

# Domain Adaptation for Intelligent Fault Diagnosis under Different Working Conditions

Weigui Li<sup>1</sup>, Zhuqing Yuan<sup>1</sup>, Wenyu Sun<sup>1</sup>, and Yongpan Liu<sup>1,\*</sup>

<sup>1</sup> Department of Electronic Engineering, Tsinghua University, Beijing, China.

**Abstract.** Recently, deep learning algorithms have been widely into fault diagnosis in the intelligent manufacturing field. To tackle the transfer problem due to various working conditions and insufficient labeled samples, a conditional maximum mean discrepancy (CMMD) based domain adaptation method is proposed. Existing transfer approaches mainly focus on aligning the single representation distributions, which only contains partial feature information. Inspired by the Inception module, multi-representation domain adaptation is introduced to improve classification accuracy and generalization ability for cross-domain bearing fault diagnosis. And CMMD-based method is adopted to minimize the discrepancy between the source and the target. Finally, the unsupervised learning method with unlabeled target data can promote the practical application of the proposed algorithm. According to the experimental results on the standard dataset, the proposed method can effectively alleviate the domain shift problem.

## 1 Introduction

Fault diagnosis is a fundamental problem in the modern machinery prognostics health management (PHM). Thanks to the widespread use of intelligent sensors, data-drive methods have received extensive attention from academia and industry departments [1]. Recently, deep learning (DL) based intelligent fault diagnosis algorithms have already prospered and found their ways into fault diagnosis [2]. Zhang et al. [3] designed an effective deep convolutional neural network with wide first layer kernels (WDCNN) to extract features and restrain high frequency noise. Guo et al. [4] utilized wavelet transform in different scales to get rid of the vibration signal for data preprocessing and put frequency features into DCNN for fault diagnosis. They both achieved success and obtained high classification accuracy in specific training conditions.

Because the operation condition of the bearing is changing with noise, loads, and other complex environmental factors in real industrial scenarios, which often produces different data distribution and then leads to the performance degradation of the model dramatically. Therefore, domain adaptation and robust algorithms are in need.

Although the existing approaches about bearing diagnosis (such as DAN [5], DANN [6]) are competitive, several issues and challenges still need to be addressed. Firstly, existing methods commonly focus on aligning single representation distribution to minimize the discrepancy, ignoring the diversity. Secondly, smaller and faster algorithm is more conducive to terminal deployment in the field of fault detection. Based on previous works, we propose a CMMD-based domain

adaptation framework by introducing Inception module and CMMD distance to improve the detection performance. The main contributions in this paper are summarized as follows:

- (1) A novel intelligent network is proposed for bearing fault diagnosis with Inception module. Different views from representations contain more feature distribution information for classification and domain adaptation.
- (2) The CMMD distance are introduced to align the feature distribution between the source and the target.

## 2 Related work

The success of DL-based intelligent fault diagnosis mainly depends on the following two aspects: (1) extensive labeled samples are available. (2) Data distribution from the training set is the same as the test set [7]. However, the test set is often newly collected vibration data with different working conditions and different distribution. Fortunately, unsupervised deep transfer learning (UDTL) is designed to deal with the this dilemma. UDTL is commonly used in the field of computer vision (CV) and natural language processing (NLP) [8]. Their strategies are mainly divided into two class: (1) embedding adaptation layers to minimize the discrepancy by the distance measure (such as MMD, mean and covariance matrix, and so on). (2) introducing adversarial training strategy to learn domain-invariant features.

Recently, UDTL also find its way into the field of intelligent fault diagnosis. Especially, domain adaptation is usually the most common method. MMD distance measure is often used to match the domain adaptation and have achieved good results in [8]. Meanwhile, recent

\* Corresponding author: [ypliu@tsinghua.edu.cn](mailto:ypliu@tsinghua.edu.cn)

researches extend the adversarial training methods to align the distributions by introducing a domain discriminator [6]. However, all these work focus on aligning distributions of domain from a single view. In this paper, we will explore multi-view domain distribution aligning. Moreover, the CMMD distance is measured through multi-scale feature space by Inception module from googLeNet [9].

### 2.1. Domain adaptation

Generally, a typical domain adaptation (DA) framework can be formulated as jointly training the source  $D_s$  data with labels and the target  $D_t$  data without labels. Assuming that the source is  $D_s = \{(x_i^s, y_i^s)\}_{i=1}^{n_s}$  and the target is  $D_t = \{(x_i^t)\}_{i=1}^{n_t}$ , where  $x$  is the sample,  $y$  is the label and  $n$  is the number of the sample.  $D_s$  and  $D_t$  are from different joint distributions respectively:  $P(x_i^s, y_i^s)$  and  $Q(x_i^t, y_i^t)$ . In most situations,  $P$  is not equal to  $Q$ . Therefore, the fault diagnosis model is designed to learn domain-invariant features and minimize the domain shift. The process is depicted in Fig. 1 (a).

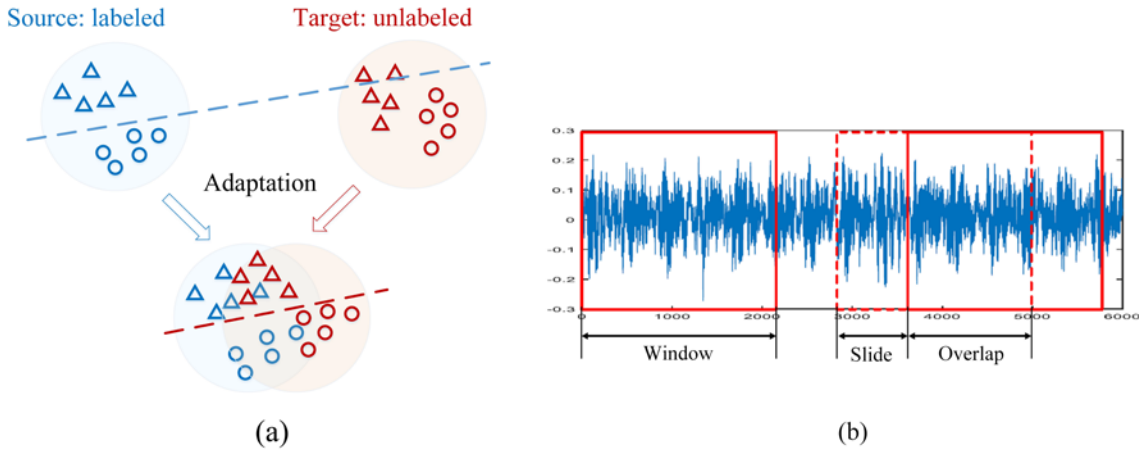


Fig. 1 (a) Illustration of domain adaptation. The top plots represent the source domain classifier; the bottom plot represents the cross-domain classifier after domain adaptation, (b) schematic diagram of single window resampling method.

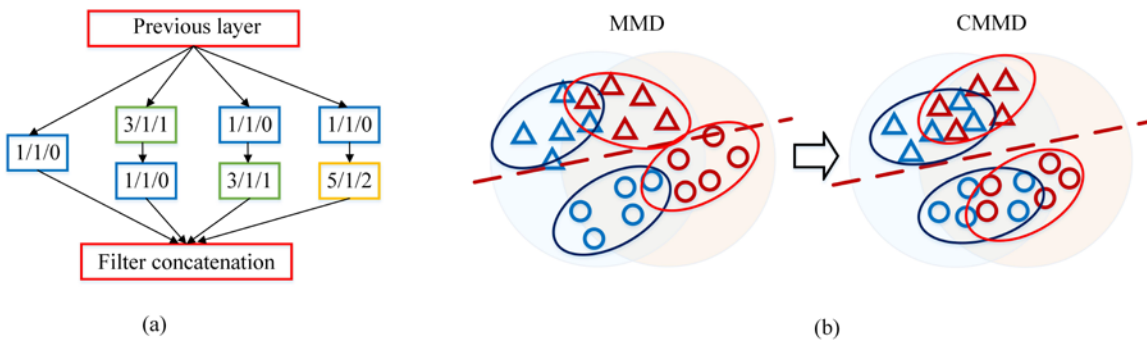


Fig. 2 (a) Inception module, the 1D convolutional parameters are denoted as follows: kernel size/stride/padding, (b) comparison of MMD and CMMD.

### 2.2 Data augmentation.

In the field of computer vision and natural language processing, the input samples are fixed. However, fault diagnosis algorithms based on the vibration signal need to resample the training samples and determine the best size of input data. In the past, the single window resampling method was widely used for data augmentation in time-series data, as shown in Fig. 1 (b). In this way, we could produce a large number of training samples without extra efforts. The existing most excellent methods tend to set the resampling size to 1024 or 2048 [2, 3, 4, 5, 6]. In this paper, the resampling size is 1024.

### 2.3 Inception module

Inception module is firstly proposed to add the width of the neural network and increase the adaptability of the network to the scale. The receptive field of different branches is different, so there is multi-scale information in it. The details of Inception module is shown in Fig. 2 (a). In addition, it is also beneficial to learn different domain-invariant representations.

### 2.4 Conditional Max Mean Discrepancy.

To approach the adaptation of unlabeled information in the target domain, most existing methods aim to bound

the discrepancy metric between the source and the target. In this paper, we focus on optimizing the multiple kernel variant of MMD (MK-MMD) proposed by Gretton et al. [10]. Denote by  $\mathcal{H}_k$  be the reproducing kernel Hilbert space (RKHS) embedded with a characteristic kernel  $k$ , where  $f$  represents a kernel function.  $D_k(x^s, x^t)$  is defined as the RKHS distance between  $P(x^s)$  and  $Q(x^t)$ .

$$D_k(x^s, x^t) := \sup_{f \in \mathcal{H}} \left( \left\| \frac{1}{m} \sum_{i=1}^m f(x_i^s) - \frac{1}{n} \sum_{i=1}^n f(x_i^t) \right\| \right) \quad (1)$$

In addition, Elhamifar and Vidal [11] have demonstrated that the same class samples may project to the same subspace even though they are from different domains. Based on this, class subspace constraint is applied to minimize the domain shift, namely CMMD, as shown in Fig. 2 (b).

$$D_k(C_i, x^s, x^t) := \sup_{f \in \mathcal{H}} \left( \frac{1}{C} \sum_{c=1}^C \left\| \frac{1}{m} \sum_{i=1}^m f(x_i^s) - \frac{1}{n} \sum_{i=1}^n f(x_i^t) \right\| \right) \quad (2)$$

Since the samples are unlabeled data in the target domain, Eq. 2 cannot be used directly. Another

hypothesis that the source and the target share the same labels, is added [12].

### 3 Proposed Framework

#### 3.1. Overview of proposed framework

The proposed framework is composed of three parts, i.e., one is for feature extraction, another is for domain adaptation, and the last is for classification. Fig. 3 shows the overview of proposed framework.

The resampling data from the source and the target respectively are put into one dimension convolutional neural network (1DCNN) for feature extraction. Next, shallow feature maps are processed to deep feature maps through Inception module. In this process, non-parametric CMMD loss is calculated and used to consist of the training loss. Finally, deep feature maps from the source are sent into classifier for prediction, along with their labels.

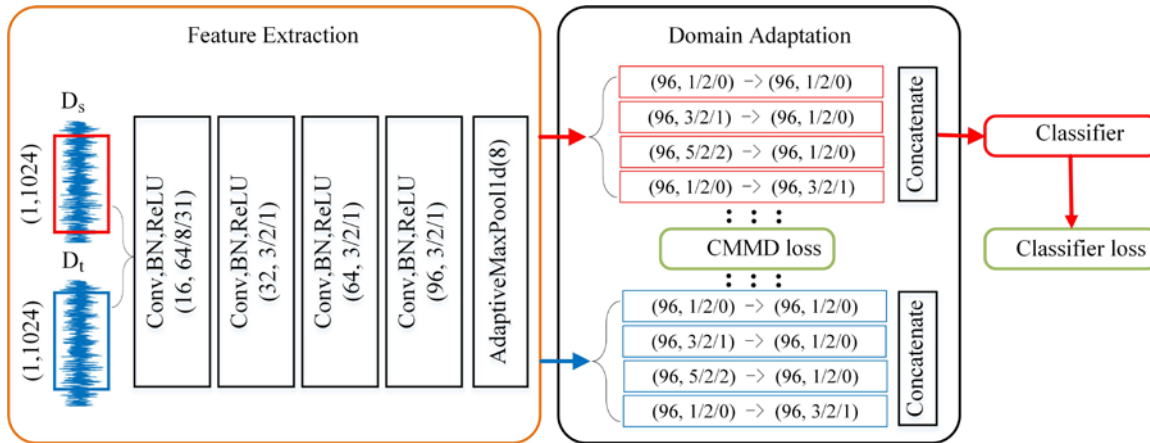


Fig. 3 Illustration of the proposed framework. The convolutional module consisted of 1d convolution, batch normalization, and LeakyReLU layers, the convolutional parameters are denoted as follows: channel number, kernel size/stride-padding.

#### 3.2. Multi-scale adaptation module.

Inspired by multi-view learning, multi-scale representation distribution aligning is also better than the single's. To learn multiple different domain-invariant features and minimize the discrepancy between class subspaces, wider and deeper representation distributions are considered to be aligned by Inception module. In addition,  $1 \times 1$  convolutions are mainly used as dimension reduction modules. In a word, Inception module is adopted to extract multi-scale features and improve the distribution aligning performance.

#### 3.3. Training strategy

To improve unsupervised domain adaptation performance, we jointly optimize the source classification error and CMMD distance. The loss of the proposed model can be formulated as:

$$L_{total} = \frac{1}{m} \sum_{i=1}^m CrossEntropy(f(x_i^s), y_i^s) + \lambda D_k(C_i, x^s, x^t) \quad (3)$$

Where the former denotes the cross-entropy between the prediction and labels, the latter is CMMD loss with trade-off parameter ( $\lambda > 0$ ). Moreover,  $\lambda$  is not fixed and changing with training epoch, as defined in Eq. 4.

$$\lambda = 2 / \left( 1 + e^{-10 \times \frac{epoch}{epochs}} \right) - 1 \quad (4)$$

When epoch is increasing,  $\lambda$  is becoming larger so that CMMD loss will take more optimization. Adam optimizer with an initial learning rate  $2e^{-4}$  is employed [13]. Moreover, multi-step learning rate scheduler is applied with 0.5 decay factor.

## 4 Experiments

### 4.1. Experimental setup

NVIDIA GeForce RTX 2080Ti is used for all experiments. The proposed framework is completed by Python and Pytorch.

#### 4.1.1 Experimental data set.

**Table 1.** Ten kinds of fault data from CWRU data set.

Class Label	0	1	2	3	4	5	6	7	8	9
Fault Location	N A	I F	B F	O F	IF	B F	O F	IF	B F	O F
Diameters (mil)	0	7	7	7	14	14	14	21	21	21

The Case Western Reserve University (CWRU) data set is a standard and recognized benchmark for bearing fault diagnosis [14]. The test rig is mainly made up of a 2hp motor, a power meter, a torque sensor, and an electronic control device. The raw vibration data was collected by the accelerometers at the frequency of 12 kHz and 48 kHz, with the loads of 0 to 3 horsepower. Following most of existing works, there are ten kinds of drive end fault data whose sampling frequency of 12 kHz, as described in Table 1. In the experiments, 1000 samples of each fault are resampled randomly from time-series data. Therefore, the data set is consisted of 10000 samples.

**Table 3.** Domain adaptation tasks from CWRU data set.

Model	Baseline(WDCNN)		DAN		DANN		Proposed	
	Mean	Max	Mean	Max	Mean	Max	Mean	Max
0→1	89.91	90.4	99.38	99.50	98.76	99.30	99.54	99.80
0→2	88.50	89.3	99.98	100.00	99.96	100.00	99.75	99.92
0→3	93.01	93.4	100.00	100.00	99.81	100.00	99.56	99.62
1→0	95.40	95.71	99.31	99.40	98.73	99.05	99.70	99.83
1→2	99.20	99.35	99.98	100.00	99.96	100.00	99.90	99.93
1→3	91.00	91.22	99.97	100.00	99.65	99.80	99.88	99.98
2→0	95.82	95.91	98.61	98.65	97.70	98.25	99.61	99.80
2→1	95.10	95.22	98.52	98.60	98.40	98.45	99.62	99.86
2→3	91.50	91.62	100.00	100.00	99.82	99.95	99.84	99.95
3→0	78.82	79.16	98.72	98.90	97.62	97.85	99.50	99.65
3→1	78.10	78.33	98.53	98.60	98.41	98.50	98.60	98.85
3→2	85.10	85.21	100.00	100.00	99.98	100.00	99.90	99.95
Average	90.12	90.40	99.42	99.47	99.07	99.26	<b>99.61</b>	<b>99.76</b>
Params	54790		1379998		2694816		119258	

#### 4.2.1 Baseline (WDCNN).

WDCNN proposed by Zhang et al. [3], is used as the baseline for comparison in many papers. To validate the effectiveness of transfer learning, the backbone of the

#### 4.1.2 Domain adaptation tasks

Working load changing is the most common phenomenon in industrial production. To validate the effectiveness of the proposed framework, we designed a series of experiments on CWRU data set. Based on sampling data from different loads, the domain adaptation tasks are defined in Table 2.

**Table 2** Domain adaptation tasks from CWRU data set.

Loads	0	1	2	3
Tasks	0→1	1→0	2→0	3→0
	0→2	1→2	2→1	3→1
	0→3	1→3	2→3	3→2

#### 4.2. Experimental results

To evaluate the proposed framework, extensive experiments are completed for comparison. Average and maximum accuracy are reported to reflect the model performance and stability in five experiments. From Table 3, it is clear that the proposed method achieved the best performance. In the following, some explanations will be presented.

proposed network also adopted the baseline module. WDCNN's experimental results are from the original paper. Comparing the baseline and the domain adaptation model, it shows that domain adaptation can not only improve the accuracy but also improve the generalization ability of the model.

#### 4.2.2 Model performance.

Without domain adaptation, the baseline can also achieve decent performance. However, it is still far from expected. In many cases, fault diagnosis methods need better performance to avoid financial losses and other safety concerns. Therefore, DL-based domain adaptation is important in modern industrial production.

To deal with unlabeled target data, UDTL methods are adopted to improve the performance. In this paper, several excellent methods are only presented for comparison, such as DAN [5], DANN [6]. Note that the comparison results are all from Wang et al.'s paper [6]. The parameters of the two experiments are very large. Although their performance are very good, they are not fit for the deployment of industrial production scenarios due to limited computational resources. Fortunately, the proposed model cannot only reach a high performance but also has fewer parameters, as shown in Table 3. The proposed CMMD distance is an advanced version of the previous MMD distance. This is the guarantee of the model's performance.

## 5 Conclusions

In this paper, a UDTL-based model of bearing fault diagnosis is proposed. Compared with previous work, we introduce Inception module and CMMD-based distance between the source and the target to realize domain adaptation in the field of bearing fault diagnosis. It also achieves better performance and has few parameters. It not only effectively alleviates the domain adaptation problem with unlabeled data but also benefit for industrial implementation of the model. In the future, we will extend it to more complicated data sets and industrial applications.

## Acknowledgments

This work was supported in part by the National Key R&D Program under Grant 2017YFB1400900, and in part by NSFC under Grant 61934005.

## References

1. D.-T. Hoang and H.-J. Kang: A survey on deep learning based bearing fault diagnosis, *Neurocomputing*, vol. 335, pp. **327 – 335**, 2019.
2. D.-T. Hoang and H.-J. Kang: Rolling element bearing fault diagnosis using convolutional neural network and vibration image, *Cognitive Systems Research*, vol. 53, pp. **42 – 50**, *Advanced Intelligent Computing*, 2019.
3. W. Zhang, G. Peng, C. Li, Y. Chen, and Z. Zhang: A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals, *Sensors*, vol. 17, no. 2, p. **425**, 2017.
4. S. Guo, T. Yang, W. Gao, and C. Zhang: A novel fault diagnosis method for rotating machinery based on a convolutional neural network, *Sensors*, vol. 18, p. **1429**, 05 2018.

5. X. Li, W. Zhang, Q. Ding, and J.-Q. Sun: Multi-layer domain adaptation method for rolling Bearing fault diagnosis, *Signal Processing*, vol. 157, pp. **180–197**, 2019.

6. Q. Wang, G. Michau and O. Fink: Domain Adaptive Transfer Learning for Fault Diagnosis, 2019 Prognostics and System Health Management Conference (PHM-Paris), Paris, France, 2019, pp. **279-285**.

7. Z. Zhao, Q. Zhang, X. Yu, C. Sun, S. Wang, R. Yan, and X. Chen: Unsupervised Deep Transfer Learning for Intelligent Fault Diagnosis: An Open Source and Comparative Study, arXiv e-prints, p. arXiv: 1912.12528, Dec.2019.

8. M. Long, Y. Cao, J. Wang, and M. I. Jordan: Learning Transferable Features with Deep Adaptation Networks, arXiv e-prints, p. arXiv: 1502.02791, Feb. 2015.

9. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. E. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich: Going deeper with convolutions, *CoRR*, vol. abs/1409.4842, 2014.

10. Gretton, A., Borgwardt, K., Rasch, M., Schölkopf, B., and Smola, A: A kernel two-sample test. *Journal of Machine Learning Research*, 13:723–773, March 2012a.

11. Elhamifar, E., & Vidal, R. (2013). Sparse subspace clustering: Algorithm, theory, and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(11), 2765–2781.

12. Yongchun Zhu, Fuzhen Zhuang, Jindong Wang, Jingwu Chen, Zhiping Shi, Wenjuan Wu, Qing He: Multi-representation adaptation network for cross-domain image classification, *Neural Networks*, Volume 119, 2019, Pages 214-221.

13. D. P. Kingma and J. Ba: Adam: A Method for Stochastic Optimization, arXiv e-prints, p. arXiv: 1412.6980, Dec 2014.

14. W. A. Smith and R. B. Randall: Rolling element bearing diagnostics using the case Western Reserve university data: A benchmark study, *Mechanical Systems and Signal Processing*, vol. 64-65, pp. **100 – 131**, 2015.