



## Article

# An Adversarial Single-Domain Generalization Network for Fault Diagnosis of Wind Turbine Gearboxes

Xinran Wang<sup>1,†</sup>, Chenyong Wang<sup>1,†</sup>, Hanlin Liu<sup>1</sup>, Cunyou Zhang<sup>1</sup>, Zhenqiang Fu<sup>2</sup>, Lin Ding<sup>3</sup>, Chenzhao Bai<sup>1</sup>, Hongpeng Zhang<sup>1</sup> and Yi Wei<sup>1,\*</sup>

<sup>1</sup> Marine Engineering College, Dalian Maritime University, No. 1 Linghai Road, Dalian 116000, China; w1120211123@dlmu.edu.cn (X.W.)

<sup>2</sup> School of Navigation and Shipping, Shandong Jiaotong University, Weihai 264200, China

<sup>3</sup> Dalian Certification, Accreditation, Inspection and Testing Supervision Office, Dalian 116000, China

\* Correspondence: weiyi@dlmu.edu.cn

<sup>†</sup> These authors contributed equally to this work.

**Abstract:** In deep learning-based fault diagnosis of the wind turbine gearbox, a commonly faced challenge is the domain shift caused by differing operational conditions. Traditional domain adaptation methods aim to learn transferable features from the source domain and apply them to the target data. However, such methods still require access to target domain data during the training process, which limits their applicability in real-time fault diagnosis. To address this issue, we introduce an adversarial single-domain generalization network (ASDGN). It relies solely on data from a single length of data acquisition in wind turbine fault diagnosis. This novel approach introduces a more flexible and efficient solution to the field of real-time fault diagnosis for wind turbines.

**Keywords:** wind turbine gearbox; fault diagnosis; domain generalization; deep learning



**Citation:** Wang, X.; Wang, C.; Liu, H.; Zhang, C.; Fu, Z.; Ding, L.; Bai, C.; Zhang, H.; Wei, Y. An Adversarial Single-Domain Generalization Network for Fault Diagnosis of Wind Turbine Gearboxes. *J. Mar. Sci. Eng.* **2023**, *11*, 2384. <https://doi.org/10.3390/jmse11122384>

Academic Editor: José António Correia

Received: 12 November 2023

Revised: 10 December 2023

Accepted: 13 December 2023

Published: 18 December 2023



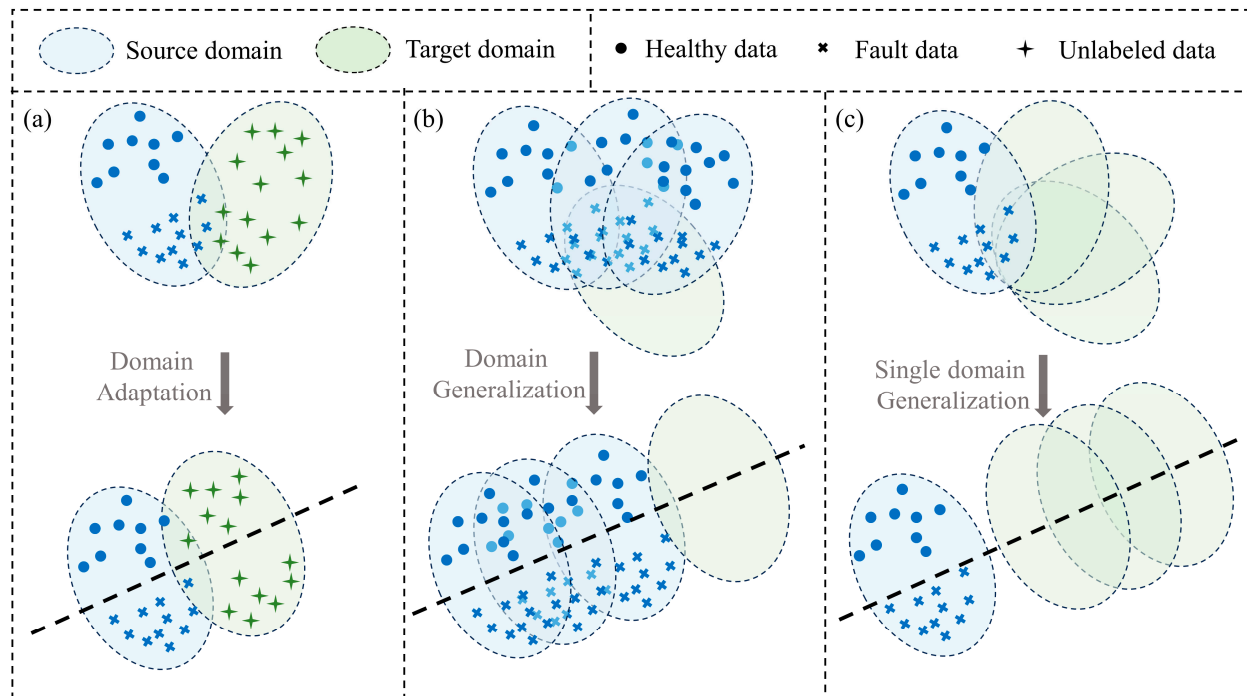
**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The gearbox is a crucial component of the wind turbine (WT) drivetrain system, operating under fluctuating load conditions for extended periods. It is characterized by a high failure rate, extended downtime during failures, and substantial consequences. Gearbox fault diagnosis is a focal point in the maintenance and operation of wind turbines [1,2]. Among various methods, fault diagnosis approaches based on deep learning (DL) have garnered significant attention due to their advantages such as not requiring specialized knowledge and achieving high diagnostic accuracy [3,4]. Zhang et al. proposed a semi-supervised learning model based on a multi-associative layer network and applied it to the fault diagnosis of planetary gearboxes [5]. This method demonstrated significant performance advantages in situations with limited labeled data. However, DL relies on a substantial amount of labeled data for training and assumes that test data and training data originate from the same distribution. In practice, wind turbine gearbox operations are often subject to changing conditions influenced by factors like wind and weather. The data collected frequently deviate from the distribution of training data, significantly diminishing the diagnostic accuracy of DL models under real operational conditions [6–8]. Domain adaptation (DA) is the primary method for addressing data distribution shift and can be employed to achieve high-precision fault diagnosis for WT gearboxes under varying operating conditions [9].

In recent years, DA has found extensive applications in domain shift fault diagnosis tasks [10,11]. Its primary objective is to reduce the distribution distance in high-order subspaces between the source domain and the target domain, extracting domain-invariant features to mitigate the effects caused by domain shift, as illustrated in Figure 1a. However, there are still limitations when applying DA methods to practical WT fault diagnosis tasks. Firstly, DA methods necessitate prior access to target domain data, which remains

challenging for WTs. Additionally, these methods can only adapt to specific target domains, lacking the required generalization capacity for unknown operating conditions. Ensuring robustness in WT fault diagnosis tasks with variable operating conditions proves to be a challenging endeavor [12].



**Figure 1.** The illustration depicts the schematics of DA, DG, and SDG. The distinct domains signify data originating from different operational conditions. (a) DA: extracting transferable features from both the source and target domains to facilitate knowledge transfer. (b) DG: extracting domain-invariant features from multiple source domains, aiming for direct application to an unseen target domain. (c) SDG: harnessing domain-invariant features solely from a single-source domain, eliminating the multi-source domain constraints inherent in DG.

In order to eliminate the dependence on target domain data, fault diagnosis methods based on domain generalization (DG) have emerged as a new research focus. DG methods involve learning from multiple source domains, extracting domain-invariant features, and consequently constructing fault diagnosis models capable of generalizing to unknown target domains, as illustrated in Figure 1b [13–15].

Compared to DA, DG exhibits the capability to generalize to unknown domains. However, DG demands a larger volume of data, necessitating a substantial number of samples from multiple source domains to achieve optimal results. In the context of WTs, collecting comprehensive fault samples from various operating conditions is a time-consuming and labor-intensive process. Therefore, this study explores a single-source domain generalization (SDG) fault diagnosis method, which alleviates the constraints associated with multiple source domains. This approach achieves generalization to unknown domains with training data sourced from a single domain, as depicted in Figure 1c.

In line with the real-time fault diagnosis needs of actual WTs, this study proposes a novel Adversarial single-domain generalization network (ASDGN). The ASDGN primarily consists of a task module and a domain augmentation module. The domain augmentation module is capable of generating augmented data that diverges from the source domain distribution. This augmented data constitute several expanded domains, each exhibiting a shift from the source data. Subsequently, the task module extracts invariant fault features from this data to enhance its generalization capabilities. To ensure that the domain augmentation module can generate data with sufficient distributional disparities, a variational

autoencoder (VAE) is employed to amplify the feature distribution difference between the augmented data and the source data. Concurrently, the Wasserstein distance is used as a metric to constrain the distributions of the two, guaranteeing intra-class compactness and semantic consistency. Through the iterative minimax game between the two, augmented domains are generated, enabling the model to achieve generalization on unknown domains.

The main contributions of this paper can be summarized into the following three points:

(1) We investigated the generalization issue of WT gearbox fault diagnosis under a single-source domain. Most existing studies related to WT gearbox fault diagnosis are conducted under the assumption of multi-source domain data. This paper studies the problem using only single-source domain data, which is more in line with actual operational conditions.

(2) We proposed an adversarial single-domain generalization model (ASDGN), which is applied to the fault diagnosis of WT gearboxes. This model exhibits commendable generalization performance in unknown domains, enabling it to adapt to the changing conditions in wind turbine scenarios.

(3) We conducted experimental validation and comparative experiments using the wind power gearbox dataset and the Southeast University gearbox dataset. The results demonstrate the efficacy and superiority of the proposed model in single-domain generalization problems.

The rest of the paper is structured as follows. Section 2 introduces related work. Section 3 elaborates on the proposed method in detail. Section 4 discusses the experiments and analyzes the results. Section 5 concludes the paper.

## 2. Related Work

Fault diagnosis methods based on DL, such as convolutional neural networks (CNNs) and Long Short-Term Memory (LSTM) networks, have garnered significant attention due to their exceptional feature extraction and fault classification capabilities. They are able to directly distill fault information from raw input data within an end-to-end framework, leading to effective fault diagnosis. Qiu et al. utilized a variety of feature extraction methods to construct a feature space comprising 52 features, which was then subjected to dimensionality reduction. This reduced feature space was employed for fault diagnosis using deep convolutional neural networks [16]. Durbhaka et al. proposed a Long Short-Term Memory (LSTM) network model. This model, optimized through multiple swarm intelligence models, more effectively captures temporal information, thus achieving high-precision gearbox fault classification [17]. Wang et al. employed kernel functions to enhance the capability of feature learning and designed a novel autoencoder known as KAE (Kernel Autoencoder). Subsequently, they constructed a DL model using one KAE and multiple traditional autoencoders to extract deeper-level features [18]. In practical engineering, due to the influence of varying working conditions, there exists domain shift between the source and target domains. DA methods can assist the model in adapting to changes across different data domains, thereby enhancing the accuracy of fault diagnosis. DA methods can generally be categorized into two types: distance metric-based domain adaptation methods and adversarial domain adaptation methods. The former primarily aligns the features of the source and target domains using methods like Maximum Mean Discrepancy (MMD). The latter introduces a domain discriminator to distinguish whether the features extracted by the feature extractor are from the source or target domain. Through adversarial training, the domain discriminator and feature extractor compete against each other, compelling the feature extractor to produce domain-invariant features, thereby enhancing the model's generalization performance in the target domain. Guo and colleagues introduced a novel intelligent fault diagnosis model, termed deep convolutional transfer learning network (DCTLN). This model adopts the minimization of MMD for domain alignment, achieving the objective of cross-domain fault diagnosis [19]. Wan et al. introduced the use of the MMD as a metric criterion. They employed both marginal distribution alignment and conditional

distribution alignment techniques in both the source and target domains. Additionally, they dynamically adjusted the relative importance between these alignment approaches [20]. Chen et al. incorporated a conditional domain discriminator in their model to differentiate whether data originated from the source domain or the target domain. Through adversarial learning between the feature extractor and the conditional domain discriminator, this approach guided the feature extractor to extract domain-invariant features [21]. An et al. introduced a domain adaptation network based on contrastive learning (DACL). It achieves domain-level MMD and conditional distribution alignment separately for the source and target domains. Additionally, they introduced an adaptive factor to dynamically adjust the relative importance of these alignment methods [22].

While DA methods have demonstrated excellent performance in fault diagnosis, they are still constrained by the availability of target domain data, making it challenging to adapt to online fault diagnosis tasks. DG fault diagnosis methods that do not require access to target domain data are gaining increasing attention as a research hotspot. Han et al. enhanced the clustering of similar faults across different domains and improved the model's domain robustness by minimizing the triplet loss [23]. Shi et al. combined multi-source augmentation with domain adversarial techniques to enhance the model's generalization ability by extracting domain-invariant features from multiple source domains [24]. Fan et al. introduced a deep mixed and domain generalization network (DMDGN) for intelligent fault diagnosis. This model adjusts inter-class and intra-class distances across multiple source domains through adversarial and divergence metric approaches. Experimental results demonstrate its strong generalization capabilities on unknown domains [25]. Additionally, domain generalization methods also encompass regularization-based DG, meta-learning-based DG, and representation learning-based DG [26–31]. These approaches have garnered extensive attention across various domains, including computer vision, natural language processing, and intelligent fault diagnosis.

However, the previously mentioned DG methods typically necessitate labeled data from multiple source domains, which still pose challenges in the context of WT fault diagnosis. This paper's primary focus lies in investigating the issue of fault diagnosis based on single-source domain generalization, with the aim of breaking free from the stringent reliance on data from multiple source domains and better catering to the requirements of WT fault diagnosis.

### 3. Proposed Method

#### 3.1. Problem Definition

This study aims to delve into the fault diagnosis techniques of WT gearboxes under various unknown operational conditions. The core objective is to develop a training model based on data from a single operational condition, ensuring its superior generalization performance across multiple conditions.

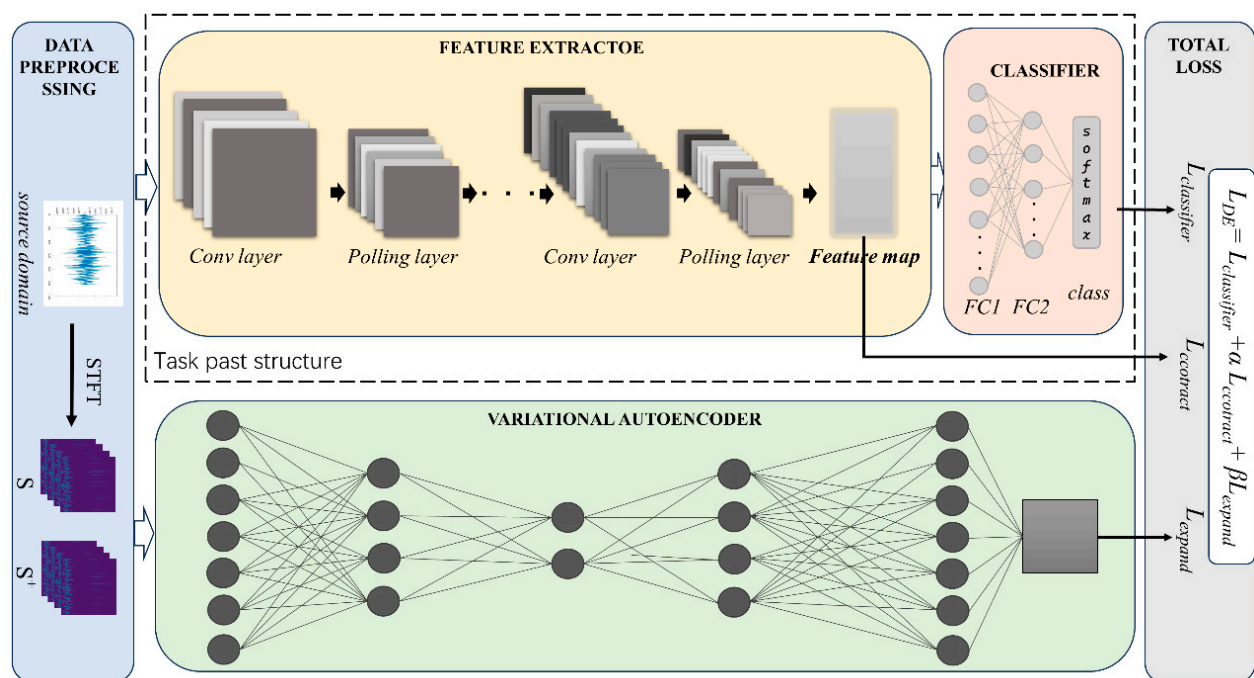
Detailed problem description: In the source domain  $D_s^j = \{[(x_{si}, y_{si})]_{i=1}^m\}_{j=1}^n$ , there are  $j$  sub-domains. Each sub-domain contains  $m$  labeled data points, with each source domain following its distinct joint distribution  $P^j(x_s^j, y_s^j)$ . For the target domain  $D_t^k = \{[(x_{tl}, y_{tl})]_{l=1}^m\}_{k=1}^n$ , the data and labels are inaccessible during training. Here,  $k$  represents the total number of target domains, with each having its unique joint distribution  $Q^k(x_t^k, y_t^k)$ . While there are variations in data feature distributions across domains, they all share a common label space. The DG method establishes a generalized model by training from multiple source domains  $D_s^j$ , enabling its application for fault diagnosis in the unknown target domain  $D_t^k$ . The main goal of the ASDGN is, given only data from a single-source domain, to train a DL model  $y = f(x)$  using labeled data from  $D_s^{j=1}$ , thereby minimizing the risk  $\varepsilon_t = E_{(x_{tl}, y_{tl}) \sim Q^k} [f(x_t^k) \neq y_t^k]$  in the target domain. Table 1 offers a comparison between this study's problem and past research studies in terms of problem settings and scope of application.

**Table 1.** Comparison of the research problem addressed in this study with previous research in terms of problem settings and the scope of application. Note: This table lists the training conditions and applicability scope of different methods. In training conditions, “✓” indicates that the data is required, and “✗” means the data is not needed. In target conditions, “✓” signifies that the model is applicable to that condition, and “✗” suggests it is not suitable for that condition.

Method	Training Conditions					Target Conditions		
	Source Domain Data	Source Domain Labels	Target Domain Data	Target Domain Labels	Multi-Source Domain for Training	Same Distribution between the Source and Target Domains	Different Distribution between the Source and Target Domains	Multi-Target Domains
DL	✓	✓	✗	✗	✗	✓	✗	✗
DA	✓	✓	✓	✗	✗	✓	✓	✗
DG	✓	✓	✗	✗	✓	✓	✓	✓
SDG	✓	✓	✗	✗	✗	✓	✓	✓

### 3.2. Structure

Figure 2 illustrates the overall structure of the ASDGN proposed in this study, which primarily consists of the task module and domain expansion module. The task model comprises a feature extractor  $F$ , defined by  $Z = F(X)$ , which encodes the model input  $X$  into feature vector  $Z$ . The classifier  $C$ , defined by  $Y = C(Z)$ , predicts the label  $Y$  through the feature vector  $Z$ . The domain expansion model, in addition to the feature extractor  $F$  and classifier  $C$ , also includes a variational autoencoder  $E$ , represented as  $X^R = E(X)$ , used for data reconstruction. Let  $x^+ \in S^+$  and  $x \in S$ . Here,  $S$  and  $S^+$  denote the source domain and the expanded domain, respectively.



**Figure 2.** Overall architecture of the ASDGN model.

For the task model, our goal is for the feature extractor  $F$  to extract domain-invariant feature vectors  $Z$  from the source domain  $S$ , which can then be accurately classified by the classifier  $C$ . Thus, we employ adversarial training techniques to generate several augmented domains  $S^+$ , with substantial domain shift [32,33]. Subsequently, we learn

domain-agnostic features from the source domain and these augmented domains to train a domain-generalized model, which is

$$\min_{\omega} E[L_{\text{classifier}}(\omega; D)] \quad (1)$$

where  $\omega$  represents the model parameters for the feature extractor and classifier.  $D = \{S, S_1^+, \dots, S_K^+\}$  denotes the set comprising the source domain and all augmented domains. In the equation above,  $L_{\text{classifier}}$  can be computed using the cross-entropy function.

$$L_{\text{classifier}}(y_i, \hat{y}_i) = -\sum_i^k y_i \log(\hat{y}_i) \quad (2)$$

where  $\hat{y}$  represents the softmax output of the model, with  $\hat{y}_i$  denoting the  $i$ th dimension of  $\hat{y}$ .  $y$  stands for the one-hot representation of the model's true label, and  $y_i$  indicates the  $i$ th dimension of  $y$ .  $K$  signifies the number of classes.

The domain expansion module functions to generate several expanded domains from the source domain, each having a distribution distinct from the source, to simulate unknown domains. In this process, we employ variational autoencoders (VAE) to amplify domain shift [34]. To prevent the expanded domains from diverging, a semantic consistency constraint is also introduced. This task is accomplished by the feature extractor  $F$ . The overall loss function for expanded domain generation can be expressed as

$$L_{DE} = L_{\text{classifier}} - \alpha L_{\text{contract}} + \beta L_{\text{expand}} \quad (3)$$

where  $L_{\text{classifier}}$  represents the classification loss and can be expressed as Equation (2),  $L_{\text{contract}}$  signifies the semantic consistency constraint and its specific implementation will be detailed in Section 3.3,  $L_{\text{expand}}$  induces domain shift and its specific implementation will be detailed in Section 3.4. The parameters  $\alpha$  and  $\beta$  are hyperparameters that balance the training. Through adversarial training between VAE and  $F$ , the objective is to generate expanded domains with the maximum domain divergence from the source domain, thereby enhancing the model's generalization capabilities in unknown domains.

Given the target function, expanded data  $x^+$  is generated in the expanded domain  $S^+$  using gradient ascent:

$$x_{t+1}^+ = x_t^+ + \mu \nabla_{x_t^+} L_{DE} \quad (4)$$

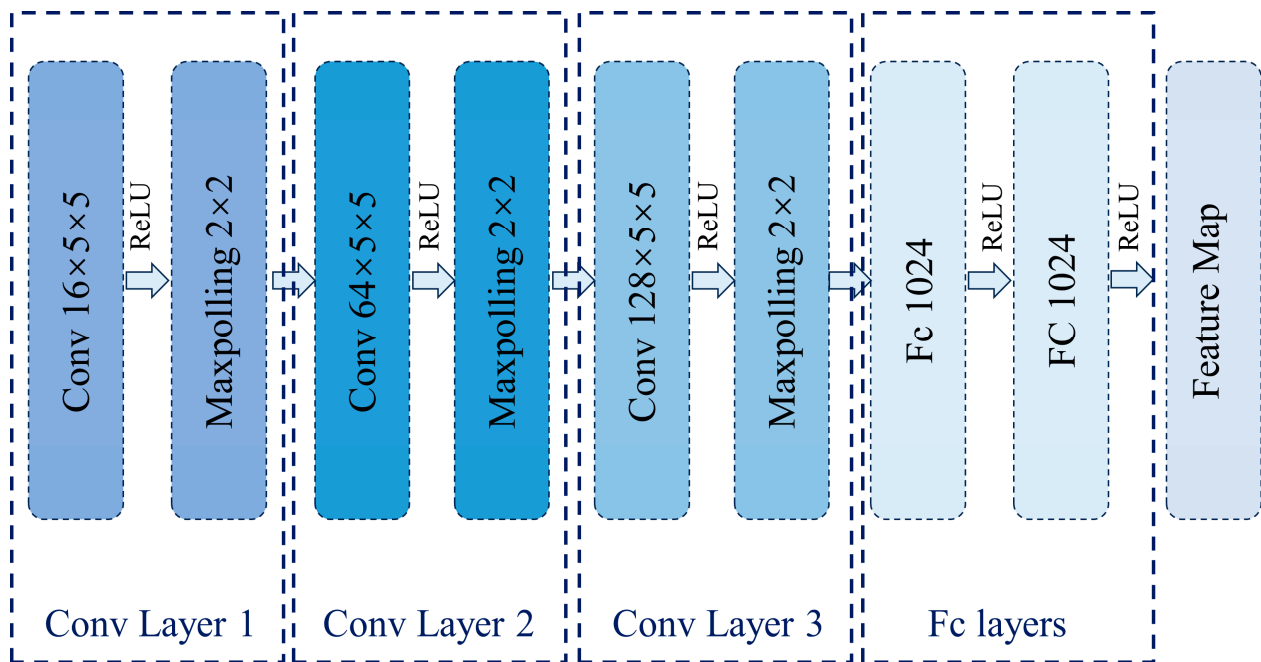
where  $\mu$  represents the learning rate. Ideal adversarial samples are generated through multiple iterations. The configuration and detailed implementation of each sub-module will be discussed subsequently.

### 3.3. Feature Extractor

Figure 3 depicts the architectural configuration of the feature extractor  $F$  [35–39]. This section primarily comprises convolutional layers and fully connected layers, ultimately producing a  $1024 \times 1$  feature vector  $Z$ . During the generation of domain-augmented data, we constrain the domain shift by measuring the Wasserstein distance between  $z$  and  $z^+$ , thereby ensuring semantic consistency.

$$L_{\text{contract}} = \|z - z^+\|_2 + \infty \cdot 1\{y \neq y^+\} \quad (5)$$

where,  $z, z^+ \in Z$ ;  $z = F(x)$  and  $z^+ = F(x^+)$ .  $\|\cdot\|$  represents the Wasserstein distance between the source domain and the augmented domain; ensures that the data originates from the same fault type.  $L_{\text{contract}}$  constrains the conditional distribution difference between the source and augmented domains, preventing the generation of ineffective samples in the augmented domain. However, this also limits the domain shift distance. Therefore,  $L_{\text{expand}}$  is introduced to amplify the domain shift, and its specific definition will be elaborated in Section 3.4.



**Figure 3.** Structure diagram of the feature extractor  $F$ .

### 3.4. Variational Autoencoder

The VAE consists of an encoder and a decoder [40–42]. The encoder maps the input data  $x$  into a latent space  $z$ , with the posterior distribution of  $z$  represented as  $q(z|x)$ . Unlike the autoencoder, the VAE maps the distribution parameters of  $z$  onto a standard distribution matching the input data, generally a Gaussian distribution. The decoder then samples from the reconstructed  $z$  to generate new samples  $\hat{x}$ . The prior probability is  $p(z)p(\hat{x}|z)$ . The objective of the VAE is to make these two distributions as close as possible, with their closeness measured by the KL divergence between them:

$$D_{KL}[q(z|x)||p(z|x)] = E_{q(z|x)} \log \frac{q(z|x)}{p(z|x)} \quad (6)$$

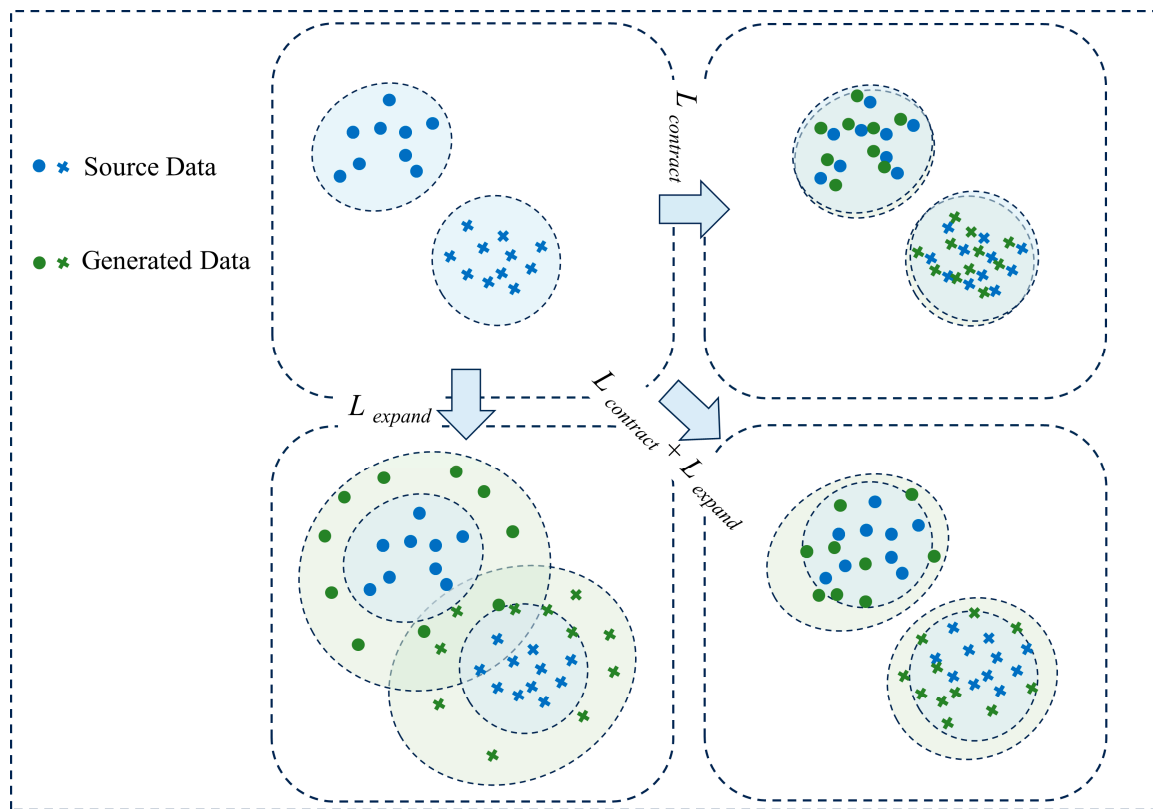
As the KL divergence is always non-negative, the VAE model can be optimized using the following equation:

$$\min_{\theta} \{-D_{KL}(q(z|x)||p(z)) + E[\log p(x|z)]\} \quad (7)$$

Utilizing the source domain data to pre-train the VAE, its parameters are then frozen. Subsequently, it is used to maximize the domain distance between the generated data and the original data:

$$L_{\text{expand}} = ||x^+ - p(q(x^+))||^2 \quad (8)$$

As illustrated in Figure 4, during this process,  $L_{\text{expand}}$  aims to expand the augmented domain, while  $L_{\text{contract}}$  attempts to contract it. Through adversarial training of the two, we acquire a generalized dataset to train a robust generalization model.



**Figure 4.** Schematic illustration of the effects of  $L_{expand}$  and  $L_{contract}$  in data generation.

### 3.5. Model Optimization

In the pre-training of VAE, Adam is employed to optimize the objective in Equation (7), leading to the following parameter update:

$$\theta_{t+1} = \theta_t - v \left( -\frac{\partial D_{KL}(q(z|x)||p(z))}{\partial \theta_t} + \frac{\partial E[\log p(x|z)]}{\partial \theta_t} \right) \quad (9)$$

where  $v$  represents the initial learning rate for pre-training, and  $\theta$  denotes the parameters of VAE.

Subsequently, based on Equations (3) and (4), augmented data is generated in the expanded domain. We employ SGD as the optimizer for adversarial training, and the parameter updates for the augmented data are as follows:

$$x_{t+1}^+ = x_t^+ + \mu \left( \frac{\partial (\text{Cross\_entropyLoss}(C(F(x^+), y)))}{\partial x_t^+} - \alpha \frac{\partial (\text{MSE}(F(x_t^+), F(x) + \infty \cdot 1\{y \neq y^+\})}{\partial x_t^+} + \beta \frac{\partial (\text{MSE}(x^+ - p(q(x))))}{\partial x_t^+} \right) \quad (10)$$

where  $\mu$  is the initial learning rate for domain expansion, and  $x_t^+$  represents the parameters of the augmented data.

Lastly, using Adam to optimize the parameters  $\omega$  within the task model, with the optimization target given by Equation (2), the specific representation for the parameter update of the task model is

$$\omega_{t+1} = \omega_t - \rho \left( \frac{\partial (\text{Cross\_entropyLoss}(C(F(x'), y)))}{\partial \omega_t} \right) \quad (11)$$

where  $\rho$  is the initial learning rate. In summary, the training process of the ASDGN is summarized in Algorithm 1.

**Algorithm 1.** ASDGN

---

```

#Pre-train VAE
Input: Source dataset  $S = \{x_i, y_i\}_{i=1}^n$ ; VAE model  $E(\theta; X)$ ; pre-train epoch E1
for i = 1 to E1 do:
    Randomly sample from S
    Forward propagation and calculation Equation (7)
    Backward propagation to update  $\theta$  by Equation (9)
end
Return: pre-trained VAE model
#Domain Augmentation
Input: Source dataset  $S = \{x_i, y_i\}_{i=1}^n$ ; pre-trained VAE model  $E(\theta; X)$ ; pre-trained feature extractor F;
classifier C; number of augmentation domains K; Adversarial train epoch E2
for i = 1 to K do:
    for i = 1 to E2 do:
        Randomly sample m data  $X = \{x_i\}_{i=1}^m$  from S;  $X^+ == X.clone()$ 
        Forward propagation and calculation Equations (2), (5) and (8)
        Calculation Equation (3)
        Backward propagation to update  $X^+$  by Equation (9)
    end
    Create  $S^+ = \{x_i^+\}_{i=1}^n$ 
end
Return:  $D = \{S, S_1^+, \dots, S_K^+\}$ 
#Domain Augmentation
Input: Dataset  $D = \{S, S_1^+, \dots, S_K^+\}$ ; feature extractor F; classifier C; Task model train epoch E3.
for i = 1 to E3 do:
    Randomly sample data from S
    Forward propagation and calculation Equation (2)
    Backward propagation to update  $\omega$  by Equation (11)
end
Return: Task model

```

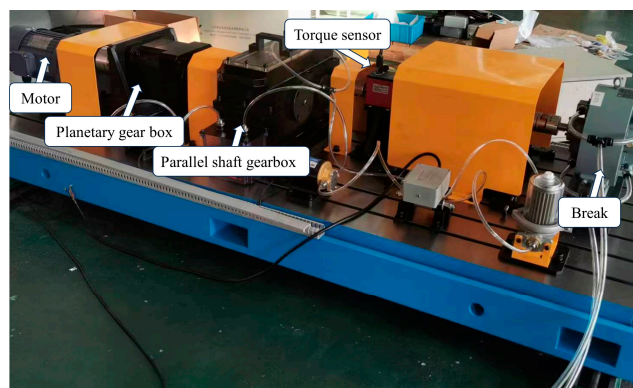
---

**4. Experiments**

To validate the effectiveness and superiority of the proposed method, experiments were conducted on two datasets and compared with other state-of-the-art fault diagnosis models. Additionally, the roles of each module were verified through ablation studies.

**4.1. Dataset Description**

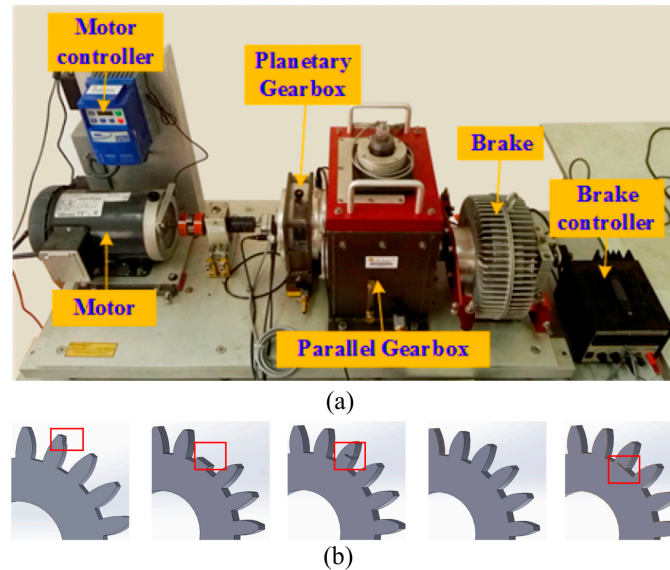
1. Dataset 1: This dataset originates from a wind turbine gearbox fault simulation test rig, and the rig's structure is depicted in Figure 5. The dataset encompasses four bearing health conditions under four loads: 0, 2, 4, and 8. The faults include Normal (N), Inner Race Fault (IR), Ball Fault (B), and Outer Race Fault (OR). The data sampling frequency is 20 kHz. This dataset provides an effective means to verify the robustness of the model.



**Figure 5.** Wind power gearbox fault testbench structure diagram.

2. Dataset 2 [43]: The gear fault data is collected from the gearbox fault simulation test rig, as shown in Figure 6a. It includes two conditions of speed–load, 20–0 and 30–2.

Under each condition, there are five types of gear fault states, health, chipped, miss, root, and surface, as illustrated in Figure 6b. The data sampling frequency is 5120 Hz. The two conditions in this dataset have significant differences, which effectively validates the generalization performance of the model.



**Figure 6.** (a) shows the structure diagram of the gearbox fault simulation testbench, and (b) illustrates the five fault states. From left to right, they are surface, miss, chipped, health, and root.

Each health condition contains 1000 training samples and 300 test samples. Each sample contains 1024 sampling points. The samples are converted into time–frequency images through continuous wavelet transformation. A total of six fault diagnosis tasks were set up to evaluate the generalization ability and robustness of the ASDGN in real-time fault diagnosis. The task settings are shown in Table 2.

**Table 2.** Task setting.

Task	Train Data		Text Data	
	Condition	Number	Condition	Number
B1	0 (Nm)	$1000 \times 4 = 4000$	0 Nm, 2 Nm, 4 Nm, 8 Nm	$300 \times 4 \times 4 = 4800$
B2	2 (Nm)	$1000 \times 4 = 4000$		
B3	4 (Nm)	$1000 \times 4 = 4000$		
B4	8 (Nm)	$1000 \times 4 = 4000$		
G1	20–0 (rpm-V)	$1000 \times 5 = 5000$	20 rpm–0 V, 30 rpm–2 V	$300 \times 5 \times 2 = 3000$
G2	30–2 (rpm-V)	$1000 \times 5 = 5000$		

#### 4.2. Comparison Experiment

To evaluate the performance of the ASDGN, we selected several outstanding fault diagnosis models for comparison.

CNN (baseline): the CNN applies the model trained on source domain data directly to the target domain data.

DT-DDG [44]: DT-DDG integrates adversarial training with dynamic weighting and batch spectral penalization regularization to extract domain-invariant and discriminative features from multiple source domains. It is an effective domain generalization (DG) method.

L2A-OT [45]: L2A-OT is a DG method based on domain augmentation. It enhances the model's generalizability by establishing a pseudo source domain through divergence maximization.

AMInet [46]: AMInet is an adversarial domain generalization network guided by mutual information (MI). It achieves domain generalization through adversarial training by maximizing the lower bound of MI and minimizing the upper bound of MI.

RTDGN [47]: RTDGN is another novel DG method. It enhances the network's generalization capabilities by mining domain-invariant information through two phases, task-agnostic domain adaptation and task-related domain generalization.

To ensure fairness in the experiments, the input data and hyperparameter settings for all comparison methods were set according to the requirements of the original papers. Additionally, adjustments were made to the experimental settings to ensure satisfactory results for our specific scenario. All experiments were conducted on a GPU with Nvidia 3070Ti, and PyTorch was used as the programming platform. The selection of hyperparameters can be guided by referencing the hyperparameter configurations of baseline models and similar tasks. Fine-tuning adjustments are then made to ensure the model converges to a satisfactory state. In the method proposed in this paper, the number of epochs for pre-training is  $E1 = 25$ , for adversarial training is  $E2 = 20$ , and for task model training it is  $E3 = 200$ . For the comparison methods, the number of training batches for the task model is also set to 200. The batch size is uniformly set to 32. To eliminate the impact of randomness, the results are presented as the average of 20 trials [48].

#### 4.3. Experimental Results

The comparative experimental results on Dataset 1 are shown in Figure 7 and Table 3. Based on the experimental results, the following conclusions can be drawn: The CNN, which does not incorporate domain generalization structure, has the lowest average accuracy of only 78.37%. This is due to the difference in distribution between the test and training data, leading to reduced model generalization performance. Our proposed model demonstrated a notable improvement over the baseline, achieving a substantial accuracy boost of 13.58%. This underscores the model's effectiveness in addressing domain generalization challenges. In the four domain generalization tasks, the model proposed in this paper consistently achieves high accuracy, indicating its robustness under different data distributions. When compared to other domain generalization models, it was observed that, compared to the baseline, both L2A-OT and AMInet effectively enhance accuracy by 6.98% and 10.45%, respectively. However, their performances are slightly inferior to the model proposed in this paper, suggesting that the VAE-based domain augmentation method is marginally superior in exploring unknown domains compared to the maximum divergence-based domain augmentation method and MI-based domain generalization method.

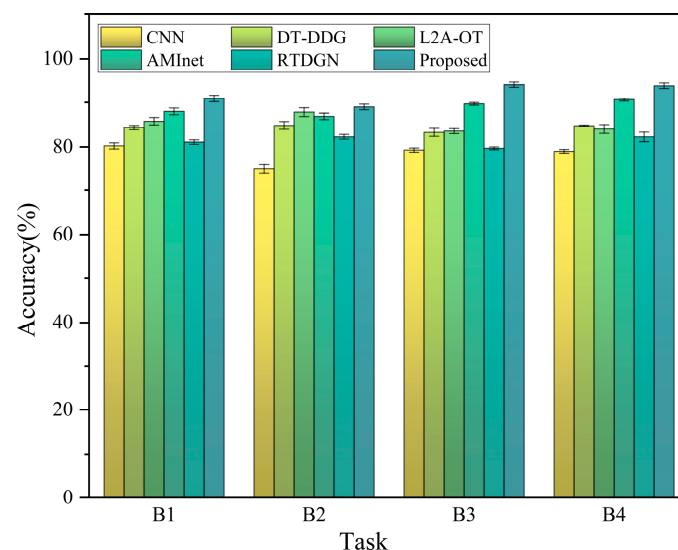


Figure 7. Histogram of the experimental results for Dataset 1.

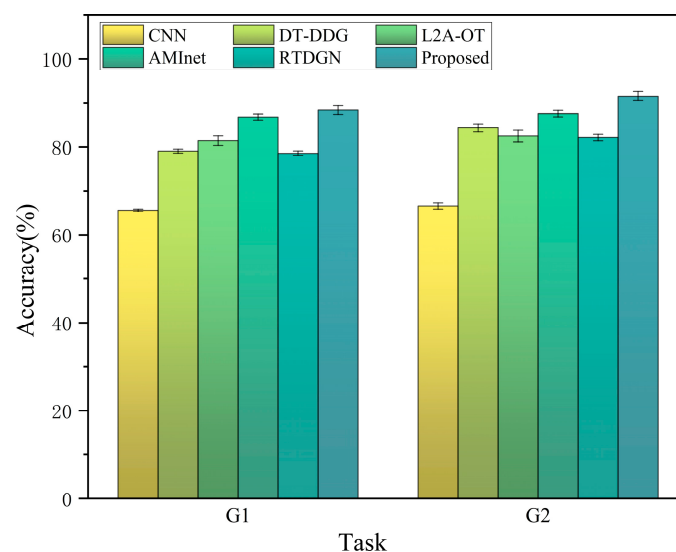
**Table 3.** Dataset 1 experimental results (accuracy %).

Method	CNN	DT-DDG	L2A-OT	AMInet	RTDGN	Proposed
B1	80.25 $\pm$ 0.71	84.42 $\pm$ 0.49	85.84 $\pm$ 0.82	88.02 $\pm$ 0.73	81.10 $\pm$ 0.51	<b>90.90 <math>\pm</math> 0.73</b>
B2	74.94 $\pm$ 0.96	84.91 $\pm$ 0.87	87.85 $\pm$ 0.95	86.91 $\pm$ 0.71	82.29 $\pm$ 0.55	<b>88.99 <math>\pm</math> 0.62</b>
B3	79.28 $\pm$ 0.49	83.32 $\pm$ 0.91	83.61 $\pm$ 0.57	89.69 $\pm$ 0.32	79.70 $\pm$ 0.33	<b>94.07 <math>\pm</math> 0.59</b>
B4	79.00 $\pm$ 0.43	84.83 $\pm$ 0.16	84.08 $\pm$ 1.00	90.64 $\pm$ 0.32	82.29 $\pm$ 1.08	<b>93.83 <math>\pm</math> 0.63</b>
Average	78.37	84.37	85.35	88.82	81.35	<b>91.95</b>

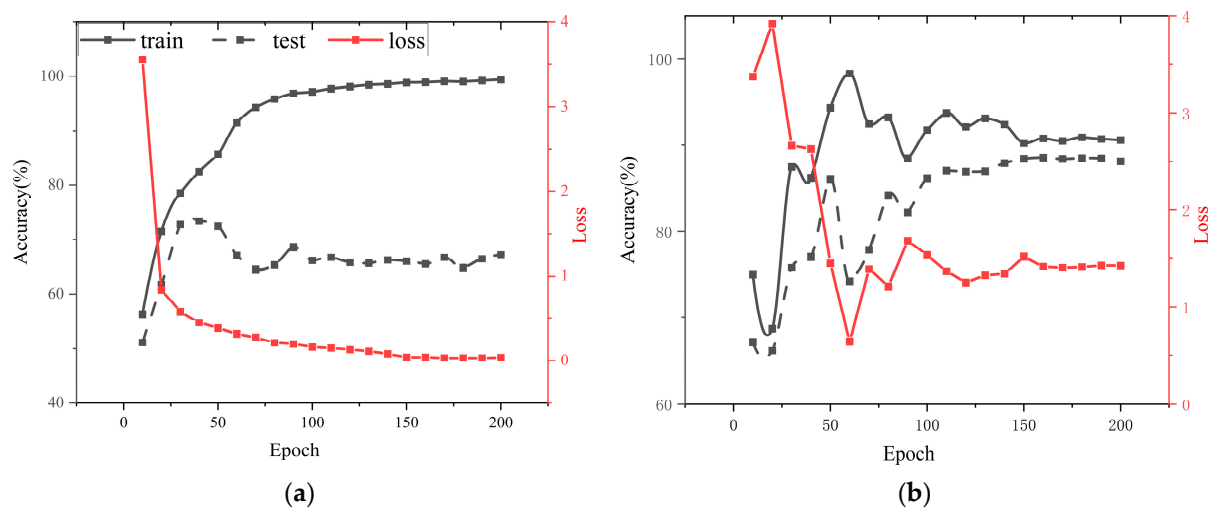
The results of the aforementioned methods on Dataset 2 are shown in Table 4 and Figure 8. It can be observed that the accuracy of the CNN on this dataset is only 66.17%. This is because the two conditions in Dataset 2 differ in both load and speed, resulting in a greater distribution difference, thereby increasing the generalization difficulty. Nevertheless, the DMDGN model proposed in this study still maintains an average accuracy of 90.04%, demonstrating its ability to maintain good generalization performance even in complex tasks.

**Table 4.** Dataset 2 experimental results (accuracy %).

Method	CNN	DT-DDG	L2A-OT	AMInet	RTDGN	Proposed
G1	65.66 $\pm$ 0.29	79.06 $\pm$ 0.46	81.42 $\pm$ 1.07	86.86 $\pm$ 0.70	78.51 $\pm$ 0.56	<b>88.43 <math>\pm</math> 1.01</b>
G2	66.67 $\pm$ 0.72	84.35 $\pm$ 0.95	82.46 $\pm$ 1.31	87.63 $\pm$ 0.74	82.11 $\pm$ 0.74	<b>91.65 <math>\pm</math> 1.08</b>
Average	66.17	81.71	81.94	87.25	80.31	<b>90.04</b>

**Figure 8.** Histogram of the experimental results for Dataset 2.

Taking task G2 in Dataset 2 as an example, Figure 9a and Figure 9b, respectively, illustrate the iteration patterns of the CNN and ASDGN during training. From the figures, it can be observed that as the iterations progress, the accuracy rate of the CNN on the training set incrementally rises. However, on the test set, after an initial rise, the accuracy starts to decline and stabilizes after 50 epochs. This may be attributed to the model's inadequacy in extracting domain-invariant features, leading to overfitting. In contrast, for the ASDGN, even though the training speed slightly decelerates due to the introduction of the expanded domain, it demonstrates stability after 100 epochs. Notably, its accuracy on the test set closely matches that of the training set, indicating that the ASDGN is more effective in extracting domain-invariant features, thereby enhancing its generalization capabilities in unknown domains.



**Figure 9.** Taking task G2 as an example, this figure contrasts the iterative curves of the baseline and proposed models. (a) The graph depicts the iterative convergence curve of the CNN; (b) the graph depicts the iterative convergence curve of the ASDGN.

In addition, we also compared the average training times of different models, as detailed in Table 5. It is observed that the ASDGN incurs a longer training time due to the necessity of domain expansion. However, considering its robust generalization performance across multiple target domains, the additional training time is acceptable.

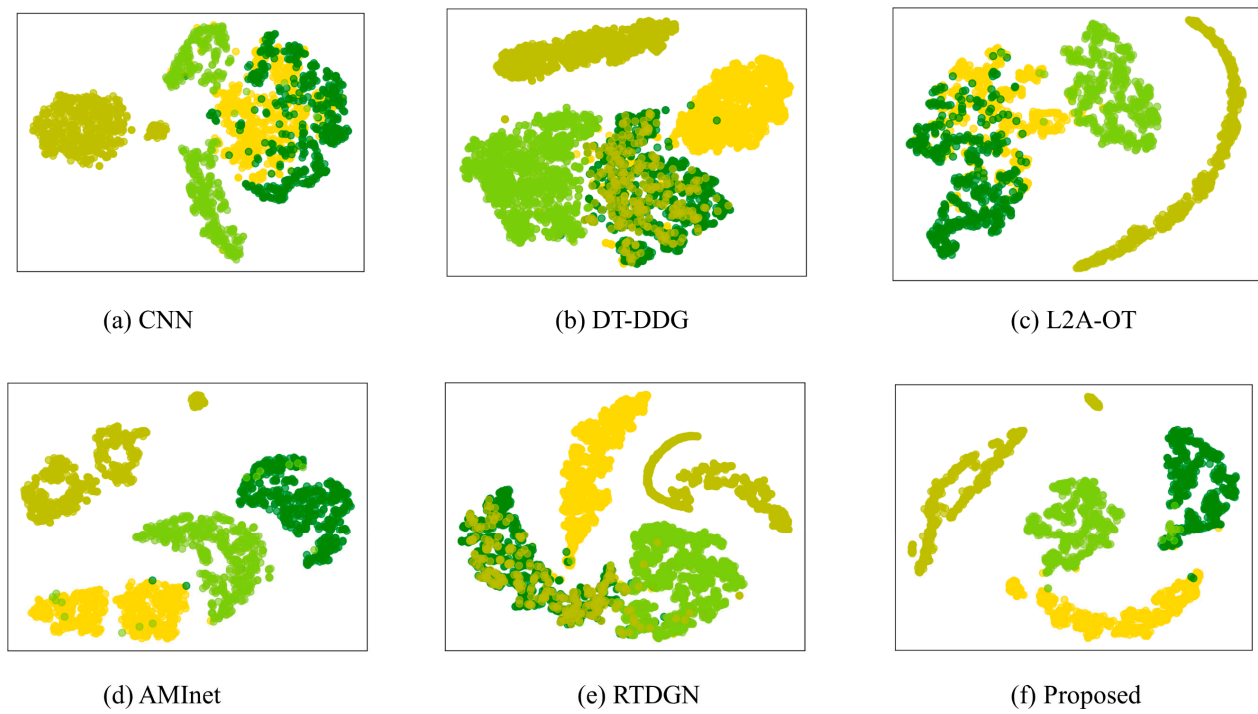
**Table 5.** The computation times of different models (taking Dataset 1 as an example).

Model	Training Time (s)
CNN	219.37
DT-DDG	311.54
L2A-OT	392.78
AMInet	413.49
RTDGN	288.15
<b>Proposed</b>	<b>506.42</b>

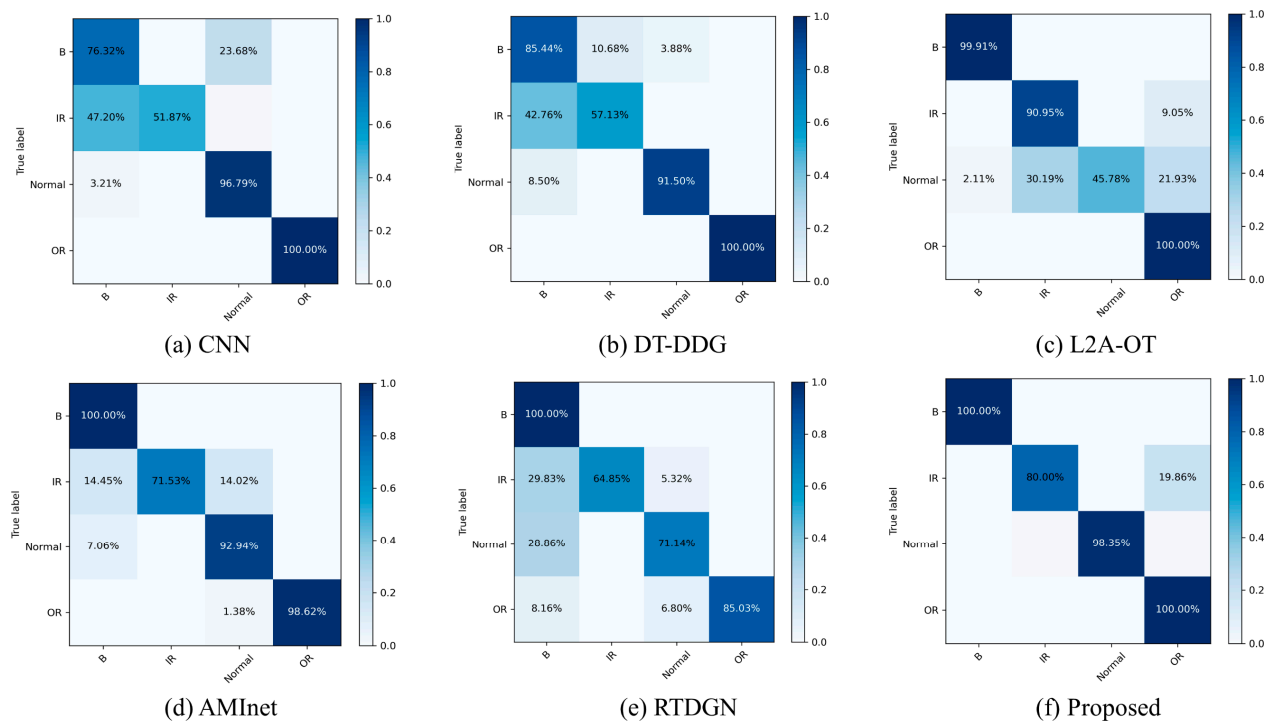
#### 4.4. Visualization Analysis

To intuitively observe the efficacy of the ASDGN in fault diagnosis, we employed the T-SNE method to visualize the feature layer. Using B3 as a reference, the feature visualization results are shown in Figure 10, with its corresponding confusion matrix presented in Figure 11.

As observed from Figure 10f, the features extracted by the ASDGN demonstrate commendable clustering and distinct inter-class boundaries. On the contrary, Figure 10a reveals that the CNN has ambiguous inter-class distances, with some features even showing signs of confusion, making accurate classification challenging. This vividly underscores the robust generalization capability of the ASDGN across multiple unknown domains.



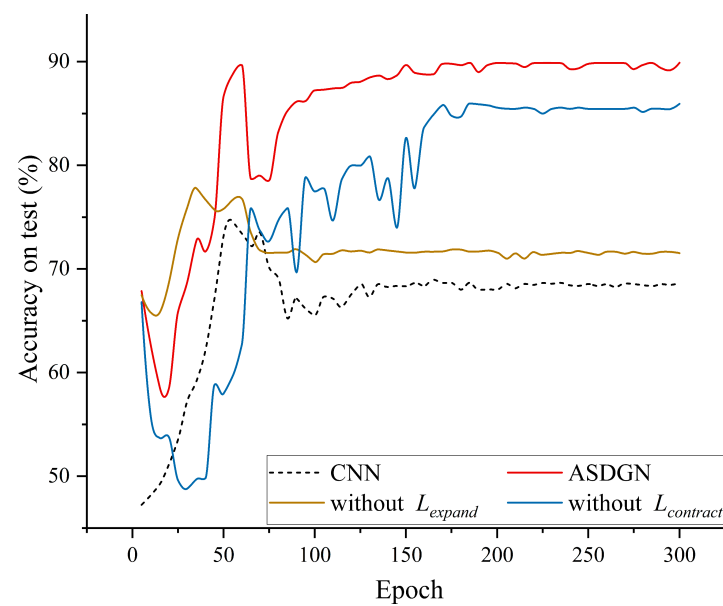
**Figure 10.** Taking B3 as an example, the figure shows the feature visualization of the feature layer for each method.



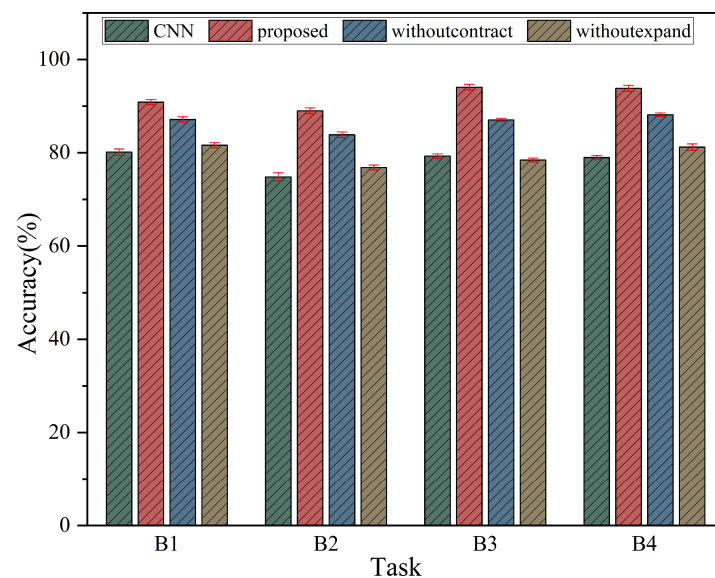
**Figure 11.** Taking B3 as an example, the figure shows the confusion matrix results for each method.

#### 4.5. Ablation Experiments

To investigate the roles of  $L_{contract}$  and  $L_{expand}$  during the domain expansion model optimization process, we conducted ablation experiments to compare the effects on model classification accuracy and the training process when either  $L_{contract}$  or  $L_{expand}$  are omitted. The experimental results are shown in Figures 12 and 13.



**Figure 12.** Taking G1 as an example, we compare the iterative processes of the CNN and ASDGN without  $L_{contract}$ , or  $L_{expand}$ .



**Figure 13.** A comparison of the accuracy rates for the CNN and ASDGN without  $L_{contract}$  or  $L_{expand}$  on Dataset 1.

From the results, it is evident that the accuracy of the ASDGN without  $L_{expand}$  drops significantly. This is because, although the expanded domain increases the number of samples, it does not cover the unknown domain, resulting in an inadequate generalization capability. The ASDGN without  $L_{contract}$  has slightly lower accuracy than the standard ASDGN and shows poorer stability during iterations. This is attributed to the absence of  $L_{contract}$ 's constraint, leading to the inclusion of some faulty samples in the generated samples. Additionally, it is worth noting that since the ASDGN pre-trains the feature extractor and classifier during the domain expansion phase, its initial accuracy when training the task model is higher than that of a CNN model trained from scratch.

## 5. Conclusions

This paper focused on the fault diagnosis of wind turbine gearboxes, considering the variability of operating conditions and the challenges in data acquisition. A single-source domain generalization model based on adversarial training was proposed. This model can be trained using labeled data from a single operating condition and can generalize across multiple unknown domains. The model employs adversarial training between the VAE and feature extractor to generate a series of expanded domains. From these domains, generalized fault features are extracted, enhancing the diagnostic accuracy in unknown domains. Extensive experiments on two datasets demonstrated the model's efficacy in improving fault diagnosis accuracy in unseen domains. Compared to the CNN, the proposed method boosted accuracy by 14.83% and 23.87% on the bearing and gear datasets, respectively. Through comparative experiments, it can be discerned that, in comparison to domain generalization methods widely employed in other scenarios, our model exhibits superior accuracy and robustness. Moreover, ablation studies further elucidate the influence and significance of the two modules in adversarial training. As the training process does not require access to target domain data and the model exhibits strong generalization across multiple unknown domains, it holds great potential for real-time fault diagnosis in wind turbine systems. In future endeavors, we aim to employ more intricate neural network architectures as feature extractors, integrating components such as self-attention mechanisms and recurrent neural networks to adapt to increasingly complex input scenarios. Furthermore, we will explore the incorporation of meta-learning techniques during the training phase to augment the model's generalization capabilities.

**Author Contributions:** Conceptualization, X.W. and H.L.; methodology, X.W.; software, X.W.; validation, X.W., C.W. and C.B.; formal analysis, C.Z.; investigation, Z.F.; resources, L.D.; data curation, C.W.; writing—original draft preparation, X.W.; writing—review and editing, H.L.; visualization, X.W.; supervision, C.B.; project administration, Y.W.; funding acquisition, H.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the Natural Science Foundation of China (Grant No. 52271303), the Dalian Science Technology Innovation Fund (Grant No. 2019J12GX023), the China Postdoctoral Science Foundation (Grant No. 2023M730454), the Fundamental Research Funds for the Central Universities (Grant No. 3132023627), the Science and Technology Innovation Fund of Dalian (Grant No. 2021JJ11CG004, 2022JJ11CG010), the Cultivation Program for the Excellent Doctoral Dissertation of Dalian Maritime University, and Innovative Projects for the Application of Advance Research on Equipment (62602010210).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data that support the findings of this study are available from the corresponding author upon reasonable request.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Zhang, F.H.; Chen, M.S.; Zhu, Y.Z.; Zhang, K.; Li, Q.A. A Review of Fault Diagnosis, Status Prediction, and Evaluation Technology for Wind Turbines. *Energies* **2023**, *16*, 1125. [\[CrossRef\]](#)
2. Liu, Z.P.; Zhang, L. A review of failure modes, condition monitoring and fault diagnosis methods for large-scale wind turbine bearings. *Measurement* **2020**, *149*, 22. [\[CrossRef\]](#)
3. Nguyen, C.D.; Prosvirin, A.; Kim, J.M. A reliable fault diagnosis method for a gearbox system with varying rotational speeds. *Sensors* **2020**, *20*, 3105. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Saucedo-Dorantes, J.J.; Delgado-Prieto, M.; Osornio-Rios, R.A.; Romero-Troncoso, R.D.J. Diagnosis methodology for identifying gearbox wear based on statistical time feature reduction. *Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci.* **2018**, *232*, 2711–2722. [\[CrossRef\]](#)
5. Zhang, K.; Tang, B.; Qin, Y.; Deng, L. Fault diagnosis of planetary gearbox using a novel semi-supervised method of multiple association layers networks. *Mech. Syst. Signal Process.* **2019**, *131*, 243–260. [\[CrossRef\]](#)

6. Hu, R.H.; Zhang, M.; Xiang, Z.Y.; Mo, J.L. Guided deep subdomain adaptation network for fault diagnosis of different types of rolling bearings. *J. Intell. Manuf.* **2023**, *34*, 2225–2240. [\[CrossRef\]](#)
7. Wu, H.; Li, J.M.; Zhang, Q.Y.; Tao, J.X.; Meng, Z. Intelligent fault diagnosis of rolling bearings under varying operating conditions based on domain-adversarial neural network and attention mechanism. *ISA Trans.* **2022**, *130*, 477–489. [\[CrossRef\]](#)
8. Wang, Q.B.; Xu, Y.B.; Yang, S.K.; Chang, J.T.; Zhang, J.G.; Kong, X.G. A domain adaptation method for bearing fault diagnosis using multiple incomplete source data. *J. Intell. Manuf.* **2023**, 1–15. [\[CrossRef\]](#)
9. Zhang, S.Y.; Su, L.; Gu, J.F.; Li, K.; Zhou, L.; Pecht, M. Rotating machinery fault detection and diagnosis based on deep domain adaptation: A survey. *Chin. J. Aeronaut.* **2023**, *36*, 45–74. [\[CrossRef\]](#)
10. Tian, M.; Su, X.M.; Chen, C.Z.; Luo, Y.Q.; Sun, X.M. Bearing fault diagnosis of wind turbines based on dynamic multi-adversarial adaptive network. *J. Mech. Sci. Technol.* **2023**, *37*, 1637–1651. [\[CrossRef\]](#)
11. Li, D.D.; Zhao, Y.; Zhao, Y. A dynamic-model-based fault diagnosis method for a wind turbine planetary gearbox using a deep learning network. *Prot. Control Mod. Power Syst.* **2022**, *7*, 14. [\[CrossRef\]](#)
12. Zhao, K.; Jia, F.; Shao, H.D. A novel conditional weighting transfer Wasserstein auto-encoder for rolling bearing fault diagnosis with multi-source domains. *Knowl. Based Syst.* **2023**, *262*, 11. [\[CrossRef\]](#)
13. Shen, Y.J.; Chen, B.; Guo, F.H.; Meng, W.C.; Yu, L. A Modified Deep Convolutional Subdomain Adaptive Network Method for Fault Diagnosis of Wind Turbine Systems. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 10. [\[CrossRef\]](#)
14. Zhao, C.; Shen, W.M. Mutual-assistance semisupervised domain generalization network for intelligent fault diagnosis under unseen working conditions. *Mech. Syst. Signal Proc.* **2023**, *189*, 18. [\[CrossRef\]](#)
15. Jiang, K.X.; Gao, X.J.; Gao, H.H.; Han, H.Y.; Qi, Y.S. VIT-GADG: A Generative Domain-Generalized Framework for Chillers Fault Diagnosis Under Unseen Working Conditions. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 13. [\[CrossRef\]](#)
16. Qiu, G.Q.; Gu, Y.K.; Cai, Q. A deep convolutional neural networks model for intelligent fault diagnosis of a gearbox under different operational conditions. *Measurement* **2019**, *145*, 94–107. [\[CrossRef\]](#)
17. Durbhaka, G.K.; Selvaraj, B.; Mittal, M.; Saba, T.; Rehman, A.; Goyal, L.M. Swarm-LSTM: Condition Monitoring of Gearbox Fault Diagnosis Based on Hybrid LSTM Deep Neural Network Optimized by Swarm Intelligence Algorithms. *CMC-Comput. Mat. Contin.* **2021**, *66*, 2041–2059.
18. Wang, F.T.; Dun, B.S.; Liu, X.F.; Xue, Y.H.; Li, H.K.; Han, Q.K. An Enhancement Deep Feature Extraction Method for Bearing Fault Diagnosis Based on Kernel Function and Autoencoder. *Shock Vib.* **2018**, *2018*, 12. [\[CrossRef\]](#)
19. Guo, L.; Lei, Y.G.; Xing, S.B.; Yan, T.; Li, N.P. 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. [\[CrossRef\]](#)
20. Wan, L.J.; Li, Y.Y.; Chen, K.Y.; Gong, K.; Li, C.Y. A novel deep convolution multi-adversarial domain adaptation model for rolling bearing fault diagnosis. *Measurement* **2022**, *191*, 17. [\[CrossRef\]](#)
21. Chen, Z.Y.; He, G.L.; Li, J.P.; Liao, Y.X.; Gryllias, K.; Li, W.H. Domain Adversarial Transfer Network for Cross-Domain Fault Diagnosis of Rotary Machinery. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 8702–8712. [\[CrossRef\]](#)
22. An, Y.Y.; Zhang, K.; Chai, Y.; Liu, Q.; Huang, X.H. Domain adaptation network base on contrastive learning for bearings fault diagnosis under variable working conditions. *Expert Syst. Appl.* **2023**, *212*, 9. [\[CrossRef\]](#)
23. Han, T.; Li, Y.F.; Qian, M. A Hybrid Generalization Network for Intelligent Fault Diagnosis of Rotating Machinery Under Unseen Working Conditions. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 11. [\[CrossRef\]](#)
24. Shi, Y.W.; Deng, A.D.; Deng, M.Q.; Xu, M.; Liu, Y.; Ding, X.; Bian, W.B. Domain augmentation generalization network for real-time fault diagnosis under unseen working conditions. *Reliab. Eng. Syst. Saf.* **2023**, *235*, 109188. [\[CrossRef\]](#)
25. Fan, Z.H.; Xu, Q.F.; Jiang, C.X.; Ding, S.X. Deep Mixed Domain Generalization Network for Intelligent Fault Diagnosis Under Unseen Conditions. *IEEE Trans. Ind. Electron.* **2023**, *71*, 965–974. [\[CrossRef\]](#)
26. Sagawa, S.; Koh, P.W.; Hashimoto, T.B.; Liang, P. Distributionally robust neural networks for group shifts: On the importance of regularization for worst-case generalization. *arXiv* **2019**, arXiv:1911.08731.
27. Zhang, M.; Marklund, H.; Dhawan, N.; Gupta, A.; Levine, S.; Finn, C. Adaptive Risk Minimization: Learning to Adapt to Domain Shift. In Proceedings of the 35th Conference on Neural Information Processing Systems (NeurIPS), Online, 6–14 December 2021; Neural Information Processing Systems Foundation, Inc. (NeurIPS): Montreal, QC, Canada, 2021.
28. Balaji, Y.; Sankaranarayanan, S.; Chellappa, R. MetaReg: Towards Domain Generalization using Meta-Regularization. In Proceedings of the 32nd Conference on Neural Information Processing Systems (NIPS), Montreal, QC, Canada, 2–8 December 2018; Neural Information Processing Systems (Nips): Montreal, QC, Canada, 2018.
29. Qiao, F.; Zhao, L.; Peng, X. Learning to learn single domain generalization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 12556–12565.
30. Muandet, K.; Balduzzi, D.; Schölkopf, B. Domain generalization via invariant feature representation. In Proceedings of the 30th International Conference on Machine Learning, Atlanta, GA, USA, 17–19 June 2013; pp. 10–18.
31. Li, H.; Pan, S.J.; Wang, S.; Kot, A.C. Domain generalization with adversarial feature learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 5400–5409.
32. Mirza, M.; Osindero, S. Conditional generative adversarial nets. *arXiv* **2014**, arXiv:1411.1784.
33. Ganin, Y.; Ustinova, E.; Ajakan, H.; Germain, P.; Larochelle, H.; Laviolette, F.; Marchand, M.; Lempitsky, V.J.T. Domain-adversarial training of neural networks. *J. Mach. Learn. Res.* **2016**, *17*, 1–35.
34. Kingma, D.P.; Welling, M.J. Auto-encoding variational bayes. *arXiv* **2013**, arXiv:1312.6114.

35. Pang, X.; Xue, X.; Jiang, W.; Lu, K. An investigation into fault diagnosis of planetary gearboxes using a bispectrum convolutional neural network. *IEEE/ASME Trans. Mechatron.* **2020**, *26*, 2027–2037. [\[CrossRef\]](#)
36. He, C.; Ge, D.; Yang, M.; Yong, N.; Wang, J.; Yu, J.J. A data-driven adaptive fault diagnosis methodology for nuclear power systems based on NSGAIL-CNN. *Ann. Nucl. Energy* **2021**, *159*, 108326. [\[CrossRef\]](#)
37. Senanayaka, J.S.L.; Van Khang, H.; Robbersmvr, K.G. CNN based Gearbox Fault Diagnosis and Interpretation of Learning Features. In Proceedings of the 30th IEEE International Symposium on Industrial Electronics (ISIE), Kyoto, Japan, 20–23 June 2021.
38. Song, B.Y.; Liu, Y.Y.; Lu, P.; Bai, X.Z. Rolling Bearing Fault Diagnosis Based on Time-frequency Transform-assisted CNN: A Comparison Study. In Proceedings of the IEEE 12th Data Driven Control and Learning Systems Conference (DDCLS), Xiangtan, China, 12–14 May 2023; pp. 1273–1279.
39. Zhao, H.; Liu, M.; Sun, Y.; Chen, Z.; Duan, G.; Cao, X.J. Fault diagnosis of control moment gyroscope based on a new CNN scheme using attention-enhanced convolutional block. *Sci. China Technol. Sci.* **2022**, *65*, 2605–2616. [\[CrossRef\]](#)
40. Yan, X.; She, D.; Xu, Y.; Jia, M.J. Deep regularized variational autoencoder for intelligent fault diagnosis of rotor-bearing system within entire life-cycle process. *Knowl. Based Syst.* **2021**, *226*, 107142. [\[CrossRef\]](#)
41. Rathore, M.S.; Harsha, S.J. Technologies, Non-linear Vibration Response Analysis of Rolling Bearing for Data Augmentation and Characterization. *J. Vib. Eng. Technol.* **2023**, *11*, 2109–2131. [\[CrossRef\]](#)
42. Smith, W.A.; Randall, R.B. Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study. *Mech. Syst. Signal Process.* **2015**, *64*, 100–131. [\[CrossRef\]](#)
43. Shao, S.; McAleer, S.; Yan, R.; Baldi, P.J. Highly accurate machine fault diagnosis using deep transfer learning. *IEEE Trans. Ind. Inform.* **2018**, *15*, 2446–2455. [\[CrossRef\]](#)
44. Shi, Y.; Deng, A.; Deng, M.; Li, J.; Xu, M.; Zhang, S.; Ding, X.; Xu, S.J. Domain Transferability-based Deep Domain Generalization Method Towards Actual Fault Diagnosis Scenarios. *IEEE Trans. Ind. Inform.* **2022**, *19*, 7355–7366. [\[CrossRef\]](#)
45. Zhou, K.; Yang, Y.; Hospedales, T.; Xiang, T. Learning to generate novel domains for domain generalization. In *Computer Vision—ECCV 2020: Proceedings of the 16th European Conference, Glasgow, UK, 23–28 August 2020*; Springer: Cham, Switzerland, 2020; pp. 561–578.
46. Zhao, C.; Shen, W.J. Adversarial mutual information-guided single domain generalization network for intelligent fault diagnosis. *IEEE Trans. Ind. Inform.* **2022**, *19*, 2909–2918. [\[CrossRef\]](#)
47. Qian, Q.; Zhou, J.; Qin, Y.J. Relationship transfer domain generalization network for rotating machinery fault diagnosis under different working conditions. *IEEE Trans. Ind. Inform.* **2023**, *19*, 9898–9908. [\[CrossRef\]](#)
48. She, D.; Chen, J.; Yan, X.; Zhao, X.; Pecht, M.J. Diversity maximization-based transfer diagnosis approach of rotating machinery. *Struct. Health Monit.* **2023**, *23*, 14759217231164921. [\[CrossRef\]](#)

**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.