entropy

MDPI

Article

A Novel Bearing Fault Diagnosis Method Based on Few-Shot Transfer Learning across Different Datasets

Yizong Zhang [1], Shaobo Li [1,2,*], Ansi Zhang [1,2], Chuanjiang Li [1] and Ling Qiu [3]

[1] School of Mechanical Engineering, Guizhou University, Guiyang 550025, China
[2] State Key Laboratory of Public Big Data, Guizhou University, Guiyang 550025, China
[3] School of Computer Science and Technology, Guizhou University, Guiyang 550025, China
* Correspondence: lishaobo@gzu.edu.cn

Abstract: At present, the success of most intelligent fault diagnosis methods is heavily dependent on large datasets of artificial simulation faults (ASF), which have not been widely used in practice because it is often costly to obtain a large number of samples in reality. Fortunately, various faults can be easily simulated in the laboratory, and these simulated faults contain a lot of fault diagnosis knowledge. In this study, based on a Siamese network framework, we propose a bearing fault diagnosis based on few-shot transfer learning across different datasets (cross-machine), using the knowledge of ASF to diagnose bearings with natural faults (NF). First of all, the model obtains a good feature encoder in the source domain, then defines a fault support set for comparison, and finally adjusts the support set with a very small number of target domain samples to improve the fault diagnosis performance of the model. We carried out experimental verification from many aspects on the ASF and NF datasets provided by Case Western Reserve University (CWRU) and Paderborn University (PU). The results show that the proposed method can fully learn diagnostic knowledge in different ASF datasets and sample numbers, and effectively use this knowledge to accurately identify the health state of the NF bearing, which has strong generalization and robustness. Our method does not need second training, which may be more convenient in some practical applications. Finally, we also discuss the possible limitations of this method.

Keywords: fault diagnosis; few-shot; transfer learning; across different datasets

check for updates

Citation: Zhang, Y.; Li, S.; Zhang, A.; Li, C.; Qiu, L. A Novel Bearing Fault Diagnosis Method Based on Few-Shot Transfer Learning across Different Datasets. *Entropy* **2022**, *24*, 1295. https://doi.org/10.3390/e24091295

Academic Editor: Yongbo Li

Received: 24 August 2022
Accepted: 11 September 2022
Published: 14 September 2022

1. Introduction

Bearings are indispensable parts of much important machinery and equipment, which may lead to serious economic losses and casualties in the event of failure [1]. Therefore, it is essential to obtain the state of the bearing quickly and accurately. In recent years, machine learning has been applied to intelligent fault diagnosis of bearings because of its powerful ability. Nowadays, many well-known machine learning methods, such as support vector machine (SVM) [2], deep Boltzmann machine (DBM) [3], convolution neural network (CNN) [4], generate adversarial network (GAN) [5], and so on, have achieved excellent results. The success of most studies, however, are heavily dependent on a large number of artificial simulated faults (ASF) data, which has the following two conditions: (1) there is a large amount of marked data with fault information; and (2) the training data and testing data come from the same probability distribution. However, for a variety of reasons [6], it is impractical to obtain a large number of actual fault data in the real world, which cannot meet the first condition; the second condition cannot be satisfied because of the great difference in the feature distribution between the ASF and natural faults (NF). Therefore, many research results are not applicable to the working environment of real machines and cannot be widely applied in industrial production.

Recently, some researchers have tried to expand the amount of data by means of data over-sampling [7,8] and data generation [9] to solve the dilemma of limited fault data. However, these methods focus on the size of fault data and the data quality cannot be guaranteed, so their contribution to improving the current intelligent fault diagnosis performance is limited. For instance, due to the marginalization of distribution, there is a strong linear relationship between the generated data samples and training samples [10]. Subsequently, some scholars focused on few-shot learning (such as Matching networks [11], Siamese networks [12] and Relation networks [13]) and transfer learning. Few-shot learning is expected to have the same ability as human beings in the process of recognizing new species, acquiring knowledge based on a few instances and guiding new tasks [14]. The diagnostic knowledge that transfer learning can learn in the source domain is used in the target domain [15]. Few-shot learning and transfer learning are considered to be the most promising fault diagnosis technologies for practical applications, and they have been the main research hotspots in recent years [16].

In the aspect of few-shot learning, Ren et al. [17] proposed a capsule automatic encoder model based on automatic encoder and capsule network. Experiments show that the model has the ability to extract a variety of important features from a small number of samples, and can identify fault categories quickly and accurately. Zhang et al. [18] proposed a Siamese network model with a first-layer wide kernel convolution network. Through experimental verification under the conditions of limited data, new fault categories and noise, good fault identification accuracy is achieved. Li et al. [19] used sparse automatic encoders based on deep non-negative constraints to perform diagnosis under the condition of a small amount of fault data, and achieved certain results, but the classification accuracy is significantly reduced in the case of very few samples. Feng et al. [20] proposed a semi-supervised attention-attracting meta-learning network, which uses unlabeled data to refine the model and accurately identify faults. Li et al. [21] proposed a new model-agnostic meta-learning method for fault diagnosis under complex working conditions, and acquired knowledge through the diagnosis task of known working conditions to quickly diagnose bearing faults under unknown operating conditions. Yu et al. [22] integrated the few-shot learning strategy into a multi-label convolutional neural network for bearing fault diagnosis, and completed the diagnosis task with limited samples. In addition, many scholars [23–26] have proposed different few-shot learning methods to achieve fault diagnosis, but most of these methods rely on appropriate laboratory artificial fault data.

In the aspect of transfer learning, Wang et al. [27] developed a novel transfer learning method based on a Siamese network, and used the label trimming method to improve the classification performance of the model under different working conditions and cross bearing positions. He et al. [28] designed a new type of deep multi-wavelet automatic encoder to extract the knowledge of the source domain which is similar to the feature distribution of the target domain for fault diagnosis in the new domain. Liu et al. [29] proposed a new adversarial network focusing on the performance of bearing fault diagnosis under different working conditions, and enhanced the domain adaptability through the conditional countermeasure mechanism to improve the diagnosis effect. In addition, some scholars have tried to transfer from ASF to NF. For example, Wu et al. [30] used six different fine-tuning-based methods and a meta-learning model to carry out artificial-natural fault transfer experiments. The results are compared, and it is concluded that meta-learning is better in the relatively simple finite sample transfer task. Wang et al. [6] proposed an artificial-natural fault transfer learning method based on the reinforcement relation network (RRN), and improved the classification performance of the network by label smoothing and AdaBound algorithm. The data used in the above two articles, however, are from the same machine. As a result, some researchers have tried to transfer diagnostic knowledge from one machine to another. Yang et al. [31] proposed a feature-based transfer neural network to reduce the distribution differences and inter-class distances of the learned transferable features through multi-layer domain adaptation and regularization conditions for pseudo-label learning, utilizing the diagnostic knowledge of laboratory machines to identify the health status of real-case machines. Liang et al. [32] proposed a depth domain adaptive transfer learning network and carried out experiments on the fault data of different machines, it is proved that it is effective to collect tagged fault data from one machine for training and to test another machine.

To sum up, although few-shot learning has achieved certain achievements in fault diagnosis with limited samples, these achievements are based on the standard dataset of ASF, and there is a great difference between real fault and simulated fault, so it cannot be directly applied to real industrial machines. In the current research on transfer learning, whether between different working conditions or between ASF and NF, the source domain and target domain of most experiments come from the same dataset (the source and target domain data is collected on the same test bench or machine) and follow the same feature distribution. This ignores a major problem: if you want to apply it to a real machine, you need to obtain a large amount of appropriate source domain data in the same real machine, which is not feasible. Therefore, some researchers try to obtain a large amount of fault data from a machine that is convenient for data collection, and extract diagnostic knowledge from it to identify the health status of another machine, and we believe that this is a feasible method to solve the problem that a large number of fault data cannot be obtained in real industrial production. The reason for this is that it is relatively easy to obtain ASF data in the laboratory, which includes the diagnostic knowledge of real machine bearings.

In this paper, we propose a bearing fault diagnosis method based on few-shot transfer learning across different datasets (cross-machine) inspired by the fine-tuning-based method. Our model is based on the framework of a Siamese network and has the ability of few-shot learning. First of all, the model is trained with the ASF data, and the available diagnosis knowledge is learned. Then, a fault support set for comparison is defined and it is assumed that a very small number of NF samples can be obtained. These few NF samples are input directly into the support set or replace the original samples to improve the generalization ability of the model. Finally, the knowledge of ASF is used to effectively identify the health state of the new machine bearings. The main innovations and contributions of this paper are as follows:

(1) In view of the problems that most of the current intelligent fault diagnosis methods cannot be directly applied to industry, a few-shot transfer learning method across different datasets is proposed, which can use the diagnostic knowledge learned from ASF data to effectively identify the health state of the new machine bearings.

(2) For the first time, a very small number of target domain samples are used to replace the original samples of the support set in fault diagnosis, which improves the generalization ability of the model, and has very high stability and accuracy even in different datasets (ASF-NF) with great differences in feature space distribution.

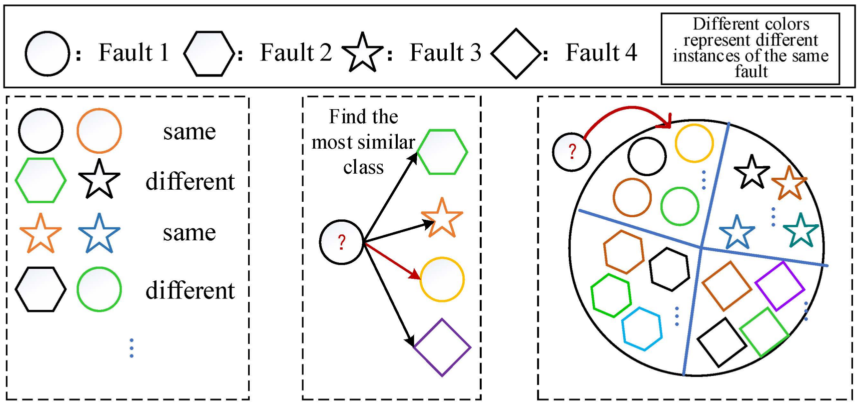
(3) Several experiments are designed to compare and verify many aspects of the proposed method, which has achieved the expected results, and our method does not need secondary training, which will be more convenient.

The structure of this paper is as follows: Section 2 introduces the theoretical background of the method proposed in this paper. Section 3 introduces the proposed method and our model. Section 4 carries on the experiment and analysis from different aspects. Section 5 gives the main conclusions.

2. Basic Theory

2.1. Few-Shot Learning Strategy

When human beings recognize a new thing, they may only need to learn knowledge from a few instances to be able to accurately identify such things. Few-shot learning is proposed in order to acquire this human skill. The general strategy of few-shot learning based on a Siamese network is shown in Figure 1. Different from the general deep learning strategy, the input during training is a pair of the same or different samples (x_1, x_2), one only needs to label the sample pairs (x_1, x_2) with the same or different class. The output is the probability of similarity between sample pairs (x_1, x_2). When testing, there are mainly two strategies: one-shot k-way and N-shot k-way. One-shot k-way refers to the *k* categories in the support set, each class has only one instance; and N-shot k-way means that there are k categories in the support set, and each class has N instances.


Fault 1
Fault 2
Fault 3
Fault 4
Different colors represent different instances of the same fault
same
different
same
different
Find the most similar class
?
?

(a) Training (b) One-shot testing (c) N-shot testing

Figure 1. General strategies of few-shot learning.

In the one-shot k-way test, a test sample $\hat{x}$ that need to be classified and a support set are given, the support set is defined as shown in Equation (1). Next, the model judges the similarity between samples $(x_1, x_2, x_3, \ldots, x_k)$ in the support set and the test sample $\hat{x}$, and selects the highest similarity as the same class of $\hat{x}$, as shown in Equation (2).

$$S = \{(x_1, y_1), \ldots, (x_k, y_k)\} \tag{1}$$

The y is the label of the class, k represents the kth fault class.

$$C(\hat{x}, S) = \text{argmax}(P(\hat{x}, x_c)), x_c \in S, \tag{2}$$

The P is the probability of similarity, C is the fault class most similar to the test sample $\hat{x}$.

In the N-shot k-way test, there are k classes in the support set, each class has N different instances, such as shown in Equation (3), and the support set is shown in Equation (4).

$$\begin{cases} H_1 = \left\{ (x_1^1, y_1^1), (x_1^2, y_1^2) \ldots, (x_1^N, y_1^N) \right\} \\ H_2 = \left\{ (x_2^1, y_2^1), (x_2^2, y_2^2) \ldots, (x_2^N, y_2^N) \right\} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ H_k = \left\{ (x_k^1, y_k^1), (x_k^2, y_k^2) \ldots, (x_k^N, y_k^N) \right\} \end{cases} \tag{3}$$

The H is a set containing multiple instances of the same class, k represents the kth fault class, N represents the Nth instance in the same fault class.

$$S_k = \{H_1, H_2 \ldots, H_k\} \tag{4}$$

The model will judge the similarity between the k*N instances of the support set and the test samples $\hat{x}$, and select the highest similarity as the same class of $\hat{x}$, as shown in Equation (5).

$$C(\hat{x}, S_k) = \text{argmax}(P(\hat{x}, x_c)), x_c \in S_k \tag{5}$$

Here P and C are the same as Equation (2), but the difference lies in the difference between S and S_k.

2.2. Fine-Tuning-Based Method

The main goal of fault diagnosis based on transfer learning is to transfer the learned knowledge from the source domain to the target domain. Among the many current transfer learning strategies, the fine-tuning-based method has been widely studied and proved to be effective. We are inspired by the fine-tuning-based method and put forward our method strategy.

The learning process of fine-tuning-based method is divided into two stages. First, the network model learns the knowledge of diagnosis in the source domain; then, fine-tuning the full connection layer in the target domain to obtain a new classifier as shown in Figure 2.

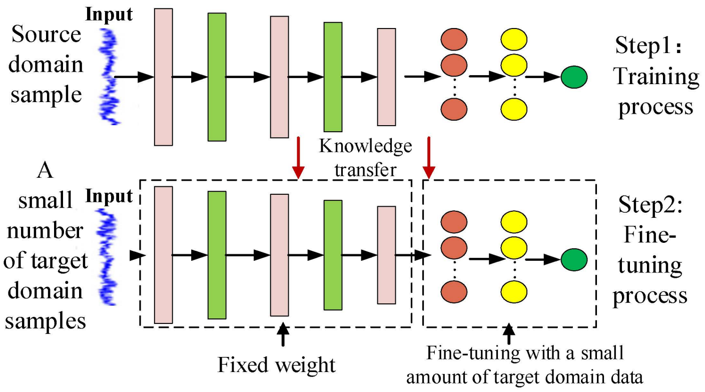

Input
Source domain sample
Step1: Training process
Knowledge transfer
A small number of target domain samples
Input
Step2: Fine-tuning process
Fixed weight
Fine-tuning with a small amount of target domain data

Figure 2. Transfer learning based on fine-tuning.

3. The Proposed Method

3.1. The Proposed Few-Shot Transfer Learning Methods

We are inspired by the fine-tuning-based method and put forward our method strategy. From Section 2, we can see that the support set plays an important reference role in the Siamese network. The test sample x examples are always compared with the samples in the support set, and the most similar examples in the support set are selected for classification, as shown in Equations (2) and (5). In the few-shot transfer learning based on a Siamese network, we assume that a small amount of target domain data has been obtained and use them to adjust the support set, as in Figure 3. The following two few-shot transfer learning methods are proposed.

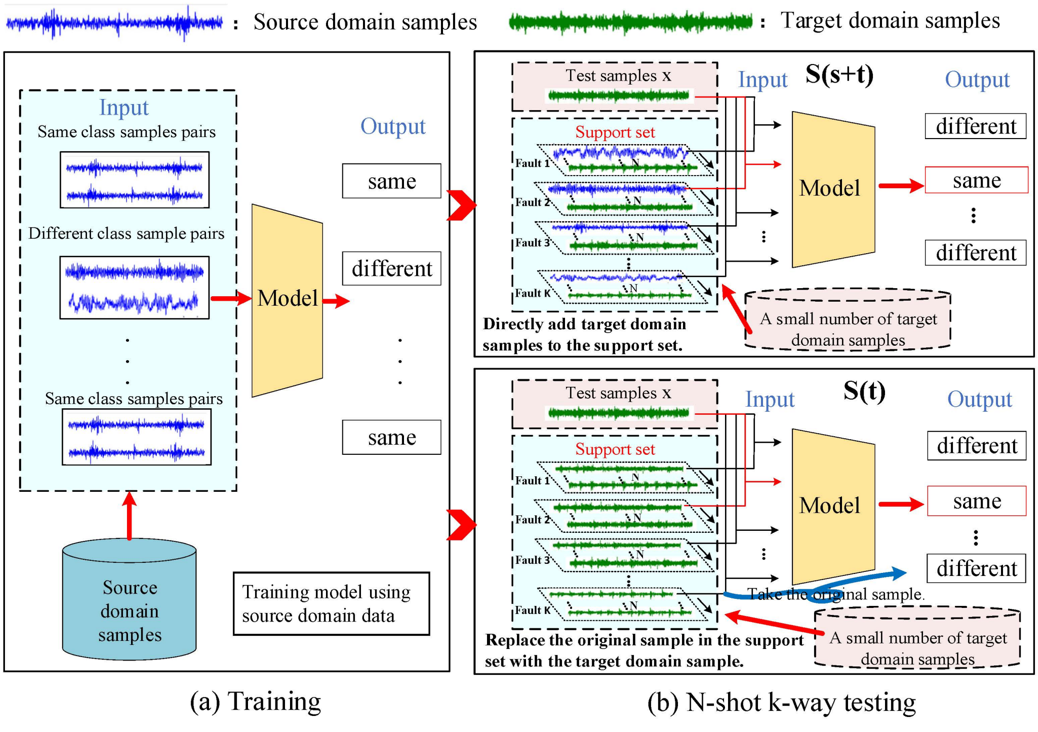


Figure 3. The proposed few-shot transfer learning methods.

(1) S(s+t): Directly add target domain samples to the support set.

This method adds a very small amount of target domain samples (x_t, y_t) to the original support set after the training with source domain data, and finally tested. In this case, the expression of the support set is Equation (6). In this paper, we uniformly use S(s+t) to denote the method of directly add target domain samples to the support set.

$$S = \{(x_{s1}, y_{s1})..(x_{sk}, y_{sk}), (x_{t1}, y_{t1})..(x_{tk}, y_{tk})\} \tag{6}$$

The s in x_s represents from the source domain, t in x_t represents from the target domain.

(2) S(t): Replace the original sample in the support set with the target domain sample.

In this method, after training the model with source domain data, a very small number of target domain samples are used to replace the original samples in the support set, and the model is finally tested. At this point, the support set is shown in Equation (7). In this paper, we uniformly use S(t) to denote the replacement of the original sample in the support set with the target domain sample.

$$S = \{(x_{t1}, y_{t1}), (x_{t2}, y_{t2}) \ldots, (x_{tk}, y_{tk})\} \tag{7}$$

The t in x_t represents the target domain.

3.2. Model

Figure 4 shows the model we use. This is a Siamese network with a deep convolution neural network with a wide first layer core (WDCNN). In this model, the two WDCNN have the same structure and parameters, and the weights are shared. The setting of the WDCNN network architecture is shown in Table 1, which is consistent with the setting in reference [33]. This design strategy is used because the vibration signal is more sensitive to the overall correlation in the time domain or frequency domain, and the useful information in the signal will be lost if the first layer core is too small, and because all layers are small cores which may be affected by high-frequency noise common in the industrial environment, resulting in poor performance of feature coding. It is proved that WDCNN with the first layer of wide kernel has good anti-noise ability, generalization ability and robustness. The model consists of a series of convolution layers, the step size of the first layer is set to 16, and the step size of the other layers is fixed to 1. In order to optimize the performance of the model, the number of convolution filters is a multiple of 16. In the previous convolution layer, the Relu activation function is used to encode the features, and the full connection layer uses the sigmoid activation function to map the features.

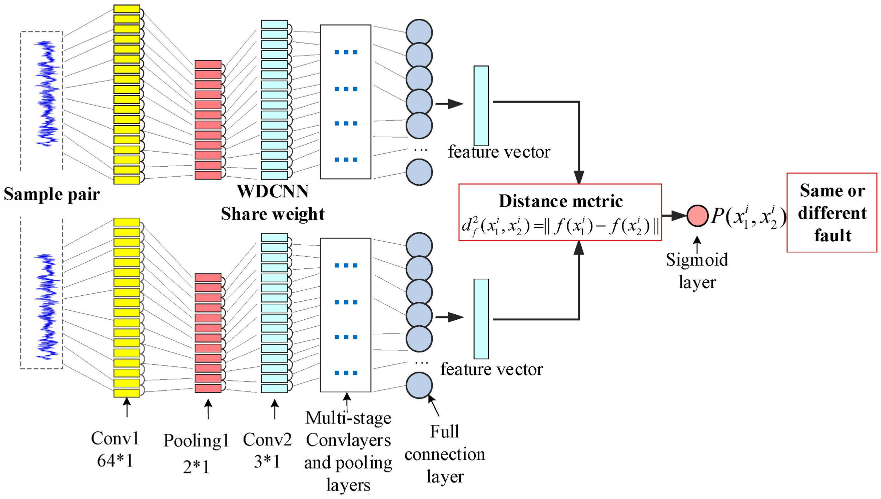


Figure 4. Few-shot learning model based on a Siamese network.

Table 1. Structural parameter of WDCNN.

No	Layer Type	Kernel Size/Stride	Kernel Number	Output Size (Width × Depth)	Padding
1	Conv1	64 × 1/16 × 1	16	128 × 16	same
2	Pooling1	2 × 1/2 × 1	16	64 × 16	valid
3	Conv2	3 × 1/1 × 1	32	64 × 32	same
4	Pooling2	2 × 1/2 × 1	32	32 × 32	valid
5	Conv3	3 × 1/1 × 1	64	32 × 64	same
6	Pooling3	2 × 1/2 × 1	64	16 × 64	valid
7	Conv4	3 × 1/1 × 1	64	16 × 64	same
8	Pooling4	2 × 1/2 × 1	64	8 × 64	valid
9	Conv5	3 × 1/1 × 1	64	6 × 64	valid
10	Pooling5	2 × 1/2 × 1	64	3 × 64	valid
11	Fully-connected	100	1	100 × 1	

Input is a pair of samples (x_1, x_2), which can be the same or different. The output is the probability of similarity between the sample pairs. Firstly, the metric distance between the outputs of the network is optimized by Equation (8), where f represents a deep convolution network. Equation (9) determines the probability of similarity, where sigm represents the Sigmoid function and FC is a dense fully connected layer.

$$d_f^2\left(x_1^i, x_2^i\right) = \|f\left(x_1^i\right) - f\left(x_2^i\right)\| \tag{8}$$

$$P\left(x_1^i, x_2^i\right) = \mathrm{sigm}(\mathrm{FC}\left(d_f^2\left(x_1^i, x_2^i\right)\right) \tag{9}$$

Let M represents the minibatch size, where i indexes the ith minibatch, let $y(i) = \left(x_1^{(i)}, x_2^{(i)}\right)$ be a length-M vector which contains the labels for the minibatch. Now we assume $y(i)$ equal to 1 when $\left(x_1^{(i)}, x_2^{(i)}\right)$ is the same class, and $y(i)$ equal to 0 when $\left(x_1^{(i)}, x_2^{(i)}\right)$ is different class. We impose a regularized cross-entropy objective on our binary classifier of the following form:

$$\begin{gathered} \mathcal{L}\left(x_1^{(i)}, x_2^{(i)}\right) = \mathrm{y(i)}\log\left(\mathrm{p}\left(x_1^{(i)}, x_2^{(i)}\right)\right) + \\ (1-\mathrm{y(i)})\log\left(1-\mathrm{p}\left(x_1^{(i)}, x_2^{(i)}\right)\right) \\ +\lambda^T|w|_2 \end{gathered} \tag{10}$$

The optimizer we chose is Adam, which calculates individual adaptive learning rates. Update parameters through Equation (11):

$$\begin{array}{rll} m_w^{(T+1)} & \mid = & \beta_1 m_w^{(T)} + (1-\beta_1)\nabla_w L^{(T)} \\ v_w^{(T+1)} & \mid = & \beta_2 v_w^{(T)} + (1-\beta_2)\left(\nabla_w L^{(T)}\right)^2 \\ \hat{m}_w & \mid = & \frac{m_w^{(T+1)}}{1-(\beta_1)^{T+1}} \\ \hat{v}_w & \mid = & \frac{v_w^{(T+1)}}{1-(\beta_2)^{T+1}} \\ w^{(T+1)} & \mid = & w^{(T)} - \eta\frac{\hat{m}_w}{\sqrt{\hat{v}_w}+\epsilon} \end{array} \tag{11}$$

where $w^{(T+1)}$ means the parameters at epoch T, $L^{(T)}$ is the loss function, β_i is the forgetting factor of the ith moment of the gradient, m and v are moving averages.

4. Experiment and Results

4.1. Data Introduction and Processing

Like most deep learning algorithms, in order to confirm our proposed transfer learning strategy, we need to prepare appropriate data samples. We selected the data provided by Case Western Reserve University (CWRU) [34] as the ASF datasets, that is, the source domain. They are collected from the experimental platform of CWRU (shown in Figure 5), and all use the single point damage of electro-discharge machining (EDM). The vibration acceleration signal of the faulty bearing is collected by the accelerometer, and the sampling frequency is 12 kHz. The bearings selected in this paper are installed at the drive end, and there are three types of bearings: inner ring fault bearing, outer ring fault bearing and normal bearing. The parameters are shown in Table 2.

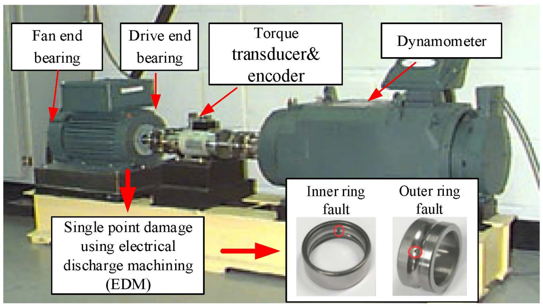

Fan end bearing
Drive end bearing
Torque transducer& encoder
Dynamometer
Inner ring fault
Outer ring fault
Single point damage using electrical discharge machining (EDM)

Figure 5. Bearing test bench of CWRU.

Table 2. Source domain data parameters.

Dataset Name	Name	Fault Location	Speed (rpm)	Loads (hp)
A	0.021-OuterRace 0.021-InnerRace Normal	Outer ring Inner ring None	1772 1772 1772	1 1 1
B	0.021-OuterRace 0.021-InnerRace Normal	Outer ring Inner ring None	1750 1750 1750	2 2 2
C	0.021-OuterRace 0.021-InnerRace Normal	Outer ring Inner ring None	1730 1730 1730	3 3 3

On the modular test bench (Figure 6), the Paderborn University (PU) researchers with 6 sets of normal bearing data, 12 sets of artificially damaged bearing data of three fault types, and 14 groups of naturally damaged bearing data caused by accelerating lifetime test [35]. Damage levels are divided according to the percentage of length of the damage relative to pitch circumference is calculated (Table 3). The vibration acceleration signal of the faulty bearing is collected by the accelerometer, and the sampling frequency is 64 kHz. We choose the natural damage dataset of PU as the target domain data, and the details of the parameters are shown in Table 4.

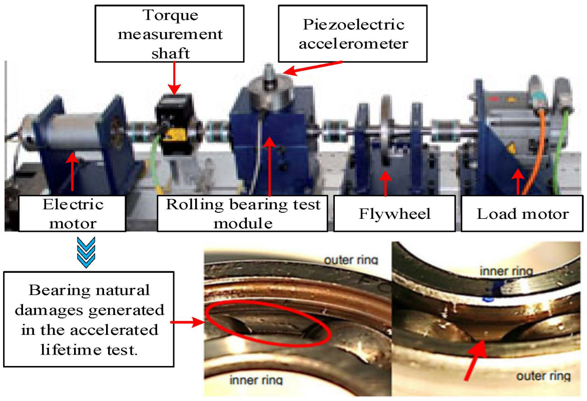

Torque measurement shaft
Piezoelectric accelerometer
Electric motor
Rolling bearing test module
Flywheel
Load motor
Bearing natural damages generated in the accelerated lifetime test.
outer ring
inner ring
inner ring
outer ring

Figure 6. Modular test bench of PU.

Table 3. Damage levels to determine the extent of damage.

Damage Level	Assigned Percentage Values	Limits for Bearing 6203
1	0–2%	≤2 mm
2	2–5%	>2 mm

Table 4. Target domain data parameters.

Dataset Name	Name	Fault Location	Damage (Main Mode and Symptom)	Damage Level	Damage Feature	Load Torque (Nm)	Speed (rpm)	Radial Force (N)
D	KI04	Inner ring	Fatigue: pitting	1	Single	0.7	1500	1000
	KA04	Outer ring	Fatigue: pitting	1	Single	0.7	1500	1000
	K005	Normal	None	None	None	0.7	1500	1000
E	KI16	Inner ring	Fatigue: pitting	2	Single	0.7	1500	1000
	KA16	Outer ring	Fatigue: pitting	2	Single	0.7	1500	1000
	K004	Normal	None	None	None	0.7	1500	1000

We sampled and processed the CWRU source domain data in Table 2, taking all 2048 data points as a sample. Because there are not enough data points in the original data, the number of samples that can be intercepted is too small, and when the number of training samples is very small, it is easy to cause over-fitting. Therefore, we use the method of overlapping sampling as shown in Figure 7a. There is a partial overlap between each sample and the subsequent sample, with an offset of 80 and the training samples are obtained. Similarly, we process the PU natural damage fault data as shown in Figure 7b. Finally, the testing samples are obtained and a small number of samples for adjustment support set (SNSASS) are obtained. It is worth noting that the testing samples and SNSASS are independent and not duplicated. SNSASS can be seen as a small number of samples of real machines that can be obtained. The experimental samples are shown in Table 5.

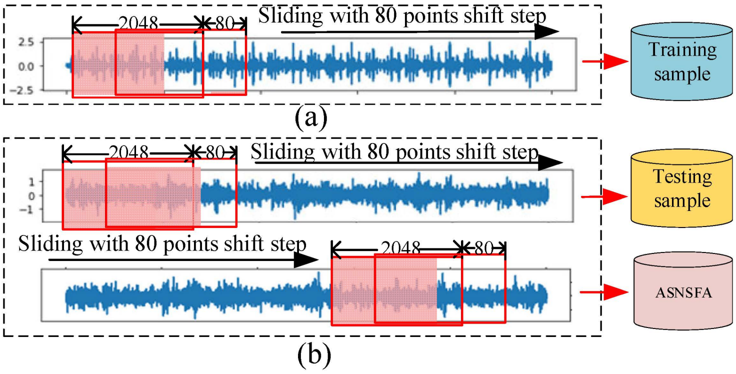

2048
80
Sliding with 80 points shift step
2.5
0.0
−2.5
Training sample
(a)
2048
80
Sliding with 80 points shift step
1
0
−1
Testing sample
Sliding with 80 points shift step
2048
80
ASNSFA
(b)

Figure 7. Data processing: (**a**) source domain data processing, (**b**) target domain data processing.

Table 5. Experimental samples.

	Sample Purpose	Inner Ring 0	Outer Ring 1	Normal 2	Total
Source domain	Training	660	660	660	1980
Target domain	SNSASS	5	5	5	15
	Testing	75	75	75	225

4.2. S(s), S(s+t) and S(t) Analysis

To verify the validity of our proposed transfer method, we performed the following three experiments as shown in Table 6.

Table 6. Eight kinds of experiments.

Number	Experiment Name	Model	Support Set
1	S(s)	Siamese network	Training sample
2	S(s+t)	Siamese network	Training sample and SNSASS
3	S(t)	Siamese network	SNSASS

(1) S(s): direct transfer method (baseline).

Direct transfer method is a simple method without any optimization and adjustment of fixed network parameters. This method uses source domain data for training and directly uses target domain data for testing. In this experiment, the support set of the direct transfer method based on a Siamese network is shown Equation (12), and the samples are all training samples from the source domain. Direct transfer method based on the Siamese network is expressed by S(s).

$$S = \{(x_{s1}, y_{s1}), (x_{s2}, y_{s2}) \ldots, (x_{sk}, y_{sk})\} \quad (12)$$

The s in x_s represents from the source domain.

In the experiment of S(s), we use the ASF data from CWRU to train and learn in the Adam optimization program, the epochs of training are 90, and batch size chooses 64, and the diagnostic knowledge learned is fixed. In the testing process, we directly input the NF samples provided by PU into the model for feature extraction, and then select the samples that are most similar to the test samples from the support set (the samples in the support set are training samples), and think that they are the same class.

(2) S(s+t): directly add target domain samples to the support set.

The training process is the same as that of (1). Before testing, however, SNSASS are added to the support set as a classification reference. The testing process is the same as that of (1), except that the support set contains both training samples and SNSASS.

(3) S(t): replace the original samples in the support set with the target domain samples.

In the experiment of S(t), the process of training stage is consistent with that of (1), and then all the samples of the original support set are replaced by SNSASS (the sample in the support set at this time is SNSASS). In the process of testing, input PU samples to test and obtain the results.

First of all, we verify the performance of S(s) (baseline), S(s+t) and S(t) in A→D, B→D, C→D, A→E, B→E and C→E transfer tasks, each experiment is carried out 10 times, and finally take the average. The experimental results are shown in Figure 8.

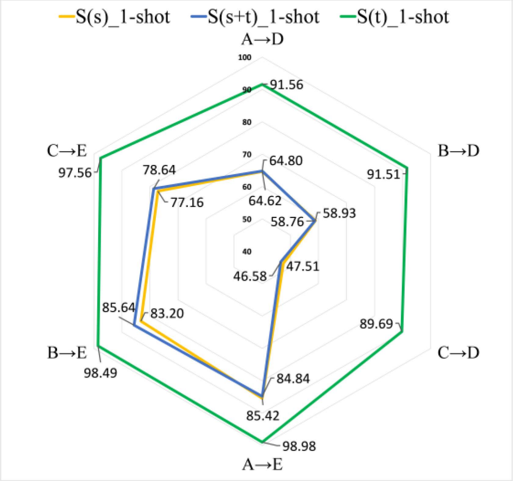


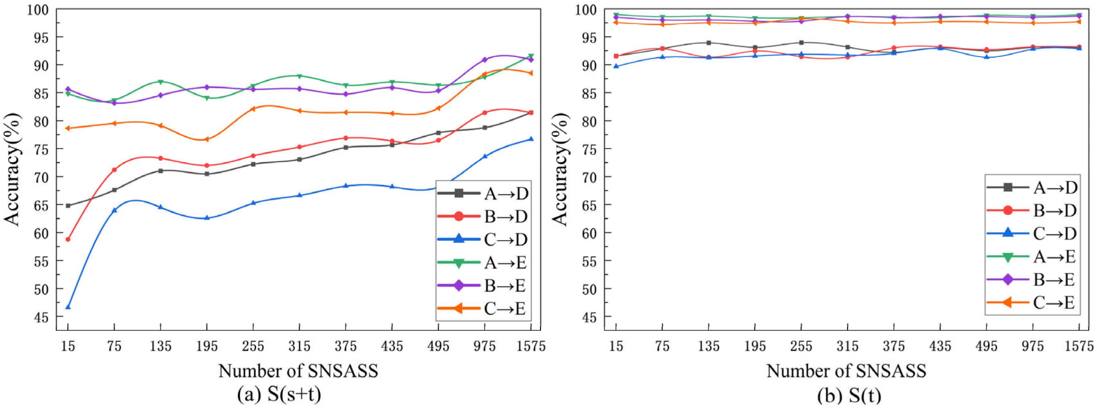
Figure 8. The results of S(s) (baseline), S(s+t) and S(t) in different transfer learning tasks.

It can be seen from Figure 8 that S(t) has an absolute advantage in all tasks. The accuracy is more than 89.69%, which is much higher than the other two methods, 42.18% higher than S(s) in C→D. This is because, based on the S(t) learning theory, the instances of the support set are all SNSASS (x_t, y_t) from the target domain, and the spatial distribution of the feature space of the test samples $\hat{x}$ that need to be classified is very similar to that of x_t, so it is easy to find similar examples in the support set and regard them as the same class of fault. The experimental results of S(s) and S(s+t) are very close, but in most cases the accuracy of S(s+t) is slightly higher than that of S(s). This is because the support set of S(s+t) has a small number of SNSASS. Based on the few-shot learning theory (see Section 2), these SNSASS can help the test sample $\hat{x}$ to better find the most similar class to itself. However, its number accounts for a small proportion (see Equation (13), $\eta = \frac{15}{1980+15} \approx 0.75\%$) in the support set, which cannot bring great performance improvement as S(t) ($\eta = \frac{15}{15} = 100\%$) does.

$$\eta = \frac{n_{SNSASS}}{n_{SNSASS} + n_{Training}} \tag{13}$$

where η is the proportion of the number of SNSASS in the total quantity. n_{SNSASS} is the number of SNSASS. $n_{Training}$ is the number of training sample.

In order to further verify the effect of η on S(s+t) and S(t), we gradually increase the number of SNSASS and repeat the experiment again, each experiment is repeated 10 times, and the result is shown in Figure 9. As can be seen from Figure 9a, with the increase in the number of SNSASS (η increase), the accuracy of S(s+t) does not increase linearly, but it shows an increasing trend as a whole, especially in A→D, B→D and C→D experiments. However, with the increase in SNSASS, the performance of S(t) has not been improved as shown in Figure 9b, and the accuracy fluctuates within an allowable error range. In other words, if we can obtain a small amount of target domain data, S(t) can give full play to its performance.



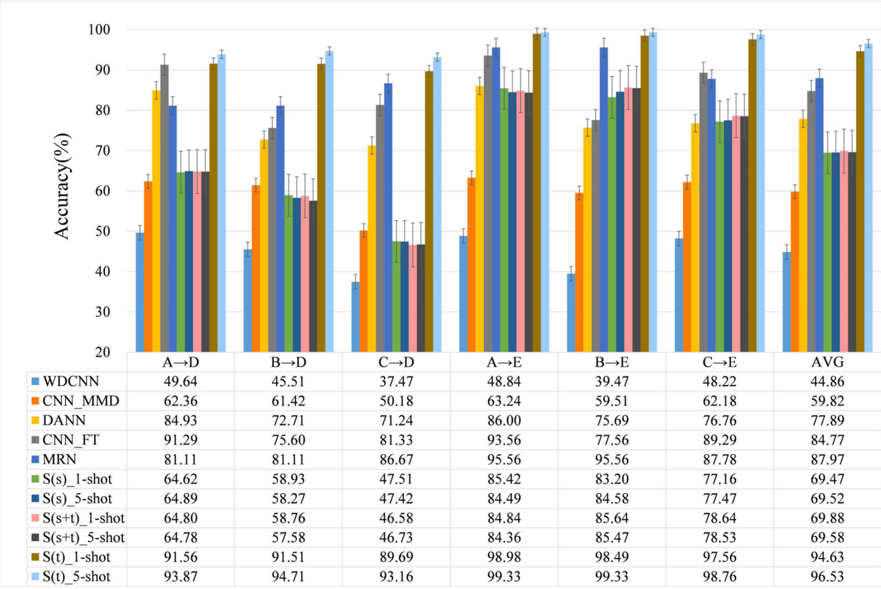
(a) S(s+t)

(b) S(t)

Figure 9. The curve of accuracy with the increase in the number of SNSASS. (**a**) the results of S(s+t). (**b**) the results of S(t).

4.3. Comparisons with Other Methods

We also contrast our method with some popular methods, which include WDCNN [18,33], CNN_MMD [36], CNN_FT [37], DANN [38] and MRN [30]. It should be noted that we set the experimental parameters to the best case according to the characteristics of each method, including data format, hyperparameters, epochs, and so on. The number of training samples is 1980, the number of SNSASS is 15, and the number of test samples is 225. Similarly, each method is tested 10 times in the A→D, B→D, C→D, A→E, B→E and C→E transfer tasks in turn, and the results are averaged. The experimental results are shown in Figure 10.



	A→D	B→D	C→D	A→E	B→E	C→E	AVG
WDCNN	49.64	45.51	37.47	48.84	39.47	48.22	44.86
CNN_MMD	62.36	61.42	50.18	63.24	59.51	62.18	59.82
DANN	84.93	72.71	71.24	86.00	75.69	76.76	77.89
CNN_FT	91.29	75.60	81.33	93.56	77.56	89.29	84.77
MRN	81.11	81.11	86.67	95.56	95.56	87.78	87.97
S(s)_1-shot	64.62	58.93	47.51	85.42	83.20	77.16	69.47
S(s)_5-shot	64.89	58.27	47.42	84.49	84.58	77.47	69.52
S(s+t)_1-shot	64.80	58.76	46.58	84.84	85.64	78.64	69.88
S(s+t)_5-shot	64.78	57.58	46.73	84.36	85.47	78.53	69.58
S(t)_1-shot	91.56	91.51	89.69	98.98	98.49	97.56	94.63
S(t)_5-shot	93.87	94.71	93.16	99.33	99.33	98.76	96.53

Figure 10. Experimental results.

Experiments show that S(t)_5-shot achieves the highest accuracy in all transfer learning tasks, with an average of 96.53%, and S(t)_1-shot ranks second with an average of 94.63%, followed by MRU, CNN_FT, DANN, S(s+t), S(s), CNN_MMD. There is no doubt that WDCNN performs the worst among all transfer learning tasks, with an average accuracy of only 44.86%. Of course, we know that it is unfair to compare WDCNN with these advanced methods, but it also reflects the difficulty of these transfer learning tasks. After all, there is a big gap between the fault features of ASF and those of NF. It is also evident from the figure that in almost all methods (except WDCNN) the results of A→D, B→D and C→D are worse than A→E, B→E and C→E, the reason is that D's lower damage (level 1 damage) level than E (level 2 damage), E's more serious damage and more obvious failure features. Learning the knowledge of A, B and C breakdown to diagnose E would be better.

Figure 11 shows the standard deviation of 10 repeated experiments for each method in the A→D, B→D, C→D, A→E, B→E and C→E transfer learning tasks. As can be seen from Figure 11, S(t)_5-shot has the smallest standard deviation among all transfer learning tasks with an average of 2.66%, followed by S(t)_1-shot with an average of 3.53%, much smaller than other methods. Except that the average standard deviation of MRN is 8.45%, the rest are more than 10%, which means that it is difficult to learn diagnosis knowledge from ASF to diagnose NF, resulting in very unstable diagnosis results of other methods. Simultaneously, it is demonstrated that S(t) has much higher stability than other methods.

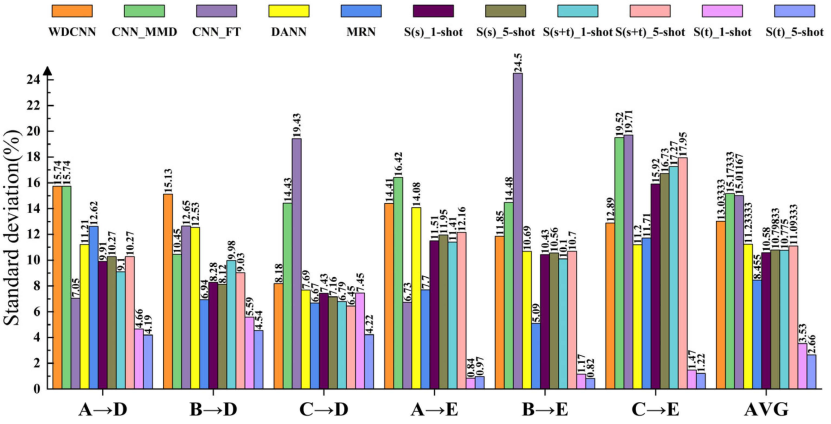


Figure 11. The standard deviation of 10 repeated experiments for each method.

4.4. The Influence of Different Source Domain and the Number of Training Samples

It is not only the quality of the data in the source domain that is very important, but the quantity of the data is also important. It is very important to select the appropriate bearing fault source domain data and quantitative training model, but in the actual industrial production, it is difficult to determine the appropriate source domain data. Therefore, in this section we discuss the harshness of the proposed method on the source domain data. We selected several relatively well-performing methods for comparison, Figure 12 is the result curve of DANN, MRN, CNN_FT and S(t)_1-shot learning fault diagnosis knowledge from datasets A, B and C, respectively, and used it to diagnose D and E. It can be clearly seen that DANN, MRN and CNN_FT learns knowledge from different source domains and fixes the model, which leads to great differences in experimental results. The reason is that there is a big gap between the working conditions of A (1772 rpm), B (1750 rpm) and C (1730 rpm) in speed. However, compared with other methods, the result of S(t)_1-shot learning from A, B and C to diagnose D and E has only a small change and a slight downward trend, indicating that S(t) has good ability to learn and can be well leverage knowledge of the source domain. The reasons for the slight downward trend with A-B-C are as follows: according to the speed of A, B and C, we think that the working condition of A is more complex than that of B, and that of B is more complex than that of C. The model can learn more obvious fault features under more complex working conditions, so as to better complete the transfer task. In [18,21,39], the authors have also obtained a similar conclusion.

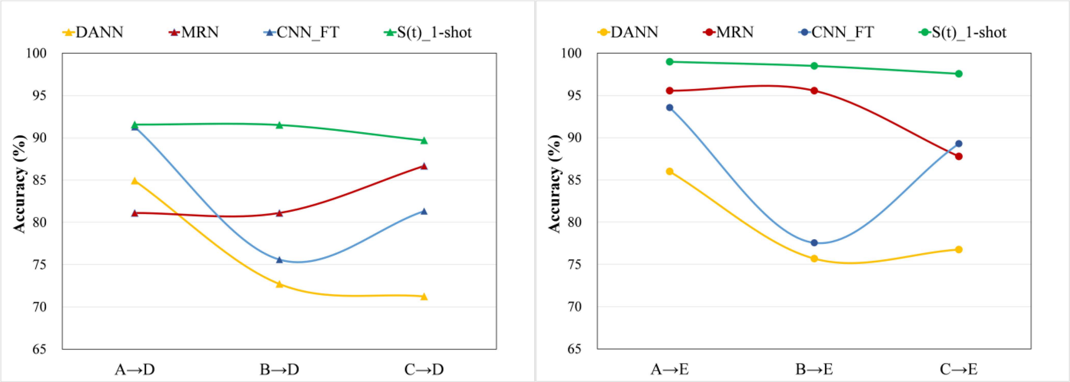
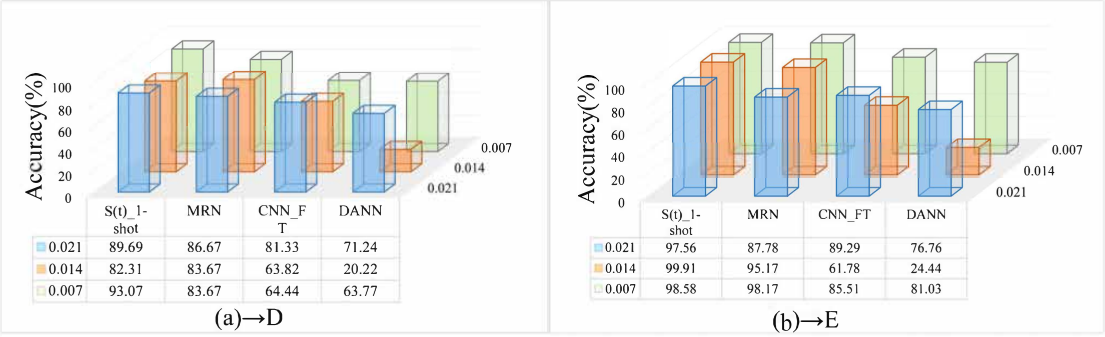


Figure 12. The variation of experimental results with different source domains.

More complex working condition will have more diagnostic knowledge. Next, we want to explore the influence of different fault diameters of the source domain bearing. Therefore, an additional small experiment was performed here. Following the principle of control variables, CWRU data (the load is 3 hp) with fault diameters of 0.021, 0.014, and 0.007 inches were used as training sets and tested in D and E, the result is shown in Figure 13.



	S(t)_1-shot	MRN	CNN_F T	DANN
🟦 0.021	89.69	86.67	81.33	71.24
🟧 0.014	82.31	83.67	63.82	20.22
🟩 0.007	93.07	83.67	64.44	63.77

(a)→D

	S(t)_1-shot	MRN	CNN_FT	DANN
🟦 0.021	97.56	87.78	89.29	76.76
🟧 0.014	99.91	95.17	61.78	24.44
🟩 0.007	98.58	98.17	85.51	81.03

(b)→E

Figure 13. Results of training in different fault degrees, (**a**) tested in D. (**b**) tested in E.

As can be seen in Figure 13, S(t) effectively learns diagnostic knowledge from different fault diameters and shows the best performance, followed by MRN, and the worst is DANN, which surprises us with just over 20% in 0.014 inches. However, we failed to find the rule that the fault diameters affect the performance of the model, which may be due to the big difference between ASF and NF.

In order to explore the performance of various methods under different sample numbers, the following groups of experiments were carried out when the number of training samples was 90, 300, 600, 1200, 1500 and 1980. As shown in Figure 14, it presents the curve of all experimental results with the number of training samples. Incredibly, the experimental results do not improve with the increase in the number of training samples, but show a special curve shape. It is because having too small a number of training samples will cause the model to learn insufficient knowledge that can be used in the target domain, resulting in poor performance when diagnosing in the target domain; and having too many training samples will cause the learned knowledge to be too focused on the source domain, which is not applicable when it is transferred to the target domain. However, compared with other methods, the results of S(t) do not change greatly with the number of data samples, which shows that the dependence of S(t) method on data samples is very small and stable. This is because the few-shot learning strategy of S(t) can learn and use knowledge in a small number of training samples and is not sensitive to the growth of data.

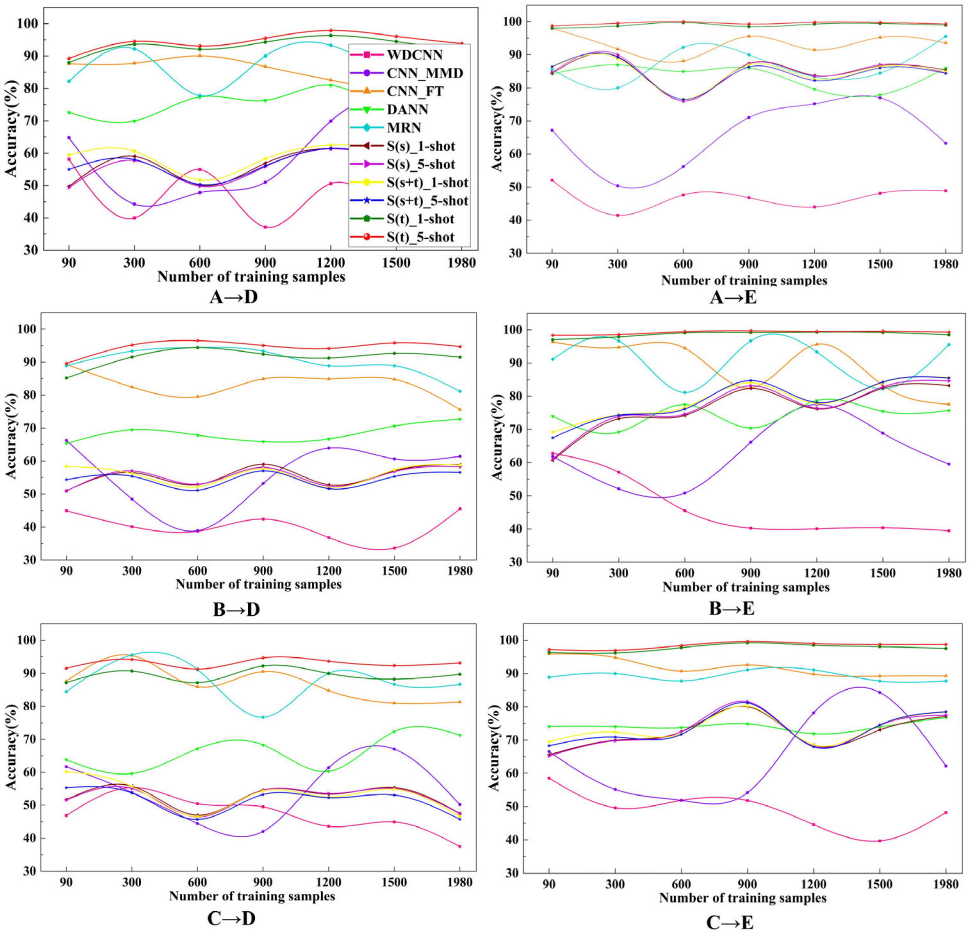


Figure 14. Variation of diagnostic results with different numbers of training samples.

Assuming that a small sample of the target domain is obtained, similarly to fine-tuning-based methods, S(t) can improve the performance of the model in transfer learning. However, after obtaining the new target domain data, fine-tuning-based methods still need to train the models that have been trained in the source domain. The S(t) method does not need a second training, but only needs to input these small target domain samples into the support set, which is more convenient than fine-tuning-based methods in some practical applications.

5. Conclusions

In this paper, it is established that there is still a long distance between the research of intelligent fault diagnosis and the practical industrial application. A bearing fault diagnosis based on few-shot transfer learning across different datasets is proposed, which uses a very small number of target domain samples to adjust the support set to improve the generalization performance of the model. Many groups of transfer experiments are carried out by using the ASF dataset of CWRU and the NF dataset of PU. The conclusions are as follows:

(1) With only a small amount of SNSASS, S(t) method greatly improves the accuracy of fault classification, and the accuracy of S(s+t) is not significantly improved, but increases with the increase in the number of SNSASS.
(2) Compared with other methods, the proposed S(t) method has the highest accuracy in all cases and is also the most stable method.
(3) S(t) can fully learn diagnostic knowledge in different source domains and sample numbers, and effectively use this knowledge to identify the health state of the target bearing, which has strong generalization and robustness. In addition, unlike the fine-tuning-based method, S(t) does not need secondary training, which is more convenient in some practical applications.

S(t) provides a feasible way to apply laboratory data knowledge to real machine fault diagnosis, solves the difficulty that a large amount of data cannot be collected in the real world, and also provides a new idea and method for transfer learning. However, obtaining a small amount of target domain data (SNSASS) is the key to the S(t) method. In some cases of actual industrial production, it is also not easy to obtain a small amount of target domain data, which is a limitation of the S(t) method. At the same time, although the difference between ASF and NF brings great challenges to the transfer learning tasks, because of the lack of available data, we were only able to perform three classification tasks. More classification experiments and verification in more datasets can be performed in the future.

Author Contributions: Y.Z.: Conceptualization; Methodology; Visualization; Data curation, Writing—Original Draft; Writing—Review and Editing; Formal analysis. S.L.: Conceptualization, Methodology, Funding acquisition. A.Z.: Conceptualization, Methodology, Data curation, Writing—review and editing, Supervision. C.L.: Writing—review and editing, Supervision. L.Q.: Writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by National Key R and D Program of China (No. 2020YFB1713300); Guizhou Science and Technology Planning Project (No. [2015]4011, No. [2017]5788); Talent Training Base Project of Colleges and Universities in Guizhou Province (No. [2020]005, No. [2020]009). Guizhou Provincial Department of Education Youth Science and Technology Talent Growth Project (No. QJH KY [2022]142); Scientific Research Project of Introducing Talents of Guizhou University (No. GDRJH (2021)74).

Acknowledgments: Thanks for the computing support of the State Key Laboratory of Public Big Data, Guizhou University.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Zhang, S.; Zhang, S.; Wang, B.; Habetler, T.G. Deep Learning Algorithms for Bearing Fault Diagnostics—A Comprehensive Review. *IEEE Access* **2020**, *8*, 29857–29881. [CrossRef]
2. Shi, H.; Fu, W.; Li, B.; Shao, K.; Yang, D. Intelligent Fault Identification for Rolling Bearings Fusing Average Refined Composite Multiscale Dispersion Entropy-Assisted Feature Extraction and SVM with Multi-Strategy Enhanced Swarm Optimization. *Entropy* **2021**, *23*, 527. [CrossRef] [PubMed]

3. Liu, D.; Wang, Q.; Tao, J.; Li, G.; Wu, J. Fault Diagnosis Method Based on Improved Deep Boltzmann Machines. In Proceedings of the 2018 IEEE 7th Data Driven Control and Learning Systems Conference (DDCLS), Enshi, China, 25–27 May 2018; pp. 458–462.
4. Zhao, B.; Zhang, X.; Li, H.; Yang, Z. Intelligent Fault Diagnosis of Rolling Bearings Based on Normalized CNN Considering Data Imbalance and Variable Working Conditions. *Knowl.-Based Syst.* **2020**, *199*, 105971. [CrossRef]
5. 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]
6. Wang, S.; Wang, D.; Kong, D.; Wang, J.; Li, W.; Zhou, S. Few-Shot Rolling Bearing Fault Diagnosis with Metric-Based Meta Learning. *Sensors* **2020**, *20*, 6437. [CrossRef]
7. Wu, Z.; Lin, W.; Fu, B.; Guo, J.; Ji, Y.; Pecht, M. A Local Adaptive Minority Selection and Oversampling Method for Class-Imbalanced Fault Diagnostics in Industrial Systems. *IEEE Trans. Reliab.* **2020**, *69*, 1195–1206. [CrossRef]
8. Zhang, Y.; Li, X.; Gao, L.; Wang, L.; Wen, L. Imbalanced Data Fault Diagnosis of Rotating Machinery Using Synthetic Oversampling and Feature Learning. *J. Manuf. Syst.* **2018**, *48*, 34–50. [CrossRef]
9. Liu, Q.; Ma, G.; Cheng, C. Data Fusion Generative Adversarial Network for Multi-Class Imbalanced Fault Diagnosis of Rotating Machinery. *IEEE Access* **2020**, *8*, 70111–70124. [CrossRef]
10. Soltanzadeh, P.; Hashemzadeh, M. RCSMOTE: Range-Controlled Synthetic Minority over-Sampling Technique for Handling the Class Imbalance Problem. *Inf. Sci.* **2021**, *542*, 92–111. [CrossRef]
11. Zhang, K.; Chen, J.; Zhang, T.; He, S.; Pan, T.; Zhou, Z. Intelligent Fault Diagnosis of Mechanical Equipment under Varying Working Condition via Iterative Matching Network Augmented with Selective Signal Reuse Strategy. *J. Manuf. Syst.* **2020**, *57*, 400–415. [CrossRef]
12. Fang, Q.; Wu, D. ANS-Net: Anti-Noise Siamese Network for Bearing Fault Diagnosis with a Few Data. *Nonlinear Dyn.* **2021**, *104*, 2497–2514. [CrossRef]
13. Lu, N.; Hu, H.; Yin, T.; Lei, Y.; Wang, S. Transfer Relation Network for Fault Diagnosis of Rotating Machinery with Small Data. *IEEE Trans. Cybern.* **2021**, 1–15. [CrossRef]
14. Mai, S.; Hu, H.; Xu, J. Attentive Matching Network for Few-Shot Learning. *Comput. Vis. Image Underst.* **2019**, *187*, 102781. [CrossRef]
15. Xiao, D.; Huang, Y.; Qin, C.; Liu, Z.; Li, Y.; Liu, C. Transfer Learning with Convolutional Neural Networks for Small Sample Size Problem in Machinery Fault Diagnosis. *Proc. Inst. Mech. Eng. C J. Mech. Eng. Sci.* **2019**, *233*, 5131–5143. [CrossRef]
16. 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.* **2021**, *119*, 152–171. [CrossRef]
17. Ren, Z.; Zhu, Y.; Yan, K.; Chen, K.; Kang, W.; Yue, Y.; Gao, D. A Novel Model with the Ability of Few-Shot Learning and Quick Updating for Intelligent Fault Diagnosis. *Mech. Syst. Signal Process.* **2020**, *138*, 106608. [CrossRef]
18. Zhang, A.; Li, S.; Cui, Y.; Yang, W.; Dong, R.; Hu, J. Limited Data Rolling Bearing Fault Diagnosis with Few-Shot Learning. *IEEE Access* **2019**, *7*, 110895–110904. [CrossRef]
19. Li, X.; Jiang, H.; Zhao, K.; Wang, R. A Deep Transfer Nonnegativity-Constraint Sparse Autoencoder for Rolling Bearing Fault Diagnosis with Few Labeled Data. *IEEE Access* **2019**, *7*, 91216–91224. [CrossRef]
20. Feng, Y.; Chen, J.; Zhang, T.; He, S.; Xu, E.; Zhou, Z. Semi-Supervised Meta-Learning Networks with Squeeze-and-Excitation Attention for Few-Shot Fault Diagnosis. *ISA Trans.* **2022**, *120*, 383–401. [CrossRef]
21. Li, C.; Li, S.; Zhang, A.; He, Q.; Liao, Z.; Hu, J. Meta-Learning for Few-Shot Bearing Fault Diagnosis under Complex Working Conditions. *Neurocomputing* **2021**, *439*, 197–211. [CrossRef]
22. Yu, C.; Ning, Y.; Qin, Y.; Su, W.; Zhao, X. Multi-Label Fault Diagnosis of Rolling Bearing Based on Meta-Learning. *Neural Comput. Appl.* **2021**, *33*, 5393–5407. [CrossRef]
23. Wang, D.; Zhang, M.; Xu, Y.; Lu, W.; Yang, J.; Zhang, T. Metric-Based Meta-Learning Model for Few-Shot Fault Diagnosis under Multiple Limited Data Conditions. *Mech. Syst. Signal Process.* **2021**, *155*, 107510. [CrossRef]
24. Pei, Z.; Jiang, H.; Li, X.; Zhang, J.; Liu, S. Data Augmentation for Rolling Bearing Fault Diagnosis Using an Enhanced Few-Shot Wasserstein Auto-Encoder with Meta-Learning. *Meas. Sci. Technol.* **2021**, *32*, 84007. [CrossRef]
25. Zhang, S.; Ye, F.; Wang, B.; Habetler, T.G. Few-Shot Bearing Fault Diagnosis Based on Model-Agnostic Meta-Learning. *IEEE Trans. Ind. Appl.* **2021**, *57*, 4754–4764. [CrossRef]
26. Feng, Y.; Chen, J.; Yang, Z.; Song, X.; Chang, Y.; He, S.; Xu, E.; Zhou, Z. Similarity-Based Meta-Learning Network with Adversarial Domain Adaptation for Cross-Domain Fault Identification. *Knowl.-Based Syst.* **2021**, *217*, 106829. [CrossRef]
27. Wang, C.; Xu, Z. An Intelligent Fault Diagnosis Model Based on Deep Neural Network for Few-Shot Fault Diagnosis. *Neurocomputing* **2021**, *456*, 550–562. [CrossRef]
28. He, Z.; Shao, H.; Wang, P.; Lin, J.; Cheng, J.; Yang, Y. Deep Transfer Multi-Wavelet Auto-Encoder for Intelligent Fault Diagnosis of Gearbox with Few Target Training Samples. *Knowl.-Based Syst.* **2020**, *191*, 105313. [CrossRef]
29. Liu, Y.Z.; Shi, K.M.; Li, Z.X.; Ding, G.F.; Zou, Y.S. Transfer Learning Method for Bearing Fault Diagnosis Based on Fully Convolutional Conditional Wasserstein Adversarial Networks. *Measurement* **2021**, *180*, 109553. [CrossRef]
30. Wu, J.; Zhao, Z.; Sun, C.; Yan, R.; Chen, X. Few-Shot Transfer Learning for Intelligent Fault Diagnosis of Machine. *Measurement* **2020**, *166*, 108202. [CrossRef]
31. Yang, B.; Lei, Y.; Jia, F.; Xing, S. An Intelligent Fault Diagnosis Approach Based on Transfer Learning from Laboratory Bearings to Locomotive Bearings. *Mech. Syst. Signal Process.* **2019**, *122*, 692–706. [CrossRef]

32. 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*, 7316–7325. [CrossRef]
33. Zhang, W.; Peng, G.; Li, C.; Chen, Y.; Zhang, Z. A New Deep Learning Model for Fault Diagnosis with Good Anti-Noise and Domain Adaptation Ability on Raw Vibration Signals. *Sensors* **2017**, *17*, 425. [CrossRef] [PubMed]
34. Case Western Reserve University Bearing Data Center Website. Available online: https://engineering.case.edu/bearingdatacenter/welcome (accessed on 8 July 2021).
35. Lessmeier, C.; Kimotho, J.K.; Zimmer, D.; Sextro, W. Condition Monitoring of Bearing Damage in Electromechanical Drive Systems by Using Motor Current Signals of Electric Motors: A Benchmark Data Set for Data-Driven Classification. In Proceedings of the PHM Society European Conference 2016, Chengdu, China, 19–21 October 2016. [CrossRef]
36. Xiao, D.; Huang, Y.; Zhao, L.; Qin, C.; Shi, H.; Liu, C. Domain Adaptive Motor Fault Diagnosis Using Deep Transfer Learning. *IEEE Access* **2019**, *7*, 80937–80949. [CrossRef]
37. Li, F.; Chen, J.; Pan, J.; Pan, T. Cross-Domain Learning in Rotating Machinery Fault Diagnosis under Various Operating Conditions Based on Parameter Transfer. *Meas. Sci. Technol.* **2020**, *31*, 085104. [CrossRef]
38. Ganin, Y.; Ustinova, E.; Ajakan, H.; Germain, P.; Larochelle, H.; Laviolette, F.; Marchand, M.; Lempitsky, V. Domain-Adversarial Training of Neural Networks. *J. Mach. Learn. Res.* **2016**, *17*, 1–35.
39. Li, C.; Li, S.; Zhang, A.; Yang, L.; Zio, E.; Pecht, M.; Gryllias, K. A Siamese Hybrid Neural Network Framework for Few-Shot Fault Diagnosis of Fixed-Wing Unmanned Aerial Vehicles. *J. Comput. Des. Eng.* **2022**, *9*, 1511–1524. [CrossRef]