work out this cross-domain adaptation issue. The essence of this work is acquiring the shared feature in different domains guided by Wasserstein distance with a discriminator. Empirical results on the domain adaptation tasks of the gearbox dataset indicate that MAAN performs better than other domain adaptation learning methods.

ACKNOWLEDGMENT

This work thankful for financial supported by the Science and Technology Research Foundation of Shenzhen (No. JCYJ20160301100921349, No. JCYJ20170817152701660)

and National Key Research and Development Program of China (No. 2017YFB0602700).

REFERENCES

[1] F. Jia, Y. Lei, J. Lin, X. Zhou, and N. Lu, "Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data," Mechanical Systems and Signal Processing, vol. 72, p. 303-315, 2016.

[2] C. Li, R.-V. Sanchez, G. Zurita, M. Cerrada, D. Cabrera, and R. E. Vásquez, "Gearbox fault diagnosis based on deep random forest fusion of acoustic and vibratory signals," Mechanical Systems and Signal Processing, vol. 76, p. 283-293, 2016.

[3] P. Wang, R. Yan, and R. X. Gao, "Virtualization and deep recognition for system fault classification," Journal of Manufacturing Systems, vol. 44, p. 310-316, 2017.

[4] W.-S. Chu, F. De la Torre, and J. F. Cohn, "Selective transfer machine for personalized facial action unit detection," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013, p. 3515-3522.

[5] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," IEEE Transactions on Neural Networks, vol. 22, no. 2, p. 199-210, 2011.

[6] M. Long, Y. Cao, J. Wang, and M. Jordan, "Learning transferable features with deep adaptation networks," in International Conference on Machine Learning, 2015, p. 97-105.

[7] M. Long, H. Zhu, J. Wang, and M. I. Jordan, "Deep transfer learning with joint adaptation networks," in International Conference on Machine Learning, 2017, p. 2208-2217.

[8] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial discriminative domain adaptation," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, p. 7167-7176.

[9] W. Lu, B. Liang, Y. Cheng, D. Meng, J. Yang, and T. Zhang, "Deep model based domain adaptation for fault diagnosis," IEEE Transactions on Industrial Electronics , vol. 64, no. 3, p. 2296-2305, 2017.

[10] L. Guo, Y. Lei, S. Xing, T. Yan, and N. Li, "Deep convolutional transfer learning network: A new method for intelligent fault diagnosis of machines with unlabeled data," IEEE Transactions on Industrial Electronics, vol. 66, no. 9, p. 7316-7325, 2018.

[11] M. Zhang, D. Wang, W. Lu, J. Yang, Z. Li, and B. Liang, "A deep transfer model with wasserstein distance guided multi-adversarial networks for bearing fault diagnosis under different working conditions," IEEE Access, vol. 7, p. 65303-65318, 2019.

[12] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in Advances in Neural Information Processing Systems, 2014, p. 2672-2680.

[13] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in International Conference on Machine Learning, 2017, p. 214-223.

[14] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of wasserstein gans," in Advances in Neural Information Processing Systems, 2017, p. 5767-5777.

[15] J. Xie, L. Zhang, L. Duan, and 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, p. 1-6.

[16] C. W. Hsu and C. J. Lin, "A comparison of methods for multiclass support vector machines," IEEE transactions on Neural Networks, vol. 13, no. 2, p. 415-425, 2002.

[17] W. Zhang, C. Li, G. Peng, Y. Chen, and Z. Zhang, "A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load," Mechanical Systems and Signal Processing, vol. 100, p. 439-453, 2018.

[18] M. Zhang, Z. Jiang, and K. Feng, "Research on variational mode decomposition in rolling bearings fault diagnosis of the multistage centrifugal pump," Mechanical Systems and Signal Processing, vol. 93, p. 460-493, 2017.

[19] L. v. d. Maaten and G. Hinton, "Visualizing data using t-sne," Journal of machine learning research, vol. 9, no. 11, p. 2579-2605, 2008.