A Transfer Learning-Based Rolling Bearing Fault Diagnosis Across Machines

Dandan Peng[1, 2], Chenyu Liu[1, 2], Alexandre Mauricio[1, 2], Wim Desmet[1, 2], Konstantinos Gryllias[1, 2]

[1]*Department of Mechanical Engineering, Faculty of Engineering Science, KU Leuven*

[2]*Dynamics of Mechanical and Mechatronic Systems, Flanders Make, Celestijnenlaan 300, BOX 2420, 3001 Leuven, Belgium*
konstantinos.gryllias@kuleuven.be

ABSTRACT

Currently, most studies focus on training an excellent deep learning model based on sufficient labeled data collected from machines. However, in real applications, it is costly or impractical to obtain massive labeled data for model training. Therefore, in this paper, a Transfer Learning (TL)-based fault diagnosis method is proposed to transfer the model learnt from one machine (source domain) to another one (target domain). In the training process, labeled source data and unlabeled target data are used, which is very promising for real industrial applications. In this frame of transfer learning-based fault diagnosis, a cyclic spectrum correlation analysis method is firstly introduced to obtain order frequency maps for removing the influence of speed variation and revealing the hidden cyclic frequency of signals. Then, the Dynamic Adversarial Adaptation Network (DAAN) is introduced to transfer label information across machines. The proposed fault diagnosis method across machines is applied on two rolling element bearing datasets collected from two different test rigs. Experimental results demonstrate the effectiveness and superiority of the proposed method compared with state-of-the-art approaches.

1. INTRODUCTION

Being key components of rotating machinery, rolling bearings are widely used in aircrafts, high-speed trains, wind turbines, etc. Once a bearing fails, the equipment is not anymore able to operate normally, and even accidents may occur. Fault diagnosis plays a crucial role in ensuring the safe operation of machinery as an important part of Prognostics and Health Management (PHM).

Bearing fault diagnosis is mainly based on vibration signals, because they are sensitive to weak faults (Yong et al., 2016). The traditional fault diagnosis chain mainly includes several steps: signal acquisition, signal denoising, feature extraction, and decision making. One well-known approach is to select the bearing fault-related frequency band for signal demodulation (Randall, 2011). Antoni et al. (2006) proposed the fast Kurtogram to select the frequency band associated with bearing faults, which has been studied and widely applied in the bearing fault diagnosis. Recently, Cyclic Spectral Correlation (CSCorr) has gained momentum in the condition monitoring community, because it is able to reveal hidden periodicities of second-order cyclostationarity (Antoni, J., 2007), such as the weak bearings signals, which are often buried in noise or masked by other stronger signals. Mauricio et al. (2018) explored the application of CSCorr in the condition monitoring of planerary gearboxes under varying speed conditions and obtained promising diagnostic results.

Thanks to the evolution of Machine Learning (ML) technology, ML-based machinery fault diagnosis methods have been significantly developed. In general, the statistical features of vibration signals are firstly extracted and then fed into ML models, such as Support Vector Machine (SVM) (Widodo et al., 2007), *k*-nearest neighbor (Lu et al., 2021), naive Bayes (Zhang et al., 2018), etc., to obtain diagnostic results. In the past decade, Deep Learning (DL) technology has attracted much attention from researchers in computer vision (Voulodimos et al., 2018), medical image segmentation (Hesamian et al., 2019), speech recognition (Nassif et al., 2019). Owing to their excellent feature extraction ability from large amount of data, DL models have been extensively studied in the field of PHM, including Convolutional Neural Network (CNN) (Wang et al., 2019; Chen et al., 2019), Long Short-Term Memory (LSTM) (Chen et al., 2021), etc.

Currently, most DL models are trained by one dataset collected from one single machine for fault diagnosis (Chen et al., 2021; Wang et al., 2019). In real life applications, it is still very impractical to train an effective deep learning model for each machine, as collecting enough labeled data covering various operating conditions and various fault types is very time-consuming and costly. Therefore, a natural idea is to leverage the label information of one machine to improve the models' diagnostic performance on other machines. However, a direct model re-application on other machines will decrease the performance considerably. The main reason is the distribution mis-match between the two machines, which is also referred as the domain shift issue (Ganin et al., 2016).

The Transfer Learning (TL) technology can address the domain shift problem and transfer label information across domains successfully (Long et al., 2017). Inspired by the adversarial training idea of Generative Adversarial Network (GAN) (Creswell et al., 2018), Ganin et al. (2016) proposed a deep transfer learning network, called Domain Adversarial Neural Network (DANN), in which the marginal distributions of the source domain and the target domain are aligned to a shared feature space for solving the domain shift issue. It is an unsupervised domain adaptation method and widely applied in the transfer learning tasks among different working conditions (Guo et al., 2018; Mao et al., 2020; Guo et al. 2021). Besides the marginal distribution alignment between domains, the conditional distribution alignment also contributes to the adaptation (Yu et al., 2019). Therefore, Yu et al. (2019) developed the Dynamic Adversarial Adaptation Network (DAAN) to extract the domain-invariant features by aligning the marginal and conditional distributions together. It has been demonstrated that these deep TL models perform well in the case of TL among different working conditions, but they may fail in the transfer learning tasks across machines.

In this paper, a novel TL-based fault diagnosis framework across machines is proposed. Firstly, the CSCorr method is performed on the vibration signal to obtain an order-frequency two dimensional (2D) map, which can not only remove the influence of speed variation, but also reveal the hidden cyclic frequency of bearing signals. Then, DAAN is introduced to align the marginal (global) and conditional (local) distributions between machines and thus successfully transfers the model trained from one machine to another one. The training data is composed of labeled source data and unlabeled target data, thus solving the difficulty of labelling data. The proposed TL-based fault diagnosis method is validated on two different bearing datasets and experimental results approve its effectiveness. The proposed transfer learning-based fault diagnosis method for rolling bearings is successfully applied to the fault diagnosis tasks across machines.

The remaining part of this paper is organized as follows. Section 2 reviews the basic theories of CSCorr and DANN. The proposed fault diagnosis framework is illustrated in Section 3. Section 4 demonstrates the effectiveness of the proposed method by excessive experiments, and finally some conclusions are given in Section 5.

2. Background Theory

This section reviews the basic theory of Cyclic Spectral Correlation (CSCorr) and Domain Adversarial Neural Network (DANN).

2.1. Cyclic Spectral Correlation

Cyclic Spectral Correlation (CSCorr) (Antoni, J., 2007) is a very powerful tool in bearing fault diagnosis field. A bearing vibration signal can be described as a second-order cyclostationary signal, defined as Eq.(1).

$$R_X(t,\tau) = R_X(t+T,\tau) = E\{x(t)x(t-\tau)^*)\} \tag{1}$$

where R_X is the autocorrelation function, E represents the expected value, $x(t)$ is the signal with the time variable t, * means conjugate, T is the cyclic period and τ is the lag parameter. The cyclic autocorrelation function describes the periodicity of the second moment of the signal.

CSCorr reveals the hidden modulation signal (the cyclic frequency α), and its carrier frequency (the spectral frequency f) (Mauricio et al., 2020). It is a bi-frequency representation of cyclic and spectral frequencies. The spectral frequency highlights the carrier component of impulses, and the cyclic frequency highlights the second-order periodicity of impulses. Specifically, it measures the correlation between two frequency components of the signal at f and $f+\alpha$. The statistical descriptor of CSCorr can be described as Eq.(2).

$$S_X(\alpha,f) = \lim_{W\to\infty} \tfrac{1}{W} E\left\{X_W^*(f)\,X_W(f+a)\right\} \tag{2}$$

$X_W(f)$ does the Fourier transform of the signal $x(t)$ over the time interval W; f is the spectral frequency dual with time t and α is the cyclic frequency dual with time-lag τ.

2.2. Convolutional Neural Networks

Convolutional Neural Networks (CNN) are mainly composed of convolution layers, pooling layers and activation functions. The convolution layer implements the convolution operation of input features via convolution kernels, whose number is determined by the one of feature maps. Then, the activation function ReLU (Nair et al., 2010) is followed to increase the nonlinear feature learning ability of the network. Subsequently, a pooling layer is stacked, which mainly contains two algorithms: Max pooling and Mean pooling. Not only can it prevent the network from overfitting, but it can also reduce feature dimensions, thereby accelerating network training. After multiple stacked convolution and pooling layers, the output features are fed to the fully connected layer and the Softmax layer. After that, all features are mapped to the range (0, 1), which is the predicted probability distribution.

2.3. Domain Adversarial Neural Network

Suppose there are two domains, the source domain $D_s = \{X_s\}$ and te target domain $D_t = \{X_t\}$. The sample in D_s is $X_s = \{x_s^1, x_s^2, ..., x_s^{n_s}\}$, where n_s is the sample number of the source domain; $Y_s = \{1, 2, ..., C\}$ are the labels corresponding to the samples, where C is the number of fault classes. The feature distribution of the source domain is $P_s(X_s)$. Then, the sample in the target domain is defined as $X_t = \{x_t^1, x_t^2, ..., x_t^{n_t}\}$, where n_t is the sample number of the target domain, and its corresponding feature distribution is $P_t(X_t)$. Since DANN is an unsupervised domain adaptation method, the samples in the target domain for training do not have to be labeled. It is worth noting that the source and the target domains share the same label space.

DANN consists of a CNN-based feature extractor G_f, a label classifier G_y and a domain discriminator G_d. During the training process, the G_f is expected to fool the G_d as much as possible, so the loss of the G_d is maximized so that G_d cannot distinguish features between the source domain and the target domain. Meanwhile, G_d aims to distinguish the features between the source domain and the target domain as much as possible, thus minimizing its loss. Through this adversarial training, the network can learn the discriminative domain-invariant features of two domains. Its overall loss function is defined as Eq.(3).

$$\begin{aligned} L(\theta_f, \theta_y, \theta_d) = & \frac{1}{n_s} \sum_{x_i \in X_s} L_y\left(G_y\left(G_f\left(x_i; \theta_f\right); \theta_y\right), y_i\right) \\ & - \frac{\lambda}{n_s + n_t} \sum_{x_i \in X_s \cup X_t} L_d\left(G_d\left(G_f\left(x_i; \theta_f\right); \theta_d\right), d_i\right) \end{aligned} \tag{3}$$

where θ_f is the parameter of G_f; θ_y and L_y are the parameter and the loss function of G_y, respectively; θ_d and d_i are the parameter and the loss function of G_d separately; d_i is the domain label (0 or 1), corresponding to the source domain or the target domain; and λ is a trade-off parameter. The training goal of DANN is to minimize the loss of the label classifier G_y so that the G_f can extract discriminative features while maximizing the loss of the domain discriminator G_d to obtain domain-invariant features.

3. Proposed Method

In fault diagnosis field, it is difficult to train a deep learning model for each machine due to the lack of labels and data. Therefore, this paper proposes a TL-based bearing fault diagnosis method across machines, as shown in Figure 1. Firstly, CSCorr is adopted as a signal preprocessing method to highlight the hidden cyclic frequencies of bearing signals and 2D order frequency maps are obtained. Then, the DAAN domain adaption method is proposed to achieve domain transfer with the input of CSCorr maps. In addition to map global feature distributions, the local subdomain (each fault class) distributions between two domains are also expected to be aligned while using the DAAN method (Yu et al., 2019). It further introduces a class-wise domain discriminator module into the original DANN network (Ganin et al., 2016) to align the local subdomain distribution. The DAAN method includes four modules, the feature extractor G_f, the label classifier G_y, the global domain discriminator G_d and the class-wise domain discriminator G_d^c.

3.1. Label Classifier

The label classifier G_y implements the classification task on the source domain dataset. In front of the label classifier, there is a feature extractor G_f, constructed by Resnet18 (Targ et al., 2016). G_f aims to extract discriminative and domain-invariant features across domains. Then, these deep features are input to the label classifier, which consists of a fully connected layer and a Softmax layer. The number of neurons in the fully connected layer is equal to the one of the bearing healthy classes C. Softmax layer is used to get the predicted probability distribution that the sample x_i belongs to class c. The loss function L_y of G_y is the cross-entropy loss function, defined as Eq.(4).

$$L_y = -\frac{1}{n_s} \sum_{x_i \in D_s} \sum_{c=1}^{C} P_{x_i \to c} log G_y\left(G_f\left(x_i\right)\right) \tag{4}$$

3.2. Global Domain Discriminator

The idea of the global domain discriminator G_d in DAAN comes from the domain discriminator in DANN. G_d is used to align the global distributions of the source and te target domains. As shown in Figure 1, the features obtained from the feature extractor G_f are input into the G_d. G_f tries to confuse G_d as much as possible, so that the global domain discriminator G_d cannot distinguish the features from the source domain or the target domain and thus the domain-invariant features between the two domains are learnt. G_d is composed of three fully connected layers, and the numbers of neurons are set to 1024, 1024 and 2, respectively. Te activation function in the global domain discriminator is the ReLU. Each fully connected layer is followed by a dropout layer to avoid network overfitting. The loss function L_g of the global domain discriminator is defined as in Eq.**Error! Reference source not found.**.

$$L_g = \frac{1}{n_s + n_t} \sum_{x_i \in D_s \cup D_t} L_d\left(G_d\left(G_f\left(x_i\right)\right), d_i\right) \tag{5}$$

where $L_d(\cdot)$ is the cross-entropy loss function, and d_i is a domain label (0 or 1), corresponding to the source or target domain.

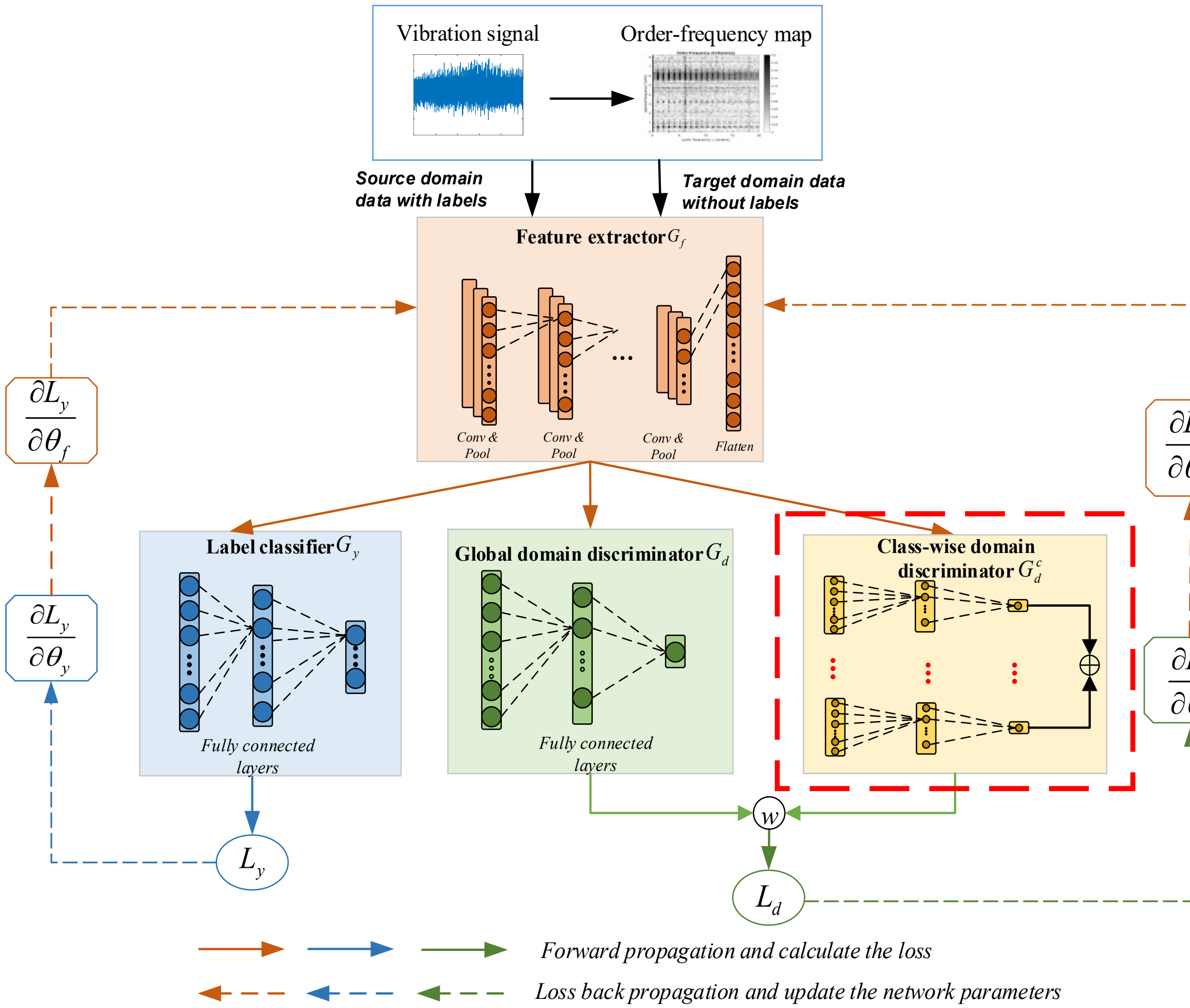


Figure 1. The fault diagnosis framework based on te CSCorr-DAAN

3.3. Class-Wise Domain Discriminator

The class-wise domain discriminator G_d^c can align the conditional distributions of the source domain and the target domain for each fault class. The G_d^c consists of C-class domain discriminators, each of which matches the features of two domains in the c_{th} class. Its network architecture is consistent with the global domain discriminator. The loss function L_l of G_d^c is defined as Eq.(6).

$$L_l = \frac{1}{n_s + n_t} \sum_{c=1}^{C} \sum_{x_i \in D_s \cup D_t} L_d^c \left(G_d^c \left(\hat{y}_i^c G_f \left(x_i \right) \right), d_i \right) \tag{6}$$

where $L_d^c(\cdot)$ is the cross entropy loss function associated with the class c; G_d^c is the c_{th} local domain discriminator; $\hat{y}_i^c$ is the predicted probability distribution over the class c of the input sample x_i; and d_i is a domain label (0 or 1), corresponding to the source domain or the target domain.

3.4. Network Training and Optimization

It can be seen from Figure 1 that the total loss function of the network includes the losses of the label classifier, the global domain discriminator, and the wise-class domain discriminator. Therefore, the total loss function of the network is defined as in Eq.**Error! Reference source not found.**.

$$\begin{aligned} L\left(\theta_f, \theta_y, \theta_d, \theta_d^c\big|_{c=1}^{C}\right) &= L_y - \lambda L_D \\ L_D &= (1-w) L_g + w L_l \end{aligned} \tag{7}$$

where λ is a trade-off parameter; L_y is the loss of the label classifier; L_D is the loss of the domain discriminator, which consists of the loss of the global domain discriminator and the wise-class domain discriminator. w is a weight parameter. When w=0, the DAAN degenerates into the DANN, and the global distribution alignment is more important. When w=1, it means that the local subdomain distribution of each fault class dominates, while the importance of the global distribution of both domains decreases. Therefore, the value of w can be dynamically adjusted according to the actual application. Then, the Adaptive Moment Estimation (Adam) (Kingma et al., 2014) optimization method is used to optimize and update the network parameters. The advantage of Adam is that each iteration of the learning rate has a certain range, making parameters relatively stable. The gradient direction in the back propagation process is required to be automatically reversed, which benefits from the Gradient Reversal Layer (GRL) (Yu et al., 2019).

4. Experimental verification and analysis

4.1. Dataset Description

Dataset I: This dataset comes from the LMSD section of KU Leuven. The test rig is shown in Figure 2. Two rolling element bearings are installed on the shaft, namely the healthy bearing and the test bearing. Accelerometers are mounted on the housing of the experimental bearing to collect vibration signals. A motor drives the shaft on which a disk is mounted. Three test bearings are considered, including a healthy bearing, a bearing with a small spall in the inner race, and a bearing with a mild spall in the inner race, as shown in Figure 3.

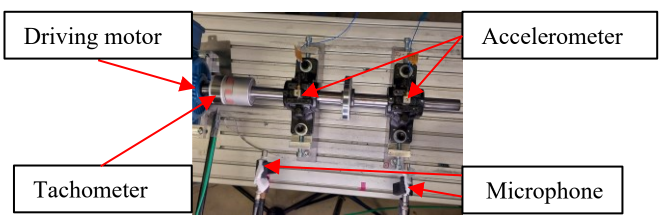


Figure 2. The bearing test rig in LMSD section of KU Leuven

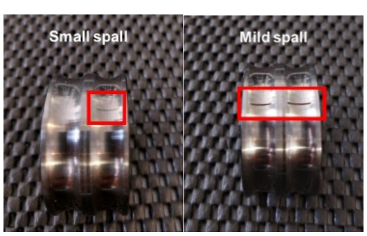


Figure 3. Fault classes: a small spall fault in the inner race and a mild spall fault in the inner race

The dataset was collected under two load working conditions, a balanced load condition and an unbalanced load condition. The balanced load condition corresponds to the use of the balanced disk. The unbalanced one corresponds to the mounting of an eccentric mass on the disk (one bolt) which causes an unbalanced load. In addition, signals are collected under six constant speed conditions, which are 3, 5, 10, 20, 30 and 40 Hz, respectively. The duration of each signal is 15 s, and the sampling frequency is 16 kHz.

Dataset II: To achieve model transfer across machines, in addition to Dataset I, a public bearing dataset (Huang et al., 2018) is introduced in the paper. The test rig is shown in Figure 4. Two bearings are mounted on the shaft: the healthy bearing and the test bearing. Accelerometers are used for the signal collection. An encoder is used to measure the rotational speed of the bearing. The dataset includes four speed conditions (the speed increases, the speed decreases, the speed first increases and then decreases, and the speed first decreases and then increases). Under each speed condition, three repeated experiments were carried out. Three test bearings (a healthy bearing, a bearing with an inner race fault and a bearing with an outer race fault) are used to do the experiments. The duration of the signal is 10s. The sampling frequency is 200 kHz.

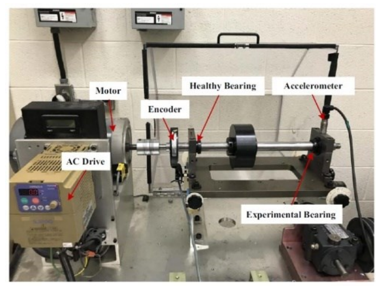


Figure 4. The bearing test rig (Huang et al., 2018)

4.2. Results and Discussion

This section validates the effectiveness of the proposed transfer learning method on the cross-load, cross-speed, and cross-machine model transfer tasks. Besides the proposed method, a CNN fault diagnosis model and a DANN transfer learning model used in (Peng et al., 2021) are considered as baseline methods.

The models are all written with Python 3.6 and the deep learning framework Pytorch and run on an Ubuntu 16.04 system with a GTX 2080 GPU. The accuracy is used to evaluate the network performance. It can be expressed as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

where TP, TN, FP, FN refer respectively to the number of true positive samples, true negative samples, false positive samples and false positive samples. The batch size of all models is 16, the epoch is 50, and the learning rate is 0.0001. Each sample is constructed with 1 s length vibrational signal for both two datasets.

4.2.1. Transfer Learning Among Load Conditions

The effectiveness of the proposed method is validated on cross-load model transfer tasks on Dataset I. 50% (681) of the samples captured under the unbalanced load condition are used for training and the rest (679) of the dataset is used for testing. Similarly, under the balanced load condition 50% (759) is used for training and the rest (757) is used for testing. The dataset from one load condition is considered as the source domain, and the dataset from the other load condition as the target domain. It is worth noting that regarding the baseline method CNN, a model is trained using the training set of the source domain, and then the well-trained model is directly applied to the test set of the target domain. For the DANN and the DAAN methods, a model is trained by the labeled training set in the source domain and the unlabeled training set in the target domain, and then the trained model is tested on the test set in the target domain.

The experimental results are shown in Table 1. Clearly, the proposed method exhibits the best transfer performance on both cross-load model transfer tasks among three methods. Specifically, when transferring from the unbalanced load to the balanced load, the accuracy of the proposed method reaches 99.33%, which is 8.49% and 7.29% higher than that of the CNN and the DANN, respectively. On another transfer learning task, the performance of the proposed method improves by 9.19% compared with that of DANN. Therefore, on the cross-load model transfer tasks, the proposed method shows a very competitive model transfer ability.

Table 1. The experimental results of three methods on TL among load conditions

Accuracy [%]	CNN	DANN	DAAN
Unbalanced load - Balanced load	90.84	92.04	**99.33**
Balanced load - Unbalanced load	88.77	89.80	**98.99**

4.2.2. Transfer Learning Among Speed Conditions

In this section, Dataset I is used to validate the effectiveness and the superiority of the proposed method on the model transfer ability among speed conditions. The dataset under one of the speed conditions is used as the source domain, and the dataset under another speed condition is used as the target domain. Similarly, in each speed condition, 50% (around 25) of the data is used for training and the remaining dataset (around 25) is for testing. The experimental results are shown in Table 2. Obviously, on each cross-speed model transfer task, the proposed method significantly improves the transfer performance of the network compared to the other two methods. In the last row of Table 2, the average diagnostic accuracy over all cross-speed transfer tasks for three methods is calculated. The average diagnostic performance of the proposed method reaches 80.20%, which is 23.81% higher than CNN and 21.49% higher than DANN. This indicates that the DAAN can learn the discriminative domain-invariant features, and thus can improve the model transfer ability among speed conditions.

4.2.3. Transfer Learning Across Machines

The effectiveness of the proposed method on the cross-machine model transfer task is validated based on Dataset I and Dataset II. One dataset is regarded as the source domain, and the other dataset as the target domain. There are 1516 samples for Dataset 1 in total and 1296 for Dataset 2. Similarly, 50% of the dataset is used for training, and the remaining 50% dataset is for testing.

Table 2. The experimental results of the three methods on TL among speed conditions

Speed	CNN [%]	DANN [%]	DAAN [%]
3-5	31.11	32.50	88.89
3-10	47.96	40.81	97.96
3-20	32.97	32.98	44.68
3-30	34.78	32.61	55.43
3-40	32.32	32.32	31.31
5-3	64.04	65.17	89.88
5-10	65.31	64.29	97.95
5-20	30.85	67.02	84.04
5-30	32.61	67.39	82.61
5-40	32.32	67.67	66.66
10-3	29.21	29.21	80.90
10-5	66.66	64.44	88.89
10-20	100.00	100.00	85.10
10-30	67.39	68.48	86.95
10-40	46.46	67.68	78.78
20-3	29.21	29.21	52.81
20-5	51.11	62.22	87.77
20-10	68.36	56.12	97.96
20-30	92.39	94.57	86.96
20-40	67.67	70.71	96.97
30-3	35.96	34.83	53.93
30-5	51.11	62.22	87.78
30-10	52.04	51.02	97.95
30-20	92.55	89.36	85.11
30-40	87.88	92.93	96.97
40-3	35.96	34.83	55.06
40-5	55.56	40.00	76.67
40-10	66.33	44.89	97.96
40-20	94.68	75.53	85.11
40-30	96.74	90.22	86.96
Mean	**56.39**	**58.71**	**80.20**

The cross-machine model transfer results of the three methods are shown in Table 3. The diagnosis results of both CNN and DANN are about 50%. Non of the two methods can successfully transfer the model across machines. To be honest, the model transfer across machines is a challenge, because there is a large difference in the feature distributions from different machines due to the following reasons. Firstly, the bearing types installed in different machines are inconsistent, so the dynamic properties of bearings are different. Secondly, the structures of machines are different. Lastly, the operating conditions and ambient environments are different. However, DAAN achieves 88.78% and 99.54% accuracy on two cross-machine transfer tasks, respectively, far exceeding the comparison methods. This indicates that the proposed method can learn discriminative domain-invariant features between two domains and successfully achieves model transfer across machines.

Table 3. The experimental results of the three methods on TL across machines

Accuracy [%]	CNN	DANN	DAAN
Huang - LVL	52.56	52.56	**88.78**
LVL - Huang	50.00	50.00	**99.54**

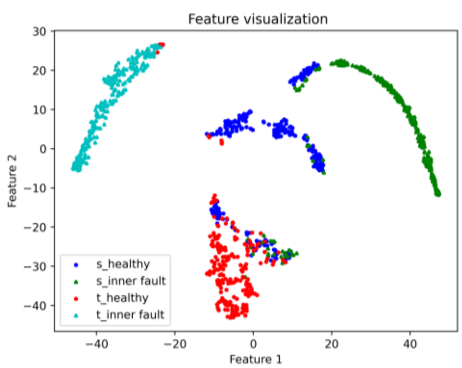


Figure 5. The feature visualization of DANN

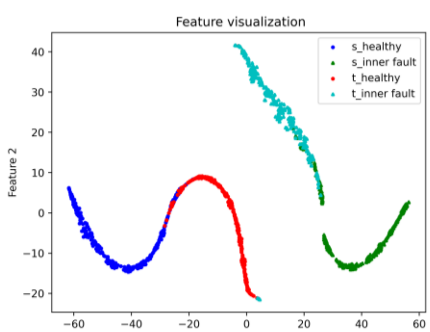


Figure 6. The feature visualization of DAAN

Figure 5 & 6 visualize the feature distribution of DANN and DAAN using *t*-SNE (Van der Maaten & Hinton, 2008). It can be clearly observed from Figure 5 that the features of the different classes are mixed together, and the features of the same class corresponding to the two domains are not mapped in the same feature space. However, regarding the DAAN method, the features of the healthy class (blue and red dots in Figure 6) of the source domain and the target domain are completely mapped to the same feature space, and the features of the inner race fault (green and cyan dots in Figure 6) of the source domain and the target domain are also completely mapped to the same feature space. Moreover, the features of the healthy class and the inner race fault class are completely separated, regardless of the source domain or the target domain. This further confirms the importance of the class-wise domain discriminator in DAAN.

5. Conclusion

This paper proposes a transfer learning-based bearing fault diagnosis method across machines. The method consists of two steps. Firstly, the vibration signal is transformed to an order-frequency map using the cyclic spectral correlation. This highlights the fault characteristic information of the measured signal. Subsequently, DAAN domain adaptation method is introduced to map the global feature distributions of two domains to a common feature space and also map the local feature distributions for each fault class to a common feature space, thereby improving the model transfer learning ability across machines. The proposed transfer learning-based fault diagnosis framework exhibits excellent diagnostic performance on cross-load, cross-speed, and cross-machine bearing fault diagnosis transfer tasks.

Acknowledgement

The authors would like to acknowledge the support of the China Scholarship Council, the Flemish Government under the “Onderzoeksprogramma Artificiële Intelligentie (AI) Vlaanderen” programme and the Research Foundation–Flanders (FWO) under the ROBUSTIFY research grant no. S006119N.

References

Lv, Y., Yuan, R., & Song, G. (2016). Multivariate empirical mode decomposition and its application to fault diagnosis of rolling bearing. *Mechanical Systems and Signal Processing*, 81, 219–234.

Randall, R. B. (2011). Vibration-based Condition Monitoring: Industrial, Aerospace and Automotive Applications, John Wiley & Sons, Ltd

Antoni, Jérôme, & Randall, R. B. (2006). The spectral kurtosis: application to the vibratory surveillance and diagnostics of rotating machines. *Mechanical Systems and Signal Processing*, 20(2), 308–331.

Antoni, J. (2007). Cyclic spectral analysis in practice. *Mechanical Systems and Signal Processing*, 21(2), 597–630.

Mauricio, A., Qi, J., Smith, W., Randall, R., & Gryllias, K.

(2018). Vibration based condition monitoring of planetary gearboxes operating under speed varying operating conditions based on cyclo-non-stationary analysis. *In International Conference on Rotor Dynamics*, 265–279.

Widodo, A., & Yang, B.-S. (2007). Support vector machine in machine condition monitoring and fault diagnosis. *Mechanical Systems and Signal Processing*, 21(6), 2560–2574.

Lu, J., Qian, W., Li, S., & Cui, R. (2021). Enhanced K-nearest neighbor for intelligent fault diagnosis of rotating machinery. *Applied Sciences*, 11(3), 919.

Zhang, N., Wu, L., Yang, J., & Guan, Y. (2018). Naive bayes bearing fault diagnosis based on enhanced independence of data. *Sensors*, 18(2), 463.

Voulodimos, A., Doulamis, N., Doulamis, A., & Protopapadakis, E. (2018). Deep learning for computer vision: A brief review. *Computational Intelligence and Neuroscience*.

Hesamian, M. H., Jia, W., He, X., & Kennedy, P. (2019). Deep learning techniques for medical image segmentation: achievements and challenges. *Journal of Digital Imaging*, 32(4), 582–596.

Nassif, A. B., Shahin, I., Attili, I., Azzeh, M., & Shaalan, K. (2019). Speech recognition using deep neural networks: A systematic review. *IEEE Access*, 7, 19143–19165.

Wang, H., Liu, Z., Peng, D., & Qin, Y. (2019). Understanding and learning discriminant features based on multiattention 1DCNN for wheelset bearing fault diagnosis. *IEEE Transactions on Industrial Informatics*, 16(9), 5735–5745.

Chen, Z., Ricardo Mauricio, A.M., Li, W., & Gryllias, K. (2019). Multi-label fault diagnosis based on convolutional neural network and cyclic spectral coherence. *Book of Proceedings Survishno*, 740–740.

Chen, X., Zhang, B., & Gao, D. (2021). Bearing fault diagnosis base on multi-scale CNN and LSTM model. *Journal of Intelligent Manufacturing*, 32(4), 971–987.

Long, M., Zhu, H., Wang, J., & Jordan, M.I. (2017). Deep transfer learning with joint adaptation networks. *In International conference on machine learning*, 2208–2217.

Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., & Bharath, A. A. (2018). Generative adversarial networks: An overview. IEEE Signal Processing Magazine, 35(1), 53–65.

Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M., & Lempitsky, V. (2016). Domain-adversarial training of neural networks. *The Journal of Machine Learning Research*, 17(1), 2030–2096.

Guo, L., Lei, Y., Xing, S., Yan, T., & Li, N. (2018). Deep convolutional transfer learning network: A new method for intelligent fault diagnosis of machines with unlabeled data. *IEEE Transactions on Industrial Electronics*, 66(9), 7316–7325.

Mao, W., Liu, Y., Ding, L., Safian, A. and Liang, X., 2020. A new structured domain adversarial neural network for transfer fault diagnosis of rolling bearings under different working conditions. *IEEE Transactions on Instrumentation and Measurement*, 70, 1–13.

Guo, L., Yu, Y., Liu, Y., Gao, H., & Chen, T. (2021). Reconstruction domain adaptation transfer network for partial transfer learning of machinery fault diagnostics. *IEEE Transactions on Instrumentation and Measurement*, 71, 1–10.

Yu, C., Wang, J., Chen, Y., & Huang, M. (2019). Transfer learning with dynamic adversarial adaptation network. *Proceedings - IEEE International Conference on Data Mining, ICDM*, 778–786.

Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. *Icml*.

Targ, S., Almeida, D., & Lyman, K. (2016). Resnet in resnet: Generalizing residual architectures. ArXiv Preprint ArXiv:1603.08029.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. ArXiv Preprint ArXiv:1412.6980.

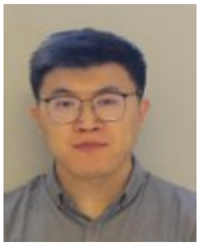
Huang, H., & Baddour, N. (2018). Bearing vibration data collected under time-varying rotational speed conditions. *Data in Brief*, 21, 1745–1749. https://doi.org/10.1016/j.dib.2018.11.019.

Peng, D., Liu, C., Desmet, W., & Gryllias, K. (2021). Deep Unsupervised Transfer Learning for Health Status Prediction of a Fleet of Wind Turbines with Unbalanced Data. *Annual Conference of the PHM Society*, 13(1).

Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9(11).

BIOGRAPHIES

Dandan Peng received the B.S. and M.S. degrees in mechanical engineering from the School of Mechanical and Electrical Engineering, University of Electronic Science and Technology of China, Chengdu, China, in 2016 and 2019, respectively. She is currently working toward the Ph.D. degree in mechanical engineering from Catholic University of Leuven, Leuven, Belgium. Her research interests include Hilbert–Huang transform, convolutional neural network, machinery condition monitoring, and fault diagnosis.



Chenyu Liu received the B.S. and M.S. degrees in aircraft manufacturing engineering and solid mechanics both from Northwestern Polytechnical University, China. He joined the noise and Vibration Research Group in the Department of Mechanical Engineering, KU Leuven, Belgium as a PhD researcher in 2017. His research interests include data driven based condition monitoring, rotating machinery prognostics, and machine learning and deep learning application.

Alexandre Mauricio received his MSc. degree in Mechanical Engineering from the Faculty of Engineering of the University of Porto (FEUP) in 2017. He is currently a PhD student at the Department of Mechanical Engineering of KU Leuven, Belgium. His main research interests focus on the diagnostics of rotating machinery operating under time varying conditions.

Wim Desmet holds a full professor position on vibro-acoustics of machines and transportation systems at the Department of Mechanical Engineering of KU Leuven, Belgium. He is also the head of noise & vibration research group. His research interests lie in the fields of numerical and experimental vibro-acoustics, uncertainty modelling of dynamic systems, aeroacoustics, active noise and vibration control, noise control engineering, multibody dynamics, dynamics of lightweight materials and systems, vehicle mechatronics and virtual prototyping.

Konstantinos Gryllias holds a 5 years engineering diploma degree and a PhD degree in Mechanical Engineering from National Technical University of Athens, Greece. He holds an associate professor position on vibro–acoustics of machines and transportation systems at the Department of Mechanical Engineering of KU Leuven, Belgium. He is also the manager of the University Core Lab Dynamics of Mechanical & Mechatronic Systems DMMS-M of Flanders Make, Belgium. His research interests lie in the fields of condition monitoring, signal processing, prognostics and health management of mech. & mechatronic systems.