Research Article

Yangshuo Liu*, Jianshe Kang, Chiming Guo, and Yunjie Bai

Diesel engine small-sample transfer learning fault diagnosis algorithm based on STFT time–frequency image and hyperparameter autonomous optimization deep convolutional network improved by PSO–GWO–BPNN surrogate model

Abstract: Aiming at the problems of complex diesel engine cylinder head signals, difficulty in extracting fault information, and existing deep learning fault diagnosis algorithms with many training parameters, high time cost, and high data volume requirements, a small-sample transfer learning fault diagnosis algorithm is proposed in this article. First, the fault vibration signal of the diesel engine is converted into a three-channel red green blue (RGB) short-time Fourier transform time–frequency diagram, which reduces the randomness of artificially extracted features. Then, for the problem of slow network training and large sample size requirements, the AlexNet convolutional network and the ResNet-18 convolutional network are fine-tuned on the diesel engine time–frequency map samples as pre-training models with the transfer diagnosis strategy. In addition, to improve the training effect of the network, a surrogate model is introduced to autonomously optimize the hyperparameters of the network. Experiments show that, when compared to other commonly used methods, the transfer fault diagnosis algorithm proposed in this article can obtain high classification accuracy in the diagnosis of diesel engines while maintaining very stable performance under the condition of small samples.

Keywords: fault diagnosis, transfer learning, surrogate model, hyperparameter optimization, small sample

1 Introduction

The condition monitoring and health management of equipment are of great significance to maintain the reliability and health of equipment; among these, fault diagnosis is one of the key technologies, and it is very important to determine the current fault state of the equipment according to the state monitoring signal in time to maintain the reliability of the equipment [1–4]. The reciprocating machinery represented by diesel engines is widely used in modern industrial equipment, and the signal monitoring of the diesel engine is mostly realized by sensors placed in the cylinder head [5–7]. However, due to the more complex structure of diesel engines, the vibration signals frequently exhibit stronger nonlinearity [8,9]. In this regard, a large number of scholars have done a lot of effective work on the fault diagnosis of diesel engines. The most widely used fault diagnosis method is a combination of signal time–frequency characteristics and machine learning. Jing *et al.* [10] extracted features, such as fractal correlation dimension, wavelet energy, and entropy, that can reflect diesel engine failure mode from the vibration signal on the surface of the diesel

* **Corresponding author: Yangshuo Liu,** Department of Equipment Command and Management, Shijiazhuang Campus, Army Engineering University of PLA, Shijiazhuang 050003, China, e-mail: liuyangshuo2022@163.com
Jianshe Kang: Department of Equipment Command and Management, Shijiazhuang Campus, Army Engineering University of PLA, Shijiazhuang 050003, China, e-mail: jskang201206@126.com
Chiming Guo: Department of Equipment Command and Management, Shijiazhuang Campus, Army Engineering University of PLA, Shijiazhuang 050003, China, e-mail: guochiming@nudt.edu.cn
Yunjie Bai: Support detachment, 66029 Unit of the Chinese People's Liberation Army, Xilinguolemeng 011200, China, e-mail: cloud_bai123@163.com

engine and input them into the fast independent compositional analysis-support vector machine classifier. Finally, high-classification accuracy is achieved in small-sample recognition; Zhang and Liu [11] proposed a signal processing method based on fully integrated intrinsic timescale decomposition. First, the diesel engine nonstationary signal is decomposed into a set of proper rotation components and a residual signal, and then the singular values of the first few proper rotation components (PRCs), the energy and energy entropy of PRCs, and the auto regressive model parameters are extracted as fault feature vectors. Finally, the failure mode of a diesel engine was accurately identified using the least squares support vector machine; Sangharatna M. Ramteke *et al.* [12] adopted the condition monitoring technology of vibration and acoustic emission analysis to obtain the fault signal of a diesel engine and analyzed the time–frequency characteristics of the signal through fast Fourier transform (FT) and short-time Fourier transform (STFT). Finally, high-precision classification of diesel engine abrasion faults is realized through the artificial neural network model (ANN); Liu *et al.* [13] used an adaptive Wigner–Ville distribution (WVD) to generate a time–frequency image of diesel engine vibration signals and extracted its four types of commonly used image features, including moment invariants, gray statistical characteristics, textural features, and the differential box-counting fractal dimension; finally, the samples are input into the relevance vector machine improved by the fast correlation filter to accurately identify the fault type of the diesel engine. Although the aforementioned methods have achieved good diagnostic results, they require the relevant personnel to have sufficient knowledge of the failure mechanism, and the selection of feature extraction also introduces greater uncertainties to the diagnostic algorithm.

Therefore, in recent years, fault diagnosis methods based on data-driven deep learning have gradually been applied to the fault diagnosis of mechanical equipment. The advantages of deep learning are that there is no need to extract relevant fault features from the original signal, which eliminates certain artificial uncertainties; another advantage is that the large parameter space of deep learning and the nonlinear learning method can autonomously learn the fault characteristics of the data from a large number of fault signals, so that different fault types can be distinguished with high accuracy. Li *et al.* [14] proposed a pattern recognition model based on bispectral and convolutional neural networks (CNNs). The size of the signal bispectral matrix is optimized by interpolation to improve the diagnostic efficiency, and the network model with the best diagnostic accuracy is obtained after comparing and analyzing different

network settings. Du *et al.* [15] built a one-dimensional convolutional network (1D-CNN), which directly used the raw vibration signal of the car engine as the input of the network to achieve end-to-end fault diagnosis; Yousef Shatnawi and Al-Khassaweneh [16] collected the sound signal of the internal combustion engine and used the wavelet packet decomposition as a feature extraction tool, finally realizing the fault identification of the internal combustion engine by using the extended ANN. Wang *et al.* [7] designed an innovative deep learning network structure, the randomized CNN (RCNN), which is constructed from several individual CNNs; and the network can automatically extract discriminative features of vibration signals using convolution computation and pooling operations. Finally, the diagnostic superiority of the proposed RCNN is demonstrated by the experimental vibration signal of a diesel engine. It can be obtained that the deep learning method based on the neural network can learn the deep-level features of fault data autonomously, avoiding the participation of human factors, but this method also has certain defects. (1) Although a large hidden layer parameter space can contain a large amount of data feature knowledge, it also means that the deep network model requires a large amount of data as support, and training a complete network from scratch requires a lot of time consumption [17]; (2) in order to improve the diagnosis efficiency and consider the small amount of fault data in industrial practice, scholars have extensively studied the diagnosis of small samples of equipment. However, deep learning models tend to overfit in the learning of small sample data, so in practical industrial applications, the diagnostic accuracy of such models is often not ideal [18,19]; (3) Deep learning models have many hyperparameters, such as learning rate, batch size, and the number of training rounds. The values of these parameters will have a significant impact on the learning performance of the model. The existing literature generally sets these parameter values based on experience or artificial adjustment, but because the single training time of the network is very long, the efficiency of artificial parameter adjustment is relatively low [20,21].

In response to the aforementioned problems, in recent years, scholars have introduced transfer learning methods into the field of fault diagnosis in an attempt to solve these problems. The biggest feature of transfer learning is to learn a lot of knowledge from the original field and then apply that knowledge to the target field that researchers are interested in. In addition, transfer learning is mostly implemented through deep networks, so this method can deeply mine data information, which

is very suitable for industrial equipment fault detection under small data samples [22,23]. Xiong *et al.* [24] proposed a variable working condition recognition method based on stacked autoencoders (SAEs) and feature transfer learning. First, SAE is used to extract a feature set sensitive to operating conditions from the diesel signal, and then a balanced assignment adaptation transfer learning method is used to map the diesel features of two different engine operating conditions to the same feature space; the migration of sensitive fault characteristics under different working conditions of two high-power diesel engines is realized; Bai *et al.* [25] introduced deep transfer learning into the fault detection of gas turbine combustors for the problem of less historical data of new gas turbine failures. The CNN network is pre-trained using data from a data-rich gas turbine and then fine-tuned to detect failures of the new gas turbine using a small amount of fault data from the new gas turbine; the effectiveness of the method is verified by experiments on two actual gas turbines. Lei *et al.* [26] combined a transfer learning method with a deep belief network, taking the fault data of a 35 kW diesel generator simulation system as the source domain data and the target domain being a 70 kW diesel generator with sparse fault data. Through the pre-training and reverse fine-tuning of the network, an accurate fault diagnosis under the condition of a small sample of diesel generators is realized. Although the existing transfer learning methods can realize the feature transfer of data in different fields, they still need to train the network model from scratch on the fault data set in a certain field, so the time cost of learning and training is very huge; however, in the actual industrial environment, the timely restoration of operation after equipment failure is critical to maintaining efficient production. Finally, for the optimization of hyperparameters of the network model, most studies take the training accuracy of the network as the optimization goal and combine it with the group optimization algorithm to optimize the hyperparameters of the network [27,28]. Zhu *et al.* [29] chose particle swarm optimization (PSO) to automatically optimize the hyperparameters of the LeNet-5 network mode. These hyperparameters include the learning rate, the number of convolution kernels, the batch size, and the number of neurons in the fully connected layer; Han *et al.* [30] evaluated the original network using a simple model with a single convolutional layer and a single fully connected layer, combined with a genetic algorithm (GA) to optimize the model's hyperparameters. Experimental results show that this method helps to reduce training time; Tong *et al.* [31] used the cuckoo search algorithm to automatically optimize the hyperparameters of the deep

autoencoder, which can effectively distinguish the fault types and severity of rolling bearings under different working conditions. Although the combination of the group optimization algorithm reduces a certain number of training times compared with the traversal manual parameter adjustment, the time cost of hyperparameter optimization under this method is still relatively large due to the long training time of the network. Therefore, improving the optimization performance of the group optimization algorithm and accelerating the calculation of the fitness function are the keys to improving the efficiency of hyperparameter optimization [32–34].

Given the above research results and inspirations, based on the vibration signal of diesel engine, a diesel engine small-sample transfer learning fault diagnosis algorithm based on the combination of diesel engine STFT time-frequency images and hyperparameter autonomous optimization deep convolutional networks improved by the PSO–gray wolf algorithm (GWO)–back propagation neural network (BPNN) surrogate model is proposed in this article. First, the cylinder head vibration signal of the diesel engine is converted into a three-channel color time–frequency map by STFT, and then the AlexNet convolutional network and ResNet-18 convolutional network trained on the ImageNet dataset are used as pre-trained models; it is to use the ImageNet dataset as a sufficient amount of source domain knowledge. Then, the shallow network parameters of the two types of convolutional network models are frozen to extract the basic features of the diesel engine time–frequency image samples; and then, the deep parameter layer of the network is randomly initialized to perform fine-tuning learning on the image samples to extract deep features. In addition, a surrogate model that combines the gray wolf optimization algorithm improved by the PSO algorithm and the BPNN is introduced to efficiently and autonomously optimize the hyperparameters of the network; so that the network can more effectively realize the transfer learning fault diagnosis of diesel engines. The main contributions of the article are as follows:

1) First, given the complex structure of the diesel engine and the cumbersome disassembly and assembly, using the vibration signal of the cylinder head of the diesel engine to realize the condition monitoring of the diesel engine, the signal acquisition is relatively simple. In addition, only STFT conversion is performed on the signal to reflect the time–frequency characteristics of the diesel engine fault signal, which greatly reduces the factors of artificial selection of fault characteristics. At the same time, simple signal preprocessing is more conducive to the network model learning and

distinguishing the category features between samples, which relieves the training pressure of the network to a certain extent and helps to improve the diagnosis efficiency.

2) Furthermore, AlexNet and ResNet-18 with sufficient source domain knowledge (ImageNet dataset) are used as transfer network models. On the one hand, it solves the problem of the small amount of diesel engine fault data, and on the other hand, it does not need to train the network outright. It only needs to use a small number of time–frequency map fault samples to fine-tune the training of the network, which greatly reduces the training cost. More importantly, the transfer learning strategy of freezing shallow parameter layers allows only part of the parameter layers of the network to participate in training, avoiding the occurrence of network overfitting. Finally, it is experimentally verified that the method has better diagnostic accuracy compared with other deep learning models.

3) Furthermore, given the low efficiency of manual adjustment of network hyperparameters and the longtime of single network training, a surrogate model combining an improved group optimization algorithm and BPNN is proposed. While improving the optimization ability of the optimization algorithm, it uses a certain number of hyperparameter combinations and actual diagnosis results to train the BPNN; then, the fast and high-precision prediction process of the BPNN is used to replace the actual training process of the network to obtain the fitness function value, which realizes the efficient and autonomous optimization of the network hyperparameters.

4) Overall, the improvement of the accuracy and efficiency of the overall fault diagnosis is the main purpose and contribution of this research. This is due to the high-precision prediction process of the surrogate model based on the combined group optimization algorithm for the actual training accuracy of the deep network model for the test sample. In this process, the combination of the group optimization algorithm greatly improves the optimization ability of the algorithm, the convolutional network-based transfer model can deeply learn and compute deep fault features of image samples in the time–frequency domain, and the network test results under a certain parameter combination need to be accurately predicted by the surrogate model. The combination of the three allows the optimal hyperparameters of the network to be searched efficiently. After a series of

optimization and nonlinear operations, the effect and efficiency of fault diagnosis have been enhanced.

2 Theoretical background

2.1 Time–frequency image generation

In the fault diagnosis process, the original equipment generally collects 1D signals, but the input type of the deep network model is generally a 2D matrix or a 3D red green blue (RGB) picture sample. The two types of convolutional networks in this article both take pictures as input, so it is necessary to convert 1D vibration signals into 3D image data. One conversion method is to directly intercept the vibration signal at equal intervals and reorganize it into a 2D matrix, which is then saved as a three-dimensional image; however, this method cannot reflect the frequency domain characteristics of the signal; another method is based on the time–frequency domain transform, which mainly includes STFT, continuous wavelet transforms, and Hilbert–Huang transform. Ahmad Taghizadeh-Alisaraei and Mahdavian [35] comparatively analyzed the effectiveness of four time–frequency representation methods, Welch test, STFT, WVD, and Choi–Williams distribution, in diesel injector fault detection. It has been verified by experiments that in the real-time performance monitoring of the engine, the STFT technology is more effective for the fault diagnosis and knock detection of fuel nozzles. In the process of studying the combustion characteristics of diesel engines for biodiesel fuel, Siavash *et al.* [36] found that STFT can effectively analyze the time–frequency information of noise other than effective acoustic signals such as piston flapping and outlet valve closing during diesel combustion.

Therefore, the STFT that is more suitable for analyzing the time–frequency characteristics of diesel engine fault signals is selected to generate time–frequency image samples. The basic idea of STFT is to localize the integration interval of the FT of the time-domain signal of the device. Suppose there is a continuous signal $y(t)$, which can be expanded in the complete quadrature signal space as [37],

$$F(j\omega) = \int_{-\infty}^{+\infty} y(t)e^{-j\omega t}dt. \tag{1}$$

Eq. (1) is the definition of FT, which is used to convert the signal from the time domain to the frequency domain. ω represents frequency, t represents time, and e^{-jwt} is a

complex function. The sufficient condition for the establishment of the formula is: in the infinite interval, $y(t)$ is integrable, which is

$$\int_{-\infty}^{+\infty} |y(t)|\mathrm{d}t < \infty. \tag{2}$$

For the nonstationary vibration signal of the diesel engine cylinder head, it is necessary to use a window function to intercept the signal to limit the time domain range of the transformation so that the intercepted samples are within a certain frequency range. Then, the FT operation is performed on the signal after the windowing, and finally, the signal after the operation is superimposed, that is, the STFT is performed on the signal. The definition formula of STFT is as follows:

$$\mathrm{STFT}_y(t, \omega) = \int_{-\infty}^{+\infty} y(t)f(t - \tau)\mathrm{e}^{-j\omega t}\mathrm{d}t, \tag{3}$$

where $f(t - \tau)$ is the window function and τ is the center of the window function. The frequency obtained after STFT can be regarded as the instantaneous frequency at that point [38].

In MATLAB, after the signal is subjected to STFT operation, the result is directly saved as a three-channel color time–frequency diagram of RGB. In this way, time–frequency picture samples of the vibration signal of the cylinder head at different faults of the diesel engine can be obtained.

2.2 Transfer learning

In recent years, transfer learning has been widely used as a frontier field of deep learning. The main idea is to use existing (source domain) knowledge to solve problems in different but related domains (target domain).

Therefore, it relaxes the stringent requirements of traditional machine learning that require a large amount of data as training samples [39,40].

Transfer learning is defined as follows: given a source domain $D_s = \{X_s, P(X_s)\}$ and a learning task $T_s\{Y_s, f_s(\cdot)\}$, a target domain $D_t = \{X_t, P(X_t)\}$ and a learning task $T_t\{Y_t, f_t(\cdot)\}$. Its purpose is to acquire knowledge in the source domain D_s and learning task T_s to help improve the learning of the prediction function in the target domain, where $D_s \neq D_t$ and $T_s \neq T_t$ [41]. Figure 1 shows the schematic of transfer learning.

In terms of technical means, transfer learning can be divided into instance-based transfer learning, feature-based transfer learning, association rule-based transfer learning, and parameter-based transfer learning. Instance-based transfer learning improves the effect and robustness of transfer learning by adjusting the weights of the parts in the source domain that are more similar to the target domain; feature-based transfer learning attempts to construct a feature subspace, which integrates the shared latent feature factors of the source domain and the target domain, which can reduce the feature difference between the two and enhance the transferability of knowledge; The purpose of transfer learning based on association rules is to find potential connections between the source domain and the target domain, focusing on the study of transferability; Parameter-based transfer learning reduces the differences between domains. The idea is to use a large amount of source domain data to train a model and then use a small amount of target domain data to fine-tune the deep parameters of the model under the transfer learning strategy of shallow parameter freezing and deep parameter learning. Make the parameters of the deep network layer more in line with the classification characteristics of the target domain data [42].

This article adopts a parameter-based transfer learning method. The source domain knowledge is a sufficient number of ImageNet datasets, and the target domain
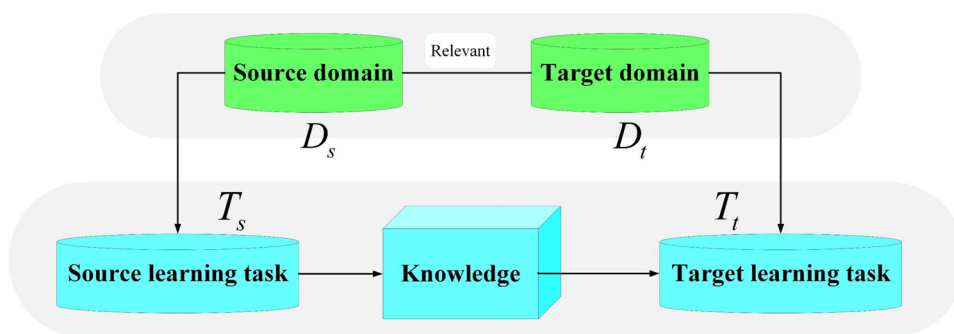

Figure 1: Schematic diagram of transfer learning.

is the time–frequency image samples of diesel engine cylinder head vibration signals. The network models are the AlexNet convolutional network and the ResNet-18 convolutional network. The migration strategy uses the freezing of the shallow network layer and the fine-tuning of the deep network layer. This approach eliminates the need for end-to-end training of the network model and the computation and reversed iteration of the difference metric between the source and target domain data at each iteration. Only a small number of samples are needed to fine-tune the deep classification parameters of the network so that the classification layer has the edge distribution characteristics of the target domain data. Then, use the deep feature extraction ability of the network to distinguish the subtle differences between the pictures to distinguish the subtle differences between the time–frequency images under different faults of the diesel engine and achieve the purpose of quickly classifying faults under the condition of small samples of equipment.

2.3 Network architecture

In recent years, high-quality deep network models such as AlexNet, Inception, GoogleNet, and ResNet have been proposed one after another, and their capabilities in image feature extraction and recognition have been effectively tested in the ImageNet image classification task competition [43,44]. Second, these network models have a certain parameter base after being trained on the ImageNet image dataset; this solves the model problem and the need for massive knowledge in the source domain for transfer learning. However, a network model with too many network layers and too large a parameter space will experience slow training speed and overfitting of training results when learning from small sample data [45]. Therefore, this article chooses to adopt a transfer learning strategy that avoids network overfitting and implements parameter-based transfer learning fault diagnosis between ImageNet image datasets and diesel engine fault data on AlexNet and ResNet-18 network models with fewer network layers.

The AlexNet network consists of five convolutional layers (Conv), three max-pooling layers, and three fully connected layers (dense), and convolutional layers and max-pooling layers are alternately arranged [46]. Figure 2(a) depicts the structure diagram of the AlexNet network, with only the convolutional layer and fully connected layer with parameter space listed, and the relevant pooling and activation layers omitted. The highlight of the AlexNet network is that it uses dual graphics processing unit

(GPUs) for network acceleration training; compared with single GPU training and learning, the learning speed is greatly improved. The activation function used by the AlexNet network is the ReLU function instead of the traditional Sigmoid function, which also speeds up the learning and solves the gradient dispersion problem well. The local response normalization of local response normalization is to establish a competition mechanism for local neurons after ReLU, so that the value with a larger response becomes relatively larger, suppressing the neuron with a smaller response and strengthening the generalization ability of the network [47]. In terms of network structure, the last two layers of the AlexNet model are changed to a fully connected layer and a Softmax classification layer corresponding to the number of diesel engine failure modes.

ResNet-18 network, as one of the typical deep residual networks, uses skipping connections by constructing residual blocks. Let the input X of the network be directly connected to the output Y of the parameterized layer through an identity map $I : X \rightarrow X$, so that the parameterized network layer learns a residual map $f : X \rightarrow Y - X$. Literature research shows that compared with convolution learning, the introduction of residual mapping can effectively reduce the learning difficulty of the network and speed up the convergence speed of the model [48]. The structure of the ResNet-18 model is mainly composed of a convolutional layer (Conv), four basic residual layers (basic layers), an average pooling layer, and a fully connected layer (dense); each residual block is, in turn, skipping connected by two convolutional layers. Different from the AlexNet model, the ResNet-18 model is a directed acyclic graph network [49]. After changing the end network layer, the global average pooling layer needs to be connected to the fully connected layer to ensure smooth network transmission. The network structure of the ResNet-18 model is shown in Figure 2(b), with only the convolutional layers and fully connected layers with parameter space listed, and the related pooling layers and activation layers omitted.

The image input dimension of the AlexNet model is $227 \times 227 \times 3$, and the input dimension of the ResNet-18 model is $224 \times 224 \times 3$. This means that the feature map needs to be adjusted to the size corresponding to the input dimension of the model before being input to the network for training; this operation is implemented in MATLAB through the augmentedImageDatastore function.

2.4 Surrogate model

In the optimization of network hyperparameters, three key hyperparameters, namely initial learning rate, batch

(a)

(b)
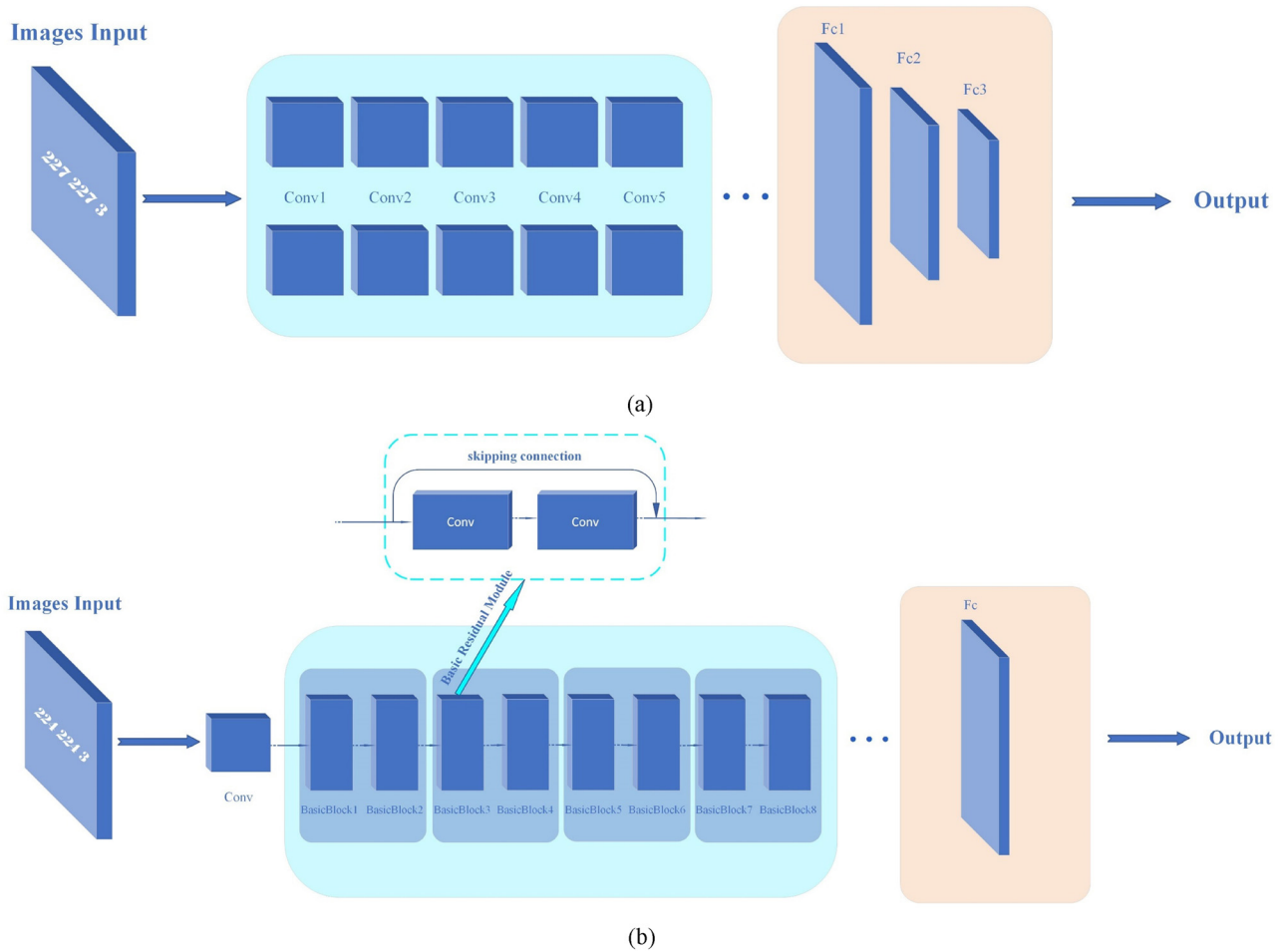
Figure 2: Network architecture diagram of (a) AlexNet and (b) ResNet-18.

size, and the maximum number of training rounds, are optimized. For supervised learning, an appropriate initial learning rate can make the objective function converge to a local minimum within the verification time; an appropriate batch size can increase the accuracy of gradient descent so that the amplitude of fluctuations during training is reduced. It is closely related to memory utilization and training speed during training; the maximum number of training rounds determines the degree of convergence of the network. Too small training rounds will cause the network to converge in advance, and too large rounds will waste time [50,51].

The surrogate model mainly includes two parts: the group optimization algorithm and the calculation of the fitness function value. The swarm optimization algorithm adopts the GWO optimized by the PSO algorithm, which improves the convergence speed and global optimization ability of the gray wolf optimization algorithm and helps optimize the hyperparameters of the network model efficiently. In terms of solving the fitness function value, first, use different hyperparameter samples to train the

network to obtain the corresponding training accuracy. Then, the trained BPNN model is embedded into the solution of the fitness function to realize the efficient and autonomous optimization of the hyperparameters of the convolution network.

2.4.1 GWO

Inspired by the way wolves hunt, in the gray wolf optimization algorithm, the variants are divided into four different groups according to the rank of wolves in nature, namely α, β, δ, and γ. Under the leadership of a wolf, the wolves continuously perform behaviors such as hunting, encircling, and attacking to realize the optimization process. The optimization diagram is shown in Figure 3 [52].

In the search process, use $|A| > 1$ to force the gray wolf to separate from its prey to determine the optimal attack target. After determining the attack target, the surrounding behavior of wolves can be expressed as follows:
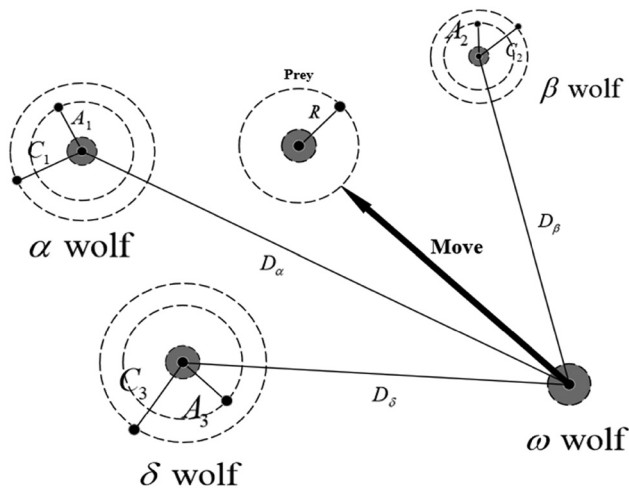
Figure 3: Schematic diagram of the GWO algorithm.

$$\begin{cases} \vec{D} = |\vec{C} \cdot \vec{X}_{P(t)} - \vec{X}(t)|, \\ \vec{X}(t+1) = \vec{X}_{P(t)} - \vec{A} \cdot \vec{D}, \\ \vec{A} = 2\vec{a} \cdot \vec{r_1} - \vec{a}, \\ \vec{A} = 2 \cdot \vec{r_2}, \end{cases} \tag{4}$$

where $\vec{D}$ is the Euclidean distance between the gray wolf and its prey; $\vec{X}(t)$ and $\vec{X}_{P(t)}$ are the position vectors of the gray wolf and the prey after moving t times, respectively; during the encirclement process, the convergence factor $\vec{a}$ decreases linearly from 2 to 0 according to the encirclement behavior of the gray wolf; the value range of the modulo of the $\vec{r_1}$ and $\vec{r_2}$ is [0,1], which represents a random change parameter with direction. After encircling the prey, consider α, β, and δ to be three potential solutions, and their positions change with the movement of the prey, such a chasing behavior can be expressed as follows:

$$\begin{cases} \vec{D_j} = |\vec{C} \cdot \vec{X_j} - \vec{X}|, \\ \vec{X_n} = \vec{X_j} - \vec{A_n} \cdot \vec{D_j}, \\ \vec{X_p}(t+1) = \dfrac{\sum_n \vec{X_n}}{3}, \end{cases} \tag{5}$$

where $j = \alpha$, β, and δ; $n = 1, 2, 3$; $\vec{D_j}$ represents the Euclidean distance from α, β, and δ to γ; $\vec{X_n}$ defines the step size and direction of γ moving toward α, β, and δ; $\vec{X_p}(t+1)$ represents the final position of γ. When the prey stops moving and the gray wolf attacks the prey, the optimal value is determined. The value of $\vec{a}$ decreases

linearly from 2 to 0, which is the core of this stage, indicating that the value of the corresponding $\vec{A}$ changes in the corresponding interval, and the next update position of the gray wolf will be closer to the prey position (optimal solution) [53,54].

2.4.2 GWO optimized by PSO (PSO–GWO)

Because the GWO algorithm does not consider the individual's own experience in the optimization process and lacks communication between the individual position and the group position, it may lead to premature convergence of the algorithm, and it is easy to fall into the local optimum. The group optimization algorithm used needs to have strong global optimization ability. So, consider improving the deficiencies of the GWO algorithm.

The essence of the PSO algorithm is that the particles continue to make directional variable-speed movements in space and find the next position through their memory and group communication to find the optimal solution. Therefore, the position update method of the particle can be used to replace the position update of the individual gray wolf, so that the GWO has memory during optimization. The update formulas for the velocity and position of the PSO algorithm are [55] as follows:

$$\begin{cases} v_i^{k+1} = v_i^k + c_1 r_1(p_{\text{best}}^k - x_i^k) + c_2 r_2(g_{\text{best}} - x_i^k), \\ x_i^{k+1} = x_i^k + v_i^{k+1}, \end{cases} \tag{6}$$

where x represents the position of the particle; v represents the flight speed of the particle; c_1 and c_2 are acceleration constants, which control the speed of the particles, represent the individual historical best position, and represent the global best position. In addition, the adjustment inertia constant γ is introduced to enhance the global search ability and local development ability of GWO, and the variation range of γ is [0.5, 1]. Therefore, the updated formula of the speed and position of wolves in the PSO–GWO algorithm is expressed as follows:

$$\begin{cases} v_i^{k+1} = \omega \cdot [v_i^k + c_1 r_1(x_1 - x_i^k) + c_2 r_2(x_2 - x_i^k) \\ \qquad\quad + c_3 r_3(x_3 - x_i^k)], \\ x_i^{k+1} = x_i^k + v_i^{k+1}. \end{cases} \tag{7}$$

The first formula in formula (5) becomes

$$\vec{D_j} = |\vec{C_n} \cdot \vec{X_j} - \omega \cdot \vec{X}|. \tag{8}$$

To verify the improvement of the improved PSO–GWO algorithm in the optimization ability, six types of optimization algorithm performance test functions were

selected to conduct simulation tests on GWO and PSO–GWO, respectively, and for each test, the number of populations and the maximum number of iterations of the algorithm were set to 30 and 500. The experimental results are shown in Figure 4. From the image of the performance test function, it can be concluded that in addition to the maximum value, the test function also has dense and continuous local minima and local maxima, which can greatly test the global optimization ability and optimization efficiency of the group optimization algorithm. From the simulation test results of the six types of test functions on the GWO and PSO–GWO algorithms, it can be concluded that the improved PSO–GWO algorithm has a better convergence effect and global optimization ability than the GWO algorithm.

2.4.3 BPNN

Use the network prediction process to replace the practical training process of convolutional networks, then the network should meet the following two requirements: (1) simple structure and fast training speed; (2) fast prediction speed and high accuracy. As a typical three-layer neural network, BPNN is composed of only input layers, hidden layers, and output layers, and the structure composition is relatively simple. In addition, BPNN is widely used in non-linear forecasts. Liang *et al.* [56] used BPNN to predict the deformation temperature of coal ash, which has higher predictive accuracy compared to traditional linear regression and Factsage calculation. Aiming at the nonstationarity and complexity of traffic flow, Peng and Xiang [57] established a prediction model through BPNN, optimized it through the GA, and finally realized an accurate prediction of traffic flow. Zhang and Lou [58] used BPNN and a deep learning fuzzy algorithm to predict stock prices and concluded that the trend prediction of stock prices by BPNN is better than that of the deep learning fuzzy algorithm. It can be seen that the BPNN has a strong predictive ability for nonlinear complex situations.

This article uses the 3-layer BPNN as a prediction model. Assuming the input super-added combination vector is $x = (x_1, x_2, x_3)$, the output of the hidden layer can be represented as follows:

$$y_i = f\left(\sum_j w_{ij}x_j - \theta_j\right), \tag{9}$$

where y_i is the output of the ith neuron in the hidden layer; x_j is the jth feature; w_{ij} is the weight value from

the jth neuron in the input layer to the ith neuron in the hidden layer; θ_j is the threshold of the jth neurons in the hidden layer; and $f(\cdot)$ is the transfer function from the input layer to the hidden layer. The output of the output layer is similar to the output of the hidden layer and can be expressed as follows:

$$z_l = g\left(\sum_i h_{li}y_i - n_l\right), \tag{10}$$

where z_l is the output of the lth neuron in the output layer; y_i is the output of the ith neuron in the hidden layer; h_{li} is the weight value from the ith neuron to the lth neuron in the hidden layer; n_l is the threshold of the lth neuron in the output layer; and $g(\cdot)$ is the transfer function from the hidden layer to the output layer [59,60].

The above process is the positive transmission process of the network. To optimize the weight and threshold of the network, the network still needs to be adjusted by feedback to minimize the error of the output layer. The schematic diagram of the BPNN algorithm is shown in Figure 5.

To enable BPNN to have the ability to predict the training accuracy of convolutional networks, first, the actual training accuracy of corresponding convolutional networks under different hyperparameter combinations is counted, thus generating a certain number of sample pairs. Let BPNN train on these sample pairs, learn the nonlinear relationship between the values of hyperparameters and the network training results, and then use them as a prediction model to predict the fitness value of the optimization algorithm.

2.4.4 PSO–GWO–BP surrogate model

The design flow of the proposed surrogate model is shown in Figure 6. It should be noted that the inverse of the output result of BPNN is used as the fitness value of PSO–GWO, that is, if the output value of BPNN is ac, then the fitness value is 1/ac. After the PSO–GWO algorithm reaches the maximum number of iterations, the values of the convolutional network's training accuracy and hyperparameter combination corresponding to the best fitness can be output.

3 Methodology

In summary, a small-sample transfer learning fault diagnosis algorithm based on the combination of diesel

(1)

(2)

(3)

(4)
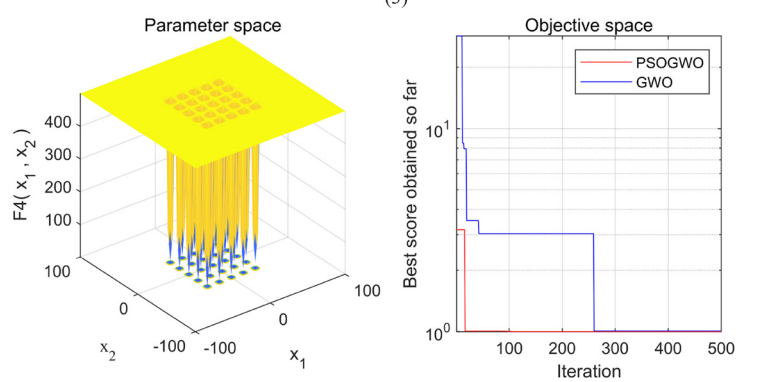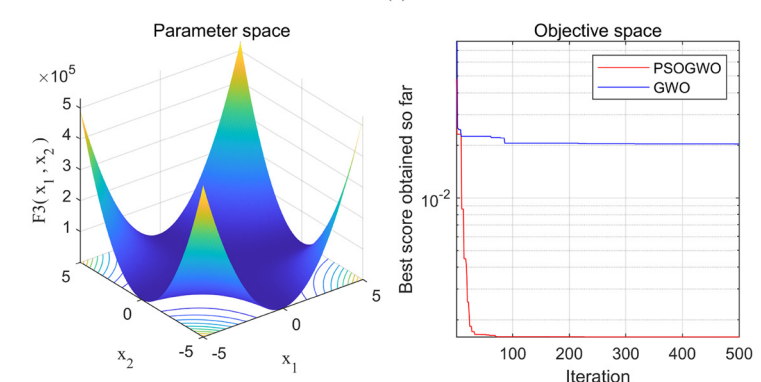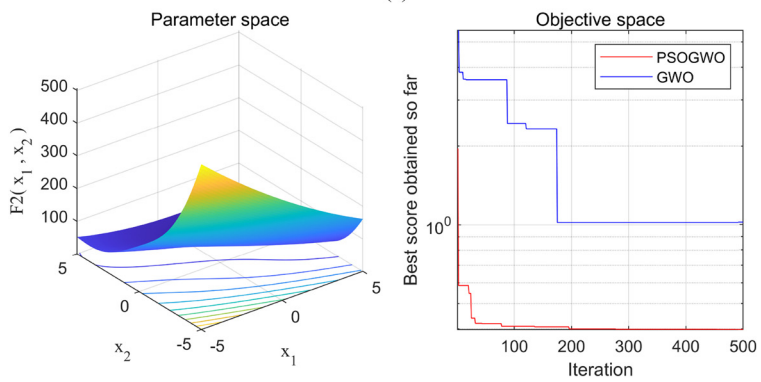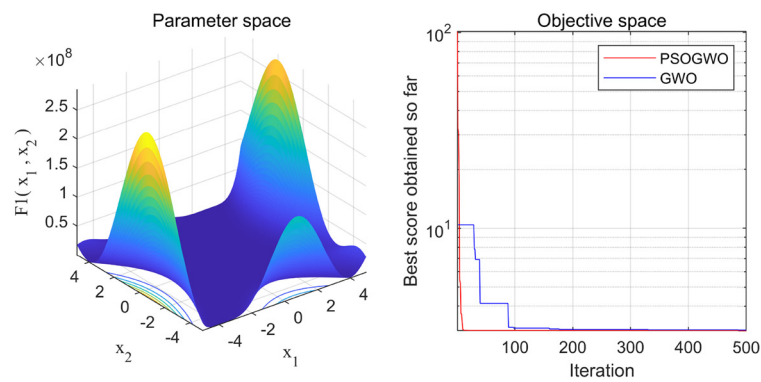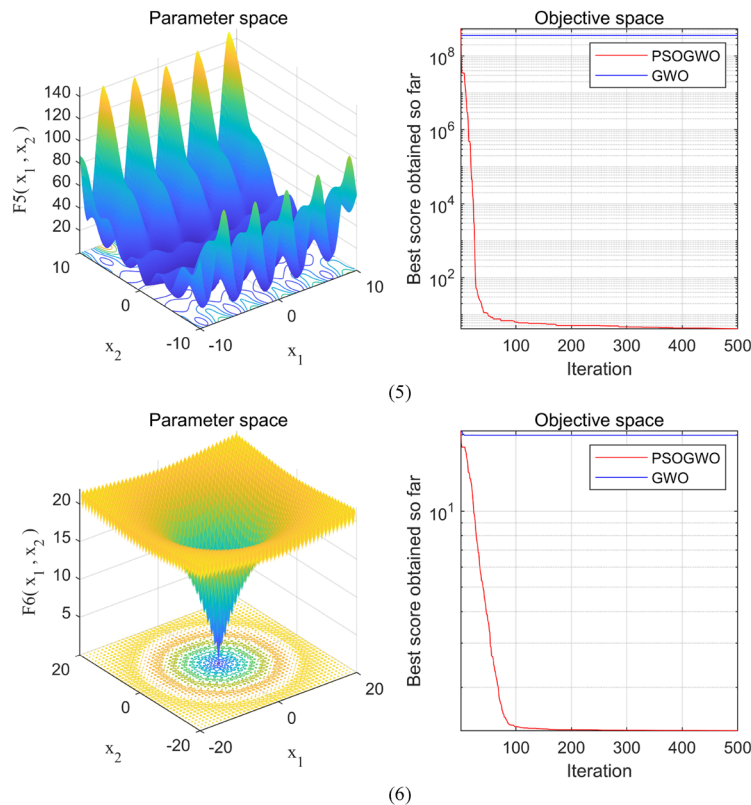
Figure 4: Simulation results of performance test functions on GWO and PSO–GWO algorithms: (1) test function F1; (2) test function F2; (3) test function F3; (4) test function F4; (5) test function F5; (6) test function F6.

engine STFT time–frequency images and hyperparameter autonomous optimization deep convolutional networks improved by the PSO–GWO–BPNN surrogate model is proposed in this article. The following problems in diesel engine fault diagnosis are attempted to be solved by this method.

1) Try to use a simple feature extraction method to reflect the fault characteristics of the signal to avoid the participation of too many human factors in the feature extraction.

2) Due to the small number of fault samples of equipment in the actual industrial environment, an attempt
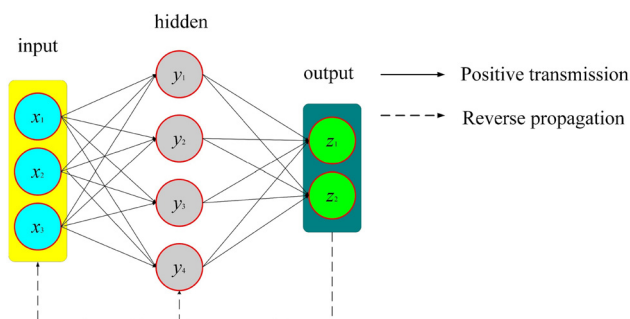
Figure 5: Schematic diagram of the BPNN.
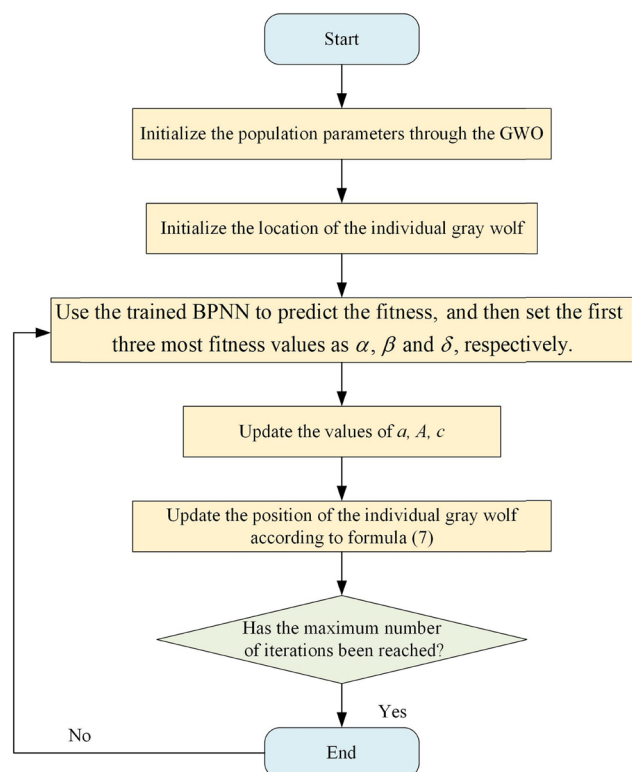
Figure 6: Surrogate model hyperparameter optimization flowchart.

is made to improve the fault diagnosis accuracy under the condition of small samples.
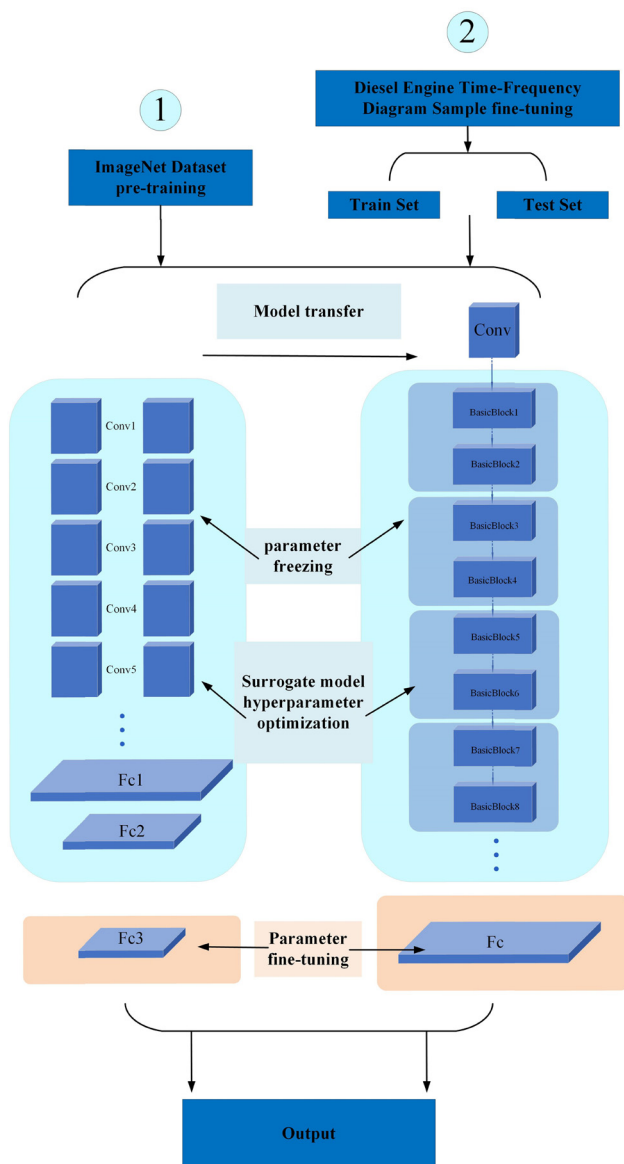
3) Attempts to improve the effectiveness and efficiency of network hyperparameter optimization.

The algorithm framework is shown in Figure 7, which includes five steps in total. The specific process includes:

1) First, the diesel engine fault preset experiment is carried out, and the vibration signal of the cylinder head is collected and preprocessed;

2) Second, the collected cylinder head vibration signal is converted into three-channel RGB time–frequency image samples through STFT and divided into a

Figure 7: The flow chart of the fault diagnosis algorithm proposed in this article.

training set and a test set according to a certain proportion;

3) Next, use the AlexNet network model and ResNet-18 network model pre-trained on the ImageNet image set as the basic migration model;

4) Furthermore, the learning rate of all network layers with parameter space before the last fully connected layer of the two-class convolutional network is set to 0. That is, these network layers are frozen, and only the parameters of the last fully connected layer are initialized so that it can learn the classification features of diesel engine failure samples;

5) Then, train the convolutional network to obtain the training accuracy and use the result set to train the BPNN. Then, use the PSO–GWO–BPNN surrogate model to optimize the hyperparameters of the two types of convolutional networks autonomously;

6) Finally, two classes of convolutional networks are trained using the optimized hyperparameters; the test sample is then diagnosed, and the classification result is obtained.

4 Experimental data collection and sample construction

Relying on the six-cylinder in-line diesel engine for testing, the condition monitoring test bench is shown in Figure 8. The entire test bench is mainly composed of a diesel engine control panel, a high-pressure common rail diesel engine, and a data acquisition system. The basic information is shown in Table 1. In terms of data acquisition, six piezoelectric vibration acceleration sensors are used to realize the multi-channel acquisition of vibration signals from the engine cylinder head, and the sampling frequency is 20,000 samples/s. A total of eight hybrid failure modes are set up in this article, see Table 2 for details. Details of the preset fault settings are shown in Figure 9.

After the conversion and effect analysis of the sensor data in the early stages, it was decided to select the fifth channel signal for research in the experiment. The sample point number was 5,000, and each of the eight types took 55 samples. The vibration signals and STFT time–frequency diagrams under eight types of faults are shown in Figure 10. It can be seen that there are subtle differences between the signals of different types of faults, whether it is the time-domain diagram or the STFT time–frequency diagram. But it is still unable to directly judge the fault type. Therefore, it is necessary to
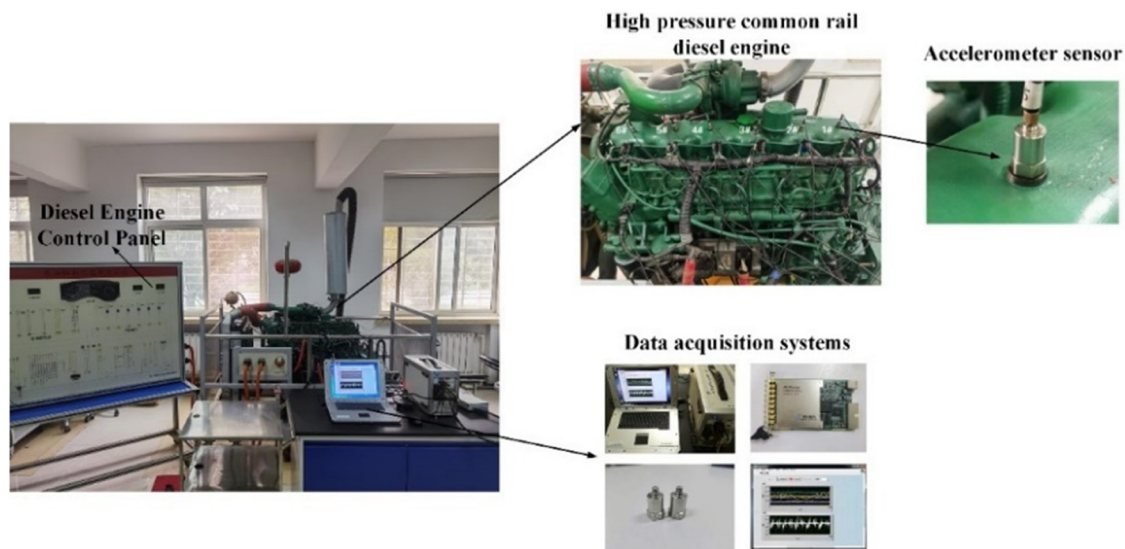
Figure 8: The condition monitoring test bench.

Table 1: Basic information on the diesel engine

Term	Information	Term	Information
Type	Six-cylinder in-line, high-press common rail	Rated power	155 kW
Model	CA6DF3-20E3	Rated speed	2,300 rpm
Size	1,330 mm × 970 mm × 1,005 mm	Net power	147 kW

Table 2: Failure mode settings

Serial number	Failure mode
M1	No failure
M2	Misfire in the first cylinder
M3	Misfire in the second cylinder
M4	Insufficient fuel supply
M5	Air filter blocked
M6	M2/M3
M7	M3/M5
M8	M3/M4/M5

rely on the powerful feature extraction and learning capabilities of the deep learning model to realize the classification and identification of diesel engine hybrid fault states.

At the same time, to verify the diagnostic effect of the proposed diagnostic method under the condition of small samples, the number of samples of each type is taken as 55, 45, 35, and 25, respectively. According to the ratio of 1:4, the samples are divided into training sets and test sets. The number of data sets under different sample size conditions is shown in Table 3. So far, the image samples for training and testing have been constructed.

5 Results and discussion

During training, the AlexNet model uses the SGDM optimizer and the ResNet-18 model uses the Adam optimizer [61]. The empirical values of the initial learning rates corresponding to the two optimizers are 1×10^{-2} and 1×10^{-3}, respectively. However, because the initial learning
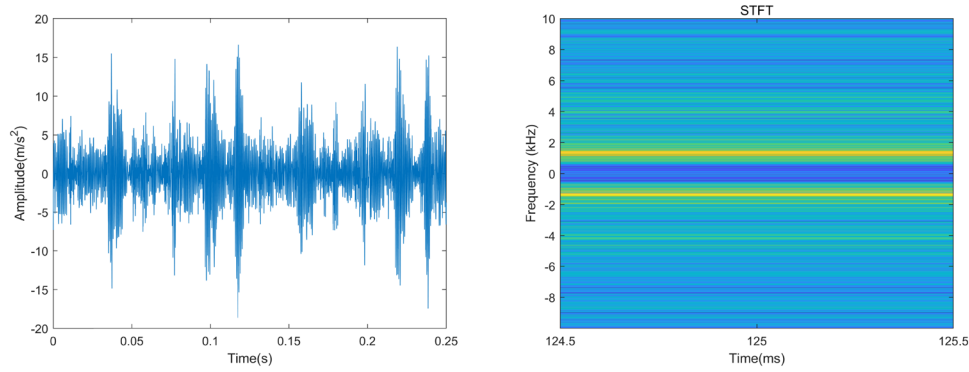

Figure 9: (a) Disconnect the startup power cord; (b) install the metal cover; and (c) replace the defective fuel injection pump.
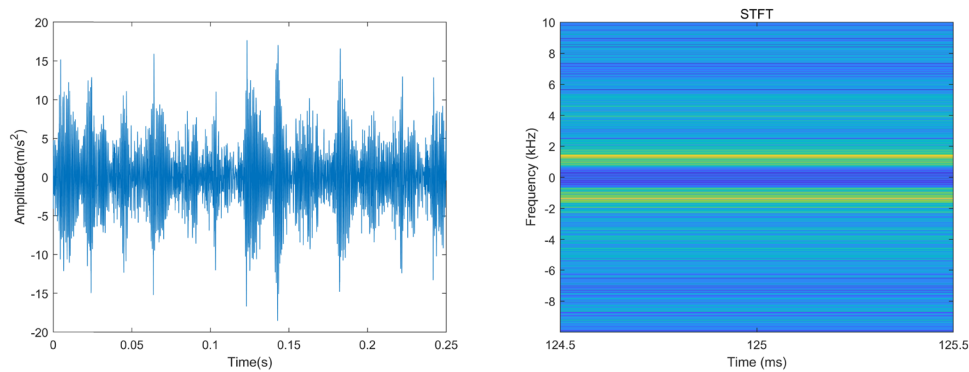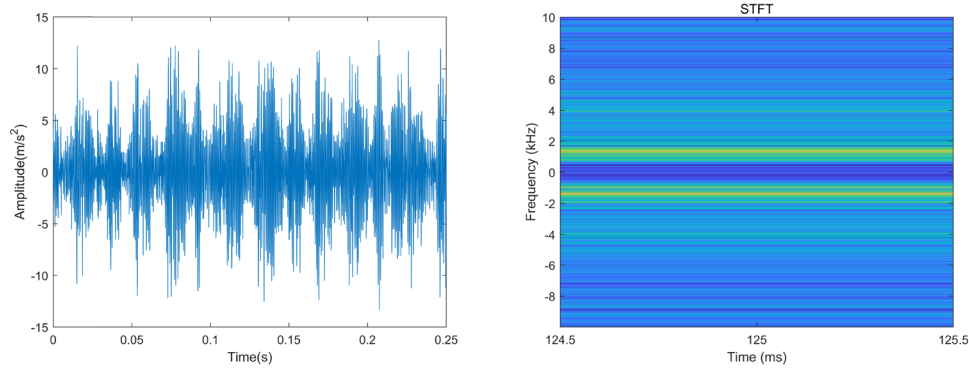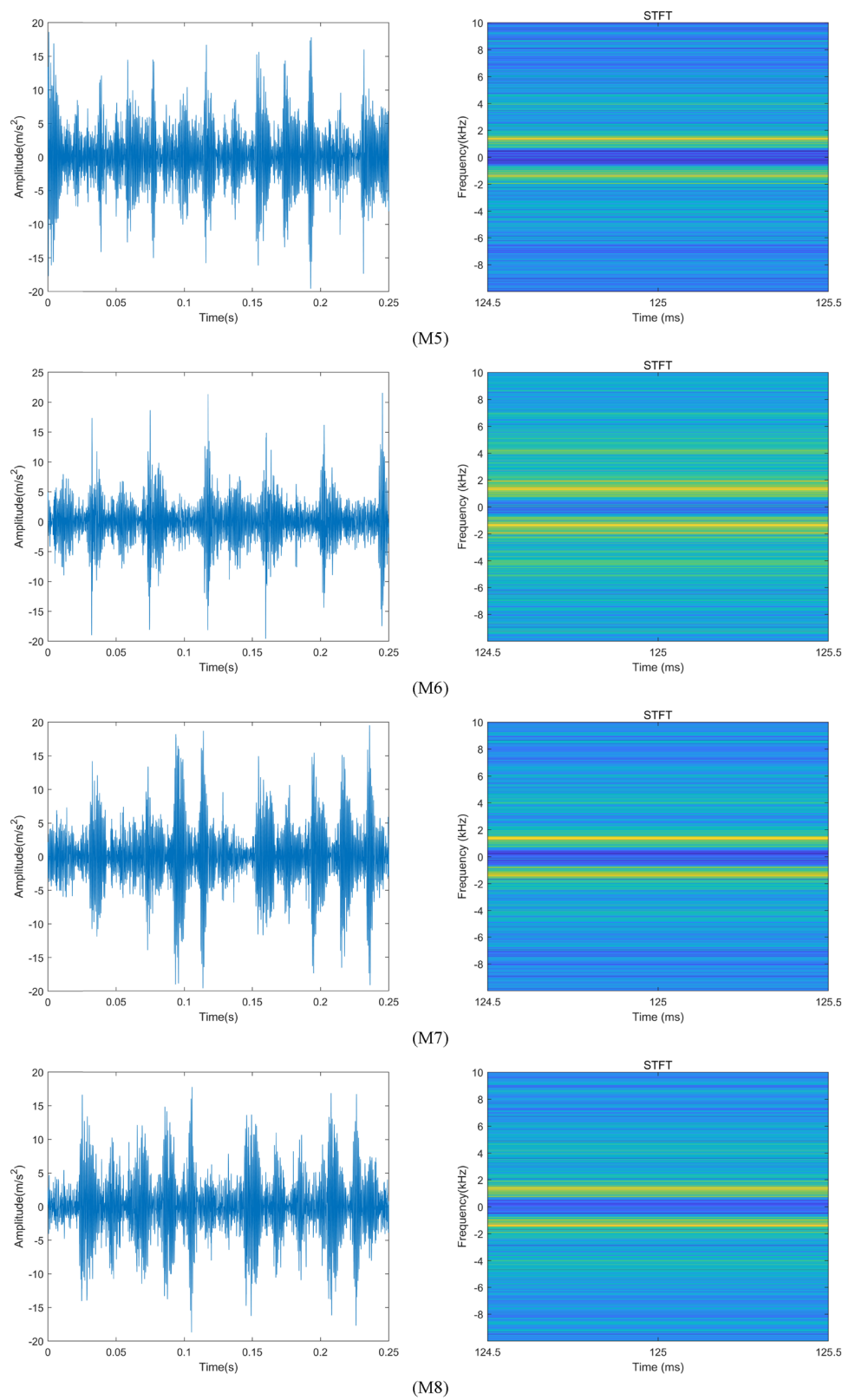
(M1)
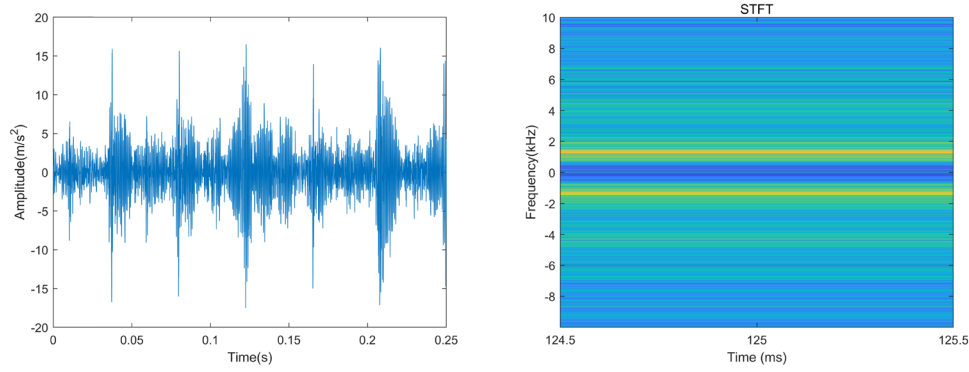
(M2)

(M3)

(M4)

Figure 10: Time domain diagram and STFT time–frequency diagram of cylinder head vibration signal under different fault conditions.

Table 3: Number of data sets under different sample size conditions

Total set	Train set	Test set
440	88	352
360	72	288
280	56	224
200	40	160

rate is too large during transfer learning, the network cannot converge, so the learning rate is usually reduced by 1–2 orders of magnitude, generally 1×10^{-4} and 1×10^{-5} [62]. In addition, in the AlexNet model, the learning rate decay strategy of decreasing 0.9 in 10 epochs is adopted, and the momentum value of stochastic gradient descent (SGD) is 0.9. At the same time, to make up for the impact of the low initial learning rate on the learning speed when the ResNet-18 model is trained under the conditions of the Adam optimizer, the weight and bias learning factors in the fully connected layer are set to a larger value of 10. These values were obtained through previous experiments. Table 4 lists the parameter settings of the network model.

5.1 Transfer learning fault diagnosis based on AlexNet network model

First, the hyperparameters of the AlexNet network model are optimized. The empirical values of the initial learning rate, the batch size, and the maximum number of training epochs of the model, and the optimization range of the surrogate model are shown in Table 5 [63,64]. Taking the total number of samples as 440, as an example, 25 sets of different hyperparameter combinations are used to train the network. The hyperparameter values and training results are shown in Table 6. The values of the hyperparameters in the table are the initial learning rate, batch size, and the maximum number of training epochs. Use these results to train the BPNN, then use the PSO–BPNN, GWO–BPNN, and the PSO–GWO–BPNN surrogate models to optimize the network's hyperparameters and finally use the optimized hyperparameter values to train the network to get the test results. In the same way, the optimization results and network training results under other sample sizes can be obtained. Table 7 and Figure 11 list the variation of the network training results with the sample size

Table 4: Model parameters and implementation platform

Training options	
Net architectures	AlexNet and ResNet-18
Optimizers	SGD (learning rate = 1×10^{-4}, momentum = 0.9) Adam (learning rate = 1×10^{-5})
Remark	The values of the learning rate, batch size, and maximum number of training rounds will be optimized

Table 5: Empirical value of hyperparameters and the optimization range of the surrogate model

Value	Initial learning rate	Batch size	Maximum number of training rounds
Experience value	1×10^{-4}	12	9
Optimization range value	$[5 \times 10^{-5}, 1.5 \times 10^{-4}]$	[6,16]	[5,10]

Table 6: Training results with different hyperparameter values

Hyperparameter values	Accuracy (%)	Hyperparameter values	Accuracy (%)
$[5 \times 10^{-5}, 6, 5]$	96.59	...	...
$[5 \times 10^{-5}, 7, 6]$	96.31	$[1.1 \times 10^{-4}, 10, 9]$	98.43
$[6 \times 10^{-5}, 7, 6]$	97.16	$[1.2 \times 10^{-4}, 15, 10]$	95.24
$[7 \times 10^{-5}, 7, 7]$	97.17	$[1.4 \times 10^{-4}, 16, 10]$	97.62
...	...	$[1.5 \times 10^{-4}, 16, 10]$	94.05

Table 7: Empirical value of hyperparameters and the optimization range of the surrogate model

Sample size	Method	Hyperparameter values	Accuracy (%)	Training time (s)
440	PSO–GWO–BPNN	$[1.1 \times 10^{-4}, 8, 8]$	98.58	24
	PSO–BPNN	$[1.3 \times 10^{-4}, 9, 8]$	98.32	32
	GWO–BPNN	$[1.2 \times 10^{-4}, 7, 8]$	98.11	30
	Experienced	$[1 \times 10^{-4}, 12, 9]$	98.00	40
360	PSO-GWO-BPNN	$[1.09 \times 10^{-4}, 7, 9]$	98.26	43
	PSO–BPNN	$[1 \times 10^{-4}, 8, 8]$	97.32	38
	GWO–BPNN	$[1.1 \times 10^{-4}, 9, 8]$	97.62	43
	Experienced	$[1 \times 10^{-4}, 12, 9]$	97.53	53
280	PSO–GWO–BPNN	$[1.1 \times 10^{-4}, 6, 9]$	96.88	34
	PSO–BPNN	$[9 \times 10^{-5}, 8, 9]$	94.88	35
	GWO–BPNN	$[1 \times 10^{-4}, 9, 9]$	93.14	33
	Experienced	$[1 \times 10^{-4}e, 12, 9]$	94.23	38
200	PSO–GWO–BPNN	$[1 \times 10^{-4}, 6, 8]$	90.00	21
	PSO–BPNN	$[8 \times 10^{-5}, 9, 9]$	86.96	30
	GWO–BPNN	$[1.02 \times 10^{-4}, 7, 7]$	86.33	42
	Experienced	$[1 \times 10^{-4}, 12, 9]$	85.23	32
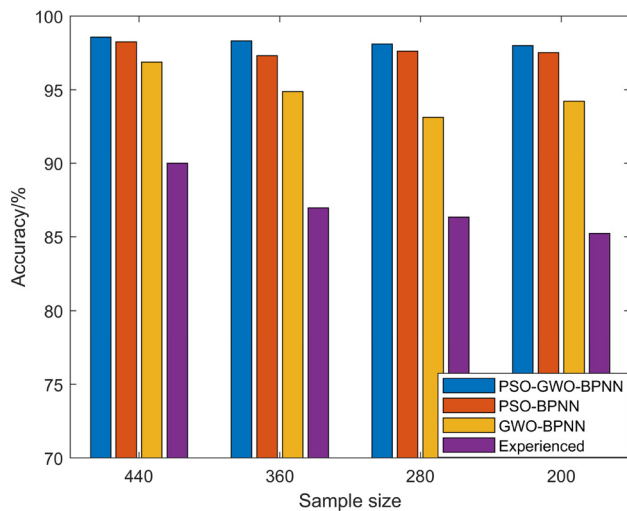
Figure 11: Comparison chart of different hyperparameter optimization methods for the AlexNet network model.
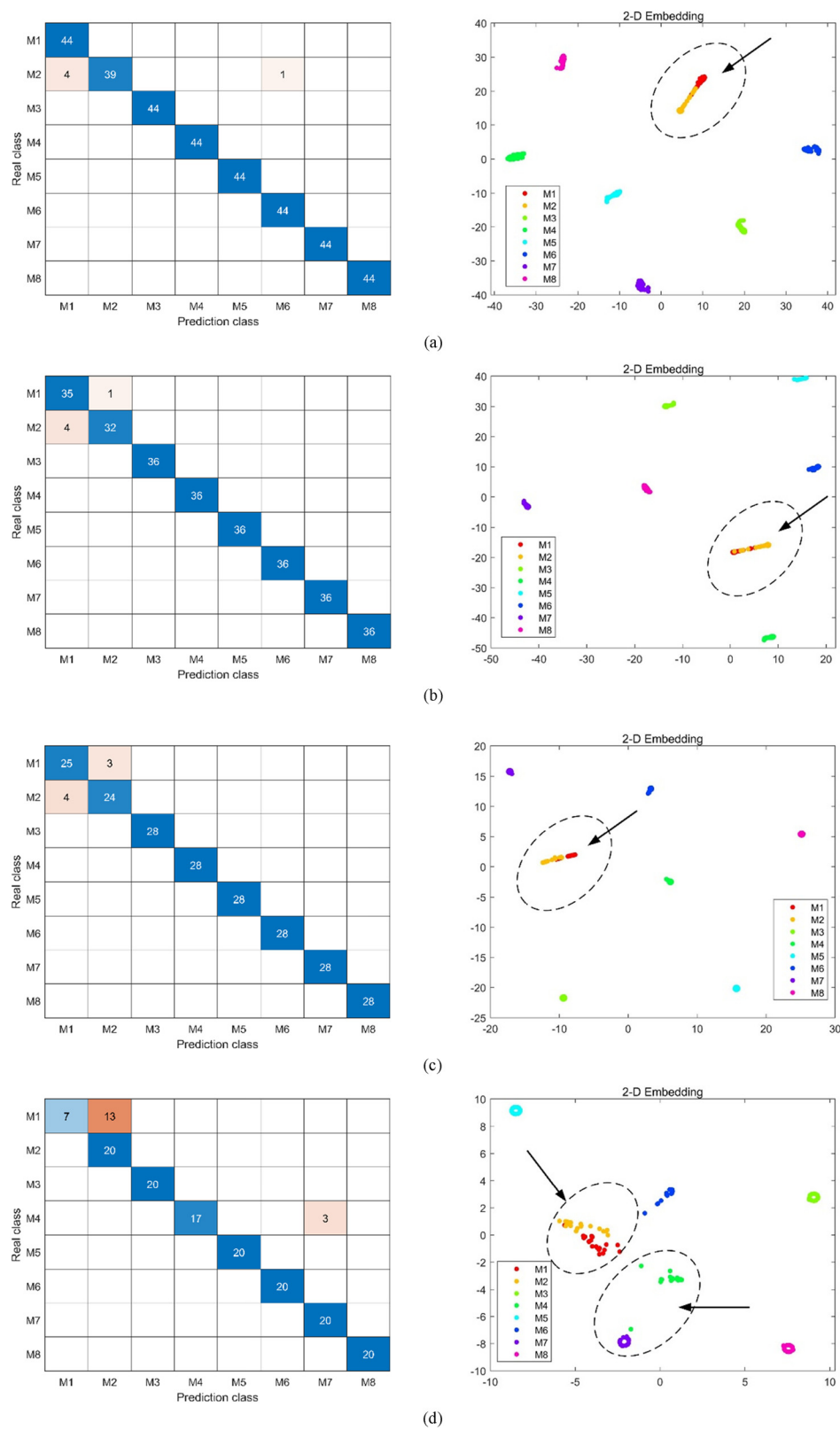
under different hyperparameter values. Figure 12 lists the confusion matrix of the network training results and the classification features calculated by the fully connected layer under different data amounts.

From the above experimental results, the following conclusions can be drawn:

1) From the training results in Table 7 and Figure 11, the PSO–GWO–BPNN surrogate model optimizes effective hyperparameters for the network. The network training results optimized by the PSO–GWO–BPNN surrogate model are better than the network optimized by the PSO–BPNN and GWO–BPNN surrogate models and the network under the empirical value; in

addition, the optimization effect of the GWO–BPNN surrogate model is not obvious compared with the empirical value, which is even lower than the empirical value in some cases. This shows that the PSO–BPNN and GWO–BPNN surrogate models do not find the globally optimal hyperparameters, and it also proves the effectiveness of the PSO–GWO–BPNN surrogate model in autonomously optimizing network hyperparameters.

2) The proposed diagnostic algorithm can achieve excellent performance in fault diagnosis under small sample conditions. First, the diagnostic accuracy of the network does not fluctuate greatly with the decrease in the sample size and remains above 90%. When the sample size is 440, the highest accuracy rate can reach 98.58%.

3) It can be concluded from Figure 12 that the proposed diagnosis algorithm can discriminate most of the eight types of mixed faults with high precision. Only M1 (normal state) faults and M2 (one cylinder misfire) faults will cause misjudgments. The reason may be that the characteristics of M1-type faults and M2-type faults are similar. In addition, when the sample size drops to 200, the phenomenon of misjudgment will be more serious, and there is a misjudgment of M4 (insufficient oil supply) faults. The reason may be that the sample size is insufficient and the data features that lead to M4-type failures are not sufficiently learned.

4) In terms of training time, the network can efficiently complete training and realize diesel engine fault diagnosis in a short time due to the strategy of freezing some parameter layers.

Figure 12: Confusion matrix of the AlexNet network training results and the classification features calculated by the fully connected layer underdifferent data amounts: (a) 440; (b) 360; (c) 280; (d) 200.

5.2 Transfer learning fault diagnosis based on ResNet-18 network model

The hyperparameters of the ResNet-18 network model are also optimized first. The empirical values of the initial learning rate of the model, the batch size, and the maximum number of training epochs, and the optimization range of the surrogate model are shown in Table 8 [65]. Taking the total number of samples as 440, as an example, 25 sets of different hyperparameter combinations are used to train the network. The hyperparameter values and training results are shown in Table 9. Table 10 and Figure 13 list the variation of the network training results with the sample size under different hyperparameter values. Figure 14 lists the confusion matrix of the network training results and the classification features calculated by the fully connected layer under different data amounts.

From the above experimental results, the following conclusions can be drawn:

1) The same experimental results as the AlexNet network model can be obtained from Table 10 and Figure 13. After the optimization of the PSO–GWO–BPNN surrogate model, the ResNet-18 network is efficiently trained, and the test results are better than the PSO–BPNN and GWO–BPNN surrogate models and

**Table 8:** Empirical value of hyperparameters and the optimization range of the surrogate model

Value	Initial learning rate	Batch size	Maximum number of training rounds
Experience value	1×10^{-5}	12	8
Optimization range value	$[5 \times 10^{-6}, 1.5 \times 10^{-5}]$	[6,16]	[5,10]

Table 9: Training results with different hyperparameter values

Hyperparameter values	Accuracy (%)	Hyperparameter values	Accuracy (%)
$[5 \times 10^{-6},6,5]$	92.35	...	...
$[6 \times 10^{-6},7,5]$	92.60	$[1.1 \times 10^{-5},11,8]$	95.88
$[6 \times 10^{-6},8,6]$	93.21	$[1.2 \times 10^{-5},12,6]$	96.02
$[9 \times 10^{-6},8,6]$	94.65	$[1.4 \times 10^{-5},15,7]$	94.48
...	...	$[1.5 \times 10^{-5},16,7]$	91.49

Table 10: Empirical value of hyperparameters and the optimization range of the surrogate model

Sample size	Method	Hyperparameter values	Accuracy (%)	Training time (s)
440	PSO–GWO–BPNN	$[1.2 \times 10^{-5},8,6]$	96.31	20
	PSO–BPNN	$[1.1 \times 10^{-5},8,8]$	92.96	32
	GWO–BPNN	$[1 \times 10^{-5},7,8]$	93.22	35
	Experienced	$[1 \times 10^{-5},12,8]$	92.65	19
360	PSO–GWO–BPNN	$[9 \times 10^{-6},6,5]$	92.01	18
	PSO–BPNN	$[1.3 \times 10^{-5},6,7]$	90.35	30
	GWO–BPNN	$[1.03 \times 10^{-5},10,9]$	89.33	32
	Experienced	$[1 \times 10^{-5},12,8]$	90.22	16
280	PSO–GWO–BPNN	$[1 \times 10^{-5},6,6]$	91.07	14
	PSO–BPNN	$[1.2 \times 10^{-5},7,9]$	88.30	19
	GWO–BPNN	$[9 \times 10^{-6},7,7]$	90.01	15
	Experienced	$[1 \times 10^{-5},12,8]$	89.60	16
200	PSO–GWO–BPNN	$[8 \times 10^{-6},6,5]$	86.25	12
	PSO–BPNN	$[9 \times 10^{-6},6,8]$	85.32	12
	GWO–BPNN	$[8 \times 10^{-6},10,6]$	84.51	16
	Experienced	$[1 \times 10^{-5},12,8]$	82.45	12
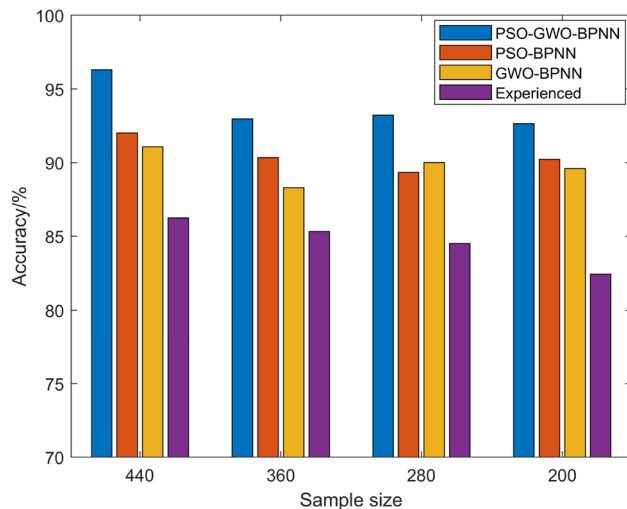
Figure 13: Comparison chart of different hyperparameter optimization methods for the ResNet-18 network model.

the empirical method. The ability of the PSO–GWO–BPNN surrogate model to well optimize the network hyperparameters is once again demonstrated.

2) The fault diagnosis algorithm based on the ResNet-18 network model can still achieve stable and excellent performance in the fault diagnosis under the conditions of small samples. The diagnostic accuracy of the network will not fluctuate greatly with the decrease in the sample size. Even when the sample size is 440, the accuracy rate can reach 96.31%; even when the sample size drops to 200, the accuracy rate can reach 86.25%. However, the overall diagnostic accuracy of the ResNet-18 network model is lower than that of the AlexNet network model, which means that the model training effect is not necessarily better with more network layers.

3) It can be concluded from Figure 14 that compared with the AlexNet network model, the ResNet-18 network model can also better identify most of the eight types of mixed faults. The misjudged fault types are also basically consistent with the AlexNet network model, including M1 (normal state) faults, M2 (first cylinder misfire), and M4 (insufficient fuel supply) faults. The difference is that when the number of samples is 360 and 200, the ResNet-18 network model has misjudged the M6 (the first- and second-cylinder misfire) faults and the M7 (the second-cylinder misfire and air filter blockage) faults. The possible reasons are that: the deep features extracted by the ResNet-18 network model and the AlexNet network model are different, and the decrease in the amount of data leads to the fact that the features of some faulty types of data are not completely learned.

4) In terms of training time, due to the introduction of residual connections in the ResNet-18 network model, it has a shorter training time cost and higher training efficiency than the AlexNet network model.

## 5.3 Validity analysis of surrogate model

To verify the prediction accuracy of the surrogate model, taking the sample size of 440 as an example, the trained BP network is used to predict the accuracy of the four hyperparameter combinations outside the optimization range of the AlexNet network and the ResNet-18 network. The comparison results with the actual training results are shown in Table 11 and Figure 15. It can be seen from the experimental results that the surrogate model can predict the actual training results of the network with high accuracy, and the error can be kept between 0.23% and 1.45%. This strongly demonstrates the effectiveness of the model. In the case of large errors, the reason may be that the number of training samples of BPNN is not enough, and the implicit relationship between the values of hyperparameters and the training results cannot be fully learned; but such errors are allowed.

In addition, based on ensuring the prediction accuracy, the role of the surrogate model is also reflected in the shortening of the time cost. Taking the training of the AlexNet network with a sample size of 440 as an example, the training time of the network is 24 s. If the traversal parameter adjustment method is used, the network needs to be trained at times. With the surrogate model, only 25 training sessions are required. The latter has a time cost of 701×24 s less than the former. Compared with such a large time gap, the iteration time of the PSO–GWO algorithm of the surrogate model is negligible. The efficiency of the PSO–GWO–BPNN surrogate model in network hyperparameter optimization time is thus demonstrated.

5.4 Performance comparison analysis

To further verify the comprehensive performance of the diagnostic algorithm proposed in this article, the time–frequency image samples of the diesel engine were input into the following models for training. These models include the AlexNet network model and the ResNet-18 network model without a freezing strategy, that is, the network layer parameters of the whole domain will be
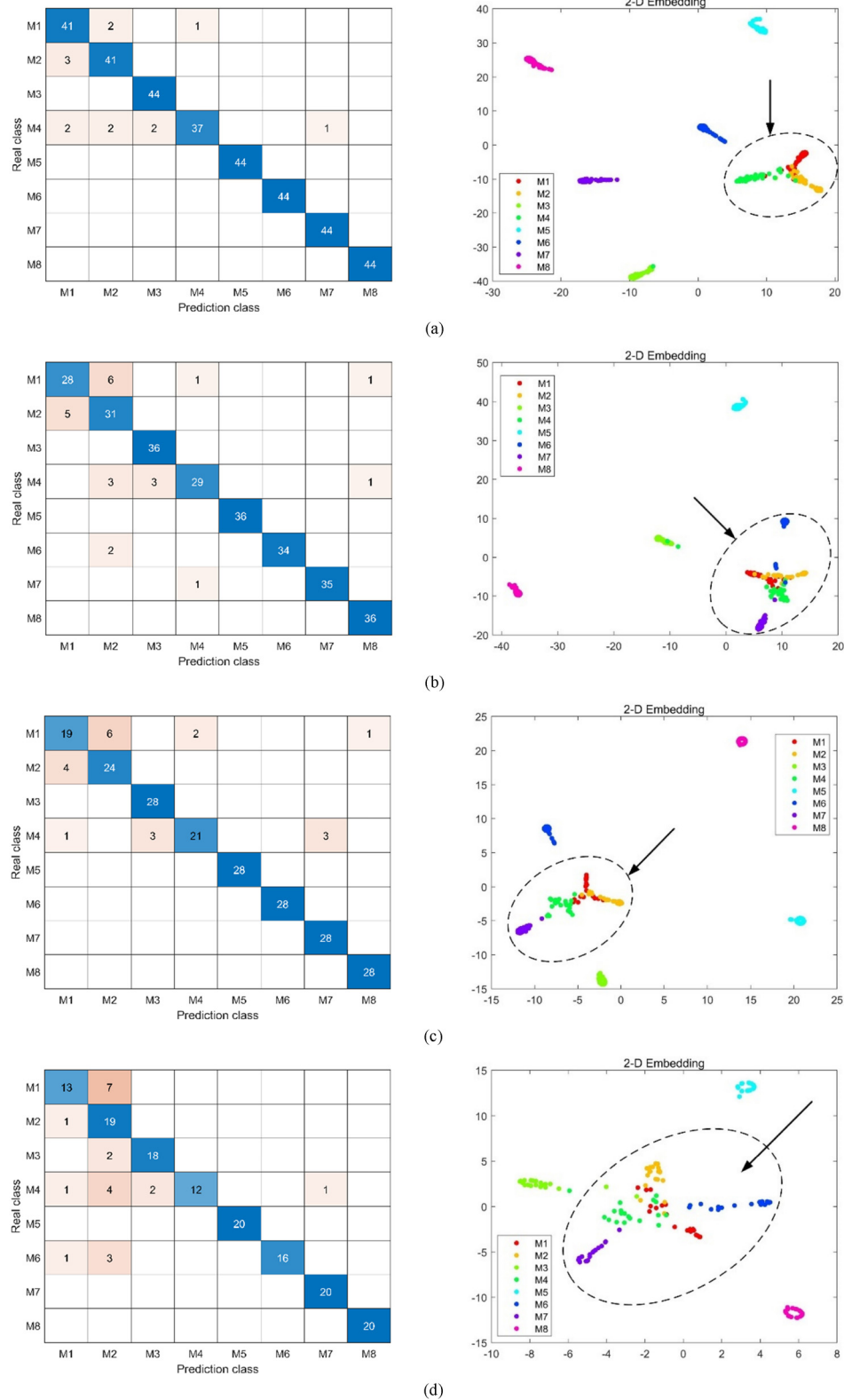
Figure 14: Confusion matrix of the ResNet-18 network training results and the classification features calculated by the fully connected layer under different data amounts: (a) 440; (b) 360; (c) 280; (d) 200.

Table 11: Predicted results of the surrogate model and the actual training values of the network

Model	Hyperparameter values	Prediction accuracy (%)	Actual accuracy (%)	Error (%)
AlexNet	$[9 \times 10^{-5},6,6]$	97.33	97.68	0.35
	$[1 \times 10^{-4},8,7]$	97.99	97.51	0.48
	$[1.1 \times 10^{-4},10,8]$	97.22	98.00	0.78
	$[1.2 \times 10^{-4},12,9]$	96.15	95.62	0.53
ResNet-18	$[9 \times 10^{-6},6,6]$	94.88	94.02	0.86
	$[1 \times 10^{-5},8,7]$	94.50	95.85	1.45
	$[1.1 \times 10^{-5},10,8]$	94.50	95.24	0.74
	$[1.2 \times 10^{-5},12,9]$	96.23	96.00	0.23
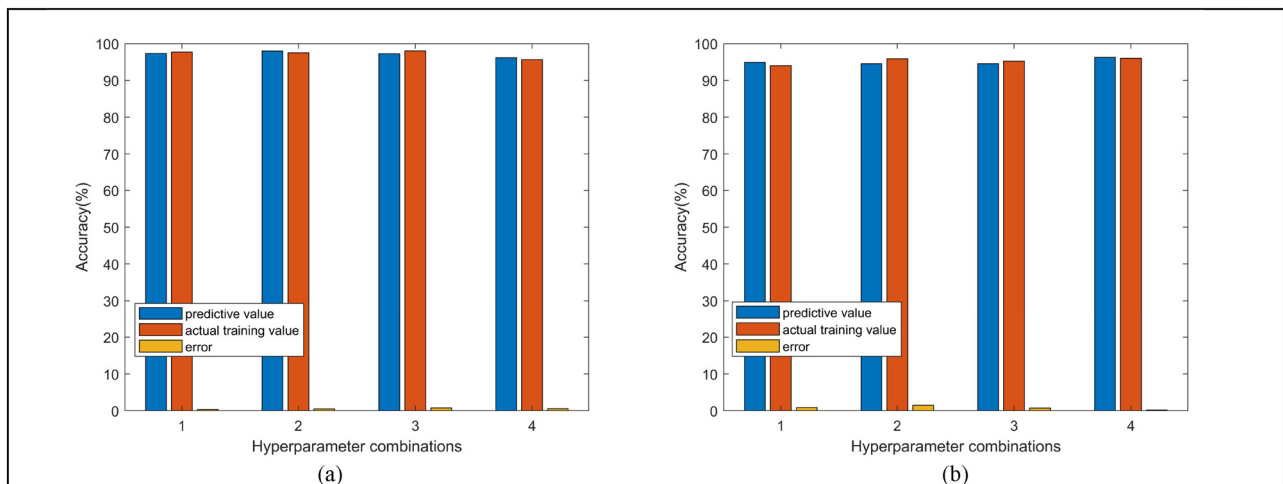
Figure 15: The predicted results of the surrogate model and the actual training values of the network. (a) AlexNet; (b) ResNet-18.

learned. These two methods are denoted by NF-AlexNet and NF-ResNet-18, respectively. Inception-v3 network with freezing strategy (F-Inception-v3) [66], VGG-16 network with freezing strategy (F-VGG-16) [67], support vector machine with PSO parameter optimization (PSO–SVM) [68], kernel learning machine with PSO (PSO–KLM) [69], and random forest with PSO (PSO–RF) [70]. The input of the SVM is the one-dimensional spectral distribution of the vibration signal after STFT calculation. The two convolutional models in this article are denoted by PSO–GWO–BPNN-AlexNet and PSO–GWO–BPNN-ResNet-18, respectively. It should also be noted that the network models used all have initial parameters, that is, they are pre-trained on the ImageNet image set. First, the sample size was set to 440, and the mean of the 10 experimental results for each model is shown in Table 12.

From the above experimental results, it can be concluded that the PSO–GWO–BPNN-AlexNet and PSO–GWO–BPNN-ResNet-18 methods proposed in this article can reach the highest 98.58% and 96.31%, respectively, surpassing other models; and the time cost is the least,

Table 12: Comparison between models

Model	Average accuracy (%)	Training time (s)
NF-AlexNet	95.84	130
NF-ResNet-18	95.66	95
F-Inception-v3	90.65	120
F-VGG-16	94.34	233
PSO–RF	76.30	93
PSO–KLM	75.23	86
PSO–SVM	71.33	63
PSO–GWO–BPNN–AlexNet	98.58	24
PSO–GWO–BPNN–ResNet-18	96.31	20

which reduces the time by 40–100 s compared with other models. This fully proves that the diagnosis method in this article can complete the hybrid fault diagnosis of a diesel engine with high precision and efficiency.

In addition, to further verify the stable performance of the diagnostic algorithm under small sample conditions. Set the sample sizes to 440, 360, 280, and 200,

respectively, and reenter the above model for training. The mean of 10 experimental results is shown in Figure 16. It can be seen from the experimental results that the PSO–GWO–BPNN-AlexNet and PSO–GWO–BPNN-ResNet-18 methods in this article perform very stable under the conditions of small samples and maintain high diagnostic accuracy. NF-AlexNet and NF-ResNet-18 perform parameter learning for all network layers, so the fault features of the data are relatively fully learned. Therefore, when the amount of data is reduced, the accuracy will not drop significantly. However, since the hyperparameters of the network cannot be optimized, the overall diagnostic accuracy is still lower than that of PSO–GWO–BPNN-AlexNet and PSO–GWO–BPNN-ResNet-18. NF-AlexNet and NF-ResNet-18 perform parameter learning for all network layers, so the fault features of the data are relatively fully learned. Therefore, when the amount of data is reduced, the accuracy will not drop significantly. However, since the hyperparameters of the network cannot be optimized, the overall diagnostic accuracy is still lower than that of PSO–GWO–BPNN-AlexNet and PSO–GWO–BPNN-ResNet-18. This shows that the deeper the number of network layers does not necessarily mean that the learning effect of the data is better, but the appropriate network model and diagnosis method need to be selected according to the data. Finally, the diagnostic effect of PSO–RF, PSO–KLM, and PSO–SVM is not very satisfactory, and the accuracy rate drops greatly when the sample size decreases. This is because traditional machine learning methods cannot learn classification features only from the spectral distribution of the data, and they do not have the powerful feature extraction ability of deep networks.
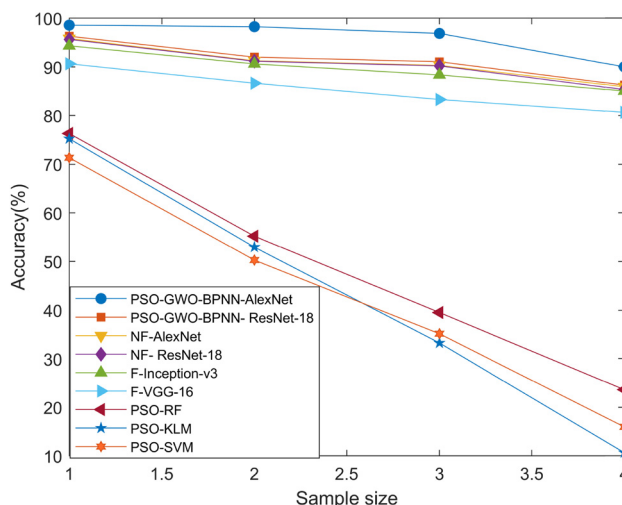
To sum up, compared with the common deep models and machine learning diagnosis algorithms, the diagnosis method proposed in this article can not only realize the hybrid fault diagnosis of a diesel engine with high accuracy and efficiency but also has very stable and good performance under the condition of small samples.

6 Conclusion

A small-sample transfer learning fault diagnosis algorithm based on the combination of diesel engine STFT time–frequency images and hyperparameter autonomous optimization deep convolutional networks improved by a surrogate model is proposed in this article. Taking the STFT time–frequency image of the vibration signal of the diesel engine cylinder head as a sample, the hyperparameters of the AlexNet and ResNet-18 networks were autonomously optimized using the PSO–GWO–BPNN surrogate model. The transfer learning from the ImageNet dataset to diesel engine fault samples is realized, and high fault diagnosis accuracy and diagnosis efficiency are obtained. The main contributions of this method are as follows:

1) The vibration signal from the cylinder head under different fault types of the diesel engine is collected by the acceleration vibration sensor. The vibration signal is then converted into a 3-channel RGB time–frequency image using STFT calculations, providing a good input for the diagnostic model. In addition, this simple feature extraction method minimizes the factors of the artificial selection of features.

2) By combining the PSO algorithm and the GWO algorithm, the PSO algorithm is used to optimize the position update method of the GWO algorithm, which enhances the communication between individuals and groups in the GWO algorithm and further improves the global optimization ability of the GWO algorithm. In addition, the PSO–GWO–BPNN surrogate model is formed by combining the PSO–GWO algorithm and BPNN, and the hyperparameters of the AlexNet and ResNet-18 networks are effectively and autonomously optimized. It is verified by the diesel engine fault diagnosis experiment that the two types of convolutional network models can efficiently classify the hybrid faults of diesel engines after being optimized by the surrogate model.

3) In this article, two network models of AlexNet and ResNet-18 are used as transfer models, and the transfer learning strategy of freezing the shallow

Figure 16: Model training results with different sample sizes.

network parameters and fine-tuning the deep network parameters realize the knowledge transfer from the ImageNet image set to the time–frequency images of diesel engine fault signals. After the fault diagnosis experiment under the condition of a small sample, it is concluded that compared with the common deep network model and machine learning algorithm, the diagnosis algorithm in this article can not only realize the hybrid fault diagnosis of a diesel engine with high precision but also has excellent and stable performance under the condition of a small sample.

To sum up, compared with the existing fault diagnosis algorithms, the algorithm proposed in this article has low requirements on the amount of data and does not require the staff to have too much professional knowledge. More importantly, the algorithm can autonomously and efficiently adjust the parameters according to the data, to better learn the knowledge and characteristics of the data. Therefore, this research can provide a theoretical reference for practical application technologies such as fault diagnosis of industrial equipment. At the same time, there are still problems and deficiencies that can be further studied in this article. For example, optimizing the deep network structure, improving the feature extraction and learning capabilities of the network, researching surrogate models with higher accuracy to improve the efficiency of parameter optimization, and finding more effective migration strategies to further reduce the data volume requirements and improve the accuracy and efficiency of fault diagnosis.

Author contributions: All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Conflict of interest: The authors state no conflict of interest.

References

[1] Deng W, Chu Z, Li ZX, Li XY, Chen HY, Zhao HM. Compound fault diagnosis using optimized MCKD and sparse representation for rolling bearings. IEEE Trans Instrum Meas. 2022;71:1–9.

[2] Cui HJ, Guan Y, Chen HY. Rolling element fault diagnosis based on VMD and sensitivity MCKD. IEEE Access. 2021;9:120297–120308.

[3] Cerrada M, Zurita G, Cabrera D, Sánchez RV, Artés M, Li C. Fault diagnosis in spur gears based on genetic algorithm and random forest. Mech Syst Signal Process. 2016;70–71:87–103.

[4] Cao R, Yunusa-Kaltungo A. An automated data fusion-based gear faults classification framework in rotating machines. Sensors. 2021;21:21.

[5] Wang R, Chen H, Guan C. A Bayesian inference-based approach for performance prognostics towards uncertainty quantification and its applications on the marine diesel engine. ISA Trans. 2021;118:159–73.

[6] Ke Y, Yao C, Song E, Dong Q, Yang L. An early fault diagnosis method of common-rail injector based on improved CYCBD and hierarchical fluctuation dispersion entropy. Digital Signal Processing: A Rev J. 2021;114:114.

[7] Wang R, Chen H, Guan C. Random convolutional neural network structure: An intelligent health monitoring scheme for diesel engines. Meas: J Int Measurement Confederation. 2021;54:171–43.

[8] Gu C, Qiao XY, Li H, Jin Y. Misfire fault diagnosis method for diesel engine based on MEMD and dispersion entropy. Shock Vib. 2021;2021:2021–14.

[9] Hou XL, Wang X. Application of fractal theory in fault diagnosis of nonlinear mechanical equipment system: A review. IOP Conference Series: Materials Science and Engineering; 2021. p. 1009.

[10] Jing YB, Liu CW, Bi FR, Bi XY, Wang X, Shao K. Diesel engine valve clearance fault diagnosis based on features extraction techniques and FastICA-SVM. Chin J Mech Eng (Engl Ed). 2017;30:991–1007.

[11] Zhang JH, Liu Y. Application of complete ensemble intrinsic time-scale decomposition and least-square SVM optimized using hybrid DE and PSO to fault diagnosis of diesel engines. Front Inf Technol Electron Eng. 2017;18:272–86.

[12] Ramteke SM, Chelladurai H, Amarnath M. Diagnosis and classification of diesel engine components faults using time–frequency and machine learning approach. J Vib Eng Technol. 2022;10:175–92.

[13] Liu Y, Zhang J, Ma L. A fault diagnosis approach for diesel engines based on self-adaptive WVD, improved FCBF and PECOC-RVM. Neurocomputing. 2016;177:600–11.

[14] Li X, Bi FR, Yang X, Tang DJ, Shen SF. Engine multiple faults detection base on bispectrum and convolutional neural network. International Conference on Sensors and Instruments 2021. Qingdao, China. 2021.

[15] Du C, Zhong R, Zhuo Y, Zhang X, Yu F, Li F, et al. Research on fault diagnosis of automobile engines based on the deep learning 1D-CNN method. Eng Res Exp. 2022;4:4.

[16] Shatnawi Y, Al-Khassaweneh M. Fault diagnosis in internal combustion engines using extension neural network. IEEE Trans Ind Electron. 2014;61:1434–43.

[17] Ioffe S, Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. Lile, France: Microtome Publishing; 2015.

[18] Zhang T, Chen J, Xie J, Pan T. SASLN: Signals augmented self-taught learning networks for mechanical fault diagnosis under small sample condition. IEEE Trans Instrum Meas. 2021;70:70–11.

[19] Yu WX, Lu Y, Wang JN. Application of small sample virtual expansion and spherical mapping model in wind turbine fault diagnosis. Expert Syst Appl. 2021;183:183.

[20] Institute of Electrical and Electronics Engineers, IEEE Signal Processing Society. 2015 IEEE International Conference on Image Processing: proceedings: ICIP; 2015. p. 7–30.

[21] Kunang YN, Nurmaini S, Stiawan D, Suprapto BY. Attack classification of an intrusion detection system using deep learning and hyperparameter optimization. J Inf Security Appl. 2021;58:58.

[22] Dong Y, Li Y, Zheng H, Wang R, Xu M. A new dynamic model and transfer learning based intelligent fault diagnosis framework for rolling element bearings race faults: Solving the small sample problem. ISA Trans. 2022;121:327–48.

[23] Zhong SS, Fu S, Lin L. A novel gas turbine fault diagnosis method based on transfer learning with CNN. Meas: J Int Measurement Confederation. 2019;137:435–53.

[24] Xiong G, Ma W, Zhao N, Zhang J, Jiang Z, Mao Z. Multi-type diesel engines operating condition recognition method based on stacked auto-encoder and feature transfer learning. IEEE Access. 2021;9:31043–52.

[25] Bai M, Yang X, Liu J, Liu J, Yu D. Convolutional neural network-based deep transfer learning for fault detection of gas turbine combustion chambers. Appl Energy. 2021;302:302.

[26] Lei X, Lu N. A deep transfer learning base fault diagnosis method for diesel generator. Jiangsu Annual Conference on Automation 2021, Jiangsu, China. 2021.

[27] Li GY, Li YX, Chen HY, Deng W. Fractional-order controller for course-keeping of underactuated surface vessels based on frequency domain specification and improved particle swarm optimization algorithm. Appl Sci. 2022;6:3139.

[28] Deng W, Zhang XX, Zhou YQ, Liu Y, Zhou XB, Chen HL, et al. An enhanced fast non-dominated solution sorting genetic algorithm for multi-objective problems. Inf Sci. 2022;585:441–53.

[29] Zhu Y, Li G, Wang R, Tang S, Su H, Cao K. Intelligent fault diagnosis of hydraulic piston pump combining improved LeNet-5 and PSO hyperparameter optimization. Appl Acoust. 2021;183:183.

[30] Han JH, Choi DJ, Park SU, Hong SK. Hyperparameter optimization using a genetic algorithm considering verification time in a convolutional neural network. J Electr Eng Technol. 2020;15:721–6.

[31] Tong J, Luo J, Pan H, Zheng J, Zhang Q. A novel cuckoo search optimized deep auto-encoder network-based fault diagnosis method for rolling bearing. Shock Vib. 2020;2020:1–12.

[32] Zhou XB, Ma HJ, Gu JG, Chen HL, Deng W. Parameter adaptation-based ant colony optimization with dynamic hybrid mechanism. Eng Appl Artif Intell. 2022;114:105139.

[33] An ZY, Wang XM, Li B, Xiang ZL, Zhang B. Robust visual tracking for UAVs with dynamic feature weight selection. Appl Intell. 2022.

[34] Wu DQ, Wu CX. Research on the time-dependent split delivery green vehicle routing problem for fresh agricultural products with multiple time windows. Agriculture. 2020;12:793.

[35] Taghizadeh-Alisaraei A, Mahdavian A. Fault detection of injectors in diesel engines using vibration time-frequency analysis. Appl Acoust. 2019;143:48–58.

[36] Siavash NK, Najafi G, Hassan-Beygi SR, Ahmadian H, Ghobadian B, Yusaf T, et al. Time-frequency analysis of diesel engine noise using biodiesel fuel blends. Sustainability (Switz). 2021;13:13.

[37] Liu H, Li L, Ma J. Rolling bearing fault diagnosis based on STFT-deep learning and sound signals. Shock Vib. 2016;2016:2016–12.

[38] 2019 Prognostics and System Health Management Conference (PHM-Qingdao). IEEE; 2019.

[39] Wen L, Gao L, Li X. A new deep transfer learning based on sparse auto-encoder for fault diagnosis. IEEE Trans Syst Man Cyber Syst. 2019;49:136–44.

[40] Guo L, Lei Y, Xing S, Yan T, Li N. Deep convolutional transfer learning network: A new method for intelligent fault diagnosis of machines with unlabeled data. IEEE Trans Ind Electron. 2019;66:7316–25.

[41] Han T, Liu C, Wu R, Jiang D. Deep transfer learning with limited data for machinery fault diagnosis. Appl Soft Comput. 2021;103:103.

[42] Weiss K, Khoshgoftaar TM, Wang DD. A survey of transfer learning. J Big Data. 2016;3:3.

[43] He K, Girshick R, Dollár P. Rethinking ImageNet Pre-training United States: IEEE Xplore; 2019.

[44] Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. ImageNet large scale visual recognition challenge. Int J Computer Vis. 2015;115:211–52.

[45] Wang C, Chen D, Hao L, Liu X, Zeng Y, Chen J, et al. Pulmonary image classification based on inception-v3 transfer learning model. IEEE Access. 2019;7:146533–41.

[46] Shi X, Cheng Y, Zhang B, Zhang H. Intelligent fault diagnosis of bearings based on feature model and AlexNet neural network. Proceedings of the Annual Conference of the Prognostics and Health Management Society. PHM, 2020, Prognostics and Health Management Society; 2020.

[47] Ghulanavar R, Dama KK, Jagadeesh A. Diagnosis of faulty gears by modified AlexNet and improved grasshopper optimization algorithm (IGOA). J Mech Sci Technol. 2020;34:4173–82.

[48] Gao M, Song P, Wang F, Liu J, Mandelis A, Qi D. A novel deep convolutional neural network based on ResNet-18 and transfer learning for detection of wood knot defects. J Sensors. 2021;2021.

[49] Yang B, Li Q, Chen L, Shen C. Bearing fault diagnosis based on multilayer domain adaptation. Shock Vib. 2020;104:2020.

[50] Xie W, Chen W, Shen L, Duan J, Yang M. Surrogate network-based sparseness hyper-parameter optimization for deep expression recognition. Pattern Recognit. 2021;111:111.

[51] Jafar A, Lee M. High-speed hyperparameter optimization for deep ResNet models in image recognition. Clust Comput. 2021

[52] Rezaei H, Bozorg-Haddad O, Chu X. Grey wolf optimization (GWO) algorithm. Studies in. Computational Intell. 2018;720:81–91.

[53] Naserbegi A, Aghaie M, Zolfaghari A. Implementation of Grey Wolf Optimization (GWO) algorithm to multi-objective loading pattern optimization of a PWR reactor. Ann Nucl Energy. 2020;148:148.

[54] Jennings NR. Software Engineering and Computer Systems (ICSECS). 2015 4th International Conference on IEEE. p. 2015.

[55] Deng W, Yao R, Zhao H, Yang X, Li G. A novel intelligent diagnosis method using optimal LS-SVM with improved PSO algorithm. Soft Comput. 2019;23:2445–62.

[56] Liang W, Wang G, Ning X, Zhang J, Li Y, Jiang C, et al. Application BP neural Netw prediction coal ash melting Charact Temp. 2020;260.

[57] Peng Y, Xiang W. Short-term traffic volume prediction using GA-BP based on wavelet denoising and phase space reconstruction. Phys A: Stat Mech Its Appl. 2020;549:549.

[58] Zhang D, Lou S. The application research of neural network and BP algorithm in stock price pattern classification and prediction. Future Gener Computer Syst. 2021;115:872–9.

[59] Yang H, Li X, Qiang W, Zhao Y, Zhang W, Tang C. A network traffic forecasting method based on SA optimized ARIMA–BP neural network. Comput Netw. 2021;193:193.

[60] Han JX, Ma MY, Wang K. Product modeling design based on genetic algorithm and BP neural network. Neural Comput Appl. 2021;33:4111–7.

[61] Nath MK, Kanhe A, Mishra M. A novel deep learning approach for classification of COVID-19 images. 2020 IEEE 5th International Conference on Computing Communication and Automation, ICCCA 2020, Institute of Electrical and Electronics Engineers Inc. Vol. 752; 2020. p. 7.

[62] Sun D, Wen H, Wang D, Xu J. A random forest model of landslide susceptibility mapping based on hyperparameter optimization using Bayes algorithm. Geomorphology. 2020;362:362.

[63] PRASA-RobMech 2017 Bloemfontein, Pattern Recognition Association of South Africa, Institute of Electrical and Electronics Engineers, Pattern Recognition Association of South Africa and Robotics and Mechatronics International Conference; 2017.

[64] Cui H, Bai J. A new hyperparameters optimization method for convolutional neural networks. Pattern Recognit Lett. 2019;125:828–34.

[65] Young SR, Rose DC, Karnowski TP, Lim SH, Patton RM. Optimizing deep learning hyper-parameters through an evolutionary algorithm. Proceedings of MLHPC 2015: Machine Learning in High-Performance Computing Environments - Held in conjunction with SC 2015: The International Conference for High Performance Computing, Networking, Storage and Analysis, Association for Computing Machinery; 2015.

[66] Liu Z, Yang C, Huang J, Liu S, Zhuo Y, Lu X. Deep learning framework based on integration of S-Mask R-CNN and Inception-v3 for ultrasound image-aided diagnosis of prostate cancer. Future Gener Computer Syst. 2021;114:358–67.

[67] Kaur T, Gandhi TK. Automated brain image classification based on VGG-16 and transfer learning. United States: IEEE Xplore; 2019.

[68] Ye M, Yan X, Jia M. Rolling bearing fault diagnosis based on VMD-MPE and PSO-SVM. Entropy. 2021;23:23.

[69] Li K, Su L, Wu J, Wang H, Chen P. A rolling bearing fault diagnosis method based on variational mode decomposition and an improved kernel extreme learning machine. Appl Sci. 2017;7(10):1004.

[70] Chen ZC, Han FC, Wu LJ, Wu JL, Cheng SY, Lin PJ, et al. Random forest based intelligent fault diagnosis for PV arrays using array voltage and string currents. Energy Convers Manag. 2018;178:250–64.