DOI: 10.19884/j.1672-5220.202307002

Bearing Fault Diagnosis Model Based on Multi-Level Domain Adaption Network

LI Wenwen*, CHEN Guangfeng
College of Mechanical Engineering, Donghua University, Shanghai 201620, China

Abstract: The complex and changeable environment in the process of bearing operation may lead to inconsistent distribution of training data and test data, and decrease the diagnosis performance of the model. Thus a bearing fault diagnosis model based on the Shuffle-CANet is proposed, and realizes bearing cross-domain fault diagnosis by improving the ShuffleNet V2 and introducing asymmetric convolution. A domain loss function is added to the model based on the idea of domain adaptation in transfer learning so that the common features of the source domain and the target domain can be extracted occasionally and the cross-domain fault diagnosis can be realized. Compared with the traditional deep learning model, this model is friendlier to mobile and embedded devices. The Shuffle-CANet is validated by different transfer tasks on two different datasets. The results show that when the source domain and the target domain are derived from the same dataset, the fault diagnostic accuracy of the model can be more than 99%. When the target domain and the source domain are derived from different datasets, the fault diagnostic accuracy of the model can be more than 95%.

Key words: bearing fault diagnosis; ShuffleNet V2; multi-level domain adaption; lightweight
CLC number: TH277; TH165.3 **Document code**: A
Article ID: 1672-5220(2024)02-0162-10

Open Science Identity (OSID)

0 Introduction

Bearing is one of the most important parts of rotating machinery[1], and its health directly affects the running state of mechanical equipment. Therefore, the fault diagnosis technology of the bearing is very important to prevent mechanical system accidents.

Wang et al.[2] proposed a bearing fault diagnosis method with the improved cepstrum analysis. Ensemble empirical mode decomposition (EEMD) was used to decompose the vibration signal and perform the cepstrum analysis on the decomposed intrinsic mode function (IMF). Yang[3] proposed a bearing fault diagnosis method based on the wavelet spectrum analysis, and used the method of wavelet decomposition and reconstruction to diagnose the bearing fault. However, the above traditional spectrum analysis methods rely on manual experience.

Saravanan et al.[4] used discrete wavelet transform to extract features, and then classified the extracted features by an artificial neural network. Li et al.[5] proposed a bearing fault diagnosis method based on multi-scale permutation entropy and an improved support vector machine binary tree. The feature of bearing vibration signal was extracted by multi-scale permutation entropy, and then the fault identification was completed by the improved support vector machine binary tree. However, these methods assume that the distribution of training data is the same as that of test data.

In the process of bearing operation, the complex working conditions may lead to inconsistent distribution of vibration data. In this case, the traditional machine learning method has limitations. However, transfer learning can apply the knowledge learned from the source domain data to the relevant domain data[6].

Hoang et al.[7] proposed a bearing fault diagnosis method based on transfer learning. The bearing fault diagnosis across working conditions was realized by fixing the parameters of the front feature extraction layer, and then fine-tuning the parameters of the later feature extraction layer and classifier. Wang et al.[8] proposed a fault diagnosis method based on deep adversarial domain adaptation. This method employed the idea of confrontation, and made the feature extractor and the domain classifier form confrontation to eliminate the data distribution difference between the source domain and the target domain and realize cross-domain fault diagnosis of bearings. Hu et al.[9] proposed a bearing fault diagnosis method based on twin domain transfer learning. The method minimized the same input distance by sharing the weight of the feature extraction layer. The method of maximizing different input distances was used to realize cross-domain fault diagnosis of bearings. The fault diagnosis of bearings at various rotational speeds was realized. However, when adversarial training was added, the model converged slowly.

Chen et al.[10] proposed a bearing fault diagnosis method based on transfer learning of deep attention mechanism at different rotational speeds. The maximum mean difference (MMD) was used to approximately

Received date: 2023-07-12
* Correspondence should be addressed to LI Wenwen, email: 1744367957@qq.com
Citation: LI W W, CHEN G F. Bearing fault diagnosis model based on multi-level domain adaption network[J]. *Journal of Donghua University (English Edition)*, 2024, 41(2): 162-171.

measure and adapt the characteristics of source domain data and target domain data by the method of domain adaptation to realize the bearing fault diagnosis at different rotational speeds. Huo et al.[11] proposed an enhanced transfer learning method for the bearing fault diagnosis based on a linear superposition network. The method used the multi-core MMD to measure and reduce the feature distribution of the two domains generated by the network, and paid more attention to the data characteristics of the target domain in the process of transfer training to realize the off-condition transfer learning of the bearing. However, for the problem of multiple category classification, the MMD is to align the global data of the source domain and the target domain, thus ignoring the alignment effect of each category.

Considering the above problems, a bearing fault diagnosis model based on a multi-level domain adaption network is proposed. Firstly, the wavelet transform is used to transform one-dimensional (1D) signal into two-dimensional (2D) time-frequency images. Then, the ShuffleNet V2[12] is improved, and the improved ShuffleNet V2 is used for feature extraction. Finally, the data distribution difference between the source domain and the target domain is measured by the multi-level domain adaption network. The bearing fault diagnosis across working conditions and equipment could be realized.

1 Model Design

This section introduces the original structure of ShuffleNet V2 and the improved ShuffleNet V2.

1.1 Theoretical basis

1.1.1 Coordinate attention

Coordinate attention mechanism[13] is a kind of attention mechanism designed for the mobile network. It imparts location information into the channel attention mechanism, taking into account not only the channel information, but also the location information related to the direction.

Specifically, for the input feature map with a size of $C\times H\times W$, where C represents the number of channels of the input feature map, and H and W represent the height and the width of the input feature map, respectively. The pooled kernels with kernel sizes of $(H, 1)$ and $(1, W)$ are used to encode each channel along horizontal and vertical coordinates, respectively. Therefore, the output $z_{hc}(h)$ of channel c with a height of h can be expressed as

$$z_{hc}(h)=\frac{1}{W}\sum_{0\leqslant i<W}x_c(h,i), \tag{1}$$

where $x_c(h,i)$ represents the original feature value of the position (h,i) on channel c, h is the index in the vertical direction, and i is the index in the horizontal direction.

Similarly, the output $z_{wc}(w)$ of channel c with a width of w can be expressed as

$$z_{wc}(w)=\frac{1}{H}\sum_{0\leqslant j<H}x_c(j,w). \tag{2}$$

The above two transformations aggregate features along two spatial directions, respectively, to get a pair of directional perceptual feature maps. The pair of directional perceptual feature maps are encoded into two attention maps, each of which contains the long-distance dependency of the input feature map along a spatial direction. The location information can therefore be saved to the generated attention map. Then the two attention maps are applied to the input feature map by multiplication to emphasize the expression of the attention area.

The structure of the coordinate attention module is shown in Fig. 1. Here, X Avg Pool and Y Avg Pool refer to 1D horizontal global pooling and 1D vertical global pooling, respectively; r is the reduction ratio for controlling the block size; Conv2d is 2D convolution; Non-linear refers to nonlinear transformation.

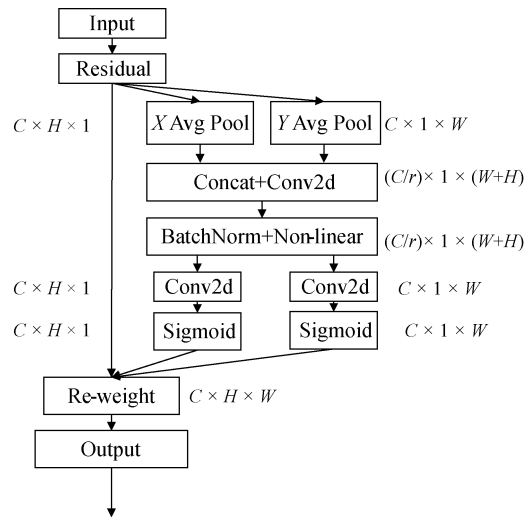

Fig. 1 Structure of coordinate attention module

1.1.2 Smooth maximum unit (SMU) activation function

The gradient of the ReLU activation function is 0 when the input value is less than 0, and the gradients of this neuron and its subsequent neurons are always 0 and do not respond to any data, which may cause the neural network to fail to converge or train slowly. In the process of network training, up to 50% of the neurons may die. The Leaky ReLU activation function initializes neurons with a small value similar to 0.01 so that ReLU tends to activate rather than die in the negative region.

The SMU-1[14] activation function f_1 uses the maximum smoothing technique to obtain the smooth approximation of the Leaky ReLU activation function:

$$f_1(x,\alpha x;\mu)=\frac{(1+\alpha)x+(1-\alpha)x\,\mathrm{erf}(\mu(1-\alpha)x)}{2}, \tag{3}$$

where $\mathrm{erf}(x)=\frac{2}{\sqrt{\pi}}\int_0^x \mathrm{e}^{-t^2}\mathrm{d}t$; α is a hyperparameter, and its size is 0.25; μ is a trainable parameter, and its initialization value in the SMU-1 activation function is

$4.352\,665\,993\,287\,951 \times 10^{-9}$.

1.1.3 Local MMD

MMD[15] is used to measure the distance between two different but related random variable distributions. Assume that a dataset satisfying P distribution is $\boldsymbol{X}_s = [x_{s1}, x_{s2}, \cdots, x_{si}, \cdots, x_{sn}]$ and a dataset satisfying Q distribution is $\boldsymbol{X}_t = [x_{t1}, x_{t2}, \cdots, x_{tj}, \cdots, x_{tm}]$. Find a mapping function f in the function domain F to map $\boldsymbol{X}_s$ and $\boldsymbol{X}_t$ to the high-dimensional space so that the mean difference of the two distributed random variables after mapping is the maximum, and this maximum value is MMD(M). It requires that F is rich enough and has enough constraints. The above two conditions can be satisfied simultaneously when F is a unit ball on the reproducing kernel Hilbert space (RKHS). Its calculation formula is

$$M^2[F, P, Q] = \left\| \frac{1}{n}\sum_{i=1}^{n} f(x_{si}) - \frac{1}{m}\sum_{j=1}^{m} f(x_{tj}) \right\|_H^2, \quad (4)$$

where $\|\cdot\|_H$ is a RKHS.

In the case of considering different sample weights, local MMD (LMMD)[16] measures the Hilbert-Schmidt norm between the kernel average embedding of the empirical distribution of the related subdomains of the source domain and the target domain.

$$\hat{d}_H(P, Q) = \frac{1}{c}\sum_{c=1}^{C} \left\| \sum_{x_{si}\in D_s} w_{sc,i}\emptyset(x_{si}) - \sum_{x_{tj}\in D_t} w_{tc,j}\emptyset(x_{tj}) \right\|_H^2, \quad (5)$$

where $w_{sc,i}$ and $w_{tc,j}$ denote the weight of x_{si} and x_{tj} belonging to class c; $\sum_{x_i\in D} w_{c,i}\emptyset(x_i)$ is a weighted sum on class c; $\hat{d}_H$ is the distributed distance measurement function.

1.2 Overall model design

1.2.1 Shuffle-CANet

ShuffleNet V2 (shown in Fig. 2) is an efficient convolution neural network proposed by Ma et al.[12], where Conv refers to convolution and DWConv refers to depthwise convolution; Concat is to connect two tensors together in a certain dimension to generate a larger tensor.

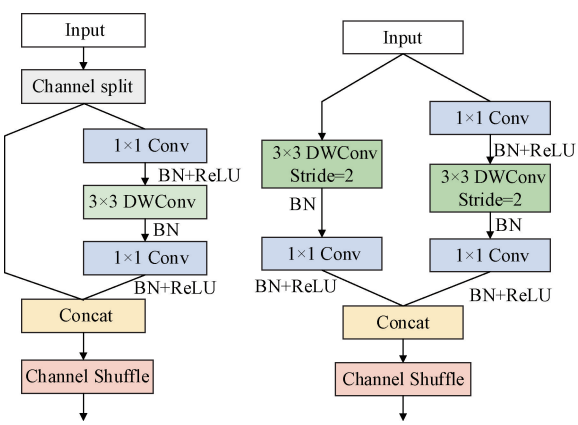

Fig. 2 ShuffleNet V2 block (on the left, the stride is 1; on the right, the stride is 2)

The Shuffle-CANet proposed in this paper is improved based on the ShuffleNet V2, as shown in Fig. 3. The coordinate attention (CoordAtt) mechanism is introduced into the ShuffleNet V2 block when the stride is 1, and the ReLU activation function is changed into SMU-1 activation function.

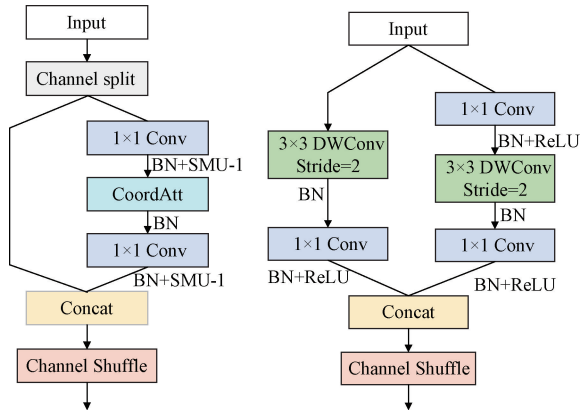

Fig. 3 Shuffle-CANet block

As shown in Fig. 4, each stage is comprised of one Shuffle-CANet block when the stride is 2 and three Shuffle-CANet blocks when the stride is 1. Then the cross stage partial (CSP)[17] operation is added to each stage. The input channel is divided into two modules: one module goes through a simple convolution operation, the other module passes through the original path. Then the results of the two modules are concated. The outputs of the feature extraction modules are expressed as $g(x_s)$ and $g(x_t)$.

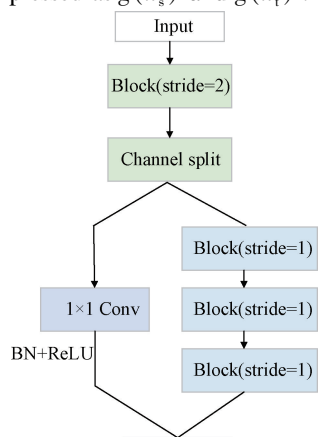

Fig. 4 Shuffle-CANet stage

1.2.2 Multi-level domain adaption network

The large convolution kernel is suitable for extracting global image features, while the small convolution kernel is suitable for extracting local image features. In order to learn more domain invariant features from the source domain and the target domain, the Inception C module in Inception V3[18] is applied to the bearing fault diagnosis model. In the Inception C module, a $n \times n$ convolution is decomposed into two convolution kernels $1 \times n$ and $n \times 1$

which can save a lot of parameters. Moreover, the receptive field of the decomposed convolution nucleus is the same as that of the pre-decomposed convolution nucleus. In each branch shown in Fig. 5, convolution kernels of different sizes or degrees of downsampling are used to extract features, which makes the features extracted by the network more abundant, and better performance can be obtained by aligning the features extracted by multiple convolution kernels with different sizes.

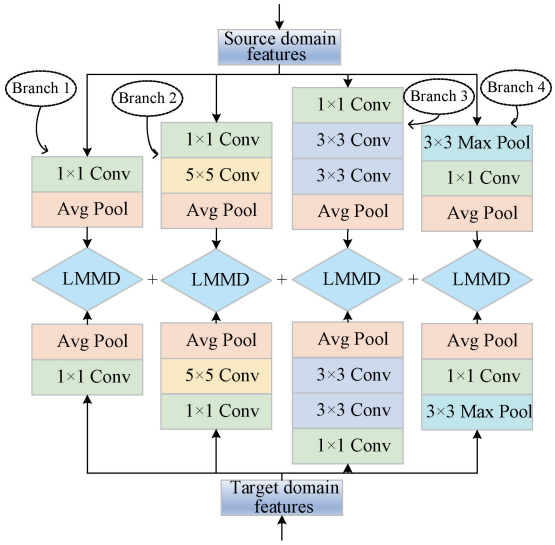

Fig. 5 Multi-level domain adaption network

2 Transfer Learning Fault Diagnosis Model

2.1 Fault diagnosis model

In order to realize bearing fault diagnosis across working conditions, a bearing fault diagnosis model based on a multi-level domain adaption network is proposed in this paper. The detailed structure of the model is shown in Fig. 6, and the detailed parameters are shown in Table 1.

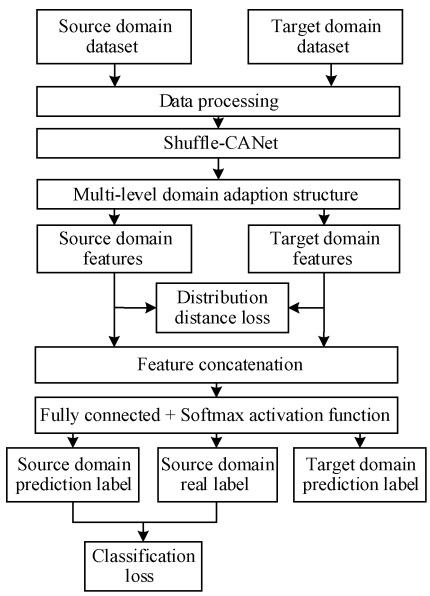

Fig. 6 Structure of proposed fault diagnosis model

Table 1 Model structure parameters

Layer	Kernel size	Stride	Repeat	Output channel
Conv 1	3×3	2	1	24
Max Pool	3×3	2	1	24
Stage 2		2 1	1 3	116
Stage 3		2 1	1 3	232
Stage 4		2 1	1 3	464
Conv 5	1×1	1	1	1 024
Branch 1	1×1	1	1	64
Branch 2	1×1 1×3 3×1	1 1 1	1 1 1	64
Branch 3	1×1 1×5 5×1	1 1 1	1 1 1	96
Branch 4	7×7 Avg Pool 1×1	1 1	1 1	64
Avg Pool	7×1	1	1	288

2.2 Optimization goal

The classifier calculates the classification loss based on the multi-scale features of the source domain data, and the domain adaption module calculates LMMD between the source domain and the target domain based on the multi-scale features. The classification loss function L_y of the fault classifier uses the cross entropy function to calculate the error between the real label and the prediction label:

$$L_y = \frac{1}{n_s}\sum_{i=1}^{n_s} L(\hat{y}_{si}, y_{si}), \tag{6}$$

where $L(\cdot)$ is cross entropy function; $\hat{y}_{si}$ is the prediction label for the ith sample; y_{si} is the real label for the ith sample.

The domain adaption module chooses LMMD[18] as the distance loss function L_d:

$$L_d = \sum_{i=1}^{n} \hat{d}_{\mathrm{H}}(h_i(g(x_s)), h_i(g(x_t))). \tag{7}$$

In order to realize cross-domain fault diagnosis of the bearing, the overall loss function L of the fault diagnosis model is obtained by combining L_y and L_d:

$$L = L_y + \beta L_d, \tag{8}$$

where $\beta = \frac{2}{1 + \exp(-5e/E)} - 1$, e is the current number of training rounds and E is the total number of training rounds.

3 Fault Diagnosis Steps

In the field of fault diagnosis, data enhancement is an important means to improve the classification accuracy. In this paper, the overlapping sampling[19] is used to divide the data, and there is a partial overlap between the samples. The divided samples are transformed into time-frequency images by wavelet transform, and the source domain dataset D_s and the target domain dataset D_t are constructed.

The input model includes D_s and D_t. The classifier calculates the classification loss based on the characteristics of the source domain data, and the domain adaption module calculates LMMD between the source domain and the target domain based on the multi-scale features. The overall loss of the model is obtained by the weighted classification loss and the weighted distance loss so as to complete the training of the model.

As a test set, the target domain data is input into the trained model to diagnose the fault category of the bearing and complete the fault diagnosis of the bearing in different environments.

The overall flowchart is shown in Fig. 7.

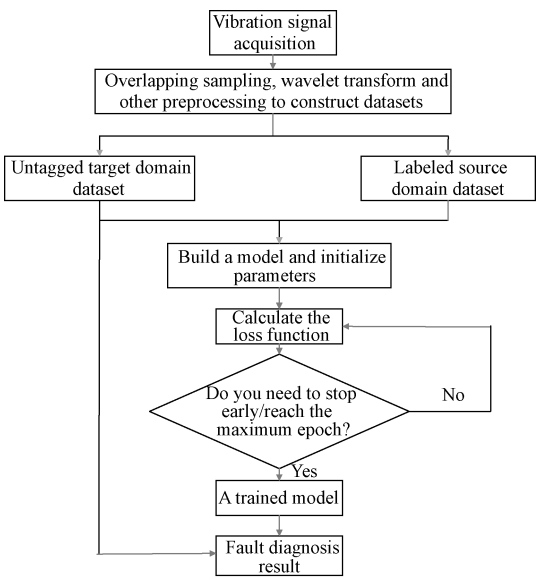

Fig. 7 Fault diagnosis model training flowchart

4 Experimental Verification

The programming language of this paper is Python 3.8 using the deep learning framework PyTorch 1.11.0. The running environment is Windows 11 64-bit operating system, the hardware environment is Intel (R) xeon (R) Platinum 8358P CPU @ 2.60 GHz, and GPU is RTX A5000 24 G.

4.1 Experiment 1: transfer experiment on the same dataset

4.1.1 Experimental data

Firstly, the bearing dataset of Case Western Reserve University (CWRU)[20] is selected to verify the proposed model. The experimental platform consists of a two-motor-drive system, a torque sensor/decoder, a power tester and an electronic controller. In this experiment, we choose the drive end (DE) bearing (SKF6205) as the bearing to be tested, and the sampling frequency is 12 kHz. Different bearing loads are set in the process of data collection, including 0, 0.746, 1.492 and 2.238 kW. As shown in Table 2, four different working conditions are set up and recorded as the dataset 0HP, 1HP, 2HP and 3HP, respectively.

Table 2 Working conditions of CWRU bearing dataset

Working condition	Load/kW	Rotational speed/(r/min)
0HP	0	1 797
1HP	0.746	1 772
2HP	1.492	1 750
3HP	2.238	1 730

As shown in Table 3, a variety of bearing states are obtained by electrical discharge machining including one normal state and nine fault states.

Table 3 Fault states and normal state of CWRU bearing dataset

	State	Pitting diameter/mm	Label
Fault	Rolling body	0.000 177 8	0
		0.000 355 6	1
		0.000 533 4	2
	Inner ring	0.000 177 8	3
		0.000 355 6	4
		0.000 533 4	5
	Outer ring	0.000 177 8	6
		0.000 355 6	7
		0.000 533 4	8
Normal		0	9

First of all, by using the method of overlapping sampling, each kind of original vibration signal under each working condition is divided into 500 samples, and each sample contains 1 024[19] data points. Then the divided samples are converted into 2D time-frequency images by continuous wavelet transform. Figure 8 shows the 2D spectra of bearings with 10 different labels (labels 0-9) under 0HP conditions.

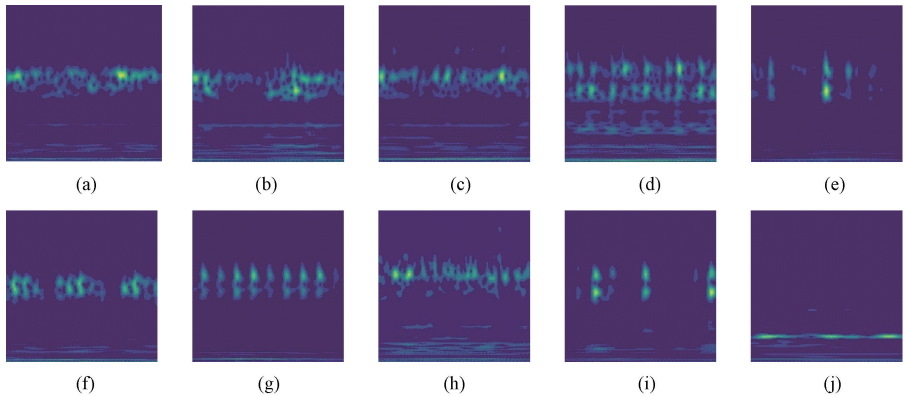

Fig. 8 2D spectra of bearings with 10 different labels under 0HP conditions: (a) label 0; (b) label 1; (c) label 2; (d) label 3; (e) label 4; (f) label 5; (g) label 6; (h) label 7; (i) label 8; (j) label 9

4.1.2 Experimental results

In order to verify the effectiveness of the proposed model, 12 transfer tasks are set in the CWRU dataset. The transfer tasks are shown in Table 4, in which 0→1 indicates that the 0HP dataset is used as the source domain data and the 1HP dataset is used as the target domain data. In the CWRU dataset, the fault diagnosis model is tested on 12 different transfer tasks and compared with other classical migration learning including deep subdomain adaptation network (DSAN)[15], deep correlation alignment[15] (Deep COARL) and deep adaptation network (DAN)[14]. The feature extraction networks are all based on the ShuffleNet V2 1x . In the training process, the batch size is 32, and the epoch is 200. In order to prevent the model from overfitting, the early stopping is set to 50, the stochastic gradient descent optimizer is selected, and the initial learning rate is set to 0.01. The experimental results are shown in Table 4.

Table 4 Diagnostic accuracy of different models on CWRU datasets

Task	Accuracy/%			
	DSAN	Deep CORAL	DAN	Proposed model
0→1	90.84	85.70	83.64	99.02
0→2	90.74	80.94	81.80	98.92
0→3	84.84	86.68	79.72	99.28
1→0	88.08	76.76	72.60	98.60
1→2	90.02	83.56	83.74	99.44
1→3	91.30	82.08	78.96	99.16
2→0	84.42	76.38	77.80	97.38
2→1	92.04	80.28	85.06	98.88
2→3	90.88	83.53	80.24	98.14
3→0	83.74	81.90	77.72	99.60
3→1	90.54	79.72	80.04	99.24
3→2	91.42	84.84	86.42	99.52

According to the results of the above experiments, it can be seen that the fault diagnosis performance of the proposed model is better than that of the compared models, and the following conclusions can be obtained.

1) By comparing DSAN, Deep CORAL and DAN models, it can be concluded that the DSAN model can most effectively reduce the data distribution difference in the source domain and the target domain, and the diagnosis performance is the best.

2) Compared with DRSN, the proposed model shows better performance in 12 transfer tasks. It is proved that the improvement of the ShuffleNet V2 and the addition of multi-domain adaption operation can improve the diagnostic accuracy of the model in the cross-domain bearing fault diagnosis.

Figures 9 and 10 are the experimental results of the proposed model in 0→1 task. In Fig. 9, it can be seen that most of the samples can be correctly identified. In addition, the training error can quickly converge and stabilize around 0.001, and the test diagnostic accuracy can quickly reach 99.00%, and stabilize at this diagnostic accuracy.

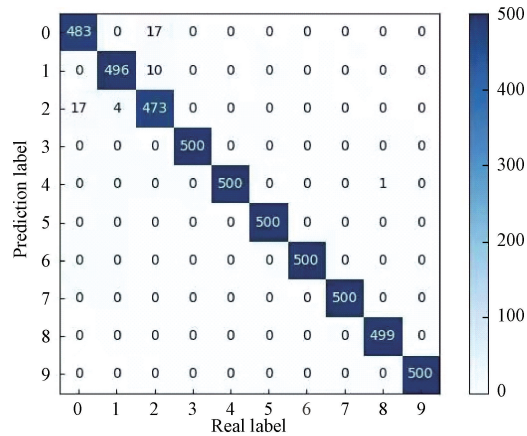

Fig. 9 Confusion matrix of diagnosis results of proposed model in 0→1 task

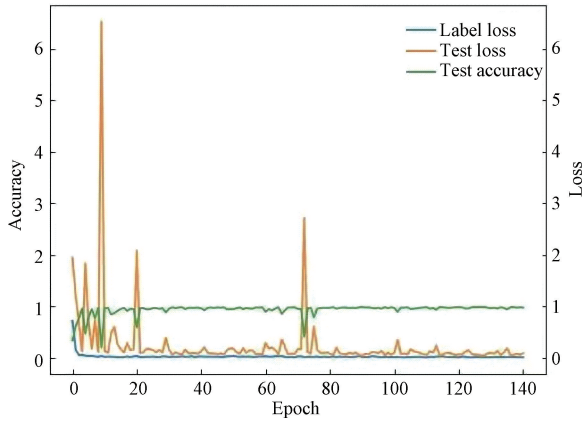

Fig. 10 Training loss curve and test diagnostic accuracy curve of proposed model in 0→1 task

In order to verify the effectiveness of the improved ShuffleNet, taking 1→0 task as the transfer task, several common networks are used as feature extraction networks to compare with the improved ShuffleNet. As can be seen from Table 5, in the same transfer task, compared with other feature extraction networks, the improved ShuffleNet not only improves the diagnostic accuracy, but also greatly reduces the parameter number of the model. On the NVIDIA RTX 3060 server, the frame per second (FPS) of the model using the improved ShuffleNet as the feature extraction network reaches 1.50. It can handle 1 536 data points per second, which is higher than other networks as the model of the feature extraction network. Compared with other feature extraction networks, the improved ShuffleNet is more suitable for mobile devices.

Table 5 Diagnostic accuracy and number of parameters of different models

Model	Accuracy/%	Parameter number/million	FPS
ResNet18	92.80	11.18	0.81
MobileNet V3 large	87.12	3.26	0.92
ShuffleNet V2	88.08	1.56	1.08
Proposed model	99.28	1.20	1.50

4.1.3 Visualization of transfer learning effect

In order to more intuitively see the effectiveness of the proposed model for cross-domain bearing fault diagnosis, the t-SNE algorithm is used to reduce the dimension of the features extracted from the network and generate a visualization graph. Taking 0→1 task as an example, as shown in Fig. 11 (a), at the beginning of model training, the data in the source domain and the target domain are far apart in space, and there is a large distribution difference between the two domains. After 150 rounds of training, as shown in Fig. 11 (b), the two domains are almost in the same spatial distribution. It can be seen that the proposed model can significantly reduce the distribution difference between the two domains.

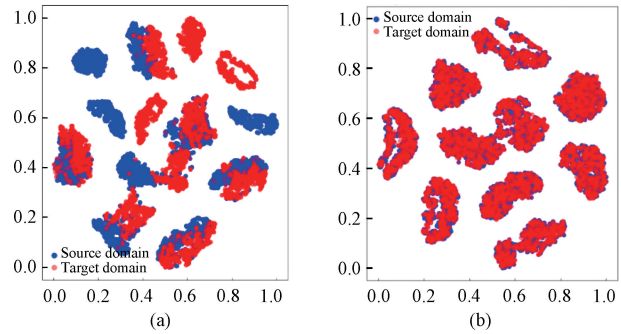

Fig. 11 Feature distribution map: (a) beginning; (b) 150 rounds of training

4.2 Experiment 2: transfer experiment on different datasets

4.2.1 Bearing dataset of the Polytechnic University of Turin

The bearing dataset of the Polytechnic University of Turin (DIRG)[21] is collected from a test-bed established in the laboratory of the Department of Mechanical and Aerospace Engineering of the Polytechnic University of Turin, and it is designed to test high-speed rotating aeronautical bearings. The dataset includes three kinds of bearing states, namely, a normal state, an inner ring fault and an outer ring fault. The load is 1 000 N and the speed is 12 000 r/min. The original vibration signals are overlapped sampled, 1 024 data points are intercepted from each sample, 500 samples are obtained for each fault type, and then each sample is converted into a 2D time-frequency image by continuous wavelet transform. The fault signals generated by bearings from different sources, types and working environments are also different. In order to verify the generalization performance of the proposed model, the CWRU bearing 0HP and 1HP (CWRU-0HP and CWRU-1HP) datasets and the DIRG bearing dataset are used as the source domain dataset and the target domain dataset, respectively. Since the DIRG dataset lacks bearing fault data related to rolling body failure, the bearing roller fault data from the CWRU dataset is excluded in experiment 2. Each dataset contains three fault states of the bearing: a normal state, an inner ring fault and an outer ring fault.

4.2.2 Experimental results

From the experimental results in Table 6, the following conclusions can be drawn. The bearing on the DIRG dataset is the high-speed rotating aeronautical bearing, and is quite different from the bearing on the CWRU dateset. Compared with the transfer task on the same dataset in experiment 1, the proposed model migrates across the datasets in experiment 2. Compared with other models, the diagnostic accuracy of the proposed model is improved obviously. This is because the proposed feature extraction network combines several

convolution kernels with different sizes. As the receptive fields of convolution kernels are different, the extracted feature information is also different. By combining the feature information in different scales, the features of each receptive field complement each other. When the feature distribution of the source domain is different from that of the target domain, the proposed model can still maintain a high diagnostic accuracy.

Table 6 Model diagnosis results on different datasets

Dataset	Accuracy/%			
	Proposed model	DSAN	DAN	Deep CORAL
CWRU-0HP→DIRG	98.19	79.24	76.05	72.43
DIRG→CWRU-0HP	96.76	79.43	83.10	78.19
CWRU-1HP→DIRG	95.86	81.09	78.98	78.90
DIRG→CWRU-1HP	96.91	82.12	80.61	80.60

4.3 Model diagnosis performance experiment

In order to verify the effectiveness of each part of the proposed model, taking the transfer task of CWRU-0HP → DIRG as an example, the performance of each improved module is tested, respectively, and the results are shown in Table 7. The basic part (base) is a ShuffleNet composed of three stages, and each stage is comprised of one block with the stride of 2 and three blocks with the stride of 1. All the models use the same model structure and loss function.

Table 7 Diagnostic accuracy of adding different modules to model

Model	Accuracy/%
Base	65.14
Base + Inception C	79.10
Base + Inception C + CSP	84.95
Base + Inception C + CSP + coordinate attention	94.81
Base + Inception C + CSP + coordinate attention+SMU-1	98.19

Firstly, the performance of the multi-scale asymmetric convolution kernel in the model is analyzed. When the Inception C module is added, the diagnostic accuracy is improved compared with the base model, indicating that adding the Inception C module and realizing multi-level domain adaptation can effectively achieve data distribution adaptation and help the model extract the shared features of the two domains.

Secondly, the influence of adding the CSP mechanism to the model is studied, and the results show that the diagnostic accuracy of the model can be improved after adding the CSP mechanism.

Then, we test the effect of the coordinate attention mechanism in the model. When the stride is 1, replacing the DWConv with the coordinate attention mechanism can improve the model diagnostic accuracy.

Finally, after replacing the activation function ReLU in the model with the SMU-1 activation function, the diagnostic accuracy of the model is higher than that of base model.

5 Conclusions

In this paper, a bearing fault diagnosis model based on a multi-level domain adaption network is proposed. The effectiveness of the proposed model is verified by using two different bearing datasets and different transfer tasks. Compared with the transfer learning models commonly used in other fault diagnosis fields, the proposed model exhibits better performance in bearing fault diagnosis.

1) The proposed model can effectively reduce the distribution difference between the two domains.

2) Adding the coordinate attention mechanism to the ShuffleNet can help the model capture the spatial information and the location information in the image better, thus improving the diagnostic accuracy of the model in the fault diagnosis.

3) Compared with the ShuffleNet V2, the improved ShuffleNet could further reduce the number of parameters, maintaining a high diagnostic accuracy while maintaining the efficiency.

References

[1] YAN R Z, ZHANG Y B. Static and dynamic characteristics of vacuum preloaded porous aerostatic bearings [J]. *Journal of Donghua University (English Edition)*, 2024, 41(1): 81-88.

[2] WANG H C, XIANG G Q, GUO Z Q, et al. Fault diagnosis of rolling bearing compound faults based on improved time-frequency spectrum analysis method [J]. *Journal of Aerospace Power*, 2017, 32(7): 1698-1703. (in Chinese)

[3] YANG X C. Rolling bearing fault diagnosis research based on wavelet spectrum analysis [J]. *Coal Mine Machinery*, 2013, 34 (1): 289-291. (in Chinese)

[4] SARAVANAN N, RAMACHANDRAN K I. Incipient gear box fault diagnosis using discrete wavelet transform (DWT) for feature extraction and classification using artificial neural network

(ANN)[J]. *Expert Systems with Applications*, 2010, 37(6): 4168-4181.
[5] LI Y B, XU M Q, WEI Y, et al. A new rolling bearing fault diagnosis method based on multiscale permutation entropy and improved support vector machine based binary tree[J]. *Measurement*, 2016, 77: 80-94.
[6] SHAO J J, HUANG Z W, ZHU J M. Transfer learning method based on adversarial domain adaption for bearing fault diagnosis[J]. *IEEE Access*, 2020, 8: 119421-119430.
[7] HOANG D T, KANG H J. A bearing fault diagnosis method using transfer learning and Dempster-Shafer evidence theory [C]// Proceedings of the 2019 International Conference on Artificial Intelligence, Robotics and Control. New York: ACM, 2019: 33-38.
[8] WANG Y, SUN X J, LI J, et al. Intelligent fault diagnosis with deep adversarial domain adaptation [J]. *IEEE Transactions on Instrumentation and Measurement*, 2021, 70: 1-9.
[9] HU X D, YANG X. Fault diagnosis method for rolling bearings based on siamese domain adversarial neural network [J]. *Bearing*, 2023(7): 79-86. (in Chinese)
[10] CHEN R X, TANG L L, HU X L, et al. A rolling bearing fault diagnosis method based on deep attention transfer learning at different rotationals[J]. *Journal of Vibration and Shock*, 2022, 41(12): 95-101, 195.
[11] HUO C R, JIANG Q S, SHEN Y H, et al. Enhanced transfer learning method for rolling bearing fault diagnosis based on linear superposition network [J]. *Engineering Applications of Artificial Intelligence*, 2023, 121: 105970.
[12] MA N N, ZHANG X Y, ZHENG H T, et al. ShuffleNet V2: practical guidelines for efficient CNN architecture design[C]//15th European Conference. New York: ACM, 2018: 122-138.
[13] HOU Q B, ZHOU D Q, FENG J S. Coordinate attention for efficient mobile network design [C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2021: 13713-13722.
[14] BISWAS K, KUMAR S, BANERJEE S, et al. SMU: smooth activation function for deep networks using smoothing maximum technique [EB/OL]. (2022-04-11) [2023-06-02]. https://arxiv.org/abs/2111.04682v2.
[15] PAN S J, TSANG I W, KWOK J T, et al. Domain adaptation via transfer component analysis [J]. *IEEE Transactions on Neural Networks*, 2011, 22(2): 199-210.
[16] ZHU Y C, ZHUANG F Z, WANG J D, et al. Deep subdomain adaptation network for image classification[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2021, 32(4): 1713-1722.
[17] WANG C Y, MARK LIAO H Y, WU Y H, et al. CSPNet: a new backbone that can enhance learning capability of CNN[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). New York: IEEE, 2020: 390-391.
[18] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2016: 2818-2826.
[19] LI Y W. Research on fault diagnosis method of rolling bearing under variable working condition based on improved CNN [D]. Harbin: Harbin University of Technology, 2019. (in Chinese)
[20] XIA Q, SHI T, WANG M Y. A level set based shape and topology optimization method for maximizing the simple or repeated first eigenvalue of structure vibration [J]. *Structural and Multidisciplinary Optimization*, 2011, 43: 473-485.
[21] DAGA A P, FASANA A, MARCHESIELLO S, et al. The Politecnico di Torino rolling bearing test rig: description and analysis of open access data [J]. *Mechanical Systems and Signal Processing*, 2019, 120: 252-273.

基于多级域自适应网络的轴承故障诊断模型

李文文*, 陈广锋
东华大学 机械工程学院, 上海 201620

摘　要: 针对在轴承运行过程中复杂多变的环境可能会导致训练数据和测试数据分布不一致、模型诊断性能下降的问题, 提出了一种基于 Shuffle-CANet 的轴承故障诊断模型。通过改进 ShuffleNet V2 并引入非对

称卷积，实现了对轴承的跨域故障诊断。根据迁移学习中的领域自适应思想，在模型中加入域损失函数来提取源域和目标域的共有特征，实现跨域故障诊断。与传统的深度学习模型相比，该模型对移动设备和嵌入式设备更加友好。在两个不同的数据集上通过不同迁移任务来验证 Shuffle-CANet 模型的故障诊断效果。研究结果表明，当源域和目标域数据来源于相同的数据集时，模型的故障诊断准确率可以超过 99%；当源域和目标域数据来源于不同的数据集时，模型的故障诊断率可以超过 95%。

关键词：轴承故障诊断；ShuffleNet V2；多级域自适应；轻量化