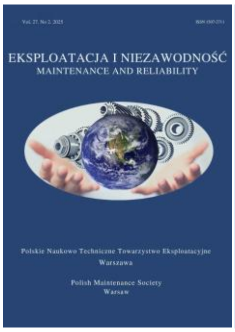


Eksploatacja i Niezawodnosc – Maintenance and Reliability

Volume 27 (2025), Issue 2

journal homepage: http://www.ein.org.pl

Article citation info:
Ge J, Zhou D, Zhuang X, Wang X, Xiang J, Integrated Multi-Scale Analysis and Advanced Prototype Model for Early Detection of Gearbox Failures Using Infrared Thermal Image Data Under Dynamic Conditions, Eksploatacja i Niezawodnosc – Maintenance and Reliability 2025: 27(2) http://doi.org/10.17531/ein/194151

Integrated Multi-Scale Analysis and Advanced Prototype Model for Early Detection of Gearbox Failures Using Infrared Thermal Image Data Under Dynamic Conditions

Indexed by: Web of Science Group

Jian Ge[a], Di Zhou[a], Xiao Zhuang[a,*], Xiaomin Wang[a], Jiawei Xiang[a]

[a] College of Mechanical and Electrical Engineering, WenZhou University, China

Highlights

- Prototype networks were first applied to infrared thermal images for fault analysis.
- A novel multi-scale module is constructed and incorporated into ProtoNet model.
- The proposed MSPNet can solve the gearbox fault diagnosis with small samples.

Abstract

To address the issues of low data quality and poor adaptability in deep learning methods for infrared image analysis in gearbox fault diagnosis, this paper introduces an enhanced deep prototype network model (MSPNet). This model employs a multi-scale strategy to improve fault diagnosis accuracy and algorithm generalization, especially with small sample sizes. First, infrared image data of six fault types under five operating conditions are collected using a rotating test bed. Gaussian noise is added to simulate real operating conditions. Next, the fault data are processed using a multiscale module to extract multiscale fault features and reduce feature value fluctuations. Finally, the proposed model is used to process the image data and is experimentally compared with five other algorithms. The experimental results demonstrate that the proposed method outperforms the other algorithms under various operating conditions.

Keywords

gearbox, fault diagnosis, infrared thermal images, small samples, variable working condition

1. Introduction

The safety and reliability of mechanical equipment is crucial in industrial production, especially for equipment with complex structure and large scale [1]. As a kind of mechanical equipment widely used in various fields, the operational performance of gearboxes largely depends on the state of gears. Therefore, in order to avoid economic losses and safety hazards caused by gearbox failures, the development of intelligent and advanced fault diagnosis systems for monitoring and diagnosing the condition of gearboxes is and its necessary [2].

Current mainstream gearbox fault diagnosis techniques include vibration detection[3], acoustic detection[4], and infrared thermal image detection. Compared to vibration and acoustic detection, infrared thermal image analysis offers significant advantages. It requires no surface preparation of the gearbox and does not necessitate sensor installation, thus avoiding potential interference and damage[5]. In contrast, vibration analysis and acoustic analysis typically require that

(*) Corresponding author.
E-mail addresses: J. Ge (ORCID: 0009-0008-7514-0450) chinagejian@outlook.com, D. Zhou (ORCID: 0000-0002-7798-3098) zhoudi@wzu.edu.cn, X. Zhuang (ORCID: 0000-0002-4131-209X) zhuangxiao@wzu.edu.cn, X. Wang (ORCID: 0009-0009-2796-3978) xmwang@stu.wzu.edu.cn, J. Xiang (ORCID: 0000-0003-4028-985X) jwxiang@wzu.edu.cn

the collected data be collected and analysed over a period of time, which may take longer to detect equipment failures. Therefore, infrared thermal image analysis is a promising technique for gearbox diagnosis and has attracted much attention in recent years[6].

Deep learning techniques have enabled the use of a wide range of models and algorithms across different fields[7-9]. Data-driven fault diagnosis methods have also emerged and are gradually being applied to gearbox fault diagnosis. However, the quantity and quality of data significantly impact the reliability of these techniques in practical applications. In real industrial environments, gearboxes typically operate normally, resulting in scarce and hard-to-obtain fault data[10, 11]. This data limitation constrains the ability to capture fault characteristics, reducing diagnostic accuracy. The problem is even more pronounced in complex and extreme industrial scenarios, where the degradation of accuracy due to data limitations is more severe[12]. Therefore, developing effective fault diagnosis methods for small-sample data remains a significant challenge.

To address this challenge, small sample methods based on transfer learning have received great attention in fault diagnosis. It is very popular for dealing with small sample problems in the fields of image classification and fault diagnosis. For example, Zhang et al. proposed a migration learning based approach to realize fault diagnosis of diesel engines [13]. Chen et al. made FRA slip phase and series migration learning methods to solve the fault problem of mechanical deformation of transformer windings [14]. Migration and sharing of knowledge by finding similarities or correlations between source and target domains is the reason why migration learning has shown excellent performance in fault diagnosis. However, in practice, this is difficult to do because it is hard to obtain enough auxiliary data to support the migration learning technique.

Generative Adversarial Networks are also commonly used to deal with small sample problems. It consists of two neural networks: a generator and a discriminator. The purpose of the generator is to generate some samples similar to real data from a random noise vector, and the purpose of the discriminator is to distinguish whether the input data is real or fake generated by the generator. For example, Zhang et al. utilized a generative adversarial network for fault detection in hot strip rolling process conditions [15]. It is worth noting that the quality of the generated data may be poor, which leads to bias under the actual working of the model.

Small-sample methods based on meta-learning are highly flexible and generalizable by using a small number of training samples to achieve relevant tasks. Meta-learning methods have great potential for engineering applications and have become a research hotspot in the field of deep learning in recent years[16]. Among the meta-learning models, the prototype network (ProtoNet) model has attracted a lot of attention from scholars in recent years[17], which classifies by learning the metric space with a metric classifier of the distance from the class prototype to the classification. ProtoNet can incorporate a priori knowledge into the nature of the embedding space, so that similar samples are clustered in the embedding space and dissimilar samples are dispersed in the embedding space, which improves the model's generalization ability and robustness. Currently, prototype network-based methods are mainly used to analyse one-dimensional data, while they are less applied in two-dimensional data. In addition, the prototype network uses tandem layers mixed with multilayer convolutional layers in processing 2D data fault feature extraction, and its multi-scale information mining ability is insufficient. In order to improve the deep information mining ability of the prototype network at different scales, as well as the ability to understand the detailed and global information of images, a new MSPNet model is proposed in this paper.

The main contributions of this paper are as follows:

(1) Aiming at the data acquisition situation with small number of labeled samples, the prototype network is introduced for fault diagnosis of gearbox based on infrared images with small samples.

(2) A multi-scale feature extraction module is introduced into the prototype network, which can simultaneously capture the details and global information of the image. By utilizing multi-scale module, the fault features in small samples can be more comprehensively extracted and analyzed.

(3) Compared with other benchmark models, the proposed MSPNet model can finally realized high precision fault diagnosis of gearbox by using infrared thermal images with

small samples.

2. Basic Theory of Prototype Network

Prototype network, as a metric-based meta-learning method for small sample learning, can reduce the influence of overfitting problem on the model, and is an effective image classification network under small sample conditions, as shown in Fig.1.

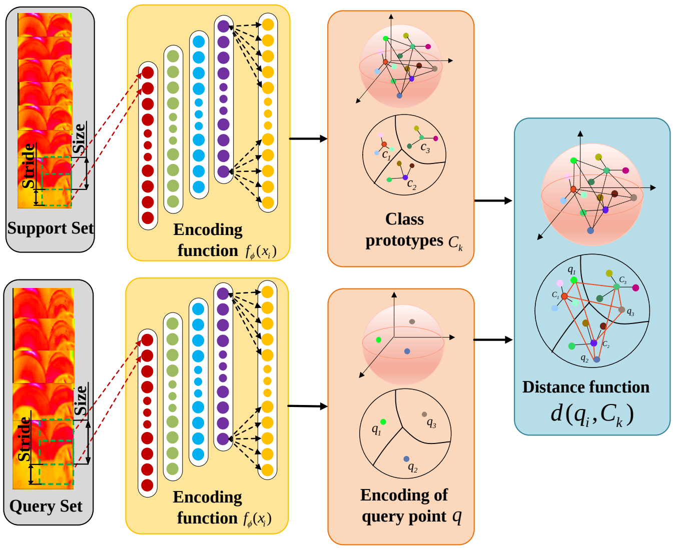


Fig. 1. Structure of the prototype network.

First, a support set containing a small number of samples from each category and a query set containing the samples to be categorized are defined. Then, the support set is passed into the neural network for feature extraction to learn the nonlinear mapping position of the support set in space. According to the encoding function $f_\phi(x_i)$, the prototype representation of each category is shown (1).

$$C_k = \frac{1}{|S_k|}\sum_{(x_i,y_i)} f_\phi(xi) \tag{1}$$

where, *k* is the class of the prototype. *x* is the data vector and *y* is the categorical label. S_k is the number of support sets. c is the position of the prototype of the class.

The prototype of each class is obtained according to Eq.1. Finally, the query set can be passed into the encoding function to obtain the mapping position and calculate the probability that the sample belongs to category *K*. The calculation process is shown (2).

$$p_\phi(y=k|x) = \frac{exp\left(-d(f_\phi(x),c_k)\right)}{\sum k'\, exp\left(-d(f_\phi(x),c_{k'})\right)} \tag{2}$$

where, *K* is the category, *d* is the Euclidean distance function.$f_\varphi(x_i)$ is the coding function to obtain the features of the sample.

3. Methodology

3.1. Multi-Scale Modules

The convolution kernel of null convolution is a modification of the standard convolution kernel, by inserting null values between neighbouring parameters to obtain a larger sensory field, without additional computational parameters and computation. Assuming that the input feature map of the previous layer is *X*, the output feature map *y* after the null convolution operation can be derived from the following (3):

$$y[i] = \sum_{k=1}^{K} x[i + d\cdot(k-1)\cdot w[k]] \tag{3}$$

where, *d* is the null rate. *w* is the convolution kernel. *w* is the parameter. *k* is the size of the convolution kernel.

The fault image data is set to $X = \{x_{1,1}, x_{1,2}, x_{1,3}, \cdots x_{i,j}\}$, $1 \le i \le D, 1 \le j \le H$. $x_{i,j}$is the pixel value at location (i,j). The input feature map is divided into four branches and feature mapping is performed on each branch individually so that it slides back and forth to extract the local features of the image in a specific step size. As shown in (4).

$$F = \sum_{k=1}^{M}(w[k]\otimes x) + b[k] \tag{4}$$

where, *W*[*k*] denotes the convolutional kernel of the *k*-th and *x* is the feature map of the input. *b*[*k*] denotes the convolutional kernel offset of the kth. is the mapping output of the multi-scale convolutional kernel on each branch.

After the convolutional layers within each branch, a batch normalization layer and an activation function are used. Thus, the feature mapping can be obtained as. As shown in (5).

$$F_i = \delta(BN(F)) \tag{5}$$

where, *δ* is the activation function. BN is the batch normalization layer.

The difference between different channels is that each individual channel is a different scale of the feature map. Large scale corresponds to global information such as contours in the image, while small scale corresponds to local information such as details in the image. The feature maps are obtained by using dilated convolutions, and different scale information is obtained by controlling different dilation rates. The reasons for using multi-scale channels to extract features is that by extracting feature information from individual channels and then integrating these different features, the proposed model is able to generate a rich set of feature maps.

This multi-scale strategy can explore the data more comprehensively and capture subtle differences that may be overlooked by single-scale analysis. The integration of features at different scales helps to understand the failure mode more carefully, thereby enhancing the diagnostic ability of our proposed model.

By performing multi-level feature fusion on each output branch, richer and more diverse feature mappings can be generated. This feature refactoring strategy enables the extraction of new features, which significantly improves the overall feature description capability of the module. Specifically, fusing features at different levels enables the model to capture feature information at different scales and levels, as shown in (6):

$$T_M = \begin{cases} T_1 = F_1 \\ T_2 = F_1 + F_2 \\ T_3 = F_1 + F_2 + F_3 \\ T_4 = F_1 + F_2 + F_3 + F_4 \end{cases}, M = 1,2,3,4 \qquad (6)$$

where, F is the feature of each branch. T is the fused feature. M is the number of each branch. Concat is a feature fusion function.

The acquired multi-level features are stitched together to form a new feature map, as shown in (7):

$$F_{output} = concat[T_1, T_2, T_3, T_4] \qquad (7)$$

Where T_1, T_2, T_3, T_4 denotes the features of the first, second, third and fourth branches respectively. F_{output} is the fused feature map. Since each level has the same importance, here, T_1, T_2, T_3, T_4 are given the same weight.

Finally, the residual connection is added, as shown in (8):

$$\tilde{F}_{output} = Add(F_{output}, F) \qquad (8)$$

where, F is the original features of the previous layer. $\tilde{F}_{output}$ is the final output fused feature map, as shown in Fig.2.

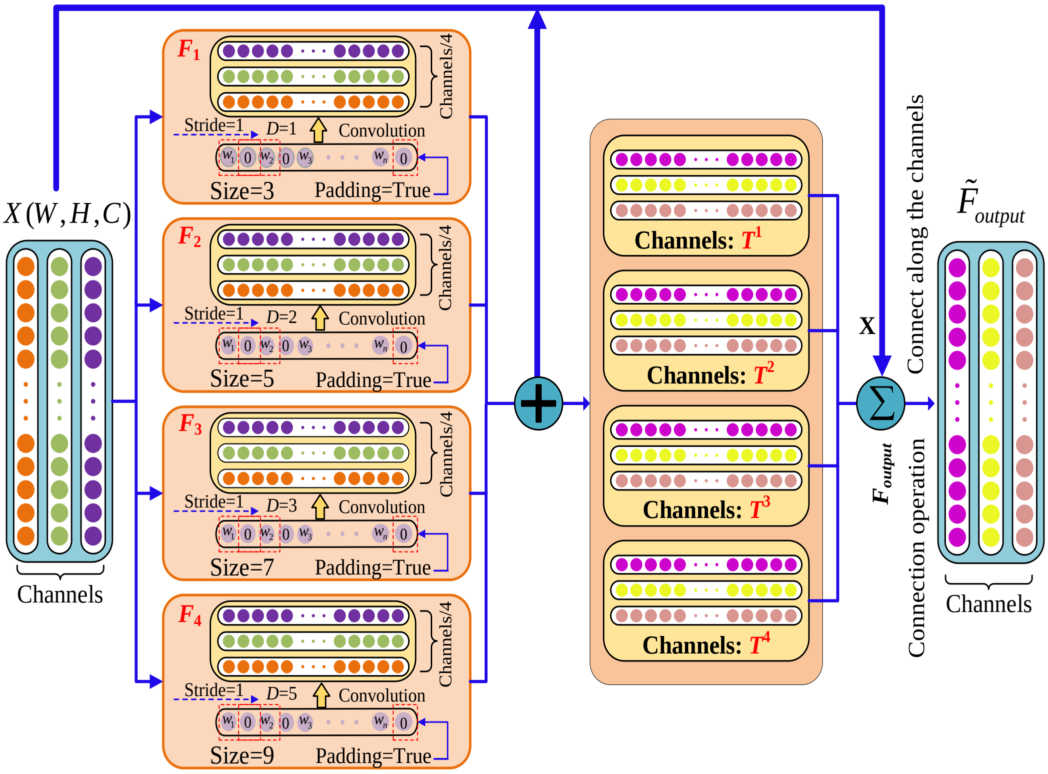


Fig. 2. Diagram of the multi-scale module.

3.2. Activation Function

In the multi-scale module, the activation function also has an effect on the features of the output, now in this paper Exponential Linear Unit (ELU) is an enhancement of RELU. ELU activation function is smooth and the negative part is in exponential form, which prevents neurons from dying, and also reduces the bias of the input distributions. ELU activation function does not saturate, the gradient does not disappear, and training is the formula of ELU is as (9):

$$f(x) = \begin{cases} a(e^x - 1) & when(x \le 0) \\ x & when(x > 0) \end{cases} \qquad (9)$$

3.3. Classification using Softmax

The Softmax function is used to calculate the weights of the weighted multi-scale feature maps with the following formula.

$$\beta_i = soft\,max(\mathrm{P}_i) = \frac{e^{P_j}}{\sum_{j=1}^{s} e^{P_j}} \tag{10}$$

$$\sum_{i=1}^{s} \beta_i = 1 \tag{11}$$

Assuming that there are n classes in each randomly drawn sample, the output probability of calculating the class as j is x. Thus, the output x is the label of maximum probability.

$$O_j = \frac{e^{(\theta(j)x)}}{\sum_{j=1}^{n} e^{(\theta(j)x)}}, j = 1,2,3, \dots n. \tag{12}$$

In the training sample, the continuously learned x is a learnable classification layer model parameter and $\sum_{j=1}^{n} O_j = 1$ is the sum of all output probabilities of 1.

3.4. Fault Diagnosis Based on MSPNet

In this paper, a gearbox fault diagnosis framework is proposed based on MSPNet and infrared images, as shown in Fig.3. It includes three steps:

(1) Data acquisition: an infrared camera is used to acquire infrared images of gearboxes, and the acquired infrared image dataset is divided into a support set, a query set and a test set. The support set and query set are used to train the fault prototype and model parameters, and the query set is used to verify the accuracy of the model.

(2) Feature extraction: the support set and query set are input into the proposed MSPNet for feature extraction, and the model is trained with forward propagation algorithm and back propagation algorithm.

(3) Pattern Recognition. The test set is nonlinearly mapped to obtain the feature space location, and the test prototype is fault classified using Softmax classifier.

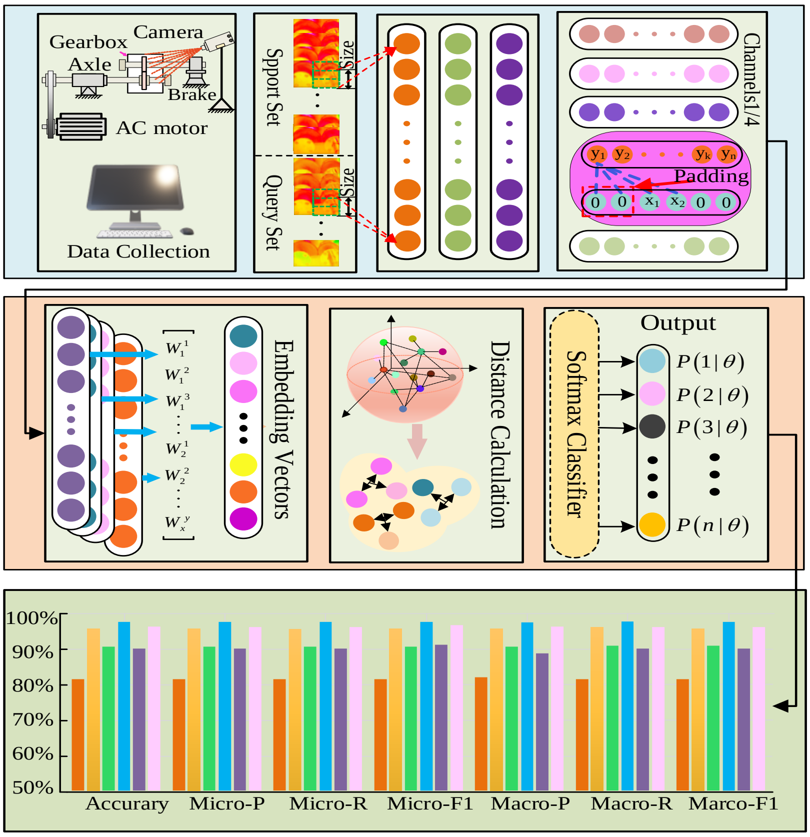


Fig. 3. Flowchart of the proposed MSPNet model for gearbox fault diagnosis.

4. Results and discussion

In this section, we conduct the gearbox failure experiments on the rotating machinery testing rig. Fig.4 shows the schematic structure of the experimental test rig. The experimental platform consists of an AC motor, a belt, a rotating shaft, a gearbox and an infrared camera. The infrared camera is mounted in front of the experimental platform aligned with the gearbox to collect infrared images, and the rotational speed is set to 1500 rpm. The main purpose of this paper is to explore the feasibility of the infrared thermal heat map and small sample model for gearbox fault diagnosis. Considering the influence of lubricant fluid during infrared images collection, the process management of the experiment is standardized as much as possible. The type of lubricant fluid used here is set as Mobile-EP 220 with a kinematic viscosity of 220 cSt@40°C in all tests. By standardizing the lubrication conditions, it is ensured that the effect of the lubricant on the gearbox performance remained consistent during different operation condition of different gearbox failure, thereby minimizing its interference with the heat map results.

In the whole experiment, loads of 0Nm, 2Nm, 3Nm, 3.7Nm and 4.2Nm are added respectively. Considering that the temperature is also influenced by the load, heat dissipation speed, ambient temperature, etc., the load in each set of experiment is set as the constant value. Besides that, each experiment is conducted in a temperature-controlled laboratory with the ambient temperature setting to 18°C. The reasons for setting the ambient temperature to 18°C is to ensure the consistency of experimental conditions, reduce thermal noise interference, and optimize equipment performance. In each set of experiment, the load and thermal conditions are kept constant to avoid influences other than the fault. Besides that, In order to minimize the influence of the energy losses and the efficiency of heat transfer on the experimental results of this paper, all experiments are performed under equal conditions. The operation of each fault is at the same environment temperature. The infrared images of gearbox are collected after the gearbox runs for 30 minutes. The internal gear parameters of the gearbox are shown in Table 1. The experiments in this paper are implemented based on Windows 64 kernel system, PyTorch framework and Python 3.8. The GPU used here is Nvidia RTX3060.

There are six types of gearbox faults including Driving Gear Tooth Break, Driving Gear Tooth Crack, Driving Gear Tooth Spalling, Driven Gear Tooth Break and Driven Gear Tooth Break, Driving Gear Tooth Spalling and Driven Gear Tooth Break, as well as a failure free (Gear without Failure). Gear infrared thermal images are acquired under five different operating conditions. The fault classification is shown in Table 3 below. In this paper, we randomly select some images under different loads to form the sample dataset. As shown in Table 4, under each operation condition, 20 images are selected as samples for each fault type, of which 16 samples are utilized as training data and 4 samples are utilized as testing data. In addition, noise is added into the images to conform to the actual engineering applications. An infrared thermal image gearbox fault dataset is formed, and the fault samples of the infrared dataset are shown in Fig.5. The temperature variation range of the faulty gearbox during experiments is shown in Table 2. The real images of the corresponding failure states are shown in Fig. 6. The training samples occupy 80% of the total number of samples. In this paper, a 5-way-5-shot sampling method is used. Each time from the six categories of species are not repeated randomly selected five classes, each class in the image from multiple working conditions in the aggregate randomly selected 5 samples as the category prototype training samples multiple cycles of sampling training.

Table 1. Specifications of experimental gearboxes.

Parameter	Value
Number of driving gear teeth	55
Number of driven gear teeth	75

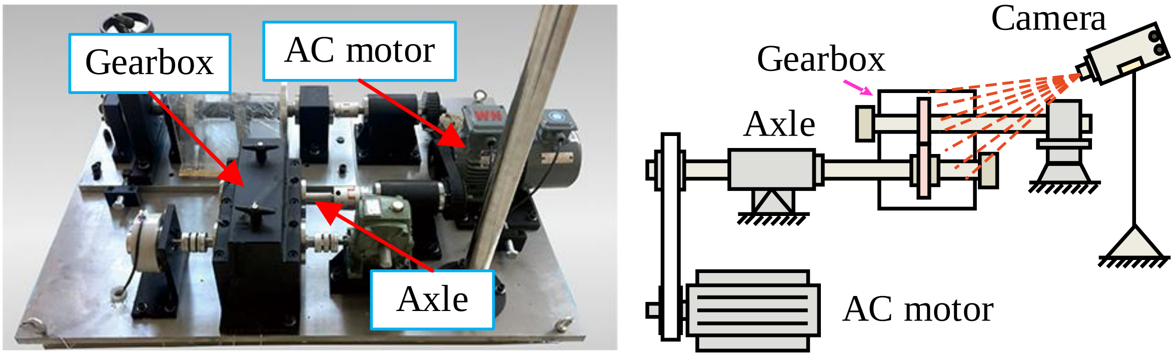


Fig. 4. (A) Rotating machinery troubleshooting test bed; (B) Simulation test bed arrangement.

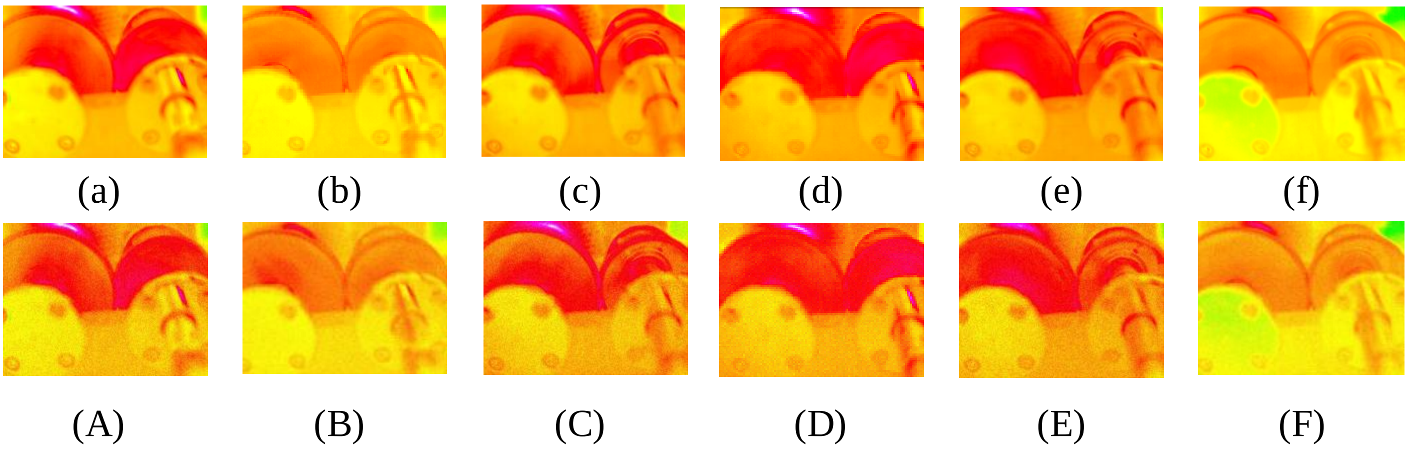


Fig. 5. Fault samples of the IR dataset. *a-f* are the original fault samples. A-F are the fault samples with Gaussian noise added.

Table 2. The Temperature Variation Range of the Faulty Gearbox during Experiments.

Health Conditions of Gearbox System	Temperature rises
Driving Gear Tooth Break	+13.73°C
Driving Gear Tooth Crack	+12.80°C
Driving Gear Tooth Spalling	+11.90°C
Driven Gear Tooth Break & Driving Gear Tooth Break	+14.50°C
Driving Gear Tooth Crack & Driven Gear Tooth Break	+13.50°C
Gear Without Failure	+9.85°C

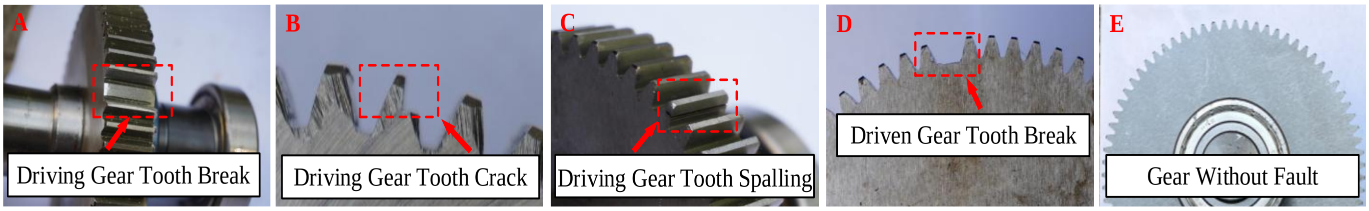


Fig. 6. The real images of different gear failures. C1 is corresponding to A; C2 is corresponding to B; C3 is corresponding to C; C4 is corresponding to A and D; C5 is corresponding to B and D; C6 is corresponding to E without fault.

Table 3. The Six Health Conditions of Gearbox System.

Health Conditions of Gearbox System	Labels of Conditions
Driving Gear Tooth Failure	Condition 1(C1)
Driving Gear Crack Failure	Condition 2(C2)
Driving Gear Pitting Failure	Condition 3(C3)
Driven Gear Tooth Breakage & Driving Gear Tooth Failure	Condition 4(C4)
Driving Gear Crack Failure & Driven Gear Tooth Breakage	Condition 5(C5)
Gear Without Failure	Condition 6(C6)

Table 4. Detailed Information of Datasets.

Type	The number of samples						Train/Test
	Mode 1	Mode 2	Mode 3	Mode 4	Mode 5	Mix All	
C1	16/4	16/4	16/4	16/4	16/4	100	80/20
C2	16/4	16/4	16/4	16/4	16/4	100	80/20
C3	16/4	16/4	16/4	16/4	16/4	100	80/20
C4	16/4	16/4	16/4	16/4	16/4	100	80/20
C5	16/4	16/4	16/4	16/4	16/4	100	80/20
C6	16/4	16/4	16/4	16/4	16/4	100	80/20

Table 5. Hyperparameters of MSPNet.

Layer name	Kernel size	Output dimension	Stride	Padding
Input	-	224×224×3	-	
Conv2d	7×7	112×112×64	2	3
MAxpool2d	3×3	56×56×64	2	1
MS1	-	56×56×256	-	-
(Conv2d BN Banch1,2,3,4 Conv2d, BN, ELU)	1×1 Batch-Norm 3,5,7,9 1×1 Batch-Norm	-	1 - 1,1,1,1 1 -	0 - 1,4,9,10 0 -
MS2	-	28×28×256	-	-
MS3	-	14×14×256	-	-
MS4	-	7×7×256	-	-
Dropout	0.5	-	-	-
Avgpool2d	-	1×1×256	1	0
Classifier		Euclidean distance Softmax		

4.1. Parameterization

In the process of constructing MSPNet, choosing appropriate hyperparameters can effectively improve the diagnosis accuracy, convergence speed and robustness. In general, the important hyperparameters in the model mainly include learning rate, convolution kernel and pooling kernel. In this paper, a technique named grid search mentioned in the literature is used to determine the architecture of MSPNet, as shown in Table 5 below. conv2d denotes the convolutional layer in the model, Maxpool2d denotes the pooling layer in the model, (7×7) denotes that the length of the convolutional kernel is 7 and the width is also 7, and (112×112×64) denotes that the length of the feature size of the output is 112, the width of 112 and dimension of 64, MSCA1 denotes the first multiscale residual module, and Banch1 is the first branch inside the first multiscale residual module. In addition, in order to prevent the explosion of the model gradient and the introduction of the BN layer, Dropout=0.5 is introduced to avoid overfitting, and the ELU activation function is used to ensure the learning efficiency. The initial learning rate of the model in this paper is 0.001, and it decreases once every 10 epochs. The decay rate is set to 0.1. The training epoch is 100.

4.2. Parameter settings of benchmark models

In this study, five different benchmark classification models were selected for comparison experiments, and the parameters of the benchmark models are shown in Table 6. The first model is ResNet, which has four residual blocks connected and each residual block consists of a convolutional layer, a normalization layer, and an activation layer[18]. The second model is EfficientNet, which consists of a stem, six main blocks, and a head, and each block contains a number of sub-blocks, each of which consists of an MBConv structure and a skip connection. The third model is ShuffleNet, which is composed of a stem, three blocks and a head, and the block part replaces the regular convolution with group convolution and deep convolution[20]. The fourth model is Vision Transformer, the main idea of ViT model is to slice the input image into 16x16 chunks (called patches) and then each patch is considered as a block which is fed into the Transformer for coding and classification[21]. The fifth model is RePnet model, which is composed of ResNet network with ProtoNet and the unprocessed Resnet is used as a feature extractor.

Table 6. Structure and parameter setting of benchmark models.

Model	Structure	Optimizer	Learn Rate
ResNet	{3,4,6,3}x{Conv2d(1,3,1)} Maxpooling2d(2),Relu	Adam	Lr=0.0001
EfficientNet	6x{MBConvBlock(3,1)} Maxpooling(2),BN(),SiLU	Adam	Lr=0.0003
ShuffleNet	3x{Conv2d(1,3,1)},max pooling(3),Conv2d(7),Relu	Adam	Lr=0.01,Cosine AnnealingLR=0.1
Vision Transformer	16xpatches,12xEncoder Block,	Adam	Lr=0.01, Cosine AnnealingLR=0.1
RePnet	{3,3,3}x{Conv2d(1,3,1)},BN(). Maxpooling2d(),Relu	Adam	Lr=0.0003,The decline rate is 0.2 for every 10 epochs

4.3. Evaluation indexes

We evaluate the performance of the proposed model in this paper through classification accuracy, precision, recall, and F1 value, and compare the proposed model with the benchmark model, which in turn verifies the validity and superiority of the proposed model where TP (True Positive), FN (False negative), TN (True negative), FP (False Positive) stand for True Positive, False Negative, True Negative and False Positive respectively. Among them, precision rate, recall rate and F1 value are calculated from partial and global indexes, respectively, and the higher value represents the better classification effect. The definitions of the above indexes are shown from (13) to (20).

$$Precision_l = \frac{TP_l}{TP_l+FP_l} \tag{13}$$

$$Recall_l = \frac{TP_l}{TP_l+FN_l} \tag{14}$$

$$Macro - R = \frac{\sum_{l=1}^{L} Recall_l}{L} \tag{15}$$

$$Macro - P = \frac{\sum_{l=1}^{L} Precision_l}{L} \tag{16}$$

$$Macro - F1 = \frac{2\cdot(Macro-P)\cdot(Macro-R)}{(Macro-P)+(Macro-R} \tag{17}$$

$$Micro - R = \frac{\sum_{l=1}^{L} TP}{\sum_{l=1}^{L} TP+\sum_{l=1}^{L} FN} \tag{18}$$

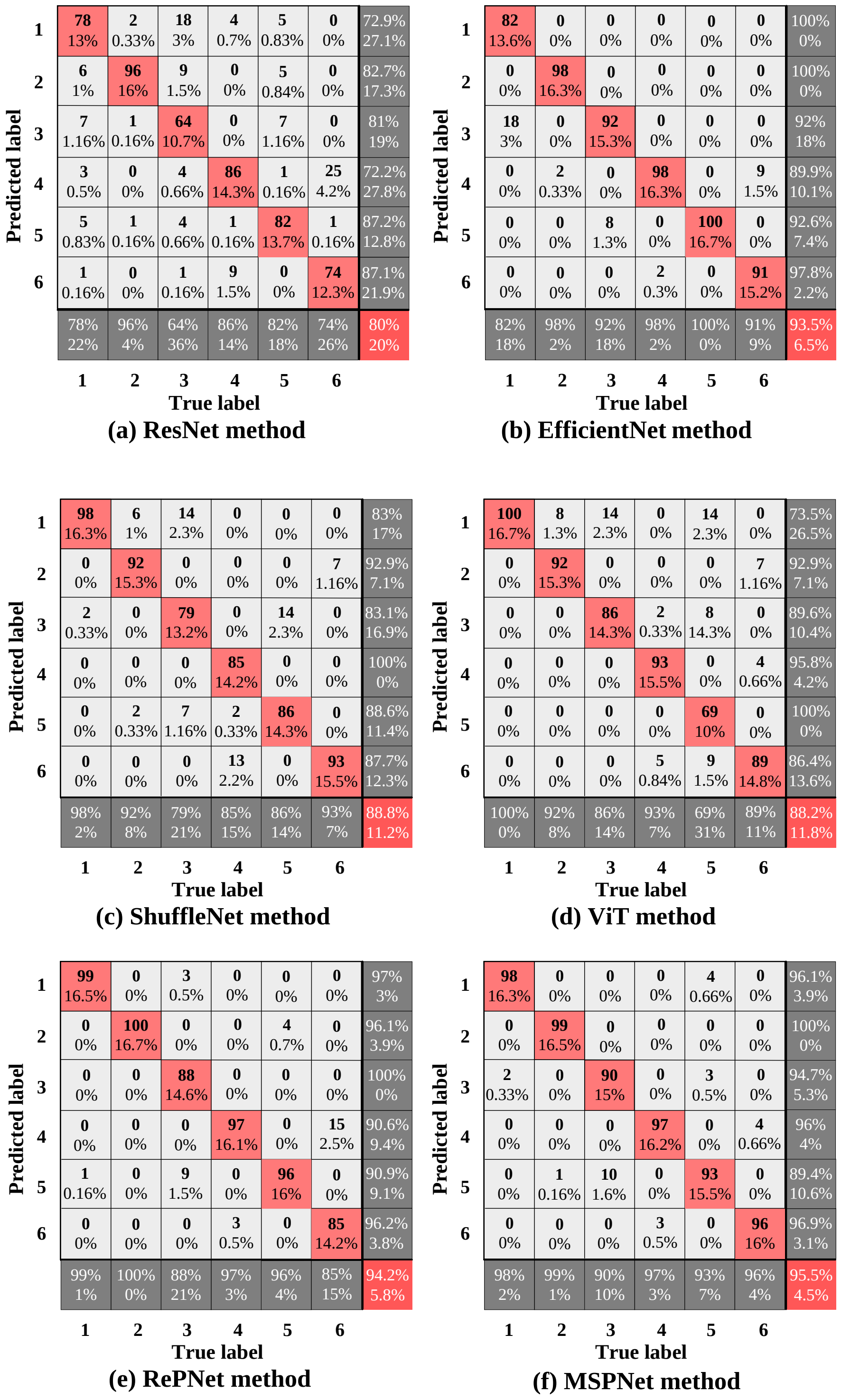
$$Micro - P = \frac{\sum_{l=1}^{L} TP}{\sum_{l=1}^{L} TP+\sum_{l=1}^{L} FP} \tag{19}$$

$$Micro - F1 = \frac{2\cdot(Micro-P)\cdot(Micro-R)}{(Micro-P)+(Micro-R} \tag{20}$$

4.4 Results analysis

In order to verify the effectiveness and superiority of the proposed method, in this section, an infrared thermal image dataset containing gearboxes with 5 different fault types is used for a fault recognition experiment using 5-way-5-shot. The infrared gearbox fault images with 5 different fault types are randomly selected as training models. In order to eliminate the influence of random factors on the diagnostic results, each method was repeated for ten experiments. The accuracy of the experimental results for ten experiments is shown in Fig.7. Fig.7 gives the confusion matrix of the optimal results of the six models to observe the detailed classification results of the proposed model for each fault category.

From Fig.7, we can see that the maximum value of the diagnosis results of the proposed method in this paper is 95.5%, and the average value is, 89.9%. The maximum accuracies of the rest of the comparative benchmark modelling methods are 80%, 93.5%, 88.8%, 88.2% and 94.2%, respectively. From these results, we can see that the fault diagnosis results of the proposed model in this paper are better than the other comparative benchmark model methods.



Predicted label \ True label	1	2	3	4	5	6	
1	78 13%	2 0.33%	18 3%	4 0.7%	5 0.83%	0 0%	72.9% 27.1%
2	6 1%	96 16%	9 1.5%	0 0%	5 0.84%	0 0%	82.7% 17.3%
3	7 1.16%	1 0.16%	64 10.7%	0 0%	7 1.16%	0 0%	81% 19%
4	3 0.5%	0 0%	4 0.66%	86 14.3%	1 0.16%	25 4.2%	72.2% 27.8%
5	5 0.83%	1 0.16%	4 0.66%	1 0.16%	82 13.7%	1 0.16%	87.2% 12.8%
6	1 0.16%	0 0%	1 0.16%	9 1.5%	0 0%	74 12.3%	87.1% 21.9%
	78% 22%	96% 4%	64% 36%	86% 14%	82% 18%	74% 26%	80% 20%

(a) ResNet method

Predicted label \ True label	1	2	3	4	5	6	
1	82 13.6%	0 0%	0 0%	0 0%	0 0%	0 0%	100% 0%
2	0 0%	98 16.3%	0 0%	0 0%	0 0%	0 0%	100% 0%
3	18 3%	0 0%	92 15.3%	0 0%	0 0%	0 0%	92% 18%
4	0 0%	2 0.33%	0 0%	98 16.3%	0 0%	9 1.5%	89.9% 10.1%
5	0 0%	0 0%	8 1.3%	0 0%	100 16.7%	0 0%	92.6% 7.4%
6	0 0%	0 0%	0 0%	2 0.3%	0 0%	91 15.2%	97.8% 2.2%
	82% 18%	98% 2%	92% 18%	98% 2%	100% 0%	91% 9%	93.5% 6.5%

(b) EfficientNet method

Predicted label \ True label	1	2	3	4	5	6	
1	98 16.3%	6 1%	14 2.3%	0 0%	0 0%	0 0%	83% 17%
2	0 0%	92 15.3%	0 0%	0 0%	0 0%	7 1.16%	92.9% 7.1%
3	2 0.33%	0 0%	79 13.2%	0 0%	14 2.3%	0 0%	83.1% 16.9%
4	0 0%	0 0%	0 0%	85 14.2%	0 0%	0 0%	100% 0%
5	0 0%	2 0.33%	7 1.16%	2 0.33%	86 14.3%	0 0%	88.6% 11.4%
6	0 0%	0 0%	0 0%	13 2.2%	0 0%	93 15.5%	87.7% 12.3%
	98% 2%	92% 8%	79% 21%	85% 15%	86% 14%	93% 7%	88.8% 11.2%

(c) ShuffleNet method

Predicted label \ True label	1	2	3	4	5	6	
1	100 16.7%	8 1.3%	14 2.3%	0 0%	14 2.3%	0 0%	73.5% 26.5%
2	0 0%	92 15.3%	0 0%	0 0%	0 0%	7 1.16%	92.9% 7.1%
3	0 0%	0 0%	86 14.3%	2 0.33%	8 14.3%	0 0%	89.6% 10.4%
4	0 0%	0 0%	0 0%	93 15.5%	0 0%	4 0.66%	95.8% 4.2%
5	0 0%	0 0%	0 0%	0 0%	69 10%	0 0%	100% 0%
6	0 0%	0 0%	0 0%	5 0.84%	9 1.5%	89 14.8%	86.4% 13.6%
	100% 0%	92% 8%	86% 14%	93% 7%	69% 31%	89% 11%	88.2% 11.8%

(d) ViT method

Predicted label \ True label	1	2	3	4	5	6	
1	99 16.5%	0 0%	3 0.5%	0 0%	0 0%	0 0%	97% 3%
2	0 0%	100 16.7%	0 0%	0 0%	4 0.7%	0 0%	96.1% 3.9%
3	0 0%	0 0%	88 14.6%	0 0%	0 0%	0 0%	100% 0%
4	0 0%	0 0%	0 0%	97 16.1%	0 0%	15 2.5%	90.6% 9.4%
5	1 0.16%	0 0%	9 1.5%	0 0%	96 16%	0 0%	90.9% 9.1%
6	0 0%	0 0%	0 0%	3 0.5%	0 0%	85 14.2%	96.2% 3.8%
	99% 1%	100% 0%	88% 21%	97% 3%	96% 4%	85% 15%	94.2% 5.8%

(e) RePNet method

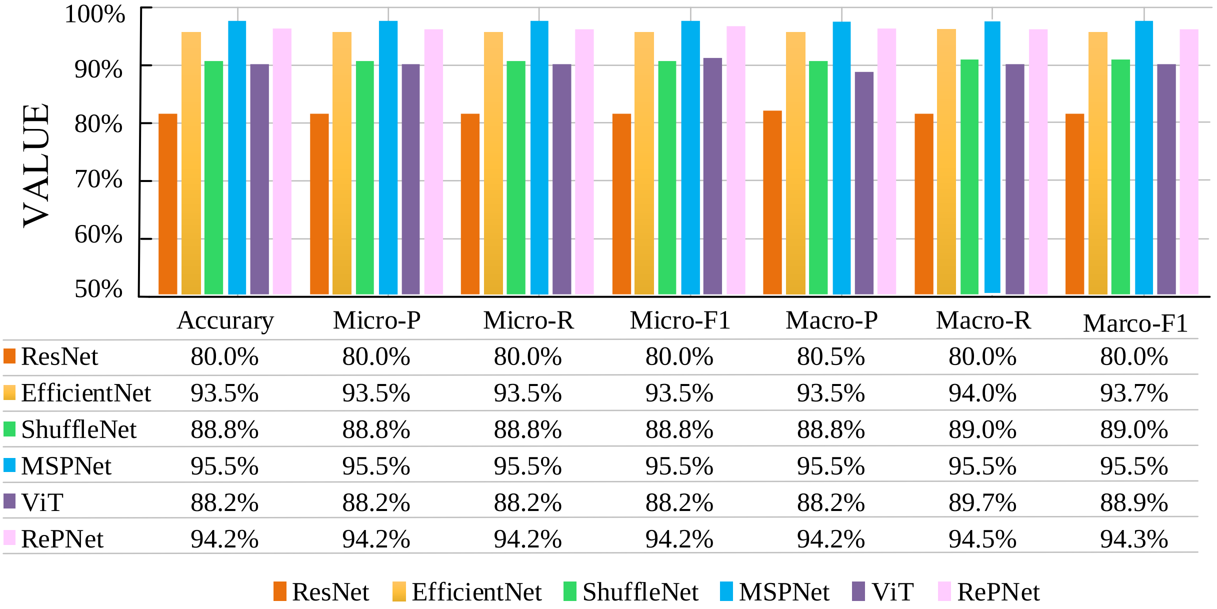
Predicted label \ True label	1	2	3	4	5	6	
1	98 16.3%	0 0%	0 0%	0 0%	4 0.66%	0 0%	96.1% 3.9%
2	0 0%	99 16.5%	0 0%	0 0%	0 0%	0 0%	100% 0%
3	2 0.33%	0 0%	90 15%	0 0%	3 0.5%	0 0%	94.7% 5.3%
4	0 0%	0 0%	0 0%	97 16.2%	0 0%	4 0.66%	96% 4%
5	0 0%	1 0.16%	10 1.6%	0 0%	93 15.5%	0 0%	89.4% 10.6%
6	0 0%	0 0%	0 0%	3 0.5%	0 0%	96 16%	96.9% 3.1%
	98% 2%	99% 1%	90% 10%	97% 3%	93% 7%	96% 4%	95.5% 4.5%

(f) MSPNet method

Fig. 7. Confusion matrix for different methods.

In order to further validate the effectiveness of the modelling, methods proposed in this paper, the accuracy, recall, and F1-Score of the experimental results of the different methods are calculated using the different evaluation metric formulas in Section 4.3, the relevant calculation results are displayed in Fig.8.

As can be seen in Fig.8, the best accuracy of MSPNet can reach 95.5%. Compared with ResNet, EfficientNet, ShuffleNet, ViT and RePNet, the accuracy of MSPNet is increased by 15.5%, 2%, 6.7%, 7.3% and 1.3%, respectively. In addition, it can be seen from Fig.8 that the Micro-F1 and Macro-F1 values of the six models are 80%, 93.5%, 88.8%, 95.5%, 88.2%, 94.2% and 80%, 93.7%, 89%, 95.5%, 88.9%, 94.3%, respectively. Analysing the F1 values, it can be seen that the accuracy of the MSPNet model proposed in this paper is improved by 15.5%, 2%, 6.7%, 7.3%, 1.3% and 15.5%, 2%, 6.5%, 6.6%, 1.2% compared with ResNet, EfficientNet, ShuffleNet, ViT, and RePNet, respectively. In addition, although RePNet also adopts the residual structure as the prototype network for the feature extractor, MSPNet still outperforms RePNet under the same conditions, which also proves the effectiveness of the multi-scale feature extraction proposed in this paper, which is able to extract more valuable fault features to be passed on to the downstream classifiers. Finally, as can be seen in Fig.8, the performance of other supervised learning networks is also lower than the MSPNet model proposed in this paper. The performance of the traditional convolutional neural network, which is based on a large amount of data and repeated training over a long period of time, is severely affected by the variation of data samples. This further validates the superiority of the prototype network under small sample conditions.



	Accurary	Micro-P	Micro-R	Micro-F1	Macro-P	Macro-R	Marco-F1
ResNet	80.0%	80.0%	80.0%	80.0%	80.5%	80.0%	80.0%
EfficientNet	93.5%	93.5%	93.5%	93.5%	93.5%	94.0%	93.7%
ShuffleNet	88.8%	88.8%	88.8%	88.8%	88.8%	89.0%	89.0%
MSPNet	95.5%	95.5%	95.5%	95.5%	95.5%	95.5%	95.5%
ViT	88.2%	88.2%	88.2%	88.2%	88.2%	89.7%	88.9%
RePNet	94.2%	94.2%	94.2%	94.2%	94.2%	94.5%	94.3%

Fig. 8. Evaluation indicators for optimal results.

5. Conclusions

In this paper, a novel few shot learning MSPNet model is proposed for intelligent fault diagnosis of gearbox by using infrared thermal images under small samples. The proposed MSPNet model introduces an innovative multi-scale module, which can effectively mine the details and global information in infrared thermal images under different scales. The proposed multi-scale strategy can explore the data more comprehensively and capture subtle differences that may be overlooked by single-scale analysis. The integration of features at different scales helps to understand the failure mode more carefully, thereby enhancing the fault diagnosis ability of the proposed MSPNet. Experimental results show that compared with other benchmark models, the proposed MSPNet can realize high precision fault diagnosis of gearbox under small samples. The proposed MSPNet model and the utilized infrared thermal images provide an effective tool for gearbox fault diagnosis under small samples.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (Nos. U62201399, U52305124), the Zhejiang Natural Science Foundation of China (Nos. LQ23E050002, LD21E050001), the Basic Scientific Research Project of Wenzhou City (Nos. G2022008, G2023028), the General Scientific Research Project of Educational Department of Zhejiang Province (Nos. Y202249008, Y202249041), the Project funded by China Postdoctoral Science Foundation (Nos. 2023M740988), the Postdoctoral Research Merit Funding Program of Zhejiang Province (Nos. ZJ2023122), the Supported by the Master's Innovation Foundation of

Wenzhou University (Nos. 3162024004106), the Master's Innovation Foundation of Wenzhou University (Nos. 3162024003067)

References

1. Gao C, He X, Dong H, Liu H, Lyu G. A survey on fault-tolerant consensus control of multi-agent systems: trends, methodologies and prospects. International Journal of Systems Science. 2022;53(13):2800-13. https://doi.org/10.1080/00207721.2022.2056772
2. Liu Y, Wang Z, Zhou D. Resilient actuator fault estimation for discrete-time complex networks: A distributed approach. IEEE Transactions on Automatic Control. 2020;66(9):4214-21. https://doi.org/10.1109/TAC.2020.3033710
3. Liang X, Zuo MJ, Feng Z. Dynamic modeling of gearbox faults: A review. Mechanical Systems and Signal Processing. 2018;98(jan.1):852-76. https://doi.org/10.1016/j.ymssp.2017.05.024
4. Wang Y, Xue C, Jia X, Peng X. Fault diagnosis of reciprocating compressor valve with the method integrating acoustic emission signal and simulated valve motion. Mechanical Systems and Signal Processing. 2015;56-57:197-212. https://doi.org/https://doi.org/10.1016/j.ymssp.2014.11.002
5. Choudhary A, Goyal D, Letha SS. Infrared Thermography-Based Fault Diagnosis of Induction Motor Bearings Using Machine Learning. IEEE Sens J. 2021;21(2):1727-34. https://doi.org/10.1109/JSEN.2020.3015868
6. Janssens O, Van de Walle R, Loccufier M, Van Hoecke S. Deep Learning for Infrared Thermal Image Based Machine Health Monitoring. Ieee-Asme Transactions on Mechatronics. 2018;23(1):151-9. https://doi.org/10.1109/tmech.2017.2722479
7. Li H, Wu P, Zeng N, Liu Y, Alsaadi FE. A survey on parameter identification, state estimation and data analytics for lateral flow immunoassay: from systems science perspective. International Journal of Systems Science. 2022;53(16):3556-76. https://doi.org/10.1080/00207721.2022.2083262
8. Liang C-M, Li Y-W, Liu Y-H, Wen P-F, Yang H. Segmentation and weight prediction of grape ear based on SFNet-ResNet18. Systems Science & Control Engineering. 2022;10(1):722-32. https://doi.org/10.1080/21642583.2022.2110541
9. Qin C, Yang R, Huang M, Liu W, Wang Z. Spatial variation generation algorithm for motor imagery data augmentation: Increasing the density of sample vicinity. IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2023;31:3675-86. https://doi.org/10.1109/TNSRE.2023.3314679
10. Dong Y, Wen C, Wang Z. A motor bearing fault diagnosis method based on multi-source data and one-dimensional lightweight convolution neural network. Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering. 2023;237(2):272-83. https://doi.org/10.1177/09596518221124785
11. Zhang C, Wen C, Liu J. Mask-MRNet: A deep neural network for wind turbine blade fault detection. Journal of Renewable and Sustainable Energy. 2020;12(5). https://doi.org/10.1063/5.0014223
12. Yang B, Lei Y, Jia F, Xing S. An intelligent fault diagnosis approach based on transfer learning from laboratory bearings to locomotive bearings. Mechanical Systems and Signal Processing. 2019;122:692-706. https://doi.org/10.1016/j.ymssp.2018.12.051
13. Zhang J, Pei G, Zhu X, Gou X, Deng L, Gao L, et al. Diesel engine fault diagnosis for multiple industrial scenarios based on transfer learning. Measurement. 2024:114338. https://doi.org/https://doi.org/10.1016/j.measurement.2024.114338
14. Chen X, Zhao Z, Guo F, Tan S, Wang J. Diagnosis method of transformer winding mechanical deformation fault based on sliding correlation of FRA and series transfer learning. Electric Power Systems Research. 2024;229:110173. https://doi.org/10.1016/j.epsr.2024.110173
15. Zhang C, Peng K, Dong J, Jiao R. A novel exergy-related fault detection and diagnosis framework with transformer-based conditional generative adversarial networks for hot strip mill process. Control Engineering Practice. 2024;144:105820. https://doi.org/https://doi.org/10.1016/j.conengprac.2023.105820
16. Vilalta R, Drissi Y. A perspective view and survey of meta-learning. Artificial Intelligence Review: An International Science and Engineering Journal. 2002;18(2):18. https://doi.org/10.1023/A:1019956318069
17. Xu W, Xian Y, Wang J, Schiele B, Akata Z. Attribute prototype network for zero-shot learning. Adv neural inf process syst. 2020;33:21969-80. https://doi.org/10.48550/arXiv.2008.08290
18. Wu Z, Shen C, van den Hengel A. Wider or Deeper: Revisiting the ResNet Model for Visual Recognition. Pattern Recognition.

2019;90:119-33. https://doi.org/10.1016/j.patcog.2019.01.006

19. Tan M, Le Q, editors. Efficientnet: Rethinking model scaling for convolutional neural networks. International conference on machine learning; 2019: PMLR. https://doi.org/10.48550/arXiv.1905.11946https://doi.org/10.48550/arXiv.1905.11946
20. Zhang X, Zhou X, Lin M, Sun J, editors. Shufflenet: An extremely efficient convolutional neural network for mobile devices. Proceedings of the IEEE conference on computer vision and pattern recognition; 2018. https://doi.org/10.48550/arXiv.1707.01083
21. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:201011929. 2020. https://doi.org/10.48550/arXiv.2010.11929