Article

A Semi-Supervised Fault Diagnosis Method Based on Improved Bidirectional Generative Adversarial Network

Long Cui ¹ , Xincheng Tian ^{1,*} , Xiaorui Shi ², Xiuqing Wang ² and Yigang Cui ²

¹ School of Control Science and Engineering, Shandong University, Jinan 250061, China; clat@mail.sdu.edu.cn

² Sinotruk Industry Park Zhangqiu, Sinotruk Jinan Power Co., Ltd., Jinan 250220, China; shixr@sinotruk.com (X.S.); wangxiuqing@sinotruk.com (X.W.); young.cui@163.com (Y.C.)

* Correspondence: txch@sdu.edu.cn; Tel.: +86-0531-88392418

Abstract: With the assumption of sufficient labeled data, deep learning based machinery fault diagnosis methods show effectiveness. However, in real-industrial scenarios, it is costly to label the data, and unlabeled data is underutilized. Therefore, this paper proposes a semi-supervised fault diagnosis method called Bidirectional Wasserstein Generative Adversarial Network with Gradient Penalty (BiWGAN-GP). First, by unsupervised pre-training, the proposed method takes full advantage of a large amount of unlabeled data and can extract features from vibration signals effectively. Then, using only a few labeled data to conduct supervised fine-tuning, the model can perform an accurate fault diagnosis. Additionally, Wasserstein distance is used to improve the stability of the model's training procedure. Validation is performed on the bearing and gearbox fault datasets with limited labeled data. The results show that the proposed method can achieve 99.42% and 91.97% of diagnosis accuracy on the bearing and gear dataset, respectively, when the size of the training set is only 10% of the testing set.

Citation: Cui, L.; Tian, X.; Shi, X.; Wang, X.; Cui, Y. A Semi-Supervised Fault Diagnosis Method Based on Improved Bidirectional Generative Adversarial Network. *Appl. Sci.* **2021**, *11*, 9401. <https://doi.org/10.3390/app11209401>

Academic Editor: Jordi Cusido

Received: 11 August 2021

Accepted: 5 October 2021

Published: 10 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 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/>).

Keywords: generative adversarial networks; semi-supervised learning; fault diagnosis; wasserstein distance

1. Introduction

As the essential part of prognostics and health management (PHM), fault diagnosis technology can enhance machinery reliability, increase operating safety, and reduce maintenance costs [1]. For example, the failure of wind turbine bearings, which are mounted at high altitudes and work in harsh conditions, will lead to higher repair costs [2]. Therefore, it is meaningful to develop reliable and accurate fault diagnosis methods for the essential machinery components of machines such as bearings and gears.

With the development of industrial information technology, data-driven methods have become mainstream of intelligent fault diagnosis. As the key process of data-driven methods, the feature extraction ability greatly affects the accuracy of diagnosis [3]. Recently, deep learning based methods have attracted attention—because its ability to automatically extract features, which are less dependent on prior knowledge about signal processing techniques and diagnostic expertise [4]. As the model deepens, the capabilities of feature extraction and abstraction are also enhanced. However, deeper models require more data for training, or the model's performance may be affected by over-fitting. Therefore, deep learning methods have higher requirements for both quality and quantity of data.

Table 1 shows the summary of the recent fault diagnosis methods based on deep learning, including supervised, unsupervised and semi-supervised learning methods. With the assumption of sufficient data, supervised deep learning based fault diagnosis methods show effectiveness. Wu et al. [5] applied the one-dimensional convolutional neural network (1-DCNN) to learn features directly from the raw vibration signals and achieved nearly 100% diagnostic accuracy of the gearbox fault dataset. Duong et al. [6] used the wavelet transform to get the 2D image of the acoustic emission signals of the bearing, then applied

a deep convolution neural network (DCNN) as a classifier and achieved an accuracy of 98.79%. Zhao et al. [7] applied the famous deep learning architecture in image processing—deep residual network (ResNet) on the 2D matrices of wavelet coefficient to improve the accuracy of gearbox fault diagnosis and reached an accuracy of 96.67%. Zhang et al. [8] proposed a gated recurrent unit (GRU) based recurrent neural network (RNN) to make full use of the temporal information from time-series signals, the model shows nearly 100% accuracy and can get 94.86% against the -4 dB signal-to-noise ratios noise. Similarly, those studies benefited from the high-quality datasets and sufficient training data.

Table 1. Summary of the recent fault diagnosis methods based on deep learning.

	Reference	Method	Objects
Supervised	[5]	One-dimensional convolutional neural network(1-DCNN)	planetary gearbox
	[6]	Deep convolution neural network (DCNN) with wavelet image	rolling bearing
	[7]	Multiple wavelet regularized deep residual network (MWR-DRN)	planetary gearbox
	[8]	Recurrent neural network (RNN) based on gated recurrent units (GRUs)	bearing and pump
Unsupervised	[9]	Local connection network (LCN) constructed by normalized sparse autoencoder (NSAE)	gearbox and bearing
	[10]	Deep sparse autoencoder (DSAE)	gear pitting fault
	[11]	Convolutional deep belief network (CDBN)	rolling bearing
	[12]	Enhanced deep convolutional generative adversarial network (DC-GAN)	gearbox and bearing
	[13]	Wavelet packet transform combined with generative adversarial networks (GAN)	robotic manipulator
Semi-supervised	[14]	Categorical adversarial autoencoder (CatAAE)	gearbox and bearing
	[15]	Consistency regularization based semi-supervised learning	rolling bearing
	[16]	semi-supervised deep sparse autoencoder (SSDSAE) with local and nonlocal information	bearing and industrial rotor
	[17]	Graph-based rebalance semi-supervised learning (GRSSL) with visibility graph feature (VGF)	rolling bearing

Although we can obtain "big data" due to the development of data acquisition technologies in real industrial scenarios, most of the available data is unlabeled raw data. On the other hand, labeling data is expensive or time-consuming. Therefore, in the case of insufficient labeled data, it is difficult for the supervised deep learning model to show excellent performance. Researchers have been seeking ways to solve this problem, such as unsupervised learning and semi-supervised learning.

Unsupervised deep learning methods such as autoencoders (AE), deep belief networks (DBN), and generative adversarial networks (GAN) have been studied to apply on fault diagnosis tasks. AE can learn features from unlabeled data by reconstructing them. Jia et al. [9] proposed a sparse autoencoder based local connection network that can mine shift-invariant fault features from vibration signals. Qu et al. [10] integrated dictionary learning in sparse coding to extract features from raw data by using a deep sparse autoencoder. DBN is another classic unsupervised learning method, Liu et al. [11] proposed a bearing fault diagnosis method based on improved convolutional DBN to extract quantitative and qualitative features from vibration signals. Recently, GAN [18] has been successfully applied in many fields because of its ability to generate realistic data. Therefore, researchers have developed some applications of GAN and its variants in the field of fault diagnosis. Data augmentation is the most common use of GAN to solve the insufficient or imbalanced problem of fault data. Wang et al. [12] applied a 1-DCNN based deep convolutional generative adversarial network (DCGAN) to enhance the GAN model, then K-means clustering algorithm and ridge regression are used to improve a CNN model for fault classification. Pu et al. [13] used GAN as an oversampling method to compensate the unbalanced dataset for fault diagnosis of an industrial robotic manipu-

lator. Liu et al. [14] combined categorical GAN and adversarial autoencoder to perform unsupervised clustering of rolling bearings.

To make more effective use of both labeled and unlabeled data, semi-supervised learning methods have attracted the attention of researchers. Yu et al. [15] proposed a semi-supervised approach based on consistency regularization principle that makes the model less sensitive to the extra perturbation imposed on the inputs. Zhao et al. [16] embedded the labeled and unlabeled data into local and nonlocal regularization terms to realize a semi-supervised deep sparse autoencoder. Chen et al. [17] proposed a graph-based rebalance semi-supervised learning method, and they mainly focus on variable conditions and imbalance unlabeled data.

This paper proposes a new semi-supervised learning method focusing on the problem of insufficient labeled data, namely Bidirectional Wasserstein Generative Adversarial Network with Gradient Penalty (BiWGAN-GP). Compare to the existing work of semi-supervised fault diagnosis methods, the proposed method makes full use of the representation learning ability of GAN. It not only can extract effective features from the unlabeled data but also can generate realistic signals for data augmentation, which makes it suitable for fault diagnosis tasks with limited labeled data. The main contributions of this paper can be summarized as follows:

- (1) By adding an encoder on the standard GAN architecture, for the first time, bidirectional GAN is introduced in the machinery fault diagnosis field, which provides a new way for automatic feature extraction from unlabeled data.
- (2) Wasserstein distance with gradient penalty is used in the unsupervised training procedure, which improves the model's stability and usability. Experimental results show the effectiveness and necessity of this improvement.
- (3) With limited labeled data to fine-tune the model's parameters, the proposed method can achieve a satisfactory diagnosis accuracy, which can reduce the cost of labeling data.

The rest of this article is organized as follows. Section 2 gives the theoretical background about GAN and bidirectional GAN, then the specific methods for improving the GAN are described. Section 3 present the detailed system framework and the training procedure of the proposed method. Experiments are conducted in Section 4 to verify the effectiveness of the proposed method for data generation and fault diagnosis. Finally, the discussion and conclusions are drawn in Sections 5 and 6, respectively.

2. Methodologies

2.1. Generative Adversarial Networks

Generative adversarial network (GAN) is a framework that can capture the real data's distribution and then generate fake data very similar to real data. It consists of two modules: generator G and discriminator D , as shown in Figure 1a.

Figure 1. Architecture of GAN and BiGAN. (a) GAN. (b) BiGAN.

Different from the traditional generative models, GAN does not compute the exact probability density function $p_X(x)$ of the data distribution (which is difficult), it models $p_X(x)$ by transforming a latent distribution $p_Z(z)$, such as a standard Normal distribution $\mathcal{N}(0, I)$. This transformation is called a generator G , which mapping from the latent feature space to data space by a feed-forward network $\tilde{x} = G(z)$, $G : \Omega_z \rightarrow \Omega_x$. Another module called discriminator D aims to distinguish between samples from the data space or generator G . Through adversarial training of G and D , $p_G(x)$ gradually approaches $p_X(x)$ and finally matches it. Therefore, GAN can learn arbitrarily complex real data distribution and then create real-like data.

The following minimax objective formulates the training of GAN:

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_X} [\log D(x)] + \mathbb{E}_{z \sim p_Z} [\log(1 - D(G(z)))] \quad (1)$$

where $D(x)$ represents the probability that discriminator will determine the data x from the true distribution p_X to be true, and $(1 - D(G(z)))$ means the probability that D will judge the sample generated by the generator $G(z)$ is false. The goal of D is to maximize the above probabilities, therefore the loss function of the discriminator is:

$$\mathcal{L}_D = -\mathbb{E}_{x \sim p_X} [\log D(x)] - \mathbb{E}_{z \sim p_Z} [\log(1 - D(G(z)))] \quad (2)$$

The goal of generator G is to minimize the probability that D successfully discriminates generated samples. The first term of Equation (1) makes no reference to G , so it can be ignored. Therefore the loss function of the generator is:

$$\mathcal{L}_G = \mathbb{E}_{z \sim p_Z} [\log(1 - D(G(z)))] \quad (3)$$

However, in the early training Equation (3) cannot provide enough gradient for the learning of G . In practice, Goodfellow proposed an alternative generator loss function:

$$\mathcal{L}_G = -\mathbb{E}_{z \sim p_Z} [\log(D(G(z)))] \quad (4)$$

By alternating training G and D , when the algorithm converges, the generated distribution of G coincides with the real data distribution, i.e., $p_G = p_X$, and D will not be able to distinguish whether the sample is real or fake, i.e., $D(G(z)) = 0.5$.

Compared with the traditional generative model, GAN is simpler to construct and more straightforward to train. However, the original GAN model can only utilize unlabeled data to perform unsupervised learning tasks such as generating fake data. It lacks the ability to perform supervised learning such as classification. For that reason, we have to modify the architecture of GAN for fault diagnosis.

2.2. Bidirectional Generative Adversarial Networks

In order to use GAN's advantages to capture data distribution for inference and classification tasks, Donahue et al. [19] proposed the bidirectional generative adversarial network (BiGAN) by introducing an additional encoder E on the basis of GAN, which gives BiGAN the ability to extract data features and to perform semi-supervised learning.

As shown in Figure 1b, the obvious difference between BiGAN and GAN is the encoder E . In addition to the generator G that maps the latent space to the data space, BiGAN introduces an encoder E to inversely map the data space to the latent space, i.e., $\tilde{z} = E(x)$. Thus $\tilde{z}$ can be used as low-dimensional features extracted from data samples by the encoder.

Besides, the structure of discriminator D needs to be modified as well. Instead of taking x or $\tilde{x} = G(z)$ as the input of D , the discriminator discriminates joint pair $(x, \tilde{z})$ or $(\tilde{x}, z)$ from data space and latent space, respectively.

Similarly to GAN, the goal of BiGAN can be defined as a minimax objective:

$$\min_{G,E} \max_D V(D, E, G) = \mathbb{E}_{x \sim p_X} [\log D(x, E(x))] + \mathbb{E}_{z \sim p_Z} [\log (1 - D(G(z), z))] \quad (5)$$

The goal of D is to maximize the value function $V(D, E, G)$, and the goal of both G and E is to minimize it. In addition, BiGAN also uses the alternative loss function of G as mentioned in Equation (4). So we can obtain the following loss functions for the training of BiGAN:

$$\mathcal{L}_D = -\mathbb{E}_{x \sim p_X} [\log D(x, E(x))] - \mathbb{E}_{z \sim p_Z} [\log (1 - D(G(z), z))] \quad (6)$$

$$\mathcal{L}_{EG} = -\mathbb{E}_{x \sim p_X} [\log (1 - D(x, E(x)))] - \mathbb{E}_{z \sim p_Z} [\log D(G(z), z)] \quad (7)$$

The encoder E helps BiGAN extract features from data to perform classification and makes BiGAN suitable for semi-supervised learning tasks.

2.3. Improve the Training of BiGAN

2.3.1. Wasserstein GAN with Gradient Penalty

Although the origin GAN model is simple to construct, a high-quality GAN model is hard to obtain. However, the training of the origin GAN may have problems such as non-convergence and mode collapse. BiGAN has the same problems because it uses similar loss functions [20]. Those issues reduce the usability, that the model not being able to generate samples of all fault types. Therefore, it is necessary to improve the training process of BiGAN.

The inappropriate measurement method causes the difficulty of training [21]. The original GAN optimizes the Jensen–Shannon (JS) divergence between the generated and

real distribution. However, when the two distributions do not have any non-negligible intersection, the JS divergence is a constant, which causes the vanishing gradient.

To solve this problem, a new measurement method is proposed, namely Earth Mover distance or Wasserstein distance [22]. Even if the two distributions do not have any overlap, this metric can still indicate the distance between them and provide effective gradients for training. The model built based on Wasserstein distance is called Wasserstein GAN (WGAN). The value function of WGAN is:

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_X}[D(x)] - \mathbb{E}_{z \sim p_Z}[D(G(z))] \quad (8)$$

which is similar to Equation (1) but without taking the logarithm, and alternative G loss is also taken.

Similarly, we can derive the loss function from Equation (8):

$$\mathcal{L}_D = -\mathbb{E}_{x \sim p_X}[D(x)] + \mathbb{E}_{z \sim p_Z}[D(G(z))] \quad (9)$$

$$\mathcal{L}_G = -\mathbb{E}_{z \sim p_Z}[D(G(z))] \quad (10)$$

The Equation (9) is also the definition of Wasserstein distance.

In addition to the loss function, the main structural modification of WGAN is about the discriminator D . It removes the sigmoid activation function at the last layer so that the output is no longer limited to (0,1). However, to enforce the Lipschitz constraint, WGAN demands clipping the weight of D with a constant c after every update step, i.e., $w_D \in [-c, c]$. This procedure may cause a phenomenon that the value of w_D always approximates c or $-c$. If c is too large or too small, it may lead to vanishing or exploding gradients.

Therefore, Gulrajani et al. [23] improved WGAN by adding a gradient penalty term in Equation (9), called WGAN-GP. It can avoid weight clipping and enforce the Lipschitz constraint. The loss function of the WGAN-GP's discriminator is:

$$\begin{aligned} \mathcal{L}_D = & -\mathbb{E}_{x \sim p_X}[D(x)] + \mathbb{E}_{z \sim p_Z}[D(G(z))] \\ & + \lambda \mathbb{E}_{\hat{x} \sim p_{\hat{x}}}[(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2] \end{aligned} \quad (11)$$

where the additional part compared to Equation (9) is the gradient penalty term. $\hat{x}$ is the linear interpolation between the real data x and the generated data $G(z)$:

$$\hat{x} = \epsilon x + (1 - \epsilon)G(z) \quad (12)$$

where λ is penalty coefficient, usually take $\lambda = 10$, and ϵ is a random number $\epsilon \sim U[0, 1]$.

2.3.2. Improve BiGAN with WGAN-GP

Eventually, we use WGAN-GP to improve the training of BiGAN. The model we got is called bidirectional Wasserstein generative adversarial networks with gradient penalty (BiWGAN-GP). The BiWGAN-GP's discriminator loss function is

$$\begin{aligned} \mathcal{L}_D = & -\mathbb{E}_{x \sim p_X}[D(x, E(x))] + \mathbb{E}_{z \sim p_Z}[D(G(z), z)] \\ & + \lambda \mathbb{E}_{\hat{x} \sim p_{\hat{x}}, \hat{z} \sim p_{\hat{z}}}[(\|\nabla_{(\hat{x}, \hat{z})} D(\hat{x}, \hat{z})\|_2 - 1)^2] \end{aligned} \quad (13)$$

where $\hat{x}$ is defined in Equation (12), and $\hat{z}$ is the linear interpolation between $E(x)$ and z :

$$\hat{z} = \epsilon E(x) + (1 - \epsilon)z \quad (14)$$

The loss function of G and E in BiWGAN-GP can obtain from Equation (8) directly by replacing the input of D with joint pairs:

$$\mathcal{L}_{EG} = \mathbb{E}_{x \sim p_X}[D(x, E(x))] - \mathbb{E}_{z \sim p_Z}[D(G(z), z)] \quad (15)$$

This equation is also the Wasserstein distance between the real distribution and the generated distribution, which can indicate the training process.

The experiment in Section 4 shows that after using WGAN with gradient penalty to improve BiGAN, the model generation result and the diagnosis accuracy in the case of a small training dataset are improved.

3. System Framework and Model Training

3.1. Model Architecture

We propose a novel approach to perform both vibration signal generation and machinery fault diagnosis, called BiWGAN-GP. It combined the architecture of BiGAN and the advanced training method of WGAN-GP. Figure 2 Shows the BiWGAN-GP's architecture. Overall, the BiWGAN-GP fault diagnosis model has three main modules: generator, encoder, and discriminator, respectively.

Figure 2. The architecture of BiWGAN-GP.

The generator consists of a multi-layer neural network with an increasing number of hidden units to mapping a low-dimension noise z into a high dimension generated signal $G(z)$. Therefore, the input dimension of the generator depends on the latent vector z , and the output dimension matches the real signal length. The number of units in the latter hidden layer is double that of the former layer. Specifically, the topology of G is (64, 128, 256, 512).

Different data preprocessing methods will affect the construction of the generator. If normalization is used, i.e., $x' = (x - x_{min}) / (x_{max} - x_{min})$, the data amplitude is limited to (0,1). Then ReLU activation function is used between the layers of G , and the SoftPlus is used in the last layer to ensure that there is no negative value in generated signals. Another commonly used data preprocessing operation is standardization, i.e., $x' = (x - \mu) / \sigma$, where μ is the mean and σ is the standard deviation of the entire dataset. Since standardization may have negative values, to avoid truncation, we use LeakyReLU as the activation function of the last layer of G to get better generation results.

The encoder is also composed of a multi-layer neural network but with a decreasing number of hidden units to extract the most salient features of the data. Specifically, the topology of E is (512, 256, 128, 64), where the number of units in the hidden latter layer is half that of the former layer. Its input and output dimensions are opposite to those of the generator. LeakyReLU is used as the activation function in the encoder.

The discriminator is divided into three parts. Two multi-layer neural networks extract features of the joint pair $(x, E(x))$ that comes from real data or $(G(z), z)$ that comes from fake data, and their topology are (256, 128, 64) and (128, 64). Then the features are concatenated together as a joint feature which size is 128. Finally, through another multi-layer neural network which topology is (128, 64, 32, 1), the discriminatory result is obtained. Also, LeakyReLU is used as the activation function.

3.2. Model Training Procedure

The training process of the BiWGAN-GP has two stages, namely the unsupervised pre-training stage and the supervised fine-tuning stage. Figure 3 shows the training procedure of the proposed fault diagnosis method.

Figure 3. BiWGAN-GP training procedure.

The unsupervised pre-training procedure is described in Algorithm 1. In this stage, all unlabeled data is used to train the model. Two Adam [24] optimizers are used to update the model parameters, one for the discriminator, corresponding to the loss function (13), and the other one is responsible for optimizing the generator and encoder parameters, corresponding to the loss function (15). Since the Wasserstein distance with gradient penalty is used to improve the loss function, there is no need to worry about the discriminator being trained too well and the convergence problem. After the unsupervised stage, the generator of BiWGAN-GP can be used to generate real-like samples for the data augmentation task.

In the supervised fine-tuning stage, we train the model for fault diagnosis using only a small amount of labeled samples. The encoder network in the BiWGAN-GP model is extracted, and a fully connected layer is appended as a classifier. An Adam optimizer is used to update the parameters of the classifier. Commonly, a typical deep learning model takes more training samples than testing samples. However, to simulate the situation where it is difficult to obtain enough high-quality labeled data for training, this article uses the way that there are far more test samples than training samples in the supervised learning stage.

Algorithm 1: The pre-training procedure of BiWGAN-GP with gradient penalty.
Default values: $\lambda = 10, \alpha = 0.0001, \beta_1 = 0.5, \beta_2 = 0.999$

Data: The gradient penalty coefficient λ , initial discriminator parameters w_0 , initial generator and encoder parameters θ_0

```

1 repeat
2    $x^{(1)}, \dots, x^{(m)} \sim P_X(x)$ 
3    $z^{(1)}, \dots, z^{(m)} \sim P_Z(z)$ 
4    $\mathcal{L}_{EG} = \frac{1}{m} \sum_{i=1}^m D(x^{(i)}, E(x^{(i)})) - \frac{1}{m} \sum_{j=1}^m D(G(z^{(j)}), z^{(j)})$ 
5    $\hat{x}^{(i)} = \epsilon x^{(i)} + (1 - \epsilon) \tilde{x}^{(i)}$ 
6    $\hat{z}^{(j)} = \epsilon z^{(j)} + (1 - \epsilon) \tilde{z}^{(j)}$ 
7    $\mathcal{L}_D = -\mathcal{L}_{EG} + \lambda \frac{1}{m} \sum_{i=1}^m ((\|\nabla_{(\hat{x}^{(i)}, \hat{z}^{(i)})} D(\hat{x}^{(i)}, \hat{z}^{(i)})\|_2 - 1)^2)$ 
8    $w \leftarrow \text{Adam}(\nabla_w \mathcal{L}_D, w, \alpha, \beta_1, \beta_2)$ 
9    $\theta \leftarrow \text{Adam}(\nabla_\theta \mathcal{L}_{EG}, \theta, \alpha, \beta_1, \beta_2)$ 
10 until convergence

```

Adequate epochs of training are necessary to learn the pattern of real data. However, the number of the unsupervised training epochs is not the more the better. To determine the appropriate number of training epochs, an experiment is designed. First, during the unsupervised pre-training, we saved the parameters of the encoder every 100 epochs. So, pre-trained models with different training epochs can be obtained. Then, fine-tuning is conducted on the saved models using the same labeled training set. Finally, the models are tested on the same testing set. Figure 4a shows the trend of diagnosis accuracy through the pre-training process. It shows that the accuracy increases rapidly at the beginning and then decreases as the number of training epochs increases. Figure 4b shows the loss of pre-training and fine-tuning with increasing pre-training epochs. The overall trend of the loss of pre-training is decreasing, but the loss of fine-tuning decreases at the beginning and then increases, which supports the trend of accuracy. Therefore, pre-training epochs around 2300 may be appropriate. One possible reason is that since the generator and encoder are optimized at the same time, an overfitting phenomenon similar to that when the autoencoder is over-trained may have occurred, resulting in a degradation of the encoder's feature representation capability.

Figure 4. Trends in diagnosis accuracy and loss through the pre-training process. (a) Trend of diagnosis accuracy. (b) Loss of pre-training and fine-tuning.

4. Experimental Verification

To verify the performance of the proposed method, two experiments are designed in this section: data generation and fault diagnosis. The proposed method is verified on two types of basic rotating machinery: bearing and gear.

4.1. Datasets Description

4.1.1. CWRU Bearing Fault Dataset

The bearing fault dataset of Case Western Reserve University (CWRU) Bearing Data Center [25] is one of the most widely used datasets in the field of fault diagnosis. This dataset collects data from a ball bearing on a test rig, which is driven by a 2 hp Reliance electric motor, and accelerometers are used to record vibration data of the bearing at the motor drive end. Using electro-discharge machining (EDM) technology, artificially created defects ranging from 0.007 inches (0.1778 mm) to 0.028 inches (0.7112 mm) in diameter located in the inner race, outer race, and rolling elements of the bearing. And vibration data is recorded for motor loads of 0 to 3 horsepower (speed 1797 rpm to 1720 rpm). The torque sensor and encoder are used to record the load and motor speed, respectively.

This article uses bearing vibration data at the drive end of the motor. As shown in Table 2, a total of 10 types of working conditions are used as classification categories, including one normal and 11 fault conditions. We use the fault data at a 12 kHz sampling rate, and the 48 kHz normal data is resampled to 12 kHz to ensure the consistency of data. First, by splitting the original data into segments of 1024 data points in the time domain, we obtained a dataset of 1000 samples for each category, for a total of 10,000 samples. Then, Fast Fourier transform (FFT) is used as a data preprocessing method to obtain frequency domain signals, and the magnitude of each sample is normalized into [0,1]. Additionally, to avoid the effect of spectral leakage, a Hamming window function is applied for each segment of the signal before FFT. Finally, due to the symmetry of the FFT results, we usually use only the first half of the results, so the length of each sample is 512 eventually.

Table 2. CWRU bearing dataset faults.

Category	Condition	Defect Diameter (Inch)	Label
1	Normal	-	Normal
2	Inner race	0.007	IR07
3	Inner race	0.014	IR14
4	Inner race	0.021	IR21
5	Rolling elements	0.007	RE07
6	Rolling elements	0.014	RE14
7	Rolling elements	0.021	RE21
8	Outer race	0.007	OR07
9	Outer race	0.014	OR14
10	Outer race	0.021	OR21

4.1.2. UoC Gear Fault Dataset

According to a benchmark study [26], the CWRU bearing fault dataset is relatively easy to diagnose. However, the University of Connecticut (UoC) gear fault dataset [27,28] is a challenge for many classical models such as autoencoder (AE) and CNN. Therefore, we use the UoC gear fault dataset to further evaluate the proposed method on different types of rotating machinery.

As shown in Figure 5, the two-stage gearbox is driven by an electric motor. The first stage consists of a 32-tooth pinion on the input shaft and an 80-tooth gear, and the second stage consists of a 48-tooth pinion and a 64-tooth gear. A magnetic brake is connected to the output shaft as the load. By replacing the pinion on the first stage input shaft, nine different working conditions are introduced into the system, including healthy condition, missing tooth, root crack, spalling, and chipping tip with five different levels of severity, as shown in Table 3 and Figure 6. An accelerometer is used to record the original vibration signals of the gears with a sampling frequency of 20 kHz, and 104 samples are recorded for each working condition, each sample containing 3600 data points.

Figure 5. Gearbox test rig and its mechanical schematics.

Table 3. UoC gear dataset faults.

Category	Condition	Severity	Label
1	Healthy	-	HT
2	Missing tooth	-	MT
3	Root crack	-	RT
4	Spalling	-	SP
5	Chipping tip	1	CT1
6	Chipping tip	2	CT2
7	Chipping tip	3	CT3
8	Chipping tip	4	CT4
9	Chipping tip	5	CT5

Figure 6. Gear working conditions.

As same as the CWRU dataset, we split the original data into segments of 1024 data points in the time domain, then FFT and standardization are performed. Eventually, the gear dataset consists of 312 samples for each category, for a total of 2808 samples.

4.2. Data Generation Experiment

Data generation is an effective application of GAN that helps to solve the problems such as lack of training samples and data imbalance. As a GAN-based model, BiWGAN-GP is also capable of generating data. Besides, a good generation result also means good feature extraction capability of the encoder [19]. Therefore, this experiment is designed to evaluate the quality of generated data.

For comparison, the origin GAN, WGAN, WGAN-GP, BiGAN, and improved BiWGAN-GP models are trained to generate vibration signals in the frequency domain. All models have the same structure of generator. The BiGAN and BiWGAN-GP have the same architecture but only different loss functions for training. The models are trained for 2000

epochs using all of the available unlabeled data. After training, 1000 unlabeled samples are generated by feeding forward random noise into the model's generator.

(1) Quantitative assessment: To quantitatively evaluate the generated data of GAN, inception score (IS) [29] and Fréchet inception distance (FID) [30] are popular evaluation metrics. However, those methods are designed for image generation tasks, and unsuitable for vibration signals. In this paper, maximum mean discrepancy (MMD) [31] and Pearson correlation coefficients (PCC) are used to quantitatively assess the quality of generated data. The MMD can be calculated directly between two sets of samples. And the PCC is used to assess the linear correlation between two specific signals. With these two methods, we can evaluate the generation effect on two different scales.

Table 4 shows the MMD and average PCC between generated data and real samples. The lower MMD means better generative result, and the closer the PCC is to 1 means the more similar the two samples are. The results on the CWRU dataset show that GAN and BiGAN have relatively higher MMD and lower PCC, and BiGAN is slightly better. With the help of Wasserstein distance, the MMD of WGAN is lower than GAN and BiGAN and the result of PCC has also improved. The results of WGAN-GP and BiWGAN-GP are much better than the other methods, which suggests that Wasserstein distance with generation quality effectively improves the model's generative capability. Furthermore, the differences between WGAN-GP and BiWGAN-GP are minor, implying that their generation capabilities are similar. As for the UoC gear dataset, the numerical difference between the various models is smaller than the results on CWRU dataset. However, the generation effect reflected in the result is similar. Although WGAN-GP reached the best MMD in the UoC dataset, the gap between BiWGAN-GP and WGAN-GP is minor. Overall, BiWGAN-GP achieves the best quantitative assessment results.

Table 4. Quantitative assessment results of the generation quality.

	CWRU Dataset		UoC Dataset	
	MMD	PCC	MMD	PCC
GAN	2.5757	0.5307	0.2706	0.8482
WGAN	0.3052	0.7640	0.1960	0.8642
WGAN-GP	0.0334	0.8270	0.0215	0.8898
BiGAN	2.0498	0.6101	0.2393	0.8557
BiWGAN-GP	0.0226	0.8277	0.0221	0.8917

(2) Visualized assessment: To visualize the quality of generated results, we show the comparison between the original and generated frequency spectrum of the CWRU dataset in Figure 7. For clarity, the comparison of the UoC dataset is shown in Figure A1 in the appendix. Since the generator is trained unsupervisedly, the generated data is also unlabeled. Therefore, by iterating through the generated samples, we find the sample that is most similar to each category, i.e., samples with the smallest mean square error to original signals. The PCC between the generated and original signal of a specific category is shown in every subfigure.

Figure 7. Comparison between original and generated signals of CWRU dataset. (a) Original and generated signals of GAN. (b) Original and generated signals of WGAN. (c) Original and generated signals of WGAN-GP. (d) Original and generated signals of BiGAN. (e) Original and generated signals of BiWGAN-GP.

The visualization supports the results of the quantitative assessment. Figure 7a,d show that mode collapse is evident in GAN and BiGAN, which should be the main reason

for their higher MMD. It is noticeable that some significant frequency points in the origin are missing, but lots of unexpected spikes show up. Figure 7b shows that WGAN does not have mode collapse, but it seems to have a filtering effect that tends not to generate spikes, which also causes the error between the generated and real samples. It may be caused by gradient clipping in WGAN. In contrast, Figure 7c,e show that the difference of signals generated by WGAN-GP and BiWGAN-GP is difficult to distinguish, which is consistent with the results in the quantitative assessment. The features of the original spectrum are well-matched in the generated signals. Figure A1 shows that the generative results of the UoC dataset are relatively hard to distinguish by human eyes. However, we can tell that the signals generated by GAN and BiGAN have some meaningless fluctuations in the high-frequency part and some spikes in the gear spalling fault are missing. The results of WGAN-GP and BiWGAN-GP are better than others, and the PCC also confirmed it. The results suggest that Wasserstein distance with gradient penalty seems to be indispensable to improve the generation quality.

(3) Analysis of the training loss: The trend of training loss may explain the difference in model generation quality. Losses are recorded during the unsupervised pre-training stage. Figure 8 shows the training loss of generator G and encoder E of BiGAN and BiWGAN-GP, and the training loss of generator G of WGAN-GP. The results show that the loss of BiGAN has an unstable trend. Under the same training conditions, BiWGAN-GP shows a different result, that the loss of E and G (also means the Wasserstein distance) decreases steadily. The loss values of WGAN-GP are much lower than those of BiWGAN-GP, but the numerical comparison between them is meaningless because of the missing encoder and the different discriminator structures. It is worth noting that they both have a steady trend. These curves demonstrate that the stability of BiWGAN-GP's training is much improved and also explain why the generative ability of BiWGAN-GP is better than BiGAN.

Figure 8. The training loss of encoder and generator of BiGAN and BiWGAN-GP, and the training loss of generator of WGAN-GP on CWRU dataset.

4.3. Fault Diagnosis Experiment

To verify the fault diagnosis performance of the method proposed in this paper, a comparative experiment is conducted in this section. Four semi-supervised fault diagnosis methods are compared: original BiGAN [19], denoising autoencoder (DAE) [32], variational autoencoder (VAE) [33], and proposed BiWGAN-GP.

(1) Experiment setup: These methods have similar usage for fault diagnosis tasks, including unsupervised pre-training and supervised fine-tuning. First, all methods use the same amount of unlabeled data for pre-training, which allows the model to perform effective feature extraction. Then, the encoder is picked out from the model as a feature extractor. To evaluating the model's unsupervised learning ability, the encoder's parameters are frozen to prevent being trained in the next stage. After adding a linear classifier layer, supervised fine-tuning is conducted using different amounts of labeled data as the

training set. Since the main idea of this experiment is to evaluate the model's performance with insufficient labeled training data, the training set is smaller than the testing set. The number of testing samples of the CWRU and UoC dataset is 4000 and 900, respectively, and the number of training samples is set to 1%, 5%, and 10% of the testing set. Specifically, there are 40, 200, and 400 training samples of the CWRU dataset, and 9, 45, 90 training samples of the UoC dataset.

(2) Comparison of the diagnosis accuracy: Table 5 shows the comparison of average diagnosis accuracy under the various size of the training set. Under the same condition of training samples, the proposed method has significantly higher accuracy than others, which implies that the BiWGAN-GP can benefit more from the unsupervised training. Additionally, the BiWGAN-GP's result is much better than BiGAN's, which indicates the improvements are necessary and effective. Compare to the CWRU dataset, the UoC gear fault diagnosis task is relatively harder, therefore the average accuracy of the UoC dataset is lower, but the proposed method also reaches an acceptable result. The results demonstrate that the amount of labeled data is a direct factor in the accuracy. However, when the available labeled data is lacking, the feature representation ability obtained from unsupervised training is the crucial factor of diagnostic accuracy.

Table 5. Comparison of diagnosis accuracy(%) on different size of training set.

Method	Number of Labeled Training Samples					
	CWRU Dataset			UoC Dataset		
	40	200	400	9	45	90
DAE	94.65	97.78	98.74	38.82	73.96	84.23
VAE	79.95	89.06	94.30	37.74	75.56	81.22
BiGAN	74.22	84.86	92.09	48.04	53.26	55.33
BiWGAN-GP	97.14	99.21	99.42	51.07	85.70	91.97

To show the comparison more clearly, accuracy curves are plotted by increasing the number of training samples from 1% to 20% of the testing set. Figures 9 and 10 plot the accuracy curves on the CWRU dataset and UoC dataset, respectively. The results show that on the condition of insufficient labeled data, the proposed method can provide reliable fault diagnosis. Meanwhile, the increase of accuracy from BiGAN to BiWGAN-GP proves the effectiveness of the improvement in Section 2.3.2. It also can be concluded that more labeled data make better diagnosis performance, which partly explains why many previous works can achieve very high accuracy on fault diagnosis. With labeled data increasing, the performance difference between BiWGAN-GP and other methods decreases, but BiWGAN-GP is still better than others.

Figure 9. CWRU bearing fault diagnosis accuracy comparison.

Figure 10. UoC fault diagnosis accuracy comparison.

(3) Visualization of the extracted features: To further evaluate the feature learning ability of BiWGAN-GP, the extracted features are visualized using t-SNE [34]. Figure 11 shows the dimensional reduction result of the features extracted by the encoder of BiWGAN-GP on the CWRU dataset. Different color represents the category of working condition, and the label's meaning is as shown in Table 2. The result shows that different types of faults can be easily separated into clusters, which implies the good performance of the model's encoder. Figure 12 shows the visualization of feature extraction of the UoC dataset, and the scatter shows a similar result.

Figure 11. Visualization of feature extraction effects using t-SNE on CWRU dataset.

Figure 12. Visualization of feature extraction effects using t-SNE on UoC dataset.

5. Discussion

The above experiment verifies that the proposed method can generate real-like vibration signals and perform accurate semi-supervised fault diagnosis. The task of generating vibration signals shows that although Wasserstein distance can solve the mode collapse problem, gradient clipping will cause distortions, and the gradient penalty can restore the original signal pattern. It indicates that the improvements of the loss function of the GAN-based model are still worthy of thorough exploration. The fault diagnosis experiments demonstrate that although the number of labeled samples is an essential factor in the performance of the semi-supervised learning model; the utilization of unlabeled samples also has a significant impact on diagnostic accuracy. Therefore, using a large amount of unlabeled data more efficiently to improve the performance of the fault diagnosis methods is worthy of further investigation. In addition, it should be pointed out that currently available public datasets in the field of fault diagnosis are limited. More and various public datasets will help the development of the predictive maintenance community.

6. Conclusions

This article proposes a novel semi-supervised machinery fault diagnosis method called BiWGAN-GP. The proposed method utilizes the information of available unlabeled data effectively and reduces the model's need for the amount of labeled data, which can save the cost of manually labeling data. By using the Wasserstein distance to improve the training procedure, the model can be trained stably and reliably. The experiments show that the proposed method can generate real-like signals using unlabeled data which is easy to obtain, and also can perform accurate fault diagnosis using only a few labeled data. Furthermore, the structure of BiWGAN-GP is simple to modify. By replacing the fully connected layers with a one-dimensional convolutional layer, the model is capable to deal with the generation and diagnosis tasks of the raw time-domain signals.

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

Funding: This research was funded by National Natural Science Foundation of China under Grant no. U20A20201, Shandong Provincial Key Research and Development Program (Major Scientific and Technological Innovation Project), under Grant no. 2019JZZY010441, and the Fund of Taishan Industry Leading Talents Project.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found here: [\[25,27\]](#).

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

Appendix A. Visualization of the Generation Results of the UoC Dataset

Figure A1. *Cont.*

(c)

(d)

(e)

Figure A1. Comparison between original and generated signals of UoC dataset. (a) Original and generated signals of GAN. (b) Original and generated signals of WGAN. (c) Original and generated signals of WGAN-GP. (d) Original and generated signals of BiGAN. (e) Original and generated signals of BiWGAN-GP.

References

1. Lee, J.; Wu, F.; Zhao, W.; Ghaffari, M.; Liao, L.; Siegel, D. Prognostics and health management design for rotary machinery systems—Reviews, methodology and applications. *Mech. Syst. Signal Process.* **2014**, *42*, 314–334. [\[CrossRef\]](#)
2. Lee, G.H.; Park, Y.J.; Nam, J.S.; Oh, J.Y.; Kim, J.G. Design of a mechanical power circulation test rig for a wind turbine gearbox. *Appl. Sci.* **2020**, *10*, 3240. [\[CrossRef\]](#)
3. Hoang, D.T.; Kang, H.J. A survey on deep learning based bearing fault diagnosis. *Neurocomputing* **2019**, *335*, 327–335. [\[CrossRef\]](#)
4. Xu, Y.; Li, Z.; Wang, S.; Li, W.; Sarkodie-Gyan, T.; Feng, S. A hybrid deep-learning model for fault diagnosis of rolling bearings. *Measurement* **2021**, *169*, 108502. [\[CrossRef\]](#)
5. Wu, C.; Jiang, P.; Ding, C.; Feng, F.; Chen, T. Intelligent fault diagnosis of rotating machinery based on one-dimensional convolutional neural network. *Comput. Ind.* **2019**, *108*, 53–61. [\[CrossRef\]](#)
6. Duong, B.P.; Kim, J.Y.; Jeong, I.; Im, K.; Kim, C.H.; Kim, J.M. A deep-learning-based bearing fault diagnosis using defect signature wavelet image visualization. *Appl. Sci.* **2020**, *10*, 8800. [\[CrossRef\]](#)
7. Zhao, M.; Tang, B.; Deng, L.; Pecht, M. Multiple wavelet regularized deep residual networks for fault diagnosis. *Measurement* **2020**, *152*, 107331. [\[CrossRef\]](#)
8. Zhang, Y.; Zhou, T.; Huang, X.; Cao, L.; Zhou, Q. Fault diagnosis of rotating machinery based on recurrent neural networks. *Measurement* **2021**, *171*, 108774. [\[CrossRef\]](#)
9. Jia, F.; Lei, Y.; Guo, L.; Lin, J.; Xing, S. A neural network constructed by deep learning technique and its application to intelligent fault diagnosis of machines. *Neurocomputing* **2018**, *272*, 619–628. [\[CrossRef\]](#)
10. Qu, Y.; He, M.; Deutsch, J.; He, D. Detection of pitting in gears using a deep sparse autoencoder. *Appl. Sci.* **2017**, *7*, 515. [\[CrossRef\]](#)
11. Liu, S.; Xie, J.; Shen, C.; Shang, X.; Wang, D.; Zhu, Z. Bearing fault diagnosis based on improved convolutional deep belief network. *Appl. Sci.* **2020**, *10*, 6359. [\[CrossRef\]](#)
12. Wang, R.; Zhang, S.; Chen, Z.; Li, W. Enhanced generative adversarial network for extremely imbalanced fault diagnosis of rotating machine. *Measurement* **2021**, *180*, 109467. [\[CrossRef\]](#)
13. Pu, Z.; Cabrera, D.; Sánchez, R.V.; Cerrada, M.; Li, C.; de Oliveira, J.V. Exploiting generative adversarial networks as an oversampling method for fault diagnosis of an industrial robotic manipulator. *Appl. Sci.* **2020**, *10*, 7712. [\[CrossRef\]](#)
14. Liu, H.; Zhou, J.; Xu, Y.; Zheng, Y.; Peng, X.; Jiang, W. Unsupervised fault diagnosis of rolling bearings using a deep neural network based on generative adversarial networks. *Neurocomputing* **2018**, *315*, 412–424. [\[CrossRef\]](#)
15. Yu, K.; Ma, H.; Lin, T.R.; Li, X. A consistency regularization based semi-supervised learning approach for intelligent fault diagnosis of rolling bearing. *Measurement* **2020**, *165*, 107987. [\[CrossRef\]](#)
16. Zhao, X.; Jia, M.; Liu, Z. Semisupervised Deep Sparse Auto-Encoder with Local and Nonlocal Information for Intelligent Fault Diagnosis of Rotating Machinery. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–13. [\[CrossRef\]](#)
17. Chen, X.; Wang, Z.; Zhang, Z.; Jia, L.; Qin, Y. A semi-supervised approach to bearing fault diagnosis under variable conditions towards imbalanced unlabeled data. *Sensors* **2018**, *18*, 2097. [\[CrossRef\]](#)
18. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **2014**, *3*, 2672–2680.
19. Donahue, J.; Krähenbühl, P.; Darrell, T. Adversarial Feature Learning. In Proceedings of the 5th International Conference on Learning Representations, Toulon, France, 24–26 April 2017.
20. Mutlu, U.; Alpaydin, E. Training bidirectional generative adversarial networks with hints. *Pattern Recognit.* **2020**, *103*, 107320. [\[CrossRef\]](#)
21. Arjovsky, M.; Bottou, L. Towards Principled Methods for Training Generative Adversarial Networks. In Proceedings of the 5th International Conference on Learning Representations, Toulon, France, 24–26 April 2017.
22. Arjovsky, M.; Chintala, S.; Bottou, L. Wasserstein generative adversarial networks. In Proceedings of the 34th International Conference on Machine Learning, Sydney, Australia, 6–11 August 2017; Volume 70, pp. 214–223.
23. Gulrajani, I.; Ahmed, F.; Arjovsky, M.; Dumoulin, V.; Courville, A.C. Improved Training of Wasserstein GANs. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 5767–5777.
24. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. In Proceedings of the 3rd International Conference on Learning Representations, San Diego, CA, USA, 7–9 May 2015.
25. Case Western Reserve University (CWRU) Bearing Data Center. Available online: <https://csegroups.case.edu/bearingdatacenter/home> (accessed on 9 October 2021).
26. Zhao, Z.; Li, T.; Wu, J.; Sun, C.; Wang, S.; Yan, R.; Chen, X. Deep learning algorithms for rotating machinery intelligent diagnosis: An open source benchmark study. *ISA Trans.* **2020**, *107*, 224–255. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Cao, P.; Zhang, S.; Tang, J. Gear Fault Data. Available online: <https://doi.org/10.6084/m9.figshare.6127874.v1> (accessed on 9 October 2021).
28. Cao, P.; Zhang, S.; Tang, J. Preprocessing-free gear fault diagnosis using small datasets with deep convolutional neural network-based transfer learning. *IEEE Access* **2018**, *6*, 26241–26253. [\[CrossRef\]](#)
29. Salimans, T.; Goodfellow, I.; Zaremba, W.; Cheung, V.; Radford, A.; Chen, X. Improved techniques for training GANs. *Adv. Neural Inf. Process. Syst.* **2016**, *29*, 2234–2242.
30. Heusel, M.; Ramsauer, H.; Unterthiner, T.; Nessler, B.; Hochreiter, S. GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 6626–6637.

-
31. Gretton, A.; Borgwardt, K.M.; Rasch, M.J.; Schölkopf, B.; Smola, A.J. A Kernel Method for the Two-Sample-Problem. In *Advances in Neural Information Processing Systems 19*; MIT Press: Cambridge, MA, USA, 2006; pp. 513–520.
 32. Lu, C.; Wang, Z.Y.; Qin, W.L.; Ma, J. Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification. *Signal Process.* **2017**, *130*, 377–388. [[CrossRef](#)]
 33. Kingma, D.P.; Welling, M. Auto-Encoding Variational Bayes. In Proceedings of the 2nd International Conference on Learning Representations, Banff, AB, Canada, 14–16 April 2014.
 34. Van der Maaten, L.; Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.