

Article

Small-Sample Bearings Fault Diagnosis Based on ResNet18 with Pre-Trained and Fine-Tuned Method

Junlin Niu ^{1,2} , Jiafang Pan ^{1,2,*} , Zhaohui Qin ^{1,2} , Faguo Huang ^{1,2}  and Haihua Qin ^{1,2} 

¹ Key Laboratory of Advanced Manufacturing and Automation Technology (Guilin University of Technology), Education Department of Guangxi Zhuang Autonomous Region, Guilin 541006, China

² Guangxi Engineering Research Center of Intelligent Rubber Equipment (Guilin University of Technology), Guilin 541006, China

* Correspondence: panjiafang@glut.edu.cn

Abstract: In actual production, bearings are usually in a normal working state, which results in a lack of data for fault diagnosis (FD). Yet, the majority of existing studies on FD of rolling bearings focus on scenarios with ample fault data, while research on diagnosing small-sample bearings remains scarce. Therefore, this study presents an FD method for small-sample bearings, employing variational-mode decomposition and Symmetric Dot Pattern, combined with a pre-trained and fine-tuned Residual Network18 (VSDP-TLResNet18). The approach utilizes variational-mode decomposition (VMD) to break down the signal, determining the k value and the best Intrinsic-Mode Function (IMF) component based on center frequency and kurtosis criteria. Following this, the chosen IMF component is converted into a two-dimensional image using the Symmetric Dot Pattern (SDP) transform. In order to maximize the discrimination between two-dimensional fault images, Pearson correlation analysis is carried out on the parameters of SDP to select the optimal parameters. Finally, we use the pre-trained and fine-tuned method combined with ResNet18 for small-sample FD to improve the diagnosis accuracy of the model. Relative to alternative approaches, the suggested method demonstrates strong performance when dealing with small-sample FD.

Keywords: residual network18; small sample; pre-trained and fine-tuned; fault diagnosis; variational-mode decomposition; symmetric dot pattern



Citation: Niu, J.; Pan, J.; Qin, Z.; Huang, F.; Qin, H. Small-Sample Bearings Fault Diagnosis Based on ResNet18 with Pre-Trained and Fine-Tuned Method. *Appl. Sci.* **2024**, *14*, 5360. <https://doi.org/10.3390/app14125360>

Academic Editor: Oscar Duque-Perez

Received: 16 May 2024

Revised: 12 June 2024

Accepted: 19 June 2024

Published: 20 June 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Rolling bearings are crucial components of contemporary mechanical equipment and are extensively employed in a variety of mechanical rotating systems, including automobiles, aircraft, wind turbines and diverse industrial machinery [1]. Their main function is to support the rotating shaft and reduce the friction coefficient in the process of movement, so as to improve the operation efficiency and prolong the service life of the equipment. Thus, implementing effective FD technology not only enhances the operational efficiency and safety of equipment but also offers significant economic savings for businesses [2,3].

Vibration analysis is an efficient and economic technology in bearing FD. It extracts features by analyzing the collected vibration signals [4]. Recently, numerous techniques have been developed to transform bearing vibration signals into two-dimensional images for analysis. Continuous wavelet transform (CWT) [5], short-time Fourier transform (STFT) [6], Gramian angular field (GAF) [7] and Wigner–Ville distribution (WVD) [8] are typically utilized to transform the time–frequency signal into a characteristic image. However, due to the fixed window, STFT cannot easily and accurately capture the signal’s local attributes. CWT’s computational requirements are larger than STFT, and the analysis process is cumbersome [9]. As a means of signal visualization, SDP is commonly employed in FD due to its straightforward processing and ability to reduce computational demands during signal conversion [10]. For example, Wang et al. [11] used the SDP method and the squeeze excitation with a convolutional neural network (SE-CNN) model for FD. Tang et al. [12]

used the SDP method and the graph cardinality preserving attention network (GCPAT) for FD. Sn et al. [13] introduced a novel FD technique using the enhanced Manhattan distance in the SDP image. Nevertheless, as mechanical vibration responses consist of overlapping multi-frequency feature information, it is crucial to isolate fault characteristics through signal breakdown and filtration for definitive fault determination [14]. Various frequency bands exhibit unique fault-associated traits. The VMD algorithm is capable of partitioning the frequency spectrum of the signal and extracting the relevant decomposition components of the designated signal [15–17]. Hence, the VMD method should be utilized for fault feature extraction. Based on this characteristic, the vibration signal can be analyzed using the VMD algorithm, and the resulting components can be converted using SDP. Next, features indicative of the petal shape are extracted through image processing techniques, and a suitable pattern recognition approach is employed to conduct FD.

In many practical applications, the data available for FD are extremely scarce because the bearing is usually in normal working condition [18]. This lack of data poses a challenge to the training of the FD model based on deep learning, due to the insufficiency of fault data, which may impair the model's ability to learn effectively, resulting in reduced generalization capabilities and FD precision. To address this issue, researchers began to explore the use of the generative confrontation network (GAN) and transfer learning (TL) to enhance the model's performance. The GAN is composed of generators and discriminators [19]. Because generators and discriminators confront each other in the training process, generators will continuously improve the quality of generated data and make it approach the real data. Therefore, GAN can usually generate very realistic data samples to address the issue of a small sample [20–22]. Although the GAN performs well in generating realistic data, it faces many challenges. Among them, unstable training, mode collapse and evaluation difficulties are the most prominent problems. Furthermore, the training process demands substantial computational resources and meticulous hyperparameter tuning, and it is highly sensitive to hyperparameter settings.

On the other hand, TL utilizes knowledge from one domain to aid tasks in another related field. In bearing FD, we can pre-train the neural network model on a large dataset and subsequently transfer it for fine-tuning in the task of FD [23–25]. This method of pre-training and fine-tuning can make the model converge faster and still maintain good performance in the case of data scarcity [26,27]. For example, Xie et al. [28] used the TL strategy for small-sample FD. Dong et al. [29] introduced a framework for FD using a dynamic model and TL to address the issue of small samples. Hu et al. [30] used TL and a measurement network to realize cross-domain small-sample FD. Zhang et al. [31] introduced an FD method based on the finite element method and TL. Yuan et al. [32] proposed a multi-channel TL network based on graph attention (GAMTLN) to address the issue of small-sample FD.

To sum up, aiming at the problem that bearing faults are limited and difficult to obtain in reality, we developed a method for FD in small-sample bearings using a two-dimensional image and ResNet18, employing a pre-trained and fine-tuned method. The primary contributions of this study are outlined as follows:

1. The original vibration signal is broken down using VMD; the k value and IMF component of VMD decomposition are selected by using the center frequency and kurtosis criteria.
2. The IMF decomposed by VMD is successively generated into a picture through SDP transformation, and the Pearson correlation analysis method is used to select the optimal SDP parameters to increase the discrimination of a bearing fault image, which further improves the conversion effect from a vibration signal into a two-dimensional image.
3. The ResNet18 model with a pre-trained and fine-tuned method is used to efficiently learn and represent the extracted two-dimensional image features, which realizes more accurate and robust capture and recognition of bearing fault features.

The remainder of this paper is structured as follows. Section 2 presents the fundamental concepts of VMD, SDP, and ResNet18. Section 3 outlines the detailed steps of the proposed FD framework. Section 4 verifies and compares some other methods through two datasets. Section 5 is the conclusion.

2. Relevant Theoretical Approaches

2.1. VMD

VMD is a method for signal processing that decomposes signals by non-recursive variational methods [33]. Unlike Empirical-Mode Decomposition (EMD), VMD can address the issue of frequency aliasing. VMD can break down time-series data into a series of band-limited IMFs and adjust the optimal center frequency and bandwidth of each IMF. In VMD, the goal is to decompose a set of modal components so that the sum of the band widths of these components is minimum, ensuring that the aggregate of all modal components can reproduce the original signal. The calculation equation is as follows [34].

$$\begin{cases} \min_{\{u_k\}, \{\omega_k\}} \left\{ \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right)^* u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \\ \text{s.t. } \sum_{k=1}^k u_k = f(t) \end{cases} \quad (1)$$

where $u_k = \{u_1, u_2, \dots, u_k\}$ represents the k modal components derived from the decomposition, $\omega_k = \{\omega_1, \omega_2, \dots, \omega_k\}$ indicates the central frequency of each component, $*$ denotes the convolution operator, $f(t)$ indicates the signal to be decomposed, i.e., the input signal.

The quadratic penalty operator α is employed to guarantee the precision of the signal reconstruction, whereas the Lagrange multiplication operator $\lambda(t)$ is utilized to uphold the stringency of the constraint.

$$L(\{u_k\}, \{\omega_k\}, \lambda) = \alpha \sum \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 + \left\| f(t) - \sum_k u_k(t) \right\|_2^2 + \lambda(t) \left\| f(t) - \sum_k u_k(t) \right\|_2^k \quad (2)$$

where α is the control parameter of frequency bandwidth or penalty factor; $\lambda(t)$ means Lagrange multiplier.

The alternating direction technique for multipliers is used to alternatively update $\hat{u}_k^{n+1}$; $\hat{\omega}_k^{n+1}$ is employed to obtain the optimal solution, which is the saddle point of the Lagrangian function in Equation (2). The Fourier transform of $\hat{u}_k^{n+1}$ corresponds to the intrinsic mode functions $\hat{\lambda}^{n+1}$, allowing the set of IMFs $\{u_k\}$ and central frequency $\{\omega_k\}$ to be expressed as follows.

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i(\omega) + \frac{\hat{\lambda}(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k)^2} \quad (3)$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |\hat{u}_k(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k(\omega)|^2 d\omega} \quad (4)$$

where $\hat{u}_k(\omega)$ and $\hat{f}(\omega)$ are the Fourier transform of the k th mode and the original signal, respectively, ω_k is the center frequency of the k th mode, n is the number of iterations.

2.2. SDP

The SDP analysis method is a technique that converts time-domain signals collected by sensors from Cartesian coordinates to polar coordinates using a specific equation. The texture features of images transformed by SDP can reflect the amplitude and frequency variations in the original vibration signal. Owing to significant differences in the SDP

images of different bearing fault types, this method allows for the visualization of faults. Figure 1 displays a schematic representation of the SDP transformation. For the vibration signal, $X = \{X_1, X_2, \dots, X_i\}$ converts the amplitude of time i in the bearing vibration signal to $S[r_i, \varphi_i, \varnothing_i]$ in polar coordinates by means of symmetric points, and the calculation equation is as follows [35].

$$r(i) = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (5)$$

$$\varnothing(i) = \theta + \frac{x_{i+l} - x_{\min}}{x_{\max} - x_{\min}} g \quad (6)$$

$$\varphi(i) = \theta - \frac{x_{i+l} - x_{\min}}{x_{\max} - x_{\min}} g \quad (7)$$

where t and Δt are the abscissa axis of time and sampling time, $x_{\min}$ and $x_{\max}$ are the minimum and maximum amplitude of vibration signal time i , $r(i)$ represents the radius of a point in polar coordinates, $\varnothing(i)$ and $\varphi(i)$ are the counterclockwise and clockwise angles of the points relative to the initial line. θ represents the degree of rotation for the mirror symmetric plane, generally 60° [36]. g is the angular amplification factor, generally $g \leq \theta$. l represents a parameter indicating a time interval. l is used to reflect the difference between vibration signals, usually $1 < l < 10$.

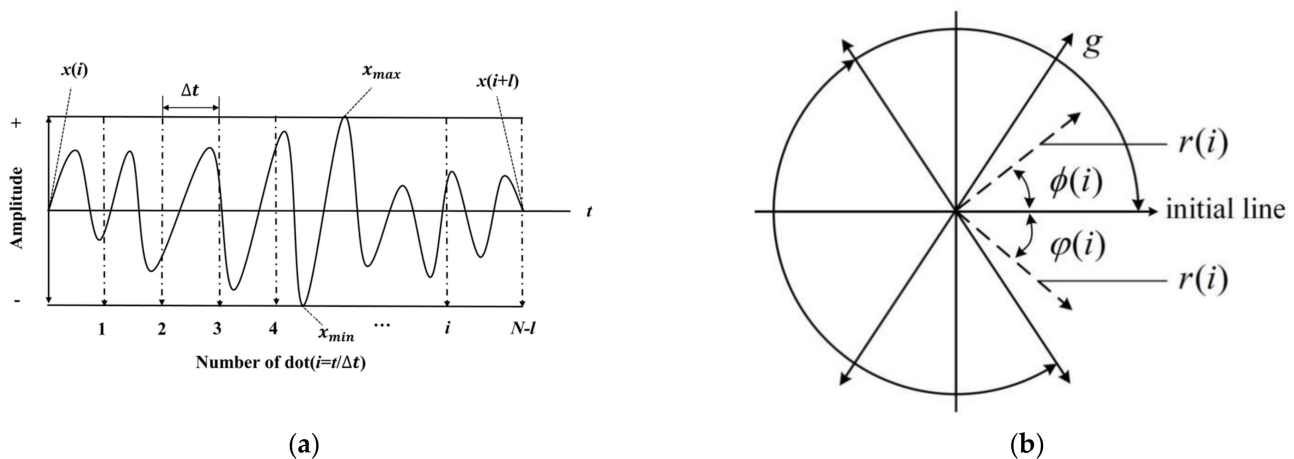


Figure 1. (a) Raw vibration signal; (b) schematic diagram of SDP transformation.

2.3. ResNet

ResNet denotes an architecture for deep neural networks. It was proposed by Kaiming He and his team in 2015 at Microsoft Research Asia [37]. The core innovation of the network design is the introduction of the concept of residual learning, which aims to address issues of gradient disappearance and performance degradation when the network depth increases in the training process. Traditional deep neural networks stack multiple layers to construct the depth model. However, with an increase in the number of network layers, the gradient gradually disappears in the process of back propagation, resulting in difficulty in training. ResNet makes it easier for the network to learn identity mapping through the residual learning mechanism; that is, the input is directly transferred to the output, thus reducing the problem of gradient disappearance. The basic building block of ResNet is the residual block. A schematic depiction of the residual block is depicted in Figure 2. Each residual block contains two convolution layers and a skip connection. By adding the input x directly to the final convolution layer output, the direct flow of information and the effective propagation of gradients are realized.

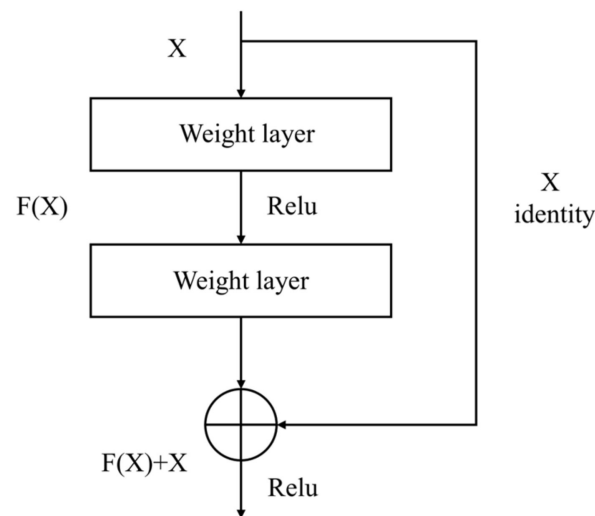


Figure 2. Residual block.

Applying the Rectified Linear Unit (Relu) activation function in the depth network can improve the training speed [38]. Due to the excellent performance of Relu in alleviating gradient disappearance and gradient explosion, improving computational efficiency, achieving sparse activation and experiments, it has become the preferred activation function of residual networks [39]. The equation of Relu is as follows [40].

$$\text{Relu}(x) = \max(0, x) \quad (8)$$

There are two types of jump connections, which mainly depend on whether the output of the residual block needs to be changed in size. As shown in Figure 3, Figure 3a illustrates the output without any processing, whereas Figure 3b requires an adjustment in the output channel size using 1×1 convolution.

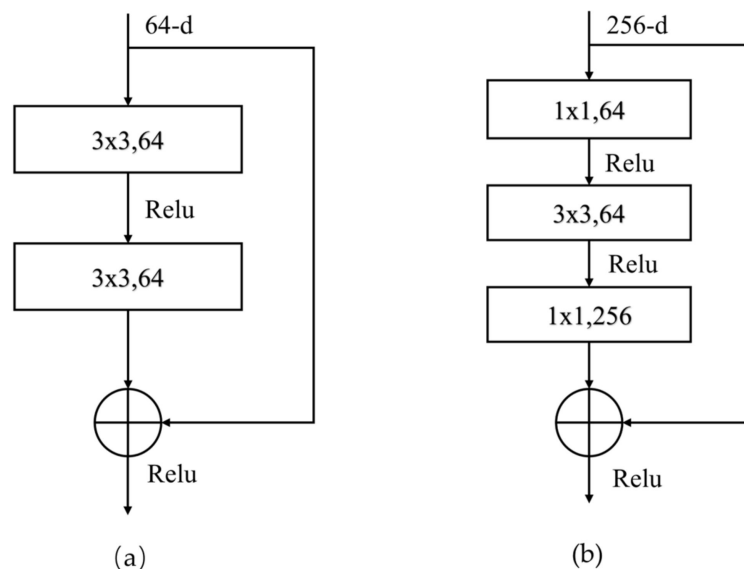


Figure 3. Two different residual structures. (a) Basic residual block (b) Bottleneck residual block.

In the variant network of ResNet, ResNet18 has good performance in image classification and other tasks. Compared with deeper networks, such as ResNet50, it usually needs less training time and fewer parameters, so as to achieve a balance between performance and efficiency to a certain extent. Therefore, this study selects the ResNet18 model for small-sample FD. The ResNet18 model is mainly composed of four groups of residual

blocks. The internal feature dimensions of each group of residual blocks gradually increase. Figure 4 illustrates the architecture of ResNet18.

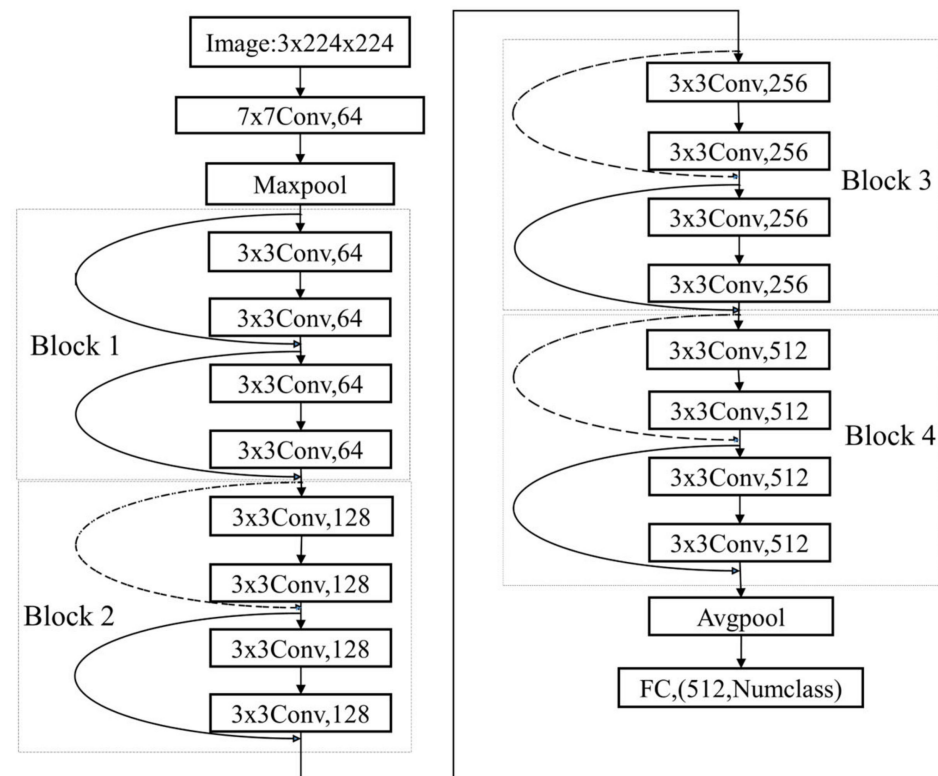


Figure 4. ResNet18 architecture.

2.4. Pre-Trained and Fine-Tuned Model

In the field of machine learning, TL can be divided into sample-based TL, feature-based TL and model-based TL [41]. Sample-based TL helps target domain learning by adjusting the use of source domain samples in the target domain, but it needs to ensure that there is a certain similarity between the source domain and the target domain; otherwise, the effect is limited. Feature-based TL transforms the features of the source domain and the target domain to make them more similar or more discriminative in the common feature space, but the transformation of features may lead to the loss of information and affect the performance of the model. Model-based TL migrates the trained model in the source domain to the target domain and uses the knowledge of the pre-training model to improve the learning effect in the target domain.

In this study, we utilized model-based TL (the pre-trained and fine-tuned method) to enhance the performance of ResNet18 in the context of small-sample FD. Therefore, we utilized a pre-trained ResNet18 model from the ImageNet dataset (source domain). The ImageNet pre-trained weights are trained on extensive-scale image categorization tasks, which makes them perform well in learning common image features. Because of the scale and variety of the ImageNet dataset, pre-trained weights can capture rich image features. In many cases, employing pre-trained weight can enhance both the performance and generalization capability of the model. First, we froze the weights of all models, except the classifier, to keep the feature representation learned by the ResNet18 model in the pre-trained phase unchanged. Next, we eliminated the initial output layer from the pre-trained model and replaced it with a new output layer, which is designed for our small-sample dataset (target domain). Finally, the model is trained on our small-sample dataset. The output layer is trained from scratch, and the parameters of other layers are fine-tuned according to the parameters of the pre-trained ResNet18 model. The principle of pre-training and fine-tuning is shown in Figure 5.

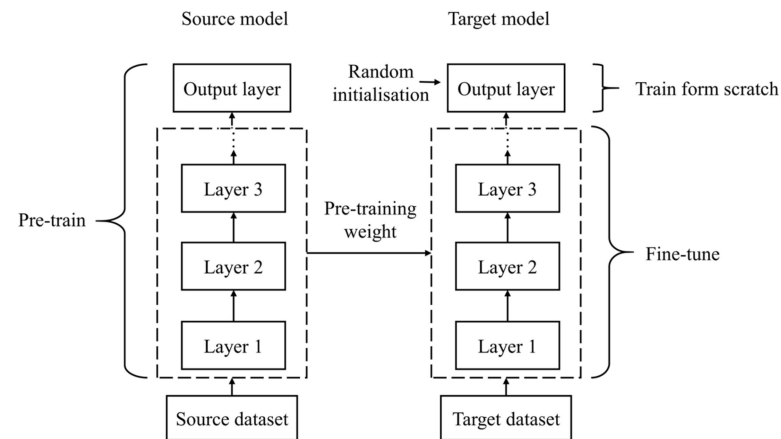


Figure 5. Principle of pre-training and fine-tuning.

3. Proposed Framework

In many practical scenarios, the bearing is in a normal working state most of the time, so the data available for FD are very limited. The limited availability of data poses challenges to training neural network-based diagnostic models. The insufficient fault data makes it challenging for the model to learn effectively, potentially impacting both its generalization ability and the accuracy of fault identification. To enhance the accuracy and adaptability of FD with small samples, this study introduces a VSDP-TLResNet18 framework for FD, employing a pre-trained and fine-tuned method, outlined in Figure 6. The methodology consists of the following steps:

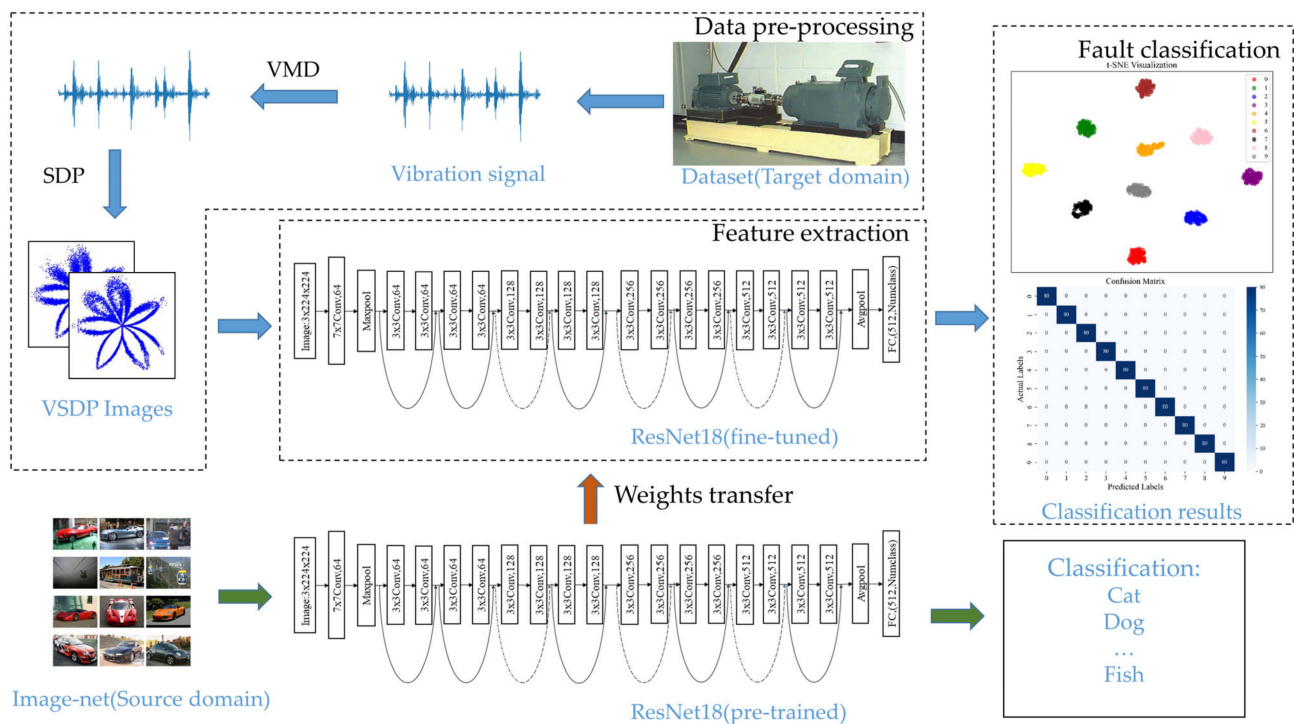


Figure 6. Diagnostic framework of the proposed method.

Initially, the original vibration signal is decomposed and processed using the VMD method. In VMD, the count of modal components k influences the outcome of the decomposition, and incorrect parameter settings can result in either incomplete or excessive decomposition. The k value is set based on the center frequency. Each IMF component's kurtosis is calculated, and the top-six IMFs are selected according to the kurtosis criterion.

To fully leverage deep learning in image processing, the six IMF components selected are successively generated into a two-dimensional image by SDP transform. Reasonable selection of parameters g and l can improve the image quality and strengthen the discrimination between fault signals, making the distinction between various fault vibration signals more obvious. In order to maximize the difference and visualize different fault images, the optimal parameters g and l of SDP are determined by Pearson correlation analysis. After generating SDP fault images, the SDP images representing various fault types are split into training and testing sets, proportionally.

The VSDP-TLResNet18 model employs the SDP image as its input. Through the pre-trained and fine-tuned method, the knowledge of the ResNet18 model pre-trained on large datasets is transferred to the task of FD; this is done to enhance the model's performance on small-sample bearing datasets. The training set is utilized for training the VSDP-TLResNet18 model. Simultaneously, the last epoch of model weight will be saved for testing, and the diagnostic results will be output.

4. Experimental Verification

To evaluate the applicability and generalization of the VSDP-TLResNet18 model, a range of tests were carried out on the Case Western Reserve University (CWRU) bearing dataset [42] and the University of Paderborn (PU) bearing dataset [43] to evaluate the model's performance. The computer hardware configuration required for the experiment is as follows: the CPU model is Intel Core i5-7300H, and the memory is 8 GB; the GPU model is NVIDIA GeForce GTX1050.

4.1. Case 1: CWRU Bearing Dataset

For the experiment, the drive end acceleration data from SKF bearings obtained from the CWRU bearing dataset were utilized, with a sampling frequency of 12 kHz. As shown in Figure 7, the faulty bearing is made by EDM. The dataset simulates four working conditions of 0 horsepower (HP), 1HP, 2HP and 3HP, corresponding to four speeds of 1797 r/min, 1772 r/min, 1750 r/min and 1730 r/min, respectively. The single-point diameter damage of the bearing selected in the experiment is 0.007 mm, 0.014 mm and 0.021 mm, respectively. Rolling element faults, inner ring faults and outer ring faults are present within each fault diameter.

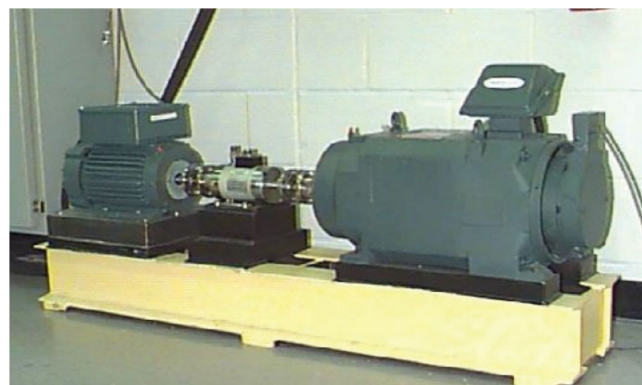


Figure 7. CWRU test bench.

The experimental dataset comprises nine sets of fault data and one set of normal data. The working condition is 0HP. Each sample is sliced into segments with a fixed length of 1024 data points, and 512 adjacent samples overlap. To facilitate feature extraction by the neural network, the samples are normalized to fall within the range of [0, 1]. There are 10 fault types in the dataset, with 100 samples for each type of fault. The dataset is divided into training and testing sets in a 2:8 ratio. The training set contains 200 samples, while the testing set includes 800 samples. The detailed dataset information is shown in Table 1.

Table 1. Dataset composition.

Fault Sizes	Fault Type	Code	Label	Train	Test
0.01778 cm	Inner ring	IR1	0	20	80
	Outer ring	RB1	1	20	80
	Ball	OR1	2	20	80
0.03556 cm	Inner ring	IR2	3	20	80
	Outer ring	RB2	4	20	80
	Ball	OR2	5	20	80
0.05334 cm	Inner ring	IR3	6	20	80
	Outer ring	RB3	7	20	80
	Ball	OR3	8	20	80
None	Normal	N	9	20	80

The chosen modal number significantly influences the effectiveness of VMD decomposition. When a small modal number is chosen, the VMD algorithm, acting like an adaptive filter bank, filters out essential information from the original signal, thus impacting the accuracy of subsequent predictions. With a large modal selection value, the center frequencies of neighboring modal components become closer, leading to the duplication of modes or introduction of additional noise. The primary distinction between different modes is the variance in central frequency. The suitable modal value is determined by examining the distribution of central frequencies across various modal numbers. As depicted in Table 2, with a modal number of 9, modal components with similar frequencies begin to appear, so k is taken as 8. When k equals 8, the central frequency resolution is preserved, and the stability and distribution uniformity of the central frequency are also maintained.

Table 2. Center frequencies of different k values.

k	Center Frequency/HZ								
5	0.0496	0.1097	0.2268	0.2877	0.3133				
6	0.0497	0.1078	0.2103	0.2369	0.2901	0.3146			
7	0.0498	0.1052	0.1576	0.2178	0.2818	0.3026	0.3268		
8	0.0037	0.0519	0.1085	0.2129	0.2368	0.2846	0.3048	0.3306	
9	0.0037	0.0518	0.1084	0.2114	0.2358	0.2818	0.3009	0.3205	0.3657

With the SDP transformation in progress θ , generally, the SDP image formed by taking 60° is a six-petal snowflake image. In order to make the IMF component decomposed by VMD correspond to the image transformed by SDP, we evaluated the kurtosis of eight IMF components decomposed by 10 kinds of faults. Next, we calculated the average kurtosis of these IMF components. Table 3 displays the average kurtosis of IMF components. Six IMF components are chosen based on the maximum kurtosis criterion. The six IMF components are successively generated on the same SDP image by SDP transformation, and the size of the generated image is $3 \times 224 \times 224$.

Table 3. Average kurtosis of 8 IMF components.

	IMF1	IMF2	IMF3	IMF4	IMF5	IMF6	IMF7	IMF8
Kurtosis average	2.9749	2.9262	3.1383	5.1418	5.3571	3.7027	3.9743	4.1558

The discrepancies in SDP diagrams among various bearing fault types are manifested in the thickness of the symmetrical arm, the degree of curvature change, the degree of dispersion and aggregation of mapping points. Our study covers g values from 0 to 50 and l values from 1 to 5. To maximize the distinction among the 10 types of bearing faults in SDP images, it is necessary to select appropriate parameters. Therefore, we used the Pearson correlation analysis method in reference [35] to determine the most appropriate parameters.

First, to avoid interference from external factors, the polar axis was eliminated in the SDP image, and a gray-scale transformation was applied, converting it into a two-dimensional numerical matrix. Subsequently, normal samples and random fault samples were selected for quantitative analysis. The average correlation coefficients of g and l under various parameter combinations were determined using the Pearson correlation analysis method. Then, the uncorrelation coefficients of g and l were obtained, and the specific data are shown in Figure 8.

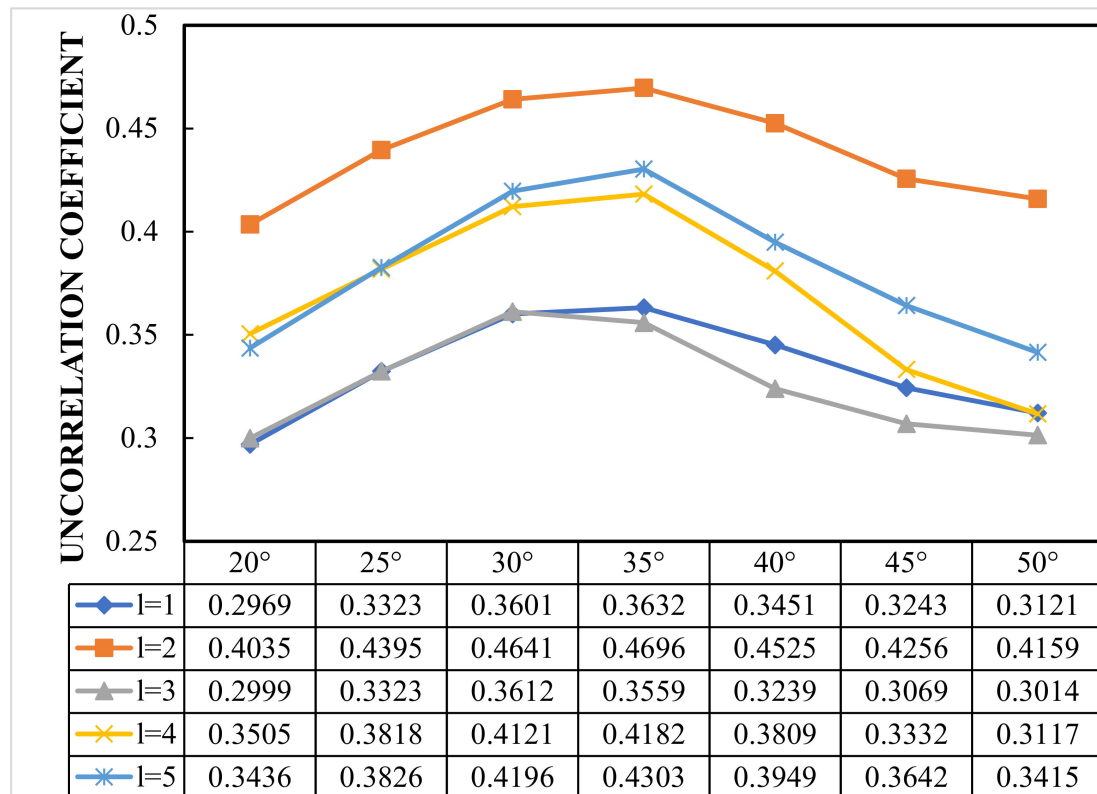


Figure 8. Uncorrelation coefficient of g and l .

The greater the value of the uncorrelation coefficient, the greater the difference between the two parameters, and this implies that the disparity among SDP bearing fault images is the greatest. As shown in Figure 8, when the horizontal label l is 2, the value of each column g is generally high and performs well. Particularly, when the vertical column label g is 35°, the value of the uncorrelation coefficient reaches 0.4696, which is the best in this row. At the same time, it also shows the highest value in all the data, with a vertical column g of 35°, which is significantly better than other values. When the parameters g and l are, respectively, set to 35° and 2, the bearing fault exhibits the greatest disparity in SDP images. When the parameter g is 20° and the parameter l is 1, the difference in the SDP bearing fault image is the smallest, and the value of the uncorrelation coefficient is 0.2969. A VSDP diagram for maximizing the difference of 10 bearing faults is shown in Figure 9 below.

Under the same SDP parameters, the differences in different bearing faults are shown in the thickness of the arm and the density of data points. As can be seen from Figure 9, different bearing fault types show strong discrimination.

4.1.1. Network Verification

The experiment's initial learning rate is configured as 0.0001, the number of epochs is set to 30 and the batch size is 8. Using the cross-entropy loss function, AdamW is utilized as the optimizer [44], and the weight attenuation coefficient is 5×10^{-5} . To further enhance the model's performance during training and ensure optimization stability, we use the cosine

annealing strategy to adjust the learning rate. The evaluation metrics include accuracy (ACC), recall (R), F1 score and precision (P). The equation for specific calculations is as follows [45]. Figure 10 illustrates the training set in the experimental results.

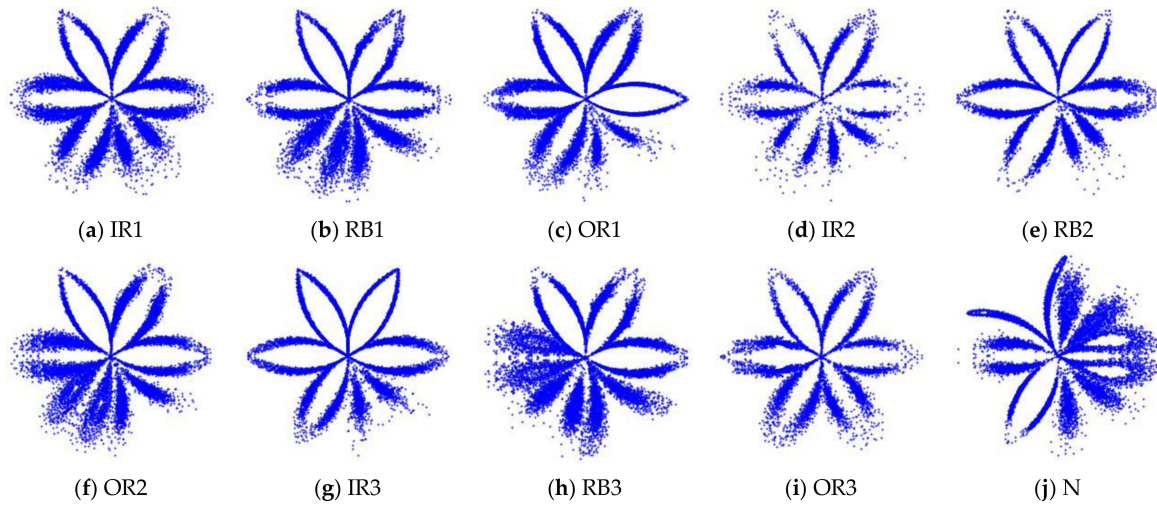


Figure 9. VSDP images of 10 faults.

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (9)$$

$$P = \frac{TP}{TP + FP} \quad (10)$$

$$R = \frac{TP}{TP + FN} \quad (11)$$

$$F1 = \frac{2(P \times R)}{P + R} \quad (12)$$

where TP is true positive; FP is false positive; TN and FN are true negative and false negative, respectively.

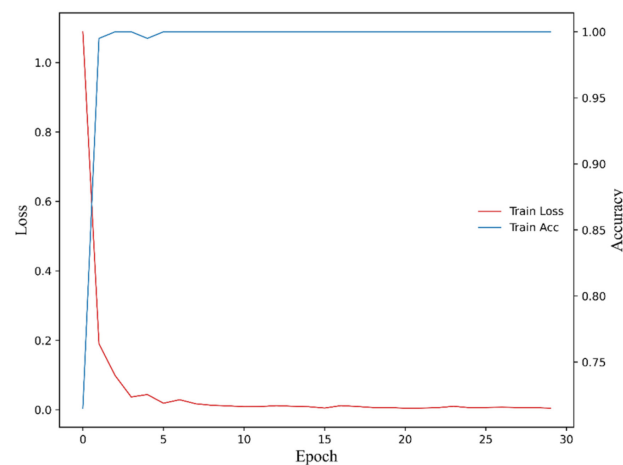


Figure 10. Training accuracy and training loss of the model under Case 1.

Figure 10 indicates a rapid initial increase in the model's training accuracy, which indicates that the initialization performance of the pre-trained and fine-tuned method is much better than that of the random initialization, and then gradually tends to be stable. The training loss decreases with the number of epochs and finally reaches a low and stable value, which reflects that the error of the model gradually decreases and tends

to be optimized. After training, save the weights of the model from the final epoch. Then, input the pictures from the testing set into the model. The confusion matrix and t-SNE visualization after the test are shown in Figure 11. The t-SNE visualization model effectively separates different types of data in the multidimensional space, indicating that VSDP-TLResNet18 based on the pre-trained and fine-tuned method achieved satisfactory results on the CWRU dataset.

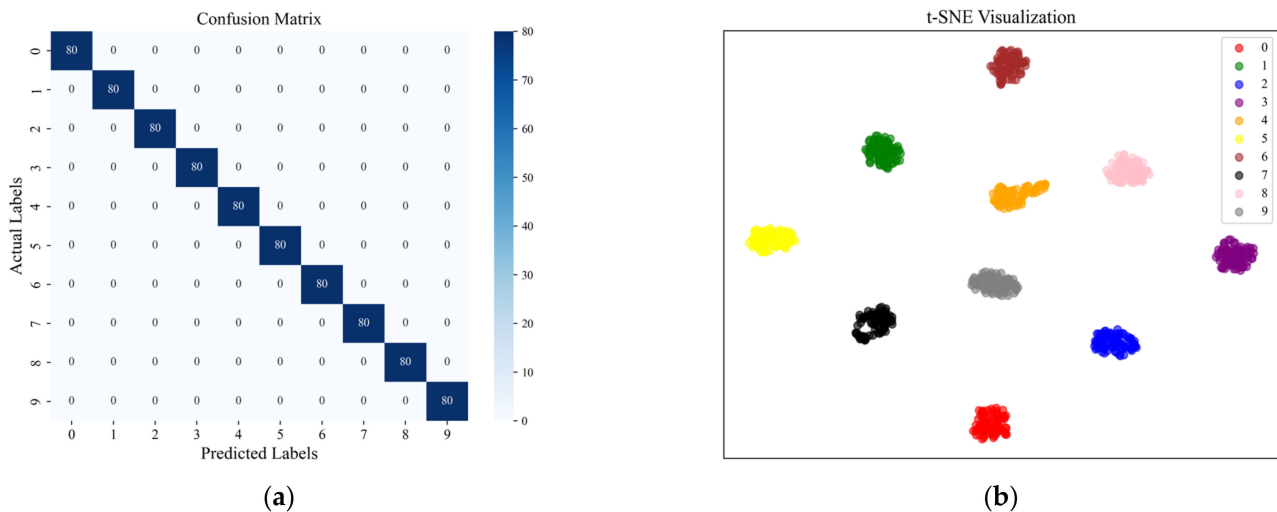


Figure 11. Validation results under Case 1: (a) confusion matrix; (b) t-SNE visualization.

To validate the impact of the training–testing data ratio on model performance, we conducted four sets of experiments using different proportions of training and testing data: A (4:6), B (3:7), C (2:8), D (1:9). The number of samples in the dataset is 100. The experimental results are shown in Table 4. As the proportion of the training set and testing set decreases, the training loss increases. It increased from 0.0024 to 0.0591; although it increased, it still remained at a low level. Figure 12 shows the experimental results of dataset D (1:9). When the experimental ratio is 1:9, the test accuracy slightly decreases to 99.67%, but it is still very close to 100%, showing good generalization ability. This shows that VSDP-TLResNet18 has strong robustness and generalization ability and can maintain high test accuracy, even with less training data.

Table 4. Experimental results of different proportions under Case 1.

	Train	Test	Train Accuracy	Train Loss	Test Accuracy
A	40	60	100%	0.0024	100%
B	30	70	100%	0.0048	100%
C	20	80	100%	0.0044	100%
D	10	90	100%	0.0591	99.67%

4.1.2. Comparison of Different Methods for Transforming Two-Dimensional Images

To demonstrate the effectiveness of VSDP in transforming two-dimensional images, we used the same TLResNet18 model to compare it with CWT, gray-scale image, WVD and SDP. The experimental results are presented in Table 5 below. The accuracy of SDP without VMD decomposition on small-sample datasets is only 83.63%, indicating that SDP transformation performs poorly in small samples. The accuracy rates of CWT and WVD are 98.38% and 97.38%, respectively, while the accuracy rate of the gray-scale image is only 92.00%, indicating that the information contained in the time–frequency image is more abundant than that in the gray-scale image. The accuracy achieved by the VSDP method is 100%, surpassing that of other methods. Recall, precision and F1 are also higher than other methods. The evaluation index reaches 100%, which shows that compared with the other

four methods of one-dimensional signal to two-dimensional image, our VSDP method can accurately identify each type of fault under the same small sample and the same network model. It also shows that the IMF component chosen after the signal is decomposed by VMD can effectively capture and utilize the key information within the data and display the key information through SDP transformation.

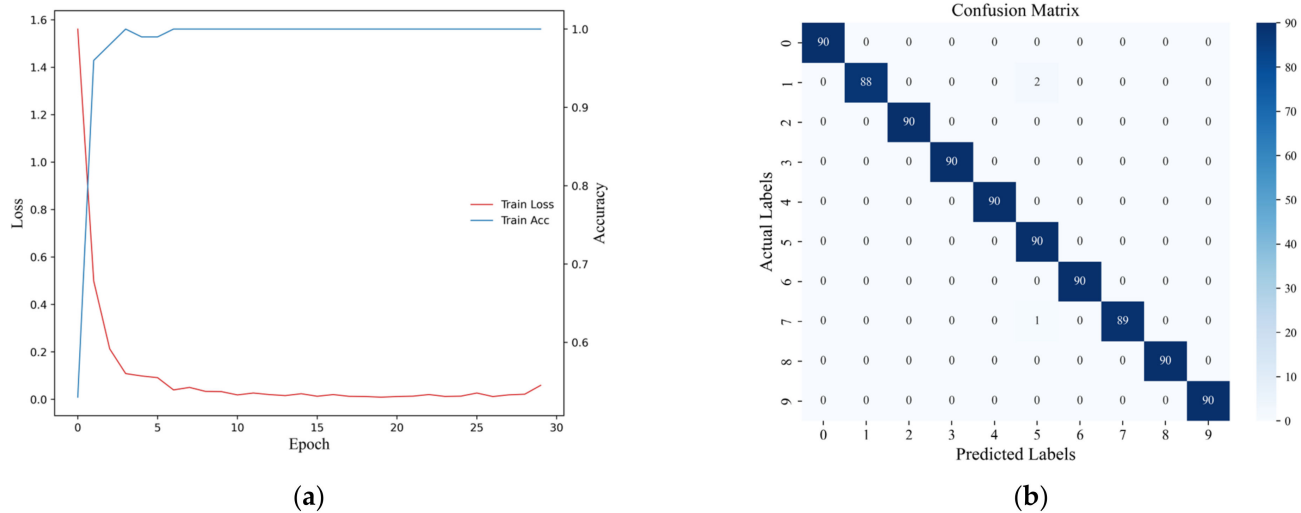


Figure 12. Experimental results of dataset D under Case 1: (a) training accuracy and training loss; (b) confusion matrix.

Table 5. Test results of different two-dimensional image methods under Case 1.

Method	Accuracy	Recall	Precision	F1
VSDP (ours)	100.00%	100.00%	100.00%	100.00%
SDP	83.63%	83.63%	82.01%	81.67%
CWT	98.38%	98.38%	98.54%	98.40%
WVD	97.38%	97.38%	97.56%	97.36%
Gray-scale image	92.00%	92.00%	92.49%	91.85%

4.1.3. Comparison with Other Networks

To demonstrate the advancement and generalization capabilities of the VSDP-TLResNet18 model for small-sample FD, utilizing the pre-trained and fine-tuned method, four datasets (A, B, C and D) were introduced for diagnostic analysis, each containing a different number of training data samples. Among them, the combined number of training samples in datasets A, B, C and D is 400, 200, 100 and 60, respectively, and each dataset contains 800 test data samples. The number of training and testing samples is uniform across all fault categories in each dataset. Under the same experimental conditions in the training set, we compared the model with the TLVgg16, TLEfficientNet and TLSqueezeNet models, popular in deep learning. The above three networks all adopted the pre-trained and fine-tuned method. We also compared it with the re-trained ResNet18.

Figure 13 presents the experimental findings. The pre-trained and fine-tuned method greatly enhances the performance of the basic model, especially for the VSDP-TLResNet18 model. This is because, through the pre-trained and fine-tuned method, the model can take advantage of the rich features learned from large-scale datasets, so as to achieve better generalization on new and relatively small-sample datasets. Among all the experimental methods, the average accuracy of VSDP-TLResNet18 based on the pre-trained and fine-tuned method is the highest, with an accuracy rate of 99.81%. It reached 100% accuracy on datasets A, B and C, which is superior to other network models. The average accuracy of the re-trained VSDP-TLResNet18 is 97.96%, which shows that the pre-trained and fine-tuned method has a strong generalization ability compared with the model re-training. The

average accuracy rates of VSDP-TLVgg16, VSDP-TLEfficientNet and VSDP-TLSqueezeNet, which also adopt the pre-trained and fine-tuned method, were 99.15%, 98.97% and 98.56%, respectively. This also proves that for the residual connection in the new task with less data, the deep network's training will not face obstacles due to gradient issues, so it can effectively migrate and use the complex features learned by the pre-trained model.

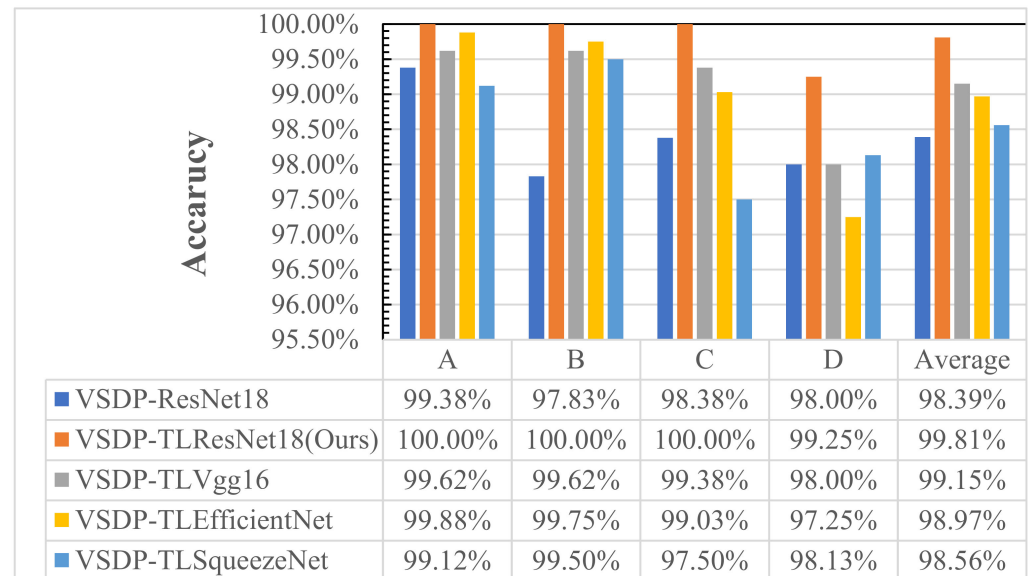


Figure 13. Verification accuracy of different datasets under Case 1.

Figure 14 displays the confusion matrix outcomes for the accuracy of five distinct models tested on validation sets within dataset D under Case 1. Figure 14 illustrates that when limited to only six training samples for each type of fault, the VSDP-TLEfficientNet model exhibits the lowest accuracy at 97.25%. VSDP-TLEfficientNet and VSDP-TLVgg16 have a little more misclassification in categories 1, 2, 3, 4 and 7, because VSDP-TLVgg16 and VSDP-TLSqueezeNet have no residual connection, so it is easy to overfit when processing small-sample data. The VSDP-TLEfficientNet model may be too complex to capture effective key information when the training data are insufficient. Compared with VSDP-ResNet18, VSDP-TLResNet18 misjudged six faults and predicted all the remaining categories correctly, showing a strong feature extraction ability on small-sample datasets.

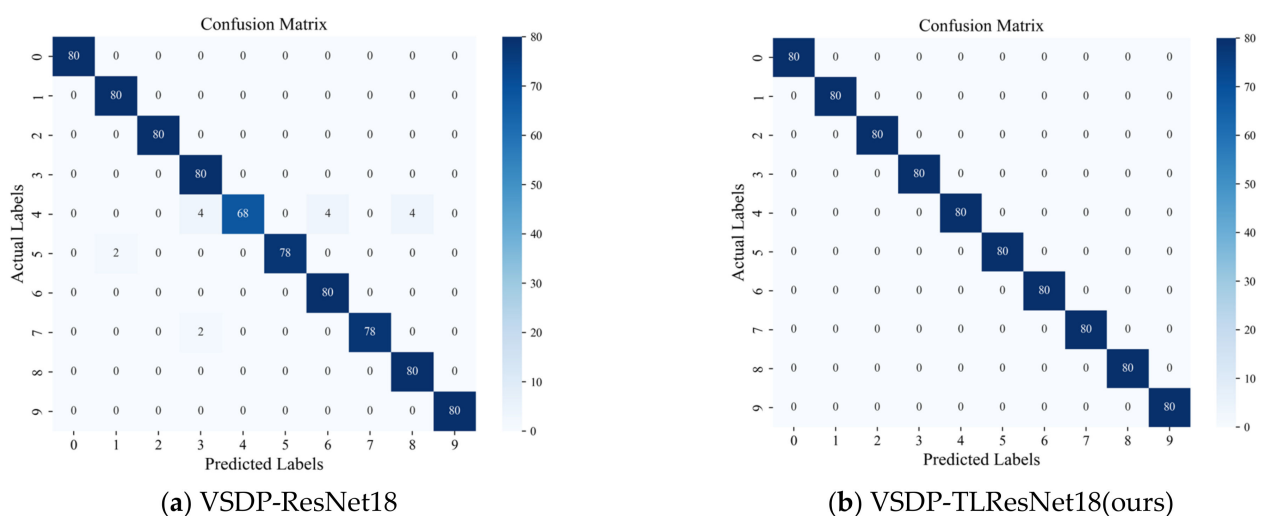
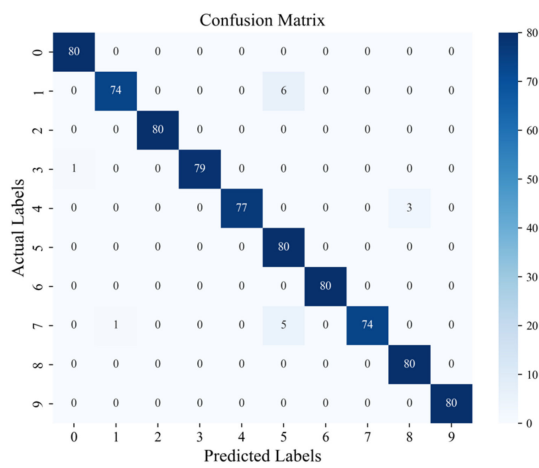
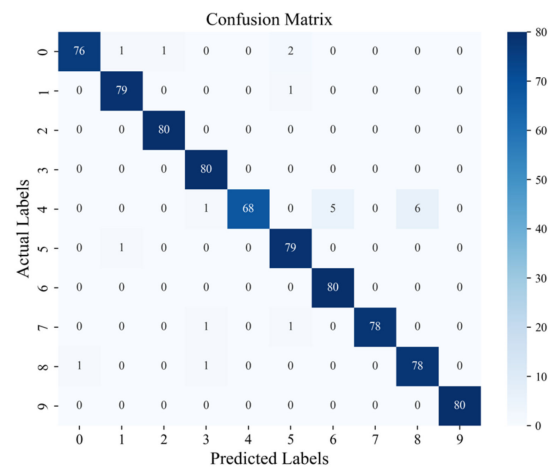


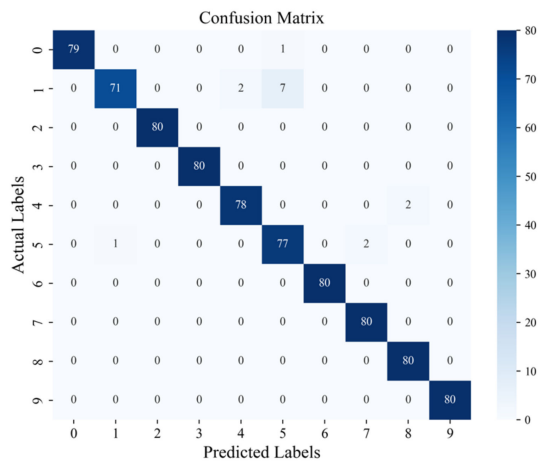
Figure 14. Cont.



(c) VSDP-TLVgg16



(d) VSDP-TLEfficientNet

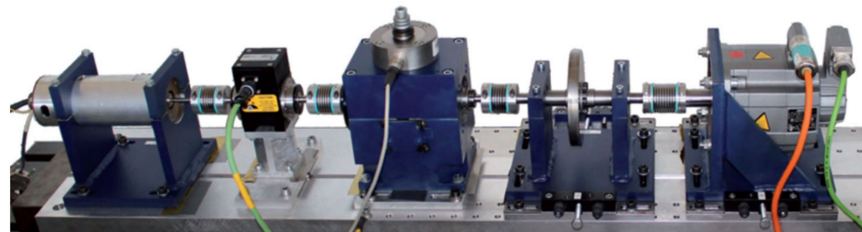


(e) VSDP-TLSqueezeNet

Figure 14. Confusion matrix of different networks on dataset D under Case 1.

4.2. Case 2:PU Bearing Dataset

The PU bearing dataset features the 6203 bearing model, encompassing both artificial and real damage types. Figure 15 displays the test bench setup, which consists of a drive motor, a torque measurement shaft, a testing module and a load motor. A piezoelectric accelerometer is employed to gather the vibration signals from the bearing pedestal, with the sampling frequency set at 64 kHz. By adjusting the speed of the drive system, the radial force on the bearing and the load torque affecting the drive system are tested.

**Figure 15.** PU test bench.

In the PU bearing dataset, we chose the subset recorded under operating conditions of 1500rpm, with a 0.7Nm load and a radial force of 1000N. In the bearing fault dataset, each fault type is represented by 100 generated samples. Using the same overlapping sampling method of 512 points as in Case 1, we slice 1024 data points with a fixed length

and distribute them into training and testing sets in a 2:8 ratio. Table 6 displays the specific fault types and their quantities within the dataset.

Table 6. Detailed composition of dataset under Case 2.

Fault Code	Fault Type	Label	Fault Source	Train	Test
K001	N	0	None	20	80
KA01	OR1	1	Artificial damage	20	80
KA04	OR2	2	Real damage	20	80
KB23	OR2+IR2	3	Real damage	20	80
KI01	IR1	4	Artificial damage	20	80
KI14	IR2	5	Real damage	20	80

The fault signal is broken down into k IMF components through VMD. The value of k is set using the same center frequency method as outlined in Case 1. As shown in Table 7, when k is 9 and k is 8, the frequency overlaps, so k is taken as 8. When k is 8, it provides a wide frequency coverage, and the frequency points are evenly distributed, which more effectively represents the important frequency components of the signal. As shown in Table 8, using the same method as in Case 1, select the best first six IMF components by averaging the kurtosis of each IMF component. Finally, each IMF component is successively generated into a picture through SDP transformation, and the SDP parameter selection is similar to Case 1. The generated fault image is shown in Figure 16.

Table 7. Center frequencies of different k values under Case 2.

k	Center Frequency/HZ								
5	0.0552	0.0908	0.1524	0.2857	0.4299				
6	0.0527	0.0712	0.1508	0.2093	0.3346	0.4447			
7	0.0250	0.0574	0.1508	0.1301	0.2476	0.3417	0.4478		
8	0.0244	0.0567	0.0925	0.1525	0.2118	0.2883	0.3547	0.4543	
9	0.0242	0.0566	0.0922	0.1525	0.2104	0.2794	0.3374	0.4035	0.4646

Table 8. Average kurtosis of each IMF component under Case 2.

	IMF1	IMF2	IMF3	IMF4	IMF5	IMF6	IMF7	IMF8
Kurtosis average	6.9350	6.5255	5.2420	11.1878	7.8981	12.6360	13.5512	12.7305

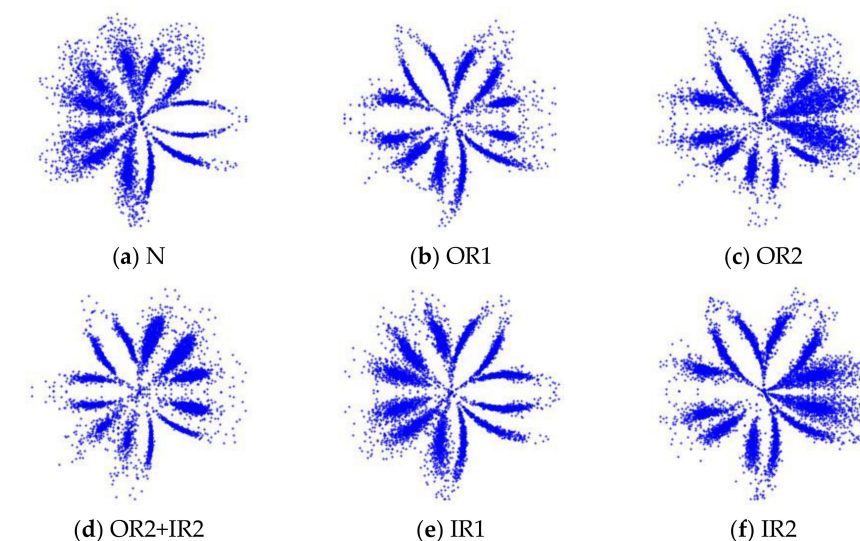


Figure 16. Fault diagram of VSDP under Case 2.

4.2.1. Model Verification

The same hyperparameters as in Case 1 are used to train the model. The training accuracy, confusion matrix and t-SNE visualization are presented in Figures 17 and 18 below. At the beginning of the training, the accuracy of the training set showed a rapid increase and tended to be stable and convergent after 10 epochs. The loss of the training set also declined rapidly in the first 10 epochs and then tended to be stable, indicating that the model effectively minimized the error in the learning process. As can be seen from Figure 18, only category 2 is incorrectly predicted as category 3 once, which indicates that an OR2 fault is incorrectly identified as an OR2+IR2 fault. Because the OR2+IR2 fault is a mixed fault generated by an OR2 fault and IR2 fault, which is very similar to a single OR2 fault, the model incorrectly identifies an OR2 fault as an OR2+IR2 mixed fault. This also shows that the classification accuracy of the model is very high in all categories. The clear clustering distribution further validated the high classification accuracy seen in the confusion matrix, demonstrating that the model can maintain good class separation, even in reduced dimensional space.

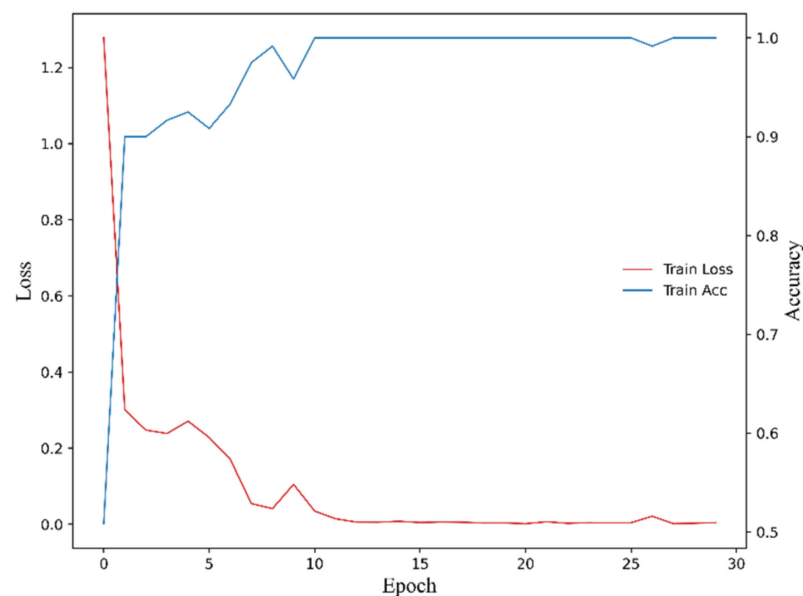


Figure 17. Training accuracy and loss under Case 2.

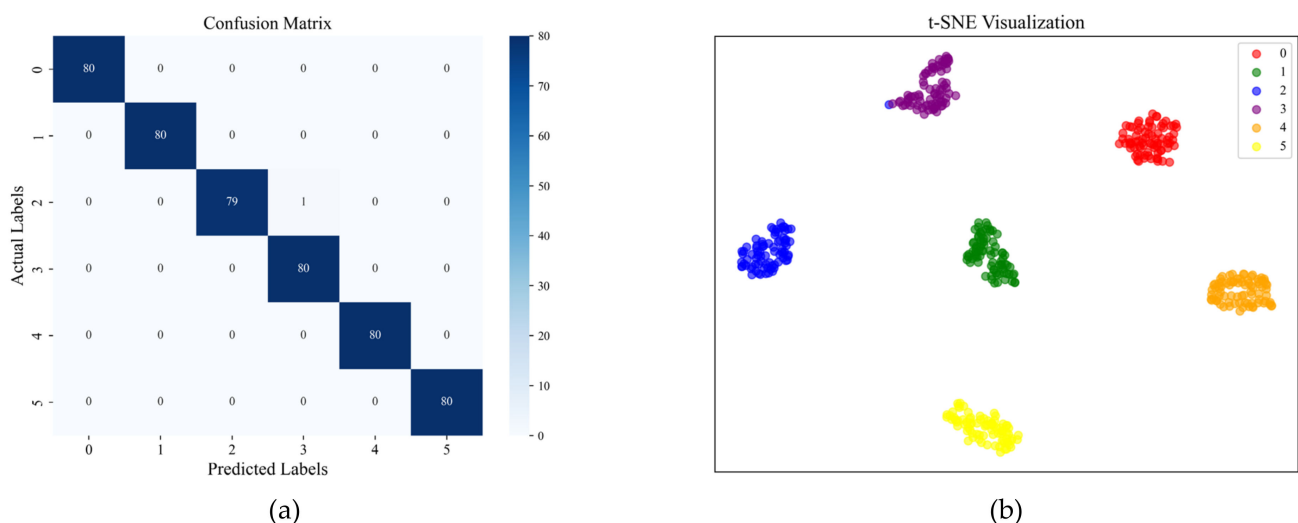


Figure 18. Validation results under Case 2: (a) confusion matrix; (b) the t-SNE visualization.

To validate the impact of the training–testing data ratio on model performance, we conducted four sets of experiments using different proportions of training and testing data: A (4:6), B (3:7), C (2:8), D (1:9). The number of samples in the dataset is 100. All experimental results are shown in Table 9. With a reduction in the experimental proportion, the training loss fluctuated slightly but changed little. The test accuracy was kept at a high level in all proportions, which is close to or higher than 99%. The experimental results of dataset D are shown in Figure 19.

Table 9. Experimental results of different proportions under Case 2.

	Train	Test	Train Accuracy	Train Loss	Test Accuracy
A	40	60	100%	0.0035	99.72%
B	30	70	100%	0.0049	99.52%
C	20	80	100%	0.0027	99.58%
D	10	90	100%	0.0058	99.26%

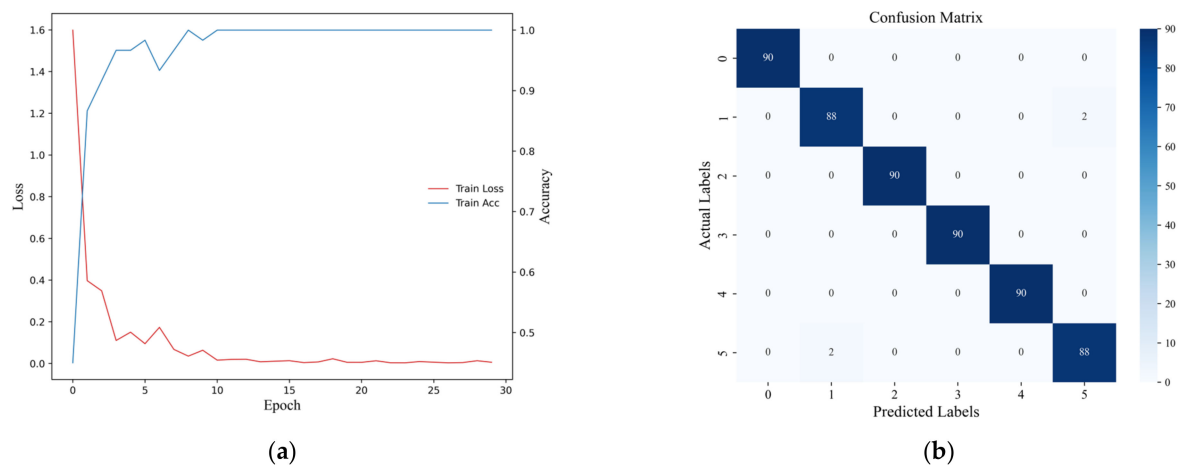


Figure 19. Experimental results of dataset D under Case 2: (a) training accuracy and training loss; (b) confusion matrix.

4.2.2. Comparison of Different Methods for Transforming Two-Dimensional Images

The experiment adopts the same comparison method as Case 1 and uses the same TLResNet18 network to compare with CWT, gray-scale image, WVD and SDP. The experimental result is shown in Table 10. The accuracy rate of SDP transformation is 70%, indicating that the independent use of SDP transformation may have limitations in small-sample feature extraction. The accuracy rates of CWT, WVD and gray-scale image are 98.12%, 94.17% and 90.00%, respectively. However, compared with the accuracy rate of VSDP, these methods are still slightly inadequate, which also shows that the combination of VMD and SDP greatly enhances the ability of feature extraction, making the classifier more accurate in identifying different categories.

Table 10. Test results of different two-dimensional image methods under Case 2.

Method	Accuracy	Recall	Precision	F1
VSDP(ours)	99.79%	99.79%	99.79%	99.79%
SDP	70.00%	70.00%	69.23%	67.17%
CWT	98.12%	98.13%	98.13%	98.12%
WVD	94.17%	94.17%	94.96%	93.94%
Gray-scale image	90.00%	90.00%	90.39%	89.84%

4.2.3. Comparison with Other Networks

Similarly, under the same experimental conditions, it is compared with the classical VSDP-TLVgg16, VSDP-TLEfficientNet, VSDP-TLSqueezeNet and VSDP-ResNet18 models. Four datasets (A, B, C and D) were introduced for diagnostic analysis, each containing a different number of training data samples. Among them, the combined number of training samples in datasets A, B, C and D is 240, 120, 60 and 36, respectively, and each dataset contains 480 test data samples. The number of training and testing samples is uniform across all fault categories in each dataset. As shown in Figure 20, the average accuracy of VSDP-TLSqueezeNet, VSDP-TLVgg16 and VSDP-TLEfficientNet is 94.32%, 97.24% and 97.18%. Although the three networks perform well on dataset A, their performance on dataset D decreases significantly, which may reflect the limitations of VSDP-TLSqueezeNet, VSDP-TLVgg16 and VSDP-TLEfficientNet in processing small-sample data. Compared with the re-trained VSDP-ResNet18, VSDP-TLResNet18 achieves an average accuracy of 99.11%. As the number of training samples decreases, the diagnostic accuracy of the four models also declines, but the diagnostic accuracy of VSDP-TLResNet18 on the four datasets is still relatively high, highlighting the effectiveness of the pre-trained and fine-tuned method in enhancing the model's generalization ability and adapting to various small-sample datasets.

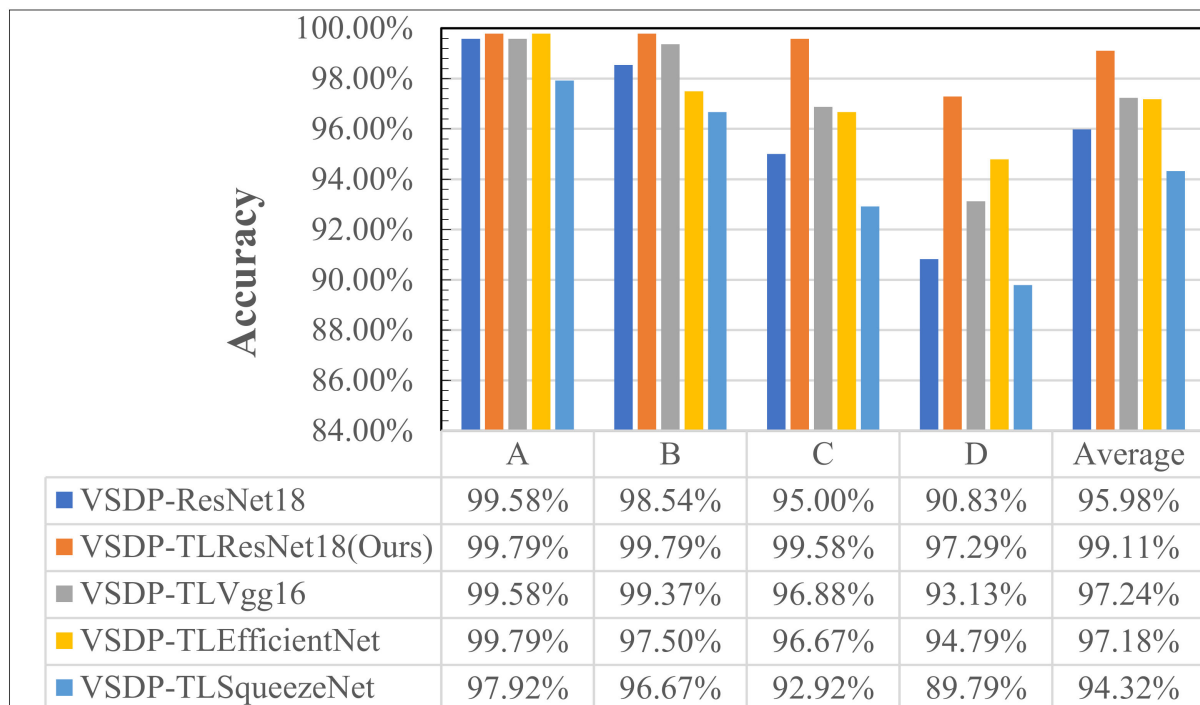


Figure 20. Verification accuracy of different datasets under Case 2.

In order to better understand the generalization of VSDP-TLResNet18, Figure 21 shows the confusion matrix of five networks on dataset D. As can be seen from the figure, compared with VSDP-ResNet18, VSDP-TLResNet18 shows better performance in most categories, is superior to other networks and the number of misclassifications is reduced, especially in categories 1 and 3. This shows that the pre-trained and fine-tuned method is helpful for the model to better adapt and identify the characteristics of each category. VSDP-TLSqueezeNet has a lot of misclassifications in categories 0 and 4; a large proportion of samples in category 0 are misclassified to category 1, which indicates that VSDP-TLSqueezeNet is not enough to capture the characteristics of these fault categories on dataset D. Both VSDP-TLVgg16 and VSDP-TLEfficientNet incorrectly identify category 5, and VSDP-TLResNet18 can better identify category 5.

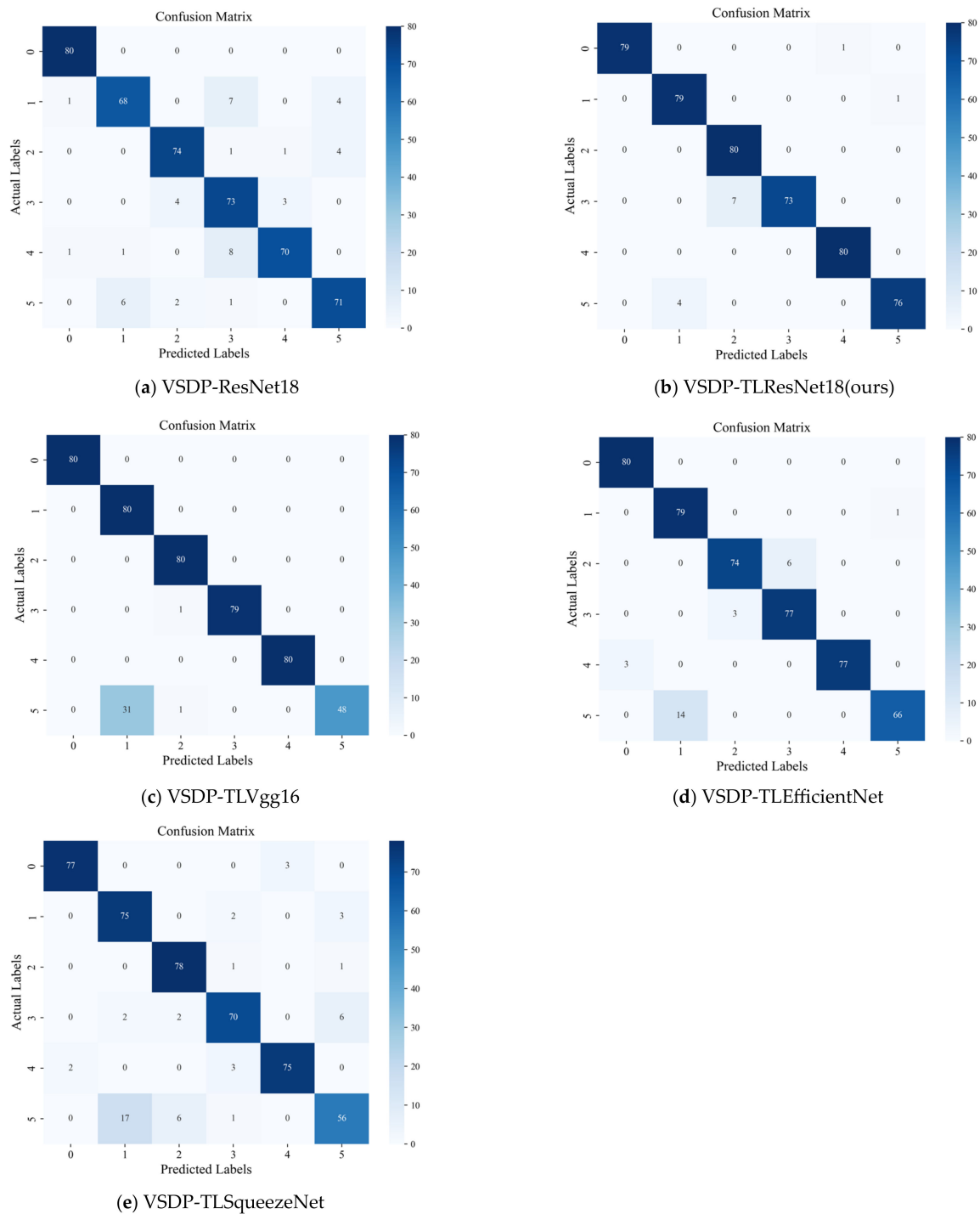


Figure 21. Confusion matrix of different networks on dataset D under Case 2.

5. Conclusions

Aiming at the problem that real bearing fault data are limited, this study proposes a small-sample FD model based on pre-trained and fine-tuned VSDP-TLResNet18. The method decomposes the rolling bearing signal by VMD technology initially and selects the k value and IMF component by using the center frequency and kurtosis criteria. Then, the SDP transform is used to map the one-dimensional signal to the two-dimensional spectrum. In order to maximize the difference in two-dimensional fault images, we used Pearson correlation analysis to select the optimal SDP parameters, ensuring that the chosen parameters can maximally distinguish differences between images. Finally, the small-

sample FD is realized by using the pre-trained and fine-tuned method with the ResNet18 model. The experimental results indicate an average accuracy of 99.81% on the CWRU dataset and 99.11% on the PU dataset, demonstrating that the proposed method exhibits strong robustness and generalization.

Although our model achieves good accuracy on the two datasets, compared with other methods, even in the case of small-scale training data, the improvement in accuracy is still low. This shows that under the existing data and methods, our model has limited room for improvement in practical applications. In addition, we only carried out FD under fixed working conditions. With a change in the actual fault mode and equipment, the model needs to be updated and re-trained regularly. How to effectively manage and update the model was not discussed in depth.

During the actual fault data collection process, some fault data may lack labels due to human or other factors. In a follow-up study, we will explore the methods to solve these problems. In the future, we also plan to explore the combination of traditional signal processing technology and methods in the field of machine vision to play a potential role in small-sample FD. Through these improvements and explorations, we hope to further improve the performance and applicability of the model.

Author Contributions: Conceptualization, J.P. and J.N.; methodology, J.N.; software, J.N. and Z.Q.; validation, J.P., J.N. and F.H.; formal analysis, J.N. and Z.Q.; resources, J.P.; data curation, F.H.; writing—original draft preparation, J.N.; writing—review and editing, J.P. and H.Q.; visualization, F.H. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Wuzhou Central Leading Local Science and Technology Development Fund Project under grant no. 202201001.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The sources of data used in this article are linked in the References section. They are publicly available.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

FD	Fault diagnosis
VMD	Variational-Mode Decomposition
IMF	Intrinsic-Mode Functions
SDP	Symmetric Dot Pattern
VSDP	VMD with SDP
TL	Transfer learning
VSDP-TLResNet18	VSDP with a pre-trained and fine-tuned Residual Network18
Resnet18	Residual Network18
CWT	Continuous wavelet transform
WVD	Wigner-Ville distribution
CWRU	Case Western Reserve University
PU	Paderborn University
t-SNE	t-distributed stochastic neighbor embedding

References

1. Neupane, D.; Seok, J. Bearing Fault Detection and Diagnosis Using Case Western Reserve University Dataset with Deep Learning Approaches: A Review. *IEEE Access* **2020**, *8*, 93155–93178. [\[CrossRef\]](#)
2. Liu, R.; Yang, B.; Zio, E.; Chen, X. Artificial intelligence for fault diagnosis of rotating machinery: A review. *Mech. Syst. Signal Process.* **2018**, *108*, 33–47. [\[CrossRef\]](#)
3. Saufi, S.R.; Ahmad, Z.A.B.; Leong, M.S.; Lim, M.H. Challenges and Opportunities of Deep Learning Models for Machinery Fault Detection and Diagnosis: A Review. *IEEE Access* **2019**, *7*, 122644–122662. [\[CrossRef\]](#)

4. Gangsar, P.; Tiwari, R. Signal based condition monitoring techniques for fault detection and diagnosis of induction motors: A state-of-the-art review. *Mech. Syst. Signal Process.* **2020**, *144*, 106908. [\[CrossRef\]](#)
5. Jia, N.; Cheng, Y.; Liu, Y.; Tian, Y. Intelligent Fault Diagnosis of Rotating Machines Based on Wavelet Time-Frequency Diagram and Optimized Stacked Denoising Auto-Encoder. *IEEE Sens. J.* **2022**, *22*, 17139–17150. [\[CrossRef\]](#)
6. He, C.; Shi, H.; Li, J. IDSNet: A one-stage interpretable and differentiable STFT domain adaptation network for traction motor of high-speed trains cross-machine diagnosis. *Mech. Syst. Signal Process.* **2023**, *205*, 110846. [\[CrossRef\]](#)
7. Zhang, B.; Pang, X.; Zhao, P.; Lu, K. A New Method Based on Encoding Data Probability Density and Convolutional Neural Network for Rotating Machinery Fault Diagnosis. *IEEE Access* **2023**, *11*, 26099–26113. [\[CrossRef\]](#)
8. Cai, J.H.; Chen, Q.Y. Bearing Fault Diagnosis Method Based on Local Mean Decomposition and Wigner Higher Moment Spectrum. *Instrum. Exp. Tech.* **2016**, *40*, 1437–1446. [\[CrossRef\]](#)
9. Li, Y.; Cheng, G.; Liu, C.; Chen, X. Study on planetary gear fault diagnosis based on variational mode decomposition and deep neural networks. *Measurement* **2018**, *130*, 94–104. [\[CrossRef\]](#)
10. Shibata, K.; Takahashi, A.; Shirai, T. Fault Diagnosis of Rotating Machinery through Visualisation of Sound Signals. *Mech. Syst. Signal Process.* **2000**, *14*, 229–241. [\[CrossRef\]](#)
11. Wang, H.; Xu, J.; Yan, R.; Gao, R.X. A New Intelligent Bearing Fault Diagnosis Method Using SDP Representation and SE-CNN. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 2377–2389. [\[CrossRef\]](#)
12. Tang, Y.; Zhang, X.; Qin, G.; Long, Z.; Huang, S.; Song, D.; Shao, H. Graph Cardinality Preserved Attention Network for Fault Diagnosis of Induction Motor Under Varying Speed and Load Condition. *IEEE Trans. Ind. Inform.* **2022**, *18*, 3702–3712. [\[CrossRef\]](#)
13. Sun, Y.; Li, S.; Wang, Y.; Wang, X. Fault diagnosis of rolling bearing based on empirical mode decomposition and improved manhattan distance in symmetrized dot pattern image. *Mech. Syst. Signal Process.* **2021**, *159*, 107817. [\[CrossRef\]](#)
14. Zhang, X.; Miao, Q.; Zhang, H.; Wang, L. A parameter-adaptive VMD method based on grasshopper optimization algorithm to analyze vibration signals from rotating machinery. *Mech. Syst. Signal Process.* **2018**, *108*, 58–72. [\[CrossRef\]](#)
15. Li, F.; Li, R.; Tian, L.; Chen, L.; Liu, J. Data-driven time-frequency analysis method based on variational mode decomposition and its application to gear fault diagnosis in variable working conditions. *Mech. Syst. Signal Process.* **2019**, *116*, 462–479. [\[CrossRef\]](#)
16. Liu, T.; Luo, Z.; Huang, J.; Yan, S. A Comparative Study of Four Kinds of Adaptive Decomposition Algorithms and Their Applications. *Sensors* **2018**, *18*, 2120. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Pandiyan, M.; Babu, T.N. Systematic Review on Fault Diagnosis on Rolling-Element Bearing. *J. Vib. Eng. Technol.* **2024**. [\[CrossRef\]](#)
18. Zhang, T.; Chen, J.; Li, F.; Zhang, K.; Lv, H.; He, S.; Xu, E. Intelligent fault diagnosis of machines with small & imbalanced data: A state-of-the-art review and possible extensions. *ISA Trans.* **2022**, *119*, 152–171. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Ruan, D.; Chen, X.; Gühmann, C.; Yan, J. Improvement of Generative Adversarial Network and Its Application in Bearing Fault Diagnosis: A Review. *Lubricants* **2023**, *11*, 74. [\[CrossRef\]](#)
20. Huang, N.; Chen, Q.; Cai, G.; Xu, D.; Zhang, L.; Zhao, W. Fault Diagnosis of Bearing in Wind Turbine Gearbox Under Actual Operating Conditions Driven by Limited Data with Noise Labels. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 3502510. [\[CrossRef\]](#)
21. Liang, P.; Deng, C.; Wu, J.; Yang, Z. Intelligent fault diagnosis of rotating machinery via wavelet transform, generative adversarial nets and convolutional neural network. *Measurement* **2020**, *159*, 107768. [\[CrossRef\]](#)
22. Zhou, F.; Yang, S.; Fujita, H.; Chen, D.; Wen, C. Deep learning fault diagnosis method based on global optimization GAN for unbalanced data. *Knowl. Based Syst.* **2020**, *187*, 104837. [\[CrossRef\]](#)
23. Chen, X.; Yang, R.; Xue, Y.; Huang, M.; Ferrero, R.; Wang, Z. Deep Transfer Learning for Bearing Fault Diagnosis: A Systematic Review Since 2016. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 3508221. [\[CrossRef\]](#)
24. Li, W.; Huang, R.; Li, J.; Liao, Y.; Chen, Z.; He, G.; Yan, R.; Gryllias, K. A perspective survey on deep transfer learning for fault diagnosis in industrial scenarios: Theories, applications and challenges. *Mech. Syst. Signal Process.* **2022**, *167*, 108487. [\[CrossRef\]](#)
25. Shao, S.; McAleer, S.; Yan, R.; Baldi, P. Highly Accurate Machine Fault Diagnosis Using Deep Transfer Learning. *IEEE Trans. Ind. Inform.* **2019**, *15*, 2446–2455. [\[CrossRef\]](#)
26. Pang, B.; Liu, Q.; Sun, Z.; Xu, Z.; Hao, Z. Time-frequency supervised contrastive learning via pseudo-labeling: An unsupervised domain adaptation network for rolling bearing fault diagnosis under time-varying speeds. *Adv. Eng. Inf.* **2024**, *59*, 102304. [\[CrossRef\]](#)
27. Song, B.; Liu, Y.; Fang, J.; Liu, W.; Zhong, M.; Liu, X. An optimized CNN-BiLSTM network for bearing fault diagnosis under multiple working conditions with limited training samples. *Neurocomputing* **2024**, *574*, 127284. [\[CrossRef\]](#)
28. Xie, J.; Tian, L.; Lin, M.; Yang, B.; Yang, J.; Wang, T. A small sample diagnosis method driven by simulation and test data: Applied to axle box bearings of high-speed train. *Meas. Sci. Technol.* **2023**, *34*, 125044. [\[CrossRef\]](#)
29. 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*, 327–348. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Hu, J.; Li, W.; Wu, A.; Tian, Z. Novel joint transfer fine-grained metric network for cross-domain few-shot fault diagnosis. *Knowl.-Based Syst.* **2023**, *279*, 110958. [\[CrossRef\]](#)
31. Zhang, Q.; He, Q.; Qin, J.; Duan, J. Application of Fault Diagnosis Method Combining Finite Element Method and Transfer Learning for Insufficient Turbine Rotor Fault Samples. *Entropy* **2023**, *25*, 414. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Yuan, Z.; Ma, Z.; Li, X.; Liu, S.; Mu, T.; Chen, Y. A Graph Attention Based Multichannel Transfer Learning Network for Wheelset Bearing Fault Diagnosis with Nonshared Fault Classes. *IEEE Sens. J.* **2024**, *24*, 1929–1940. [\[CrossRef\]](#)

33. Liu, H.; Xu, Q.; Han, X.; Wang, B.; Yi, X. Attention on the key modes: Machinery fault diagnosis transformers through variational mode decomposition. *Knowl.-Based Syst.* **2024**, *289*, 111479. [CrossRef]
34. Dragomiretskiy, K.; Zosso, D. Variational Mode Decomposition. *IEEE Trans. Signal Process.* **2014**, *62*, 531–544. [CrossRef]
35. Li, H.; Wang, W.; Huang, P.; Li, Q. Fault diagnosis of rolling bearing using symmetrized dot pattern and density-based clustering. *Measurement* **2020**, *152*, 107293. [CrossRef]
36. Gu, Y.; Zeng, L.; Qiu, G. Bearing fault diagnosis with varying conditions using angular domain resampling technology, SDP and DCNN. *Measurement* **2020**, *156*, 107616. [CrossRef]
37. He, K.; Zhang, X.; Ren, S.; Sun, J.; Recognition, P. Deep Residual Learning for Image Recognition. *arXiv* **2015**, 770–778. [CrossRef]
38. Rezaeian, N.; Gurina, R.; Saltykova, O.A.; Hezla, L.; Nohurov, M.; Reza Kashyzadeh, K. Novel GA-Based DNN Architecture for Identifying the Failure Mode with High Accuracy and Analyzing Its Effects on the System. *Appl. Sci.* **2024**, *14*, 3354. [CrossRef]
39. Qin, G.; Zhang, K.; Lai, X.; Zheng, Q.; Ding, G.; Zhao, M.; Zhang, Y. An Adaptive Symmetric Loss in Dynamic Wide-Kernel ResNet for Rotating Machinery Fault Diagnosis Under Noisy Labels. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 3517512. [CrossRef]
40. Yarotsky, D. Error bounds for approximations with deep ReLU networks. *Neural Netw.* **2017**, *94*, 103–114. [CrossRef]
41. Misbah, I.; Lee, C.K.M.; Keung, K.L. Fault diagnosis in rotating machines based on transfer learning: Literature review. *Knowl.-Based Syst.* **2024**, *283*, 111158. [CrossRef]
42. Case Western Reserve University. Available online: <https://engineering.case.edu/bearingdatacenter/welcome> (accessed on 10 May 2024).
43. Paderborn University. Available online: <http://groups.uni-paderborn.de/kat/BearingDataCenter> (accessed on 10 May 2024).
44. Loshchilov, I.; Hutter, F. Decoupled weight decay regularization. *arXiv* **2017**. [CrossRef]
45. Guan, Y.; Meng, Z.; Li, J.; Cao, W.; Sun, D.; Liu, J.; Fan, F. A novel diagnostic framework based on vibration image encoding and multi-scale neural network. *Expert Syst. Appl.* **2024**, *251*, 124054. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.