Journal of Computers Vol. 33 No. 3, June 2022, pp. 135-146
doi: 10.53106/199115992022063303011

Fault Diagnosis under Varying Working Conditions with Domain Adversarial Capsule Networks

Feiyan Fan, Jiazhen Hou, Tanghuai Fan*

School of Information Engineering, Nanchang Institute of Technology, Nanchang, 330099, China
fantanghuai@nit.edu.cn,

Received 11 August 2021; Revised 13 December 2021; Accepted 13 January 2022

Abstract. Most existing studies that develop fault diagnosis methods focus on performance under steady operation while overlooking adaptability under varying working conditions. This results in the low generalization of the fault diagnosis methods. In this study, a novel deep transfer learning architecture is proposed for fault diagnosis under varying working conditions. A modified capsule network is developed by combining the domain adversarial framework and classical capsule network to simultaneously recognize the machinery fault and working conditions. The novelty of the proposed architecture mainly lies in the integration of the domain adversarial mechanism and capsule network. The idea of the domain adversarial mechanism is exploited in transfer learning, which can achieve a promising performance in cross-condition fault diagnosis tasks. With the novel architecture, learned features exhibit identical or very similar distributions in the source and target domains. Hence, the deep learning architecture trained in one working condition can be applicable to discriminative conditions without being hindered by the shift between the two domains. The proposed method is applied to analyze vibrations of a bearing system acquired under different working conditions, i.e., loads and rolling speed. The experimental results indicate that the proposed method outperforms other state-of-the-art methods in fault diagnosis under varying working conditions.

Keywords: rolling bearing, fault diagnosis, domain adversarial capsule network, domain adaptability

1 Introduction

Bearing machinery is the foundational element in power, manufacturing, and transportation infrastructures [1]. Intelligent fault diagnosis of bearing machinery has been widely researched in recent years. Various machine learning methods have been used for fault diagnoses of bearing machinery and have shown promising performance in automatic diagnosis, such as frequency analysis, time-frequency analysis, and statistic models. Recently, deep learning architectures have attracted attention from the research community because these architectures, in contrast to shallow models, exhibit better representability of faults [2]. Extremely high accuracy scores are generated by state-of-the-art deep learning methods, such as the deep belief network (DBN) [3], stacked auto-encoder (SAE) [4], convolutional neural network (CNN) [5], and long short-term memory (LSTM) [6]. However, an important issue for existing deep learning architectures is the over-fitting problem that can result in the inadaptability of fault diagnosis methods such that sophisticated models can hardly adapt to new conditions. Hence, it is necessary to train models to the specific working conditions on hand.

In recent years, capsule networks have been used for fault diagnosis [7]. In contrast to the traditional CNNs, capsule networks are equipped with a dynamic routing mechanism [8] and measure similarity between hierarchy capsules to estimate connection weights. Hence, capsule networks can classify faults using the structural relationship between features and provide improved results for fault diagnosis. However, problems persist for current capsule network–based fault diagnosis, such as model adaptability. The accuracy of the capsule network is significantly degenerated when machinery condition changes. An intuitive solution for handling the issue involves collecting samples under all possible working conditions; however, this is an excessively expensive method. Theoretically, discriminative fault representations of one fault category under various working conditions is the basic reason for the degenerated adaptability of fault diagnosis.

Considering the above factors, it is important to improve the adaptability of fault diagnosis; this is the key research problem of the present study. Domain adversarial strategy is a popular technique for enhancing model adaptability for fault diagnosis [9]. It simultaneously classifies the fault and the corresponding domain via a double-flow architecture. In general, the domain adversarial strategy can be realized in almost all classical deep learning architectures. A maximum mean difference (MMD) measurement block is inserted in the original architecture

* Corresponding Author

to decrease domain divergence between the source and target domains. Many classical deep learning architectures have been updated with the MMD measurement block. For example, modified transfer CNNs have been evaluated in cross-load fault recognition [10]. In this study, aiming to solve adaptability degeneration in fault diagnosis, the advantages of the domain adversarial strategy and capsule network are jointly exploited to achieve transfer learning by the domain adversarial capsule network (DACN) architecture. Specifically, the proposed method can realize multisource/multitarget-domain transfer tasks. Furthermore, the decoder in the original capsule network is updated with the gradient reversal layer, and the original reconstruction process is replaced with the domain classification process. In contrast to current architectures for fault diagnosis, the novelty of this study can be explained through two points. In terms of methodology, our proposed architecture integrates the domain adversarial mechanism and capsule network architecture, which achieves a new transfer learning architecture for diagnosing faults. In practice, our proposed model can correctly diagnose faults under changing working conditions without model retraining.

The contributions of our proposed DACN are as follows:

1) A two-flow hierarchical architecture is established to extract and correlate high-representative features between the source and target domains. This architecture can technically achieve the transfer learning between domains.

2) The capsule network and domain adversarial concepts are integrated to accurately and adaptively recognize faults in various working conditions. This integrated architecture can serve fault diagnosis in varying working conditions, significantly improving the generalization of the fault diagnosis technology.

3) Experiments are performed on three fault datasets, and the results demonstrate that the proposed method can realize accuracies of 98.46%, 96.71%, and 93.15% on these datasets, respectively. These experimental results comprehensively demonstrate the promising performance of our DACN in fault diagnosis under diverse working conditions.

The remainder of this paper is organized as follows. In Section 2, the brief theory of the classical capsule network and domain-adversarial training is reviewed. Our proposed architecture is described in Section 3. In Section 4, we present our experimental analysis. Finally, the conclusions and summary of the study and directions for future studies are presented in Section 5.

2 Related Studies

2.1 Fault Diagnosis

Owing to the functional similarity between fault diagnosis and image classification, various deep learning architectures have been introduced for classifying faults. Fourier transform and wavelet transform have been used to extract two-dimensional spectral maps such that mature image classification architectures can be transplanted into the study of fault diagnosis. For example, the CNN and capsule networks have been exploited for fault diagnoses, demonstrating excellent performance [11].

To realize end-to-end fault diagnoses, one-dimensional architectures have been recently developed to automatically learn one-dimensional features from raw vibration signals. Hence, fault diagnoses are independent of hand-crafted efforts [12]. An important common deep-learning architecture is the deep-belief network (DBN), which extracts highly abstract representations of faults using multiple layers of restricted Boltzmann machines [13]. Additionally, various hierarchical deep architectures have been developed by combining pairs of deep learning blocks with specific functions. Typical examples include combinations of stacked autoencoders and DBN blocks [13] and squeeze-and-excitation and CNN blocks [14].

Recently, multi-stream architectures have attracted increasing research attention because they can provide more evidence for fault classification in contrast to a single-stream flow. For example, frequency and time-frequency features were jointly input into deep learning architectures and fused at the deepest layers [15]. Other multi-stream architectures focused on the multi-scale factor of the raw signal and features with different temporal scales were extracted in multiple flows [16].

In general, the state-of-the art deep learning architectures have been highly successful in the field of fault diagnosis. However, most fault diagnosis models are evaluated under a single working condition such that training and testing samples are collected under one machinery load and rolling speed. When training and testing samples are taken from diverse working conditions, the accuracy of fault diagnosis is significantly degenerated. Hence, a more adaptive deep learning-based method is desired for fault diagnosis.

2.2 Capsule Network

Theoretically, the capsule network was developed to solve the issue caused by the max pooling operation, which allows neurons in one layer to ignore all other neurons with the exception of the most active neurons in a local pool in the previous layer [17]. Hence, the correlation factor between features is missed for pattern recognition, thereby resulting in performance degeneration. To address the issue, a dynamic routing mechanism is designed to update the connection weights with the similarity measurement of layerwise capsules [8]. The capsule is the basic element for feature representation; it is a vector whose length and orientation denote the probability of the existence of the entity and characteristics of the entity, respectively. Each dimension of the capsule vector presents a characteristic of the feature, and a high-dimensional vector can present features from multiple aspects. To classify the capsule vectors efficiently, the "squashing" function is used during the model training to normalize the vector length from 0 to 1. A capsule network is established with layerwise capsules. Lower layers include more capsules with low dimensions, whereas higher layers are constituted by a smaller number of high-dimensional capsules. For the low-layer capsule, the spatial characteristic of a small area is presented by capture vectors. Conversely, a larger area is presented by capsules at higher layers to provide a better representation to the samples. Therefore, an increase in dimensionality is observed from low-layer capsules to high-layer capsules.

The dimensional increment of layers determines the relation between capsules and the architecture of the capsule networks. The layerwise capsule mapping is realized by weighted summation and weights are updated via dynamic routing. In addition to featuring extraction layers at the top, the capsule network includes three other components:

1) **Primary capsule layer:** This is constituted to partition the feature maps from primary convolution layers into capsules of size $6 \times 6 \times 32 \times 16$.

2) **Digital capsule layer:** In this layer, "squashing" and dynamic routing are applied to capsules. This layer has two branches, one for classification and the other for graph reconstruction.

3) **Decoder:** In this, all capsules are reshaped to one-dimensional vectors, which are then input into the fully connected layers to complete the classification.

2.3 Domain Adversarial Network

The development of the domain adversarial networks is driven by the domain adaptation issue, which transfers knowledge from the labeled source domain to the unlabeled target domain. The data distribution in the source domain and target domain presents a domain shift; the objective of domain adaptation is to reduce this feature distribution shift across domains. Domain adversarial networks are generally proposed in the field of image processing [18]. Subsequently, various domain adversarial networks have been developed for intelligent fault diagnosis to solve the problem caused by the discrepancy between different working conditions. For example, deep neural networks have been combined with MMD to learn the frequency spectrum acquired from different working loads [19]. The MMD measurement has been combined with the generative model to improve the performance of the cross-domain diagnosis. Another strategy for realizing domain adaptability is based on generative adversarial networks (GANs) [20]. The Wasserstein distance-guided multi-adversarial networks have been developed to handle the cross-domain issue [21]. Furthermore, classic CNN architectures were updated with the GAN framework to improve diagnosis adaptability. In addition to the study on fault diagnosis under varying working conditions, a recent study attempted to solve another more difficult issue with respect to the task across machinery. In this area, deep convolution architectures were updated by the domain adversarial strategy.

2.4 Motivations and Novelties

Representability and adaptability are the two most important points in the performance evaluation of fault diagnosis methods. However, these points are to some extent mutually contradictory. A high representability is generated by a sufficient or overt fitting effect on the fault feature distribution, which in turn significantly decreases the generalization of the diagnosis methods. Alternatively, adaptability improvement can decrease the representability of the deep learning features, thereby resulting in the degeneration of the diagnosis performance. The method proposed in this study is intended to determine a balance between the representability and adaptability of fault diagnosis.

Comparing with current fault diagnosis methods, the methodological novelty of the proposed method mainly lies in the integration of the domain adversarial framework and capsule network, such that a novel DACN is established. In terms of functionality, our proposed DACN can achieve the transfer learning to diagnose cross-condition faults with single-condition training data alone. Specifically, the classical capsule network is modified such that the graph reconstruction branch in the original architecture is replaced with domain identification modules to label the target domain.

3 Proposed Fault Diagnosis Scheme

3.1 Adversarial Capsule Networks

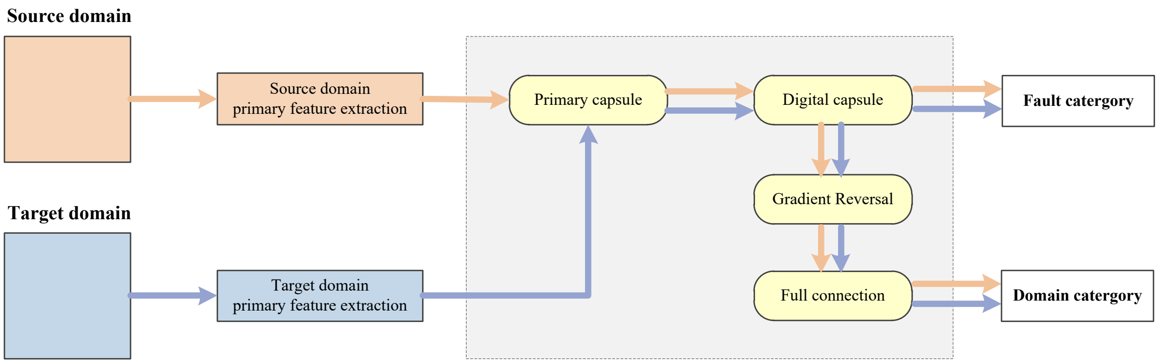


Fig. 1. Domain adversarial capsule network (DACN)

The aim of this study is to realize a good performance in fault representability and diagnosis adaptability via modifying the classical capsule network wherein the domain adversarial concept is introduced to generate a novel adversarial capsule network. The framework of the adversarial capsule network is shown in Fig. 1. In contrast to other deep architectures, the advantages of the DACN can be explained through two points. First, a high-order vector feature of capsules can comprehensively represent the sample characteristic and improve the accuracy of the fault diagnosis. Second, the use of the gradient reversal calculation transforms the original reconstruction branch into the domain classification branch such that the domain shift is reduced in high-order vector features. The property enables our proposed method to perform the diagnosis task under varying working conditions without requiring model retraining.

Mathematically, the proposed DACN is presented as follows. The whole DACN includes three parts: a feature extractor *G*, domain classifier *D*, and fault classifier *C*. The model parameter set of these three components correspond to θ_G, θ_D , and θ_C. The optimization process can minimize the error of the *C* while maximizing the error of the *D* that indicates a minimum shift between the vector features across domains as follows:

$$\left(\hat{\theta}_G, \hat{\theta}_C\right) = \arg\min_{\theta_G, \theta_C} L_0\left(\theta_G, \hat{\theta}_D, \theta_C\right), \tag{1}$$

$$\hat{\theta}_D = \arg\max_{\theta_D} L_0\left(\hat{\theta}_G, \theta_D, \hat{\theta}_C\right), \tag{2}$$

where L_0 denotes the error measurement function that quantitatively presents the difference between the predicted label and true label, and $\hat{\theta}_G, \theta_D$ and θ_C denote optimization results of $\theta_G, \theta_D, \theta_C$, respectively. After the adversarial training process for the fault classifier and domain classifier, the representative features in the source and target domains exhibit the same or similar data distribution. Hence, the trained model can adapt to the unlabeled data in the target domain.

By assuming that the labeled data in the source domain is x^s and unlabeled data in the target domain is x^t , the vector features after multilevel mapping is extracted as $f_h^s = G_s(x^s)$ and $f_h^t = G_s(x^t)$. The maximum of the vector length $p_c = \text{squash}(\text{length}(f_h^{s,t})$ is used for the fault classification. Alternatively, the domain classification is realized via a gradient reversal layer, two levels of full connection, and the softmax function as follows:

$p_d = \text{softmax}(W(f_h^{s,t})$.

Based on optimization functions provided in Eqs. 1 and 2, the comprehensive margin loss function includes two terms with respect to the fault classification:

$$L_c = T_c \max(0, m^+ - p_c)^2 + \lambda(1-T_c)\max(0, p_c - m^-)^2 , \tag{3}$$

where c denotes the label; p_c denotes the activity level of the capsule c in the backend level; T_c denotes the output of the one-hot encoded label; λ denotes the non-negative hyperparameter, and m+ and m− denote boundaries; the number of capsules in the backend level should be identical to the number of categories. Specifically, our proposed adversarial capsule networks are trained with a margin loss of m+ = 0.9, m− = 0.1, and λ = 0.5.

The domain classification is a binary classification task, and thus the cross-entropy loss function is used to measure the classification loss as follows:

$$L_d = -\frac{1}{n_s}\sum_{i=1}^{n_s}\log\frac{e^{x^s_{d,i,1}}}{e^{x^s_{d,i,1}}+e^{x^s_{d,i,2}}} - \frac{1}{n_t}\sum_{i=1}^{n_t}\log\frac{e^{x^t_{d,i,2}}}{e^{x^t_{d,i,1}}+e^{x^t_{d,i,2}}} , \tag{4}$$

where $x^s_{d,i,1}$ and $x^s_{d,i,2}$ denote first and second dimensions of the feature vector of the ith sample in the source domain, respectively, and $x^t_{d,i,1}$ and $x^t_{d,i,2}$ denote their counterparts in the target domain, respectively.

To accelerate the model training process, a two-stage parameter optimization strategy is proposed. The first stage involves optimizing the fault classifier, whereas the second stage involves optimizing the domain classifier. The optimization of the fault classifier is as follows:

$$\hat{\theta}_{G_s} = \arg\min L_c(\theta_{G_s}, \hat{\theta}_{G_{cap}}, \hat{\theta}_C) ; \tag{5}$$

$$\hat{\theta}_{G_{cap}} = \arg\min L_c(\hat{\theta}_{G_s}, \theta_{G_{cap}}, \hat{\theta}_C) ; \tag{6}$$

$$\hat{\theta}_C = \arg\min L_c(\hat{\theta}_{G_s}, \hat{\theta}_{G_{cap}}, \theta_C) , \tag{7}$$

where $\hat{\theta}_{G_s}, \hat{\theta}_{G_{cap}}, \hat{\theta}_C$ denote optimization results of $\theta_{G_s}, \theta_{G_{cap}}, \theta_C$, respectively.

The second stage focuses on the optimization of the domain classifier. In a manner different from the fault classifier, the optimization result of the domain classifier involves decreasing the feature representation shift between the source and target domains:

$$\hat{\theta}_{G_t} = \arg\max L_d(\hat{\theta}_{G_s}, \theta_{G_t}, \hat{\theta}_{G_{cap}}, \hat{\theta}_D) ; \tag{8}$$

$$\hat{\theta}_D = \arg\min L_d(\hat{\theta}_{G_s}, \hat{\theta}_{G_t}, \hat{\theta}_{G_{cap}}, \theta_D) , \tag{9}$$

where $\hat{\theta}_{G_t}, \hat{\theta}_D$ denote optimization results of θ_{G_t}, θ_D , respectively.

3.2 DACN-based Fault Diagnosis Method

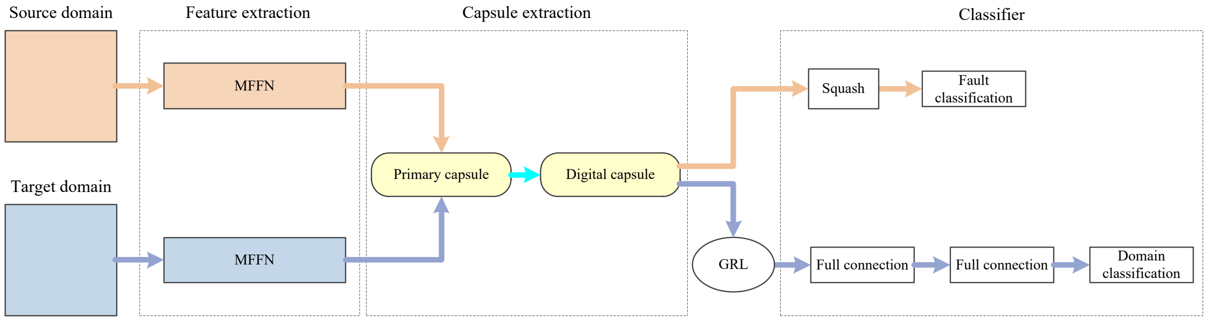


Fig. 2. DACN-based fault diagnosis

The entire architecture of the DACN-based fault diagnosis method is shown in Fig. 2. Our previously proposed multiple-feature fusion network (MFFN) [15] is employed to provide sufficient evidence from one-dimensional and two-dimensional features from the raw vibration signal. Thereafter, the fused features are further handled to extract the capsule vectors. These capsule vectors are processed by the gradient reversal layer and a carefully designed training process to remove the cross-domain shift.

4 Experimental Analyses

4.1 Setup

A. Model parameters

The parameters for our proposed model are listed in Table 1.

Table 1. Parameters of the proposed model

Module	Name	Size/Step/Number	Expansion	Parameter size	Output size
1D-CNN	Input_1	-	-	0	(None,4096,1)
	ResBlock_1	3/1/16	1	912	(None,4096,16)
	ResBlock_2	3/1/16	2	1568	(None,4096,16)
	ResBlock_3	3/1/4	4	441	(None,4096,4)
	Max_Pooling	2/1/-	-	0	(None,2048,4)
2D-CNN	Input_2	-	-	0	(None,128,128,3)
	Conv2D_1	30/5/256	-	691200	(None,20,20,256)
	Conv2D_2	6/2/256	-	2359552	(None,8,8,256)
	Inception_1	(1,2,3,4,5,6,7,8)/1/32	-	1672448	(None,8,8,256)
	Reshape_1	-	-	0	(None,2048,8)

SCE	Global_Average_Pooling_1	-	-	0	(None,4)
	Global_Average_Pooling_2	-	-	0	(None,8)
	Global_Average_Pooling_3	-	-	0	(None,4)
	Concatenate_1	-	-	0	(None,2048,16)
	Concatenate_2	-	-	0	(None,16)
	Dense_1	8/-/-	-	136	(None,8)
	Dense_2	16/-/-	-	144	(None,16)
	Multiply	-	-	0	(None,2048,16)
DACN	PrimaryCaps	-	-	0	(None,2048,16)
	DigitCaps	-	-	2621400 1048576 786432	(None,10,8), (None,4,8), (None,3,8)
	Lambda_1	-	-	0	(None,10), (None,4)(None,3)
	Dense_3	100/-/-	-	500	(None,100)
	Dense_4	2/-/-	-	202	(None,2)

B. Dataset

To evaluate the performance of the proposed DACN, experiments were performed on defective bearing datasets obtained from the Case Western Reserve University Bearing Data Center (Dataset A) [22], Jiangnan University (Dataset B) [23], and Paderborn University (Dataset C) [24]. The details of three datasets are shown in Table 2 to Table 4. The samples in the three databases are collected under varying working conditions, which support the performance evaluation under varying working conditions.

Table 2. Dataset A

Working condition	A1	A2	A3
Speed (rpm)	1772	1750	1730
Load (hp)	1	2	3
Fault category	10	10	10
Training sample	12000	12000	12000
Testing sample	4000	4000	4000

Table 3. Dataset B

Working condition	B1	B2	B3
Speed (rpm)	600	800	1000
Fault category	4	4	4
Training sample	4800	4800	4800
Testing sample	1600	1600	1600

Table 4. Dataset C

Working condition	C1	C2	C3
Speed (rpm)	1500	1500	1500
Load (hp)	0.7	0.1	0.7
Radial force (N)	1000	1000	400
Fault category	3	3	3
Training sample	3600	3600	3600
Testing sample	1200	1200	1200

4.2 Experimental Results

A. Experimental Results under Varying Working Conditions

The experimental results under varying working conditions are shown in Table 5 to Table 7. In the tables, the first row denotes the source domain, whereas the first column denotes the target domain. For example, the value at the second row and third column denotes the correctness provided by the model that is trained by the samples in the A2 domain and tested by the samples in the A1 domain. This can be abbreviated as "A2→A1."

Table 5. Performance under varying working conditions on dataset A

Working condition	A1	A2	A3
A1	99.85%	96.88%	97.90%
A2	99.47%	99.79%	99.51%
A3	98.23%	98.76%	100.0%

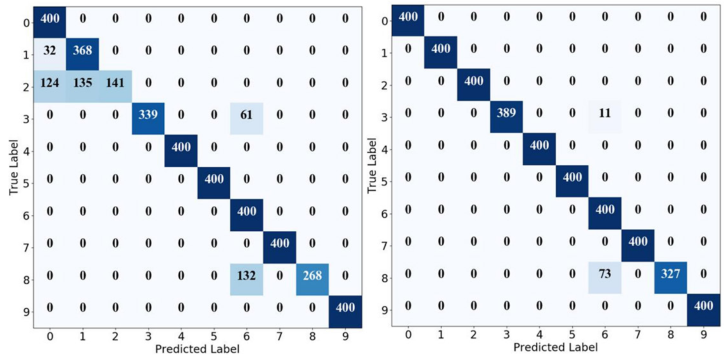
Table 6. Performance under varying working conditions on dataset B

Working condition	B1	B2	B3
B1	99.81%	94.16%	95.34%
B2	97.41%	100.0%	96.40%
B3	98.23%	98.72%	99.77%

Table 7. Performance under varying working conditions on dataset C

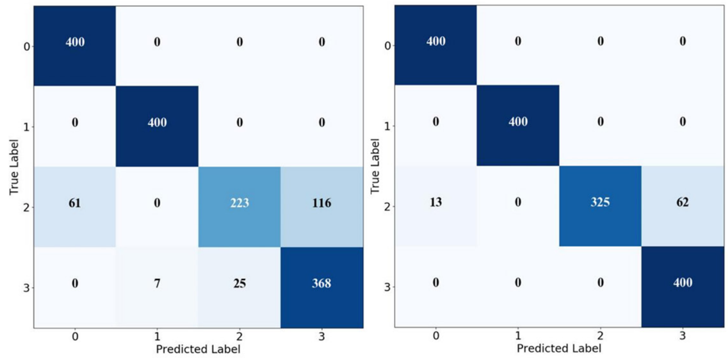
Working condition	C1	C2	C3
C1	99.97%	93.74%	92.48%
C2	92.69%	99.91%	93.12%
C3	91.66%	95.20%	99.89%

We can conclude the following from the aforementioned results. First, performance degeneration occurs for a task under varying working conditions in contrast to a task performed under a single-working condition. For example, the scores along the diagonal line are significantly larger than those of other elements, which is caused by the intra-class bias between working conditions. Second, although transfer learning is conducted by our DACN, we can find a significant performance variation across varying working conditions. An optimal correctness of 99.51% is generated in the case of "A3→A2," whereas the correctness decreases to 91.66% in the case of "C1→C3." This can be attributed to different degrees of the domain shift. Third, despite relative performance degeneration and variation, our method generates satisfying results because its correctness always exceeds 90% under various working conditions, which clearly demonstrates the contributions of the transfer learning mechanism in our DACN.



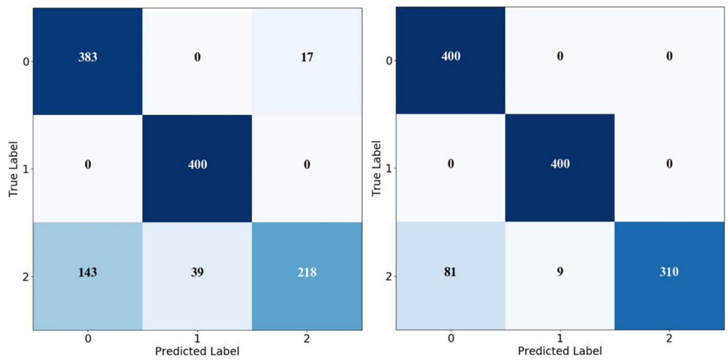
(a) (b)

Dataset A



(c) (d)

Dataset B



(e) (f)

Dataset C

Fig. 3. Confusion matrix of the ablation study: (a) MFFN on dataset A, (b) DACN on dataset A, (c) MFFN on dataset B, (d) DACN on dataset B, (e) MFFN on dataset C, and (f) DACN on dataset C

B. Ablation study

An ablation study was conducted by removing the domain adversarial module to experimentally compare the performance between MFFN and DACN. The experimental comparison results are shown in Fig. 3, where the confusion matrix is obtained from three datasets. The performance degeneration is evident after removing the domain adversarial module. For dataset A, 259 samples under working condition 2 are misclassified by MFFN, whereas all samples are correctly recognized by DACN on the same dataset. A similar phenomenon is noted for datasets B and C. In contrast to the single working condition, varying working conditions, such as different loads and speed, would generate discrepant but correlated patterns. The domain adversarial module in DACN establishes the correlation across working conditions, such that transfer learning can be achieved by DACN to transfer valuable clues from a domain/working condition to other domain/working conditions. This is a desirable property for fault diagnosis, as it implies a possible strategy for fault diagnosis under small datasets.

4.3 Experimental Comparison

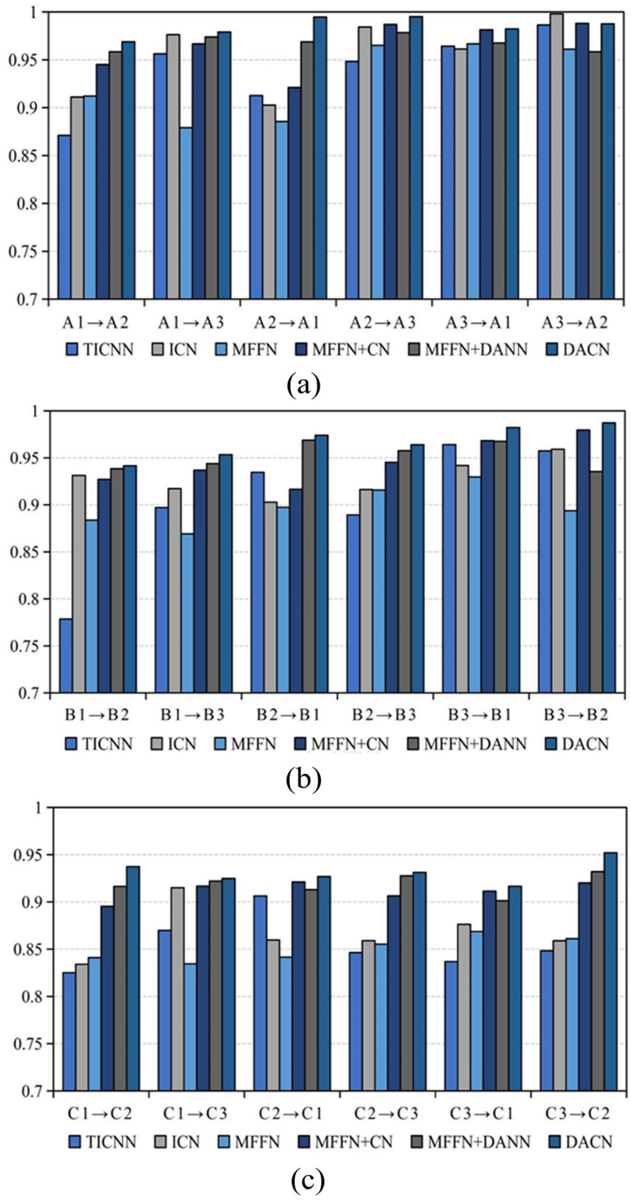


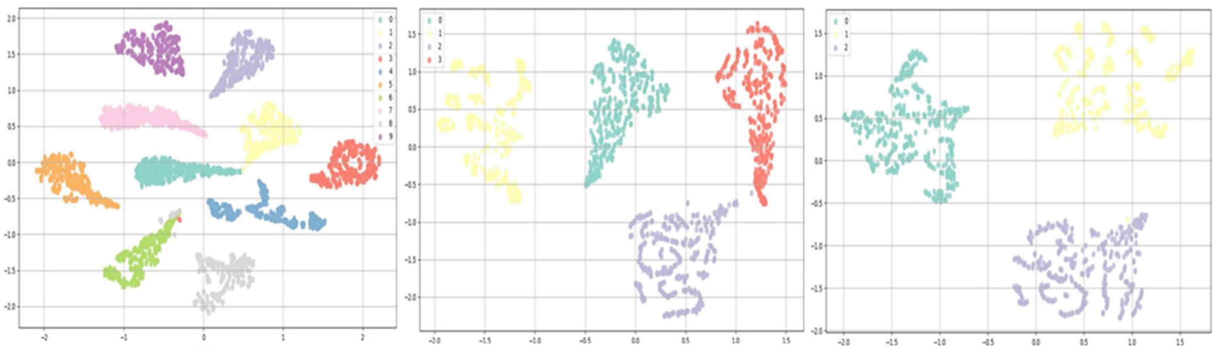
Fig. 4. Experimental comparisons on the (a) dataset A, (b) dataset B, and (c) dataset C

Three state-of-the-art methods, the temporal information convolutional neural network (TICNN) [25], inception CapsNet (ICN) [26], and MFFN [15], are included for experimental comparisons. Additionally, typical hierarchical frameworks, such as MFFN+CN (CapsNet) and MFFN+DANN (adversarial domain neural network), are included. Fig. 4 presents comparison results. The correctness of our proposed DACN always exceeds 96%, which is a satisfying performance for fault diagnosis. Conversely, DACN outperforms other methods in terms of average correctness although it displays the second-best average correctness under varying working conditions in the case of "A3→A2." Moreover, the average correctness of compared methods and DACN under varying working conditions is shown in Table. 8. The correctness of MFFN is 92.83%, which is lower than those of TICNN and ICN. Alternatively, when a DACN is added to establish DACN, the fault diagnosis correctness significantly increases to be the best. These results clearly demonstrated the contributions of the domain adversarial module, which can achieve the pattern transferring between different domains, exploiting more valuable clues for fault diagnosis.

Table 8. Comparison of average correctness under varying working conditions

Model	Average correctness		
	A	B	C
TICNN	93.99%	90.35%	85.55%
ICN	95.56%	92.81%	86.72%
MFFN	92.83%	89.82%	85.03%
MFFN+CN	96.49%	94.56%	91.19%
MFFN+DANN	96.75%	95.20%	91.87%
DACN	98.46%	96.71%	93.15%

4.4 Visualization



(a) (b) (c)

Fig. 5. Feature visualization with the learned feature on (a) dataset A, (b) dataset B, and (c) Dataset C

To better understand the benefits of the proposed DACN, the t-SNE technique was applied to decrease the dimensionality of the learned features into two dimensions for map generation. The two-dimensional (2D) feature maps are shown in Fig. 5, where different colors denote different faults or normal categories. After DACN feature learning, a clustering effect of a fault category is observed under varying working conditions. From the feature maps, clear margins can be seen between fault categories, which enables simple classification behavior. A shallow backend architecture is sufficient for fault classification.

5 Conclusion

We proposed a novel DACN–based fault diagnosis method for diagnosing faults under varying working conditions. The novelty of our proposed DACN lies in that it integrates the domain adversarial mechanism and capsule network to achieve transfer learning for fault diagnosis under varying working conditions. In terms of functionality, the major contribution of our proposed DACN is that it can decrease the domain shift between working conditions via domain adversarial training. Theoretically, consistency under varying domains reflects the type of invariance that approaches intrinsic attributes of faults. Our proposed DACN can determine a good tradeoff between feature representability and adaptability. This acts as a foundation for better fault diagnosis results of DACN in contrast to previous architectures.

Different patterns of faults under varying conditions pose challenges for fault diagnosis. This study intends to address the specific issues caused by changing working conditions, such as loads and speed. Another significant issue in the current study is the fault diagnosis under different machinery, i.e., the model is trained by the data on one machinery and evaluated on another machinery. This problem is not considered in this study and motivates our future work. Moreover, a potential limitation of DACN is its large model scale, which may reduce the model feasibility in practice, especially the online applications. Hence, we will try to reduce the model scale of DACN to make it more efficient in our future works.

6 Acknowledgement

This study is supported by the Key Research and Development Program of Jiangxi Province, China (No. 20192BBE50076; 20203BBGL73225) and the research of face image recognition based on deep learning (GJJ211921).

References

[1] S. Zhang, S. Zhang, B. Wang, T.G. Habetler, Deep learning algorithms for bearing fault Diagnostics-A comprehensive review, IEEE Access 8(2020) 29857-29881.
[2] J. Schmidhuber, Deep learning in neural networks: An overview, Neural Networks 61(2015) 85-117.
[3] W. Huang, G. Song, H. Hong, K. Xie, Deep architecture for traffic flow prediction: deep belief networks with multitask learning, IEEE Transactions on Intelligent Transportation Systems 15(5)(2014) 2191-2201.
[4] W. Sun, S. Shao, R. Zhao, R. Yan, X. Zhang, X. Chen, A sparse auto-encoder-based deep neural network approach for induction motor faults classification, Measurement 89(2016) 171-178.
[5] S. Tang, S. Yuan, Y. Zhu, Convolutional Neural Network in Intelligent Fault Diagnosis Toward Rotatory Machinery, IEEE Access 8(2020) 86510-86519.
[6] J. Lei, C. Liu, D. Jiang, Fault diagnosis of wind turbine based on Long Short-term memory networks, Renewable Energy 133(2019) 422-432.
[7] L. Chen, N. Qin, X. Dai, D. Huang, Fault Diagnosis of High-Speed Train Bogie Based on Capsule Network, IEEE Transactions on Instrumentation and Measurement 69(9)(2020) 6203-6211.
[8] S. Sabour, N. Frosst, G.E. Hinton, Dynamic routing between capsules, in: Proc. Advances in neural information processing systems, 2017.
[9] F. Li, T. Tang, B. Tang, Q. He, Deep Convolution Domain-adversarial Transfer Learning for Fault Diagnosis of Rolling Bearings, Measurement 169(2021) 108339.
[10]A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, in: Proc. Advances in Neural Information Processing Systems 25, 2012.
[11]Z.J. Wang, L.K. Zheng, W.H. Du, W.N. Cai, J. Zhou, J.T. Wang, X.F. Han, G.F. He, A Novel Method for Intelligent Fault Diagnosis of Bearing Based on Capsule Neural Network, Complexity (2019) 6943234.
[12]T. Ince, S. Kiranyaz, L. Eren, M. Askar, M. Gabbouj, Real-Time Motor Fault Detection by 1-D Convolutional Neural Networks, IEEE Transactions on Industrial Electronics 63(11)(2016) 7067-7075.
[13]Z. Chen, W. Li, Multisensor Feature Fusion for Bearing Fault Diagnosis Using Sparse Autoencoder and Deep Belief Network, IEEE Transactions on Instrumentation and Measurement 66(7)(2017) 1693-1702.
[14]H. Wang, J. Xu, R. Yan, R.X. Gao, A New Intelligent Bearing Fault Diagnosis Method Using SDP Representation and SE-CNN, IEEE Transactions on Instrumentation and Measurement 69(5)(2020) 2377-2389.
[15]Z. Chen, H. Lu, S. Tian, J. Qiu, T. Kamiya, S. Serikawa, L. Xu, Construction of a Hierarchical Feature Enhancement Network and Its Application in Fault Recognition, IEEE Transactions on Industrial Informatics 17(7)(2021) 4827-4836.
[16]G. Jiang, H. He, J. Yan, P. Xie, Multiscale Convolutional Neural Networks for Fault Diagnosis of Wind Turbine Gearbox, IEEE Transactions on Industrial Electronics 66(4)(2019) 3196-3207.
[17]M.E. Paoletti, J.M. Haut, R. Fernandez-Beltran, J. Plaza, A. Plaza, J. Li, F. Pla, Capsule Networks for Hyperspectral Image Classification, IEEE Transactions on Geoscience and Remote Sensing 57(4)(2019) 2145-2160.
[18]Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, V. Lempitsky, Domain-adversarial training of neural networks, The Journal of Machine Learning Research 17(1)(2016) 2030-2096.
[19]C.C. Che, H.W. Wang, X.M. Ni, Q. Fu, Domain adaptive deep belief network for rolling bearing fault diagnosis, Computers & Industrial Engineering 143(2020) 106427.1-106427.9.
[20]H. Alqahtani, M. Kavakli-Thorne, G. Kumar, Applications of Generative Adversarial Networks (GANs): An Updated Review, Archives of Computational Methods in Engineering 28(2)(2021) 525-552.
[21]C. Cheng, B.T. Zhou, G.J. Ma, D. Wu, Y. Yuan, Wasserstein distance based deep adversarial transfer learning for intelligent fault diagnosis with unlabeled or insufficient labeled data, Neurocomputing 409(2020) 35-45.
[22]Case Western Reserve University Bearing Data Centre Website. <http://csegroups.case.edu/bearingdatacenter/pages/download-data-file.> (accessed 17.10.08)
[23]K. Li, Jiangnan University Bearing Dataset. <http://mad-net.org:8765/explore.html.> (accessed 18.09.15)
[24]Paderborn University Website. <https://mb.uni-paderborn.de/en/kat/main-research/datacenter/bearing-datacenter/data-sets-and-download> (accessed 20.12.01)
[25]W. Zhang, C.H. Li, G.L. Peng, Y.H. Chen, Z.J. 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 100(2018) 439-453.
[26]Z.Y. Zhu, G.L. Peng, Y.H. Chen, H.J. Gao, A convolutional neural network based on a capsule network with strong generalization for bearing fault diagnosis, Neurocomputing 323(2019) 62-75.