An Anti-Noise Gearbox Fault Diagnosis Method based on Multi-Scale Transformer Convolution and Transfer Learning

JINLIANG WU, XIAOYANG ZHENG, XINGLONG PEI
School of Artificial Intelligence,
Chongqing University of Technology,
Pufu Avenue, Longxing Town, Yubei District, Chongqing,
CHINA

Abstract: - Gearbox fault diagnosis methods based on deep learning usually require a large amount of sample data for training, and these data are usually ideal experimental data without noise. However, due to the influence of complex environmental factors, a large number of effective fault samples may not be available and the sample data can be interfered with by noise, which affects the identification accuracy of fault diagnosis methods and the stability of diagnosis results. To improve the resistance to noise while achieving high diagnosis accuracy, a multi-scale Transformer convolution network (MTCN) based on transfer learning is proposed in this paper. Concretely, a multi-scale coarse-grained procedure is incorporated to capture different and complementary features from multiple scales and filter random noises to some extent. Meanwhile, the Transformer composed of an attention mechanism is utilized to extract high-level and effective features and the transfer learning strategy is applied to overcome the limitation of insufficient fault samples for model training. Finally, the experiments are conducted to verify the effectiveness of the proposed method. The results show that the proposed method has higher accuracy and robustness under noisy environments compared with previous methods.

Key-Words: - Gearbox, Fault diagnosis, Transformer, Convolution Network, Transfer Learning.

Received: May 27, 2021. Revised: September 6, 2022. Accepted: October 8, 2022. Published: November 29, 2022.

1 Introduction

Fault diagnosis is an essential and indispensable measure to realize the health monitor and ensure the safe operation of mechanical systems in modern society. As an important and common component of mechanical equipment, the gearbox has been extensively applied in various fields such as transportation, agriculture production, space flight, aviation, and so on, [1]. Due to the influence of various working environmental conditions and load intensity, the gearbox is vulnerable to various faults which could lead to accidents and loss, [2]. Consequently, it is of critical significance to research to realize high-accuracy fault diagnosis for gearboxes.

At present, gearbox intelligent fault diagnosis methods are mainly divided into two categories: traditional fault diagnosis method and deep learning-based fault diagnosis method. In the traditional fault diagnosis methods, features are usually extracted manually from original vibration signals and then fed into machine learning models to obtain fault diagnosis results. For example, empirical mode decomposition and support vector machine are combined to diagnose the faults of gear reducer in [3]. To realize gearbox fault diagnosis, deep Boltzmann machines are developed for deep representations of the statistical parameters of the wavelet packet transform in [4]. To solve the problem with non-linearity and high dimension, a wavelet support vector machine and immune genetic algorithm are developed in [5]. To obtain quantitative indicators of gear shaft deterioration, the wavelet transform technique is used and EM algorithm and optimal Bayesian method are used for model parameters estimation in [6]. However, owing to the limitations of shallow structure and manual feature extraction, the traditional fault diagnosis method is hard to learn complex nonlinear relations and intrinsic fault features.

Deep learning methods can solve the problem of low nonlinear performance of the shallow neural network and have superior generalization ability. Due to its unique advantages and potential in automatic feature extraction and pattern recognition, more and more deep learning methods have been widely applied to the research of gearbox fault diagnosis. Qiu et al., [7], used a deep convolutional neural network (DCNN) to extract the features from both vertical and horizontal vibration signals of five

different degradation states, achieving higher accuracy with lower computational time cost compared with traditional diagnosis methods. Zhang et al., [8], introduced a novel method based on recurrent neural network (RNN) to exploit the temporal information of time-series data and learn representative features from constructed images which are converted into two-dimensional images from one-dimensional time-series vibration signals and used a multilayer perceptron (MLP) to implement fault recognition. Yu et al., [9], proposed a new deep belief network (DBN), which inserts confidence and classification rules into the deep network structure to enable the model to have good pattern recognition performance and can adaptively determine the network structure and obtain a good understanding of the features learned by the deep network.

The method based on deep learning mentioned above has achieved certain success in some applications, but there are still certain limitations. Owing to the structure of RNN, it's hard to realize the parallelization between training samples and has the problem of long memory loss, [10]. Although CNN has a superior extraction ability for local features, it's not good at learning features from a long-time sequence of signal data, [11]. In the case of large-scale and complex datasets, it's difficult for DBN to exhibit a satisfactory performance due to its structural characteristics, [10]. Different from the structure of traditional CNN and RNN, Transformer has a new special network architecture solely based on attention mechanisms and abandons recurrence and convolutions structure entirely, [12]. To be specific, all the related operations in the Transformer network are order-independent and parallelizable. With the rapid development of the Transformer network, it has been widely applied and demonstrated its outstanding performance in many fields, such as image processing, [13], pose recognition, [14], autonomous driving, [15], and natural language processing, [16], [17].

Moreover, owing to the interaction and coupling effects among different components and subsystems of the gearbox, the measured vibration signals collected from sensors installed on the house usually contain multiple intrinsic oscillatory modes, [18]. Consequently, the vibrations signals contain complex patterns at multiple time scales and inherent multi-scale characteristics, [19], [20]. Due to the limitation of the inherent structure, it's hard for traditional CNN or RNN to capture multi-scale features from vibration signals.

Furthermore, there is usually an assumption that the training data and test data have the same or similar distribution when training a deep learning network model for fault diagnosis, [21]. However, due to the complex working environment and variable load conditions in real industrial applications, it's probably hard to obtain adequate fault data for training network models. Although Transformer has shown more powerful learning ability, it will also suffer from performance decline when there are insufficient fault samples for training models. To overcome this problem, an alternative approach is to use the transfer learning strategy to train the network model. By using the transfer learning strategy, the network model can be pre-trained in the source domain dataset which provides a massive amount of prior knowledge for training, and then the model can be fine-tuned in target domain dataset whose data amount is small and effective fault data is insufficient. Therefore, this paper proposed a multi-scale Transformer convolution network (MTCN) based on transfer learning for gearbox fault diagnosis. The main contributions of this paper are summarized as follows:

1) A novel MTCN model is proposed by combining multi-scale coarse-grained processing, Transformer and CNN.
2) The proposed method can capture different and complementary features from vibration signals at multiple scales in parallel.
3) The use of transfer learning strategy enables the proposed model to be trained on a target dataset with limited effective fault samples.

The rest of this paper is organized as follows. The related method theories are introduced in Section 2. The details of the proposed MTCN architecture is elaborated in Section 3. The results of the comparative experiment to evaluate the proposed model against some other common methods is analyzed in Section 4. The conclusion of this paper is drawn in Section 5.

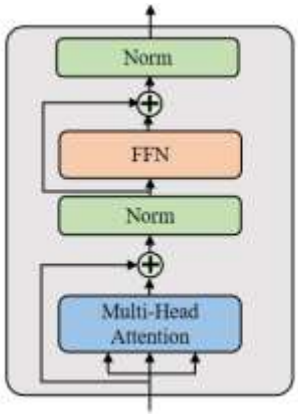


Fig. 1: The structure of transformer encoder layer

2 Related Theories

2.1 Scaled Dot-Product Attention

As a core module of the Transformer encoder, scaled dot-product attention is used to map a query vector and a set of key-value vector pairs to an output vector, [12]. To be specific, the input of scaled dot-product attention is composed of the query and keys of dimension d_k and values of the dimension d_k. The query with all keys is computed by dot-product and the dot-product result is scaled by dividing $\sqrt{d_k}$. After that, it's optional to add a mask on the scaled result. Finally, the output computed as a weighted sum of the values is obtained by applying a softmax function, which is calculated as:

$$Attention(Q,K,V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \quad (1)$$

where Q, K, V are queries, keys, and values matrices respectively, and $\sqrt{d_k}$ is the scaling factor.

2.2 Scaled Dot-Product Attention

Multi-head attention applied in the transformer enables the model to focus on the feature information from different representation sub-spaces at different positions, [12], which enhances the expression capability of each attention layer. Concretely, the queries, keys, and values are linearly projected h times to obtain different linear projection results. Then, the attention function is performed on each linear projected version of queries, keys, and values in parallel. At last, the final values are obtained by projecting the concatenated result from each calculated output of the attention function. It can be described as:

$$MultiHead(Q,K,V) = Concat(head_1,...,head_h)W^O \quad (2)$$

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \quad (3)$$

where $W_i^Q \in \mathbf{R}^{d_{\text{model}} \times d_k}$, $W_i^K \in \mathbf{R}^{d_{\text{model}} \times d_k}$, $W_i^V \in \mathbf{R}^{d_{\text{model}} \times d_v}$, $W^O \in \mathbf{R}^{hd_v \times d_{\text{model}}}$ are the parameter matrices of linear projections and h is the number of parallel attention layers.

2.3 Transformer Encoder

In the architecture of the Transformer, the encoder is composed of a stack of identical layers. As shown in Fig. 1, there are two sublayers in the structure of the transformer encoder. The first sublayer is a multi-head self-attention mechanism and the second sub-layer is a simple, position-wise fully connected feed-forward network. Around each of the two sub-layers, a residual connection and layer normalization are employed to mitigate the degradation of the network and enhance the robustness separately.

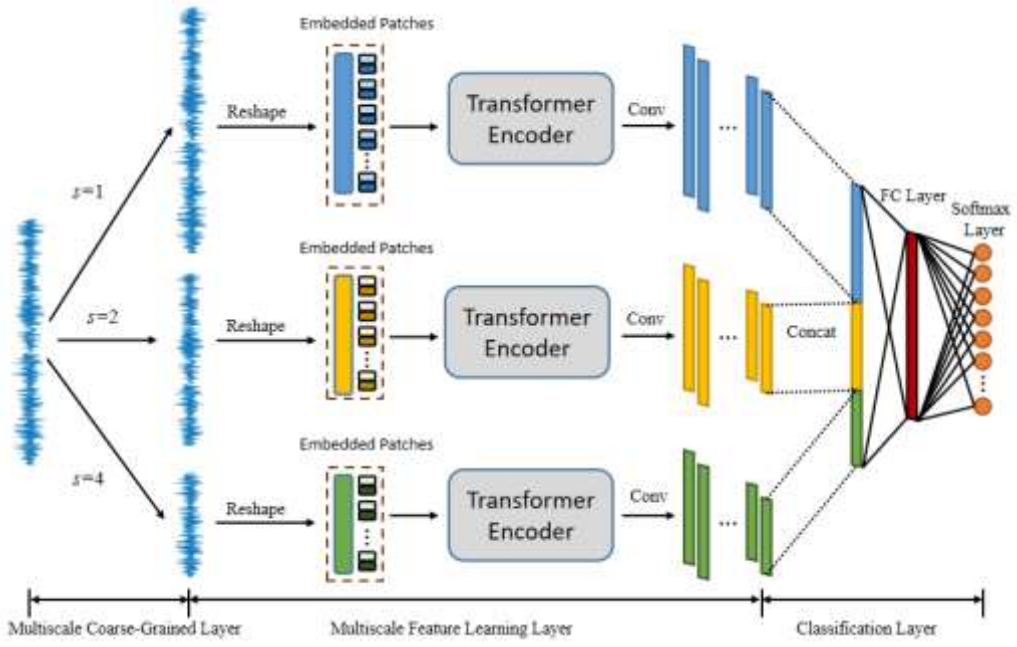


Fig. 2: The structure of proposed MTCN

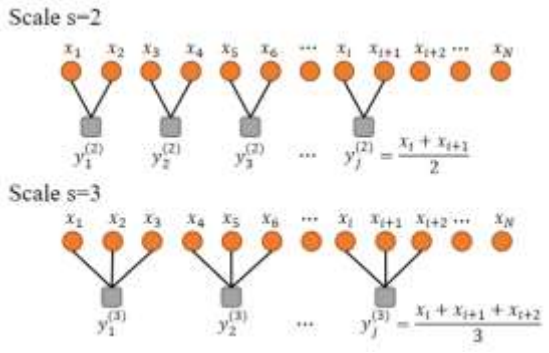


Fig. 3: Illustration of the coarse-grained procedure for s = 2, 3

3 Proposed Method

To realize the high-accuracy fault diagnosis even under strong noisy environments and to solve the limitation of the rare effective fault data in certain cases, a multi-scale transformer convolution network based on transfer learning is proposed creatively in this paper. As shown in Fig. 2, the structure of the proposed model consists of three parts.

In order to enable the proposed model to learn multi-scale features, a simple coarse-grained procedure is incorporated in the structure, which has the advantages of lower complexity and computational cost compared with traditional multi-scale transformation, [18]. Concretely, for a given vibration signal series $x=\{x_1,...,x_i,...,x_N\}$, the computed output result of a multi-scale coarse-grained procedure is calculated as:

$$y_j^{(s)}=\frac{1}{s}\sum_{n=(j-1)s+1}^{js}x_i, 1\le j\ge\frac{N}{s} \tag{4}$$

where s is the scale factor, N is the length of input data. As shown in Fig. 2, the scale factors are set as 1, 2, 4 respectively in this paper.

Moreover, the original vibration signals can be smoothed and down-sampled by the coarse-grained operation. This operation is a kind of simple low-pass filtering process through moving average with a non-overlapped window, resulting in filtering certain high-frequency perturbations and random noises to some extent, [18].

In the multi-scale feature learning layer, multiple transformer encoders and convolution layers are employed to learn complementary and high-level features from the coarse-grained signals with different time scales in a parallel manner. Since the multiple stacked attention layers in the transformer have a strong learning ability and parallel ability, the signals after being processed by the coarse-grained procedure are sent into the transformer encoder to obtain high-level features. However, the form of the input of the standard transformer encoder is a token embedding of 1D sequence, which is not suitable for vibration signals. To solve this, a given signal sample $x=[x_1,...,x_i,...,x_M]\in\mathbf{R}^M$, it is reshaped into 2D patches $x_p=\{x_p^1;x_p^2;...,x_p^N\}\in\mathbf{R}^{N\times L}$, where L is the length of each patch and $N=\lfloor M/L\rfloor$ is the length of the 2D patches. To acquire the patch embedding $z\in^{\sim\ N\times d_{\text{model}}}$, the 2D patches x_p is calculated by a learnable linear projection as:

$$z=x_pE \tag{5}$$

where $E\in\mathbf{R}^{L\times d_{\text{model}}}$ is the linear projection and d_{model} is the dimension of the vector.

Additionally, to retain the positional information about the relative or absolute position of the patches in sequence, the positional encoding is computed as follows:

$$PE_{(pos,2i)}=\sin(pos/10000^{2i/d_{\text{model}}}) \tag{6}$$

$$PE_{(pos,2i+1)}=\cos(pos/10000^{2i/d_{\text{model}}}) \tag{7}$$

where i is the dimension, pos is the position, and PE_{pos+k} represents a linear function of PE_{pos} for any fixed offset k.

The resulting sequence of the patch embeddings with positional information serves as the input of the Transformer encoder which is composed of a sole layer in this paper. The operation of the Transformer layer can be expressed as:

$$z'=LN(MSA(z)+z) \tag{8}$$

$$z''=LN(FFN(z')+z') \tag{9}$$

where MSA is the muti-head self-attention, FFN is a simple fully connected feed-forward network, and LN is the layer normalization.

For the Transformer encoder in this paper, the layer number is set as 1, the number of heads in the multi-head attention is 4 and the dimension of the feed-forward network is 200. Followed by the Transformer encoder, a convolution block that contains a convolution layer, a batch normalization layer, and a max pooling layer is employed to process the output sequence. In the convolution layer, the kernel size is 3, the padding is 1 and the stride is 1. The activation function is ReLU and the pool size and the stride are 4 and 2 in the max

pooling layer. For the output from each scale, the final outputs are concatenated together to be fed into the classification layer to finish fault diagnosis.

To solve the problem that a large number of effective fault samples for training are not available in certain cases, the proposed method MTCN is trained by transfer learning. Concretely, the MTCN is pre-trained with randomly initialized parameters on the source domain dataset which processes rich fault samples. Then the classifier of the MTCN is modified to adapt to the new target task and the parameters of the MTCN are fine-tuned in the target domain dataset with a limited number of training samples. Fig.4 presents the flowchart of the proposed model for gearbox fault diagnosis and the general procedure can be concluded as the following six steps:

Step1: The source domain dataset and target domain datasets are collected from different experimental facilities.

Step2: The training dataset and testing dataset are obtained by dividing the source domain dataset and target domain datasets and each sample is processed through FFT.

Step3: The MTCN model is constructed with a multi-scale coarse-grained layer, Transformer encoder, CNN, and classification layer.

Step4: The MTCN model with randomly initialized parameters is pre-trained and verified in the source domain dataset.

Step5: The classifier of the MTCN model is modified to adapt to the target task.

Step6: The parameters of the MTCN model after pre-trained are fine-tuned in the training dataset in the target domain and verified its performance in the testing dataset.

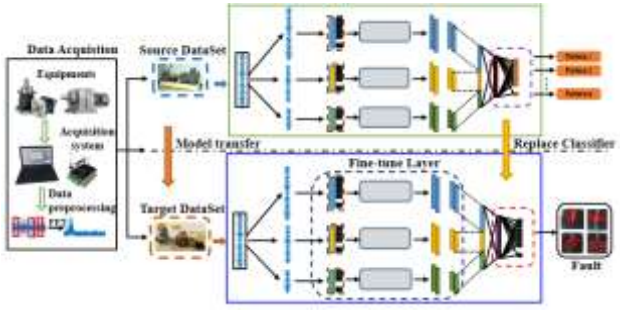


Fig. 4: The flowchart of the proposed model for gearbox fault diagnosis

(a) (b)

Fig. 5: Experiment platform of CWRU bearing dataset. (a) Experiment rig; (b) Schematic illustration

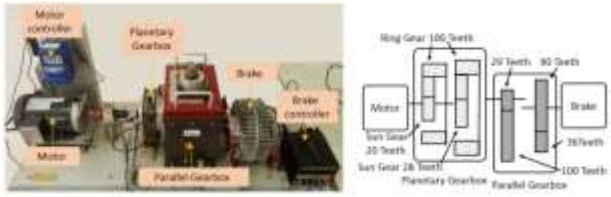
4 Experiment

In this section, the experiments to verify the performance and effectiveness of the proposed model and the analysis of the comparative results are elaborated on in detail.

4.1 Data description and Experiment Setup

The experiment data is from two datasets of different mechanical facilities, including the Case Western Reserve University (CWRU) bearing dataset, [22] and the Southeast University (SEU) gearbox dataset, [23].

The CWRU bearing dataset is a benchmark dataset that has been extensively applied in the area of fault diagnosis. As shown in Fig. 5, the experiment platform of the CWRU bearing dataset is mainly composed of a motor, torque transducer, and dynamometer. Motor is used to provide power and change workloads. Torque transducer converts the physical change of torque into an accurate electrical signal. Dynamometer is used to measure power. The accelerometer is installed on the drive end and fan end of the motor housing and the vibration signal data is collected with a sampling frequency of 12kHz. Through electric discharge machining (EDM), the single-point faults are introduced on the bearing inner ring, bearing outer ring, and rolling elements with fault diameters of 0.007 inches, 0.014 inches, and 0.021 inches, respectively. According to different fault locations, bearing fault types can be divided into an inner-race fault (IF), Outer race fault (OF), and rolling body fault (Ball fault). All bearings were tested at four different motor loads (0, 1, 2, and 3hp) corresponding to bearing speeds of 1797, 1772, 1750, and 1730rpm, respectively. Therefore, there are ten types for each load condition, including nine fault types and one normal type.



(a) (b)

Fig. 6: Experiment platform of SEU gearbox dataset. (a) Experiment rig; (b) Schematic illustration

The gearbox dataset provided by Shao et al., [23] is acquired through Spectra Quest's Drivetrain Dynamic Simulator (DDS). As shown in Fig. 6, the experiment platform is composed of a brake, motor, parallel gearbox, planetary gearbox, motor controller, and brake controller. The brake and brake controller are used for braking. The motor and motor controller are used to change the running load state. For the parallel gearbox, each gear rotates with its fixed central axis, while the planetary gearbox is composed of a planetary carrier, inner ring gear, sun gear, and several planetary gears. To measure the load, a torque sensor is installed between the motor and the planetary gearbox. On the surface of the DDS experimental platform, seven vibrating sensors are installed to collect the vibration signal data. Six sensors are used to measure the vibration of the planetary gearbox and the parallel gearbox in x, y, and z directions respectively and one sensor is adopted to measure the driving motor. Moreover, the speed and load configuration are set as 20Hz-0V and 30Hz-2V. As listed in Table 1, the gearbox has two fault components which are bearing and gear and they all have four fault types separately. As a result, there are eight fault types and one health type in the gearbox dataset.

Table 1. Fault type information of gearbox dataset

Fault component	Fault type	Description
Gear	Chipped	Crack occurs in the feet.
	Miss	One of feet is missed.
	Root	Crack roots in the feet.
	Surface	Wear occurs in the surface.
Bearing	Ball	Crack occurs in the ball.
	Outer	Crack occurs in outer ring.
	Inner	Crack occurs in inner ring.
	Combo	Crack occurs in inner and outer ring.

In order to train the proposed model by using transfer learning strategy, the dataset mentioned above is applied as the source domain dataset and target domain dataset respectively. To be specific, the CWRU-bearing dataset is set as the source domain dataset. The amount of each fault type in the CWRU bearing dataset is 2400 and 1600 samples of it are randomly selected to use as a training set and the rest samples are adopted as a testing set. The SEU gearbox dataset is set as the target domain dataset. For the 9 types in the SEU gearbox dataset, each of them has 300 samples. 200 samples are randomly selected to constitute the training set and the rest are used as the testing set. For each data sample in the dataset, it contains 1024 data points which is the half result obtained through FFT.

In the pre-training stage, the Adam optimizer with a learning rate of 1e-4 is adopted to train the proposed model in the source domain dataset. Moreover, the batch size is set as 600 and the training epoch is set as 200. In the fine-tuning stage, the classifier of the proposed model is modified to adapt to the target task according to the number of fault type in the target domain dataset. Similarly, the Adam optimizer is used to fine-tune the parameters of the proposed model and the learning rate is set as 1e-4. The train epoch is set as 50 and the batch size is set as 100 in this stage. Additionally, to avoid the impact of the difference from the running environment, all computations in the experiments are performed on the same device condition with Windows 10, Intel core I7-11700K, 32GB RAM, RTX 3080 Ti GPU 12G, Pytorch 1.10.0 and Python 3.6.

4.2 Diagnosis Results Analysis

In order to verify the performance and effectiveness of the proposed model and to avoid the randomness and occasionality of the experiment, ten trials are carried out for the proposed model to obtain the diagnosis results.

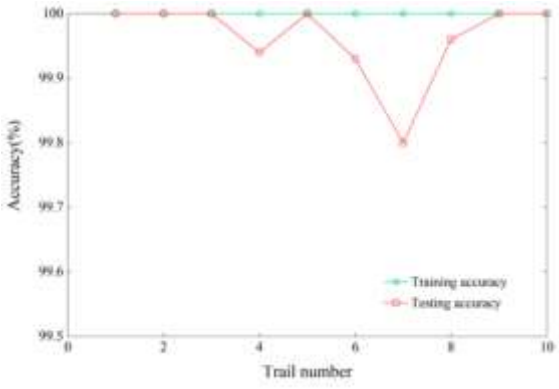


Fig. 7: Diagnosis results of 10 trials using the proposed model

Fig. 7 shows the details of the training and testing diagnosis results from the ten trials in the

target domain dataset of the transfer experiment. To be specific, the training diagnosis accuracy of the ten trials all reach 100% and the smallest diagnosis accuracy of the testing is still up to 99.80%. It can be clearly seen that the proposed model obtains outstandingly high accuracy both in training and testing in the target domain dataset.

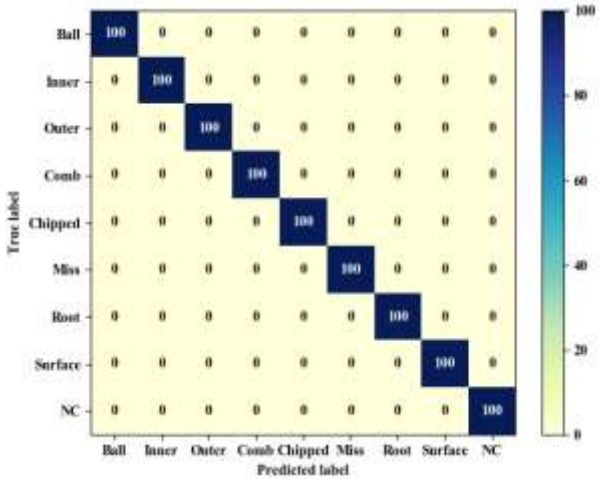


Fig. 8 Confusion matrix of the trial with the smallest testing accuracy

Moreover, the confusion matrix with the classification accuracy and misclassification error of each fault type from the trial with the smallest testing accuracy is presented in Fig.8. Its horizontal axis refers to the predicted label and the ordinate axis represents the actual label of the testing sample. The result shows that the classification accuracy of each fault type is up to 100% and there is no misclassification error, which demonstrates that the proposed model processes excellent recognition capability for fault patterns.

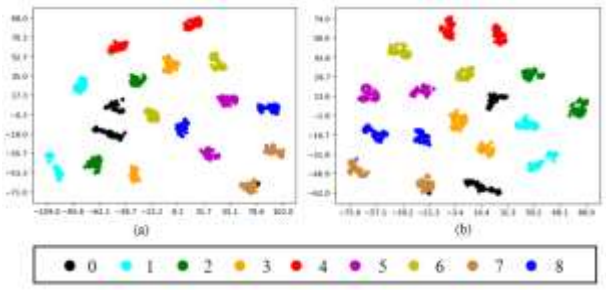


Fig. 9: Features visualizations of the feature learning layer of the proposed model via the t-SNE. (a) Transformer encoder layer (b) Convolution layer

Furthermore, in order to visually explain the adaptive feature learning ability of the proposed model, the learned features are visualized by using the t-distributed stochastic neighbor embedding (t-SNE) which can effectively project the dimension of high-level features into lower dimensions and realize feature visualization. Fig. 9 presents the feature visualization results of the transformer encoder layer and convolution layer. There are nine colors in the figure, representing nine fault types of the gearbox respectively. Because each fault type comes from two different working conditions, there are two clusters for each fault type. It can be seen that the output features of the Transformer encoder layer are enough to be divisible and distinguishable and the convolution layer enhances the effect, which confirms that the proposed model has an excellent ability to extract useful features and can distinguish different fault types effectively.

Table 2. Results with different parameter settings

No.	batch size	Training epoch	Learning rate	Training accuracy (%)	Testing accuracy (%
1			0.0001	100	99.56
2	500	150	0.001	100	99.25
3			0.01	99.99	98.94
4			0.0001	100	99.97
5	600	200	0.001	100	99.70
6			0.01	99.96	99.48
7			0.0001	100	100
8	700	250	0.001	100	99.70
9			0.01	99.93	99.42

On the other hand, in order to illustrate the influence of different parameter settings on the diagnosis accuracy of the proposed method, nine groups of different experiments are carried out on the proposed method. As shown in Table 2, the training accuracy is relatively high under different parameter settings, but when the training batch size, epoch, and learning rate are relatively small, the test accuracy will be reduced due to its influence. Appropriately increasing the batch size and epoch of training can improve the test accuracy, but when it is increased to a certain extent, it will have little impact on the test accuracy. For example, with the same learning rate, the training batch size and epoch of No.7 have increased a lot compared with No.5, but the final test accuracy has not been improved.

Table 3. The average diagnosis accuracy(%) and the standard deviation with different SNR values.

Methods	SNR(dB)							
	-6	-4	-2	0	2	4	6	Not added
WDCNN[24]	61.07±8.30	72.99±6.31	80.98±6.49	87.21±5.99	91.34±4.94	94.85±3.08	97.41±1.98	99.05±0.05
TCNN[25]	69.54±9.18	79.13±8.15	87.23±6.89	91.62±5.26	95.22±3.28	97.37±2.02	98.43±1.49	99.91±0.05
MSCNN[18]	82.57±1.41	88.48±1.48	92.48±1.02	95.14±0.56	97.03±0.35	98.15±0.30	98.57±0.29	99.21±0.03
TCN[10]	91.17±0.65	95.41±0.48	97.57±0.37	98.88±0.29	99.45±0.18	99.78±0.10	99.88±0.05	99.98±0.02
The proposed method	**92.98±0.81**	**96.17±0.36**	**98.06±0.24**	**99.01±0.15**	**99.42±0.18**	**99.53±0.23**	**99.77±0.10**	**99.96±0.06**

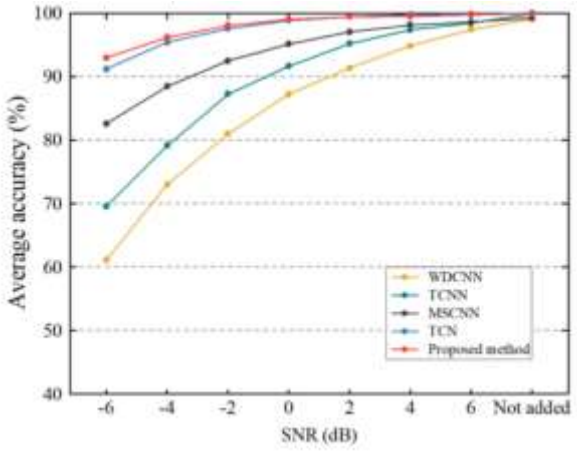


Fig. 10: Comparisons of diagnosis accuracy with different SNR value.

In the end, to study the robustness and effectiveness of the proposed model under a noise environment, eight groups of control experiments are carried out. To simulate the noisy sample, each vibration signal sample is added with the additive Gaussian white noise of different signal-to-noise ratio(SNR) which is defined as follows:

$$SNR = 10\log_{10}(\frac{P_{signal}}{P_{noise}}) \quad (10)$$

where P_{signal} and P_{noise} are the power of the signal and noise respectively. Generally, the smaller the SNR value, the greater the impact of noise on the signal, so the worse the effect of fault diagnosis. Noise with SNR ranging from - 6 to 6 is added to the vibration signal samples of the control experiment from the first group to the seventh group, and the last group without adding noise to represent the fault diagnosis under normal conditions.

In the comparative experiments, four different excellent fault diagnosis methods are adopted to compare with the proposed method. The details of these four methods can be summarized as: (1) WDCNN, [24],: Raw vibration signals are used as input and the wide kernels are applied in the first convolutional layer for extracting features and suppressing high-frequency noise. (2) TCNN, [25],: A modified version of WDCNN, where dropout techniques, kernel numbers, and fully-connected layers are added. (3) MSCNN, [18]: Fault features are extracted from raw vibration signals at different scales in a parallel way by combining multi-scale learning with CNN. (4) TCN, [10]: Transformer encoder with two layers is combined with three same CNN blocks to realize fault diagnosis.

For each method, it is executed 10 times, and the average value of the 10 trials is taken as the final result. As shown in Table 3 and Fig. 10, the proposed model and the comparative model all perform outstanding results reaching to 99% when the vibration signals are not added with noise. However, as the SNR value gradually decreases, the performance difference between models becomes more and more obvious. The diagnosis accuracy of WDCNN and TCNN declines remarkably. The MSCNN performs well when SNR is - 4 dB and - 2 dB, but it still performs poorly when SNR is - 6 dB. The proposed method and MSCNN all use multi-scale learning, but the accuracy of the proposed model doesn't change significantly with the decrease of SNR. Compared with the state-of-art fault diagnosis model TCN, the proposed method has fewer layers and blocks and it obtains superior performance of 92.98% accuracy in extremely noisy conditions of -6 dB which is higher than TCN. As a result, the comparison experiment demonstrates that the proposed model not only possesses high-accuracy fault diagnosis for the gearbox but also can effectively extract fault features under a noise environment.

5 Conclusion

To realize the high accuracy of gearbox fault diagnosis and to deal with the limitation of the rarity of fault samples in certain cases, a novel multi-scale Transformer convolution network based on a transfer learning strategy named MTCN is proposed creatively in this paper. The coarse-grained procedure incorporated in the proposed model can not only enable the MTCN to learn complementary

multi-scale features but also can filter high-frequency perturbations and random noises to some extent. The proposed MTCN can extract rich fault features and perform accurate pattern recognition through the superior learning capability of the Transformer and enhance the efficiency of training by using a transfer learning strategy. Through the analysis of the comparative experiments, the result demonstrates that the proposed model achieves a higher fault diagnosis accuracy and is robust to noise interference. Since the fault samples in practice are probably unbalanced, the future work of this research is to expand the capability of the proposed MTCN on unbalanced datasets.

References:

[1] Y. G. Lei, J. Lin, M. J. Zuo, Z. J. He, Condition monitoring and fault diagnosis of planetary gearboxes: A review, *Measurement*, Vol. 48, 2014, pp. 292-305.

[2] X. H. Liang, M. J. Zuo, Z. P. Feng, Dynamic modeling of gearbox faults: A review, *Mechanical Systems and Signal Processing*, Vol. 98, 2018, pp. 852-876.

[3] Z. J. Shen, X. F. Chen, X. L. Zhang, Z. J. He, A novel intelligent gear fault diagnosis model based on EMD and multi-class TSVM, *Measurement*, Vol. 45, 2012, pp. 30-40.

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

[5] F. F. Chen, B. P. Tang, R. X. Chen, A novel fault diagnosis model for gearbox based on wavelet support vector machine with immune genetic algorithm, *Measurement*, Vol. 46, 2013, pp. 220-232.

[6] R. Jiang, J. Yu, V. L. Makis, Optimal Bayesian estimation and control scheme for gear shaft fault detection, *Computer & Industrial Engineering*, Vol. 63, pp. 754-762.

[7] G. Q. Qiu, Y. K. Gu, Q. Cai, A deep convolutional neural networks model for intelligent fault diagnosis of a gearbox under different operational conditions, *Measurement*, Vol. 145, 2019, pp. 94-107.

[8] Y. Zhang, T. Zhou, X. Huang, L. Cao, and Q. Zhou, Fault diagnosis of rotating machinery based on recurrent neural networks, *Measurement*, Vol. 171, 2021, pp. 108774.

[9] J. B. Yu, G. L. Liu, Knowledge extraction and insertion to deep belief network for gearbox fault diagnosis, *Knowledge-Based Systems*, Vol. 197, 2020, pp. 105-118.

[10] X. L. Pei, X. Y. Zheng, J. L. Wu, Rotating Machinery Fault Diagnosis Through a Transformer Convolution Network Subjected to Transfer Learning, *IEEE Transactions on Instrumentation and Measurement*, Vol. 70, 2021, pp. 1-11.

[11] X. Y. Zheng, Z. Y. Ye, J. L. Wu, A CNN-ABiGRU method for Gearbox Fault Diagnosis, *International journal of circuits, systems and signal processing*, Vol. 16, 2022, pp. 440-446.

[12] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, Attention Is All You Need, arXiv preprint arXiv:1706.03762, 2017.

[13] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, N. Houlsby, An image is worth 16x16 words: transformers for image recognition at scale, arXiv preprint arXiv:2010.11929, 2020.

[14] K. Lin, L. J. Wang, Z. C. Liu, End-to-End Human Pose and Mesh Reconstruction with Transformers, arXiv preprint arXiv:2012.09760, 2021.

[15] A. Prakash, K. Chitta, A. Geiger, Multi-Modal Fusion Transformer for End-to-End Autonomous Driving, arXiv preprint arXiv:2104.09224, 2021.

[16] Z. Dai, Z. Yang, Y. Yang, J. Carbonell, Q. V. Le, R. Salakhutdi, Transformer-xl: Attentive language models beyond a fixed-length context, arXiv preprint arXiv:1901.02860, 2019.

[17] J. Devlin, M. W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, arXiv preprint arXiv:1810.04805, 2019.

[18] G. Jiang, H. He, J. Yan and P. Xie, "Multiscale Convolutional Neural Networks for Fault Diagnosis of Wind Turbine Gearbox," *IEEE Transactions on Industrial Electronics*, Vol. 66, 2019, pp. 3196-3207.

[19] H. Liu and M. Han, A fault diagnosis method based on local mean decomposition and multi-scale entropy for roller bearings, *Mechanism and Machine Theory*, Vol. 75, 2014, pp. 67-78.

[20] J. Zheng, H. Pan, and J. Cheng, Rolling bearing fault detection and diagnosis based on composite multiscale fuzzy entropy and ensemble support vector machines, *Mechanical Systems and Signal Processing*, Vol. 85, 2017, pp. 746-759.

[21] Z. Chen, K. Gryllias, W. Li, Intelligent Fault Diagnosis for Rotary Machinery Using Transferable Convolutional Neural Network, *IEEE Transactions on Industrial Informatics*, Vol. 16, 2020, pp. 339-349.

[22] W. A. Smith, R. B. Randall, Rolling element bearing diagnostics using the case western reserve university data: A benchmark study, *Mechanical Systems and Signal Processing*, Vol. 64-65, 2015, pp. 100-131.

[23] S. Shao, S. McAleer, R. Yan, P. Baldi, Highly accurate machine fault diagnosis using deep transfer learning, *IEEE Transactions on Industrial Informatics*, Vol. 15, No. 4, 2019, pp. 2446-2455.

[24] W. Zhang, G. Peng, C. Li, Y. Chen, Z. Zhang, A New Deep Learning Model for Fault Diagnosis with Good Anti-Noise and Domain Adaptation Ability on Raw Vibration Signals, *Sensors*, Vol. 17, No. 3, 2017, 425-445.

[25] Z. Chen, K. Gryllias, W. Li, Intelligent Fault Diagnosis for Rotary Machinery Using Transferable Convolutional Neural Network, *IEEE Transactions on Industrial Informatics*, Vol.16, No.1, 2020, 339-349.