COMPLEX SYSTEM MODELING AND SIMULATION
ISSN 2096-9929 05/06 pp 55–64
Volume 1, Number 1, March 2021
DOI: 10.23919/CSMS.2021.0006

A Threshold-Control Generative Adversarial Network Method for Intelligent Fault Diagnosis

Xinyu Li, Sican Cao, Liang Gao, and Long Wen*

Abstract: Fault diagnosis plays the increasingly vital role to guarantee the machine reliability in the industrial enterprise. Among all the solutions, deep learning (DL) methods have achieved more popularity for their feature extraction ability from the raw historical data. However, the performance of DL relies on the huge amount of labeled data, as it is costly to obtain in the real world as the labeling process for data is usually tagged by hand. To obtain the good performance with limited labeled data, this research proposes a threshold-control generative adversarial network (TCGAN) method. Firstly, the 1D vibration signals are processed to be converted into 2D images, which are used as the input of TCGAN. Secondly, TCGAN would generate pseudo data which have the similar distribution with the limited labeled data. With pseudo data generation, the training dataset can be enlarged and the increase on the labeled data could further promote the performance of TCGAN on fault diagnosis. Thirdly, to mitigate the instability of the generated data, a threshold-control is presented to adjust the relationship between discriminator and generator dynamically and automatically. The proposed TCGAN is validated on the datasets from Case Western Reserve University and Self-Priming Centrifugal Pump. The prediction accuracies with limited labeled data have reached to 99.96% and 99.898%, which are even better than other methods tested under the whole labeled datasets.

Key words: generative adversarial network; limited labeled data; discriminator; fault diagnosis

1 Introduction

With the rapid development of smart manufacturing, the machine complexity is growing increasing, which makes it to be a hard task for the maintenance. One tiny fault in the machine may cause a chain reaction, finally leading to the breakdown of the whole system[1]. To avoid this kind of situation, fault diagnosis has become an indispensable way to keep the stability and reliability of machinery[2].

Traditional ways to address the problem of fault diagnosis have aroused great attentions in the past decades[3]. However, it is getting hard for the traditional model based methods to extract the decisive features and then to build feature models. Thus, the demand for fault diagnosis meets data-driven methods, which have received more attentions in the smart manufacturing.

Among fault diagnosis methods, deep learning (DL) method becomes attractive for its characteristic of auto-feature extraction, which means that the features can be selected by the networks automatically instead of by hand[4]. Another reason for its popularity lies on the ability of data processing promoted by the hardware improvement, making it possible for DL methods to handle with massive high-dimensional data. Several DL methods have been successfully applied for fault diagnosis[5].

• Xinyu Li, Sican Cao, and Liang Gao are with the State Key Laboratory of Digital Manufacturing Equipment & Technology, Huazhong University of Science and Technology, Wuhan 430074, China. E-mail: lixinyu@mail.hust.edu.cn; 792286300@qq.com; gaoliang@mail.hust.edu.cn.
• Long Wen is with the School of Mechanical Engineering and Electronic Information, China University of Geosciences, Wuhan 430074, China. E-mail: wenlong@cug.edu.cn.
∗ To whom correspondence should be addressed.
Manuscript received: 2021-03-17; revised: 2021-04-09; accepted: 2021-04-11

The DL method often requires abundant data to address the fault diagnosis. It is widely acknowledged that the increase of labeled data will contribute to a promotion on the performance of the DL network. Even the data can be now easily collected in a large amount, but the labeled data are still hard and costly to obtain voluminously in practice as the labeling is usually tagged by hand. As the result, the number of labeled data is relatively low. Therefore, how to improve the final accuracy with limited labeled data for DL model becomes a very important problem.

This research proposes a new threshold-control generative adversarial network (TCGAN) method, which is based on generative adversarial network (GAN). The idea is to generate pseudo samples which have the similar distribution as input labeled data, thus the training dataset can be enlarged and the performance of classification can be improved further. Firstly, in the data pre-processing process, the 1D vibration signals will be transformed into 2D images, which would be fed as the input of TCGAN. Secondly, in the training process, the pseudo data with similar distribution of the input 2D images can be generated by generator, and the fault features would be extracted by discriminator from the enlarged dataset. The learned fault features can be classified by a softmax layer followed by the discriminator. Both discriminator and generator are designed as convolutional neural network (CNN) models for their powerful ability of feature extraction. Thirdly, a threshold-control method is proposed to adjust the relationship of synchronization between discriminator and generator dynamically and automatically in order to decrease the instability causing by the limitation of data. Finally, several experiments are conducted on bearing dataset from Case Western Reserve University (CWRU) and Self-Priming Centrifugal Pump (SPCP) to test the performance of TCGAN under limited labeled samples. The comparison of TCGAN with other methods under the whole labeled datasets is conducted, and the results show the effectiveness of TCGAN.

The rest of the research is composed as follows: Section 2 presents the related work. Section 3 shows the introduction of GAN. Section 4 contains the specific steps of TCGAN for fault diagnosis. Section 5 shows the experiment results. Conclusions and future works are presented in Section 6.

2 Literature Review

With the growing development of hardware and software, fault diagnosis has achieved increasing improvement in this field.

Statistical methods for fault diagnosis are applied to extract the most essential features under a statistical frame and are generally combined with machine learning methods. Tsai et al.[6] used independent component analysis to get independent basis images from solar cell sub images of solar modules defect detection, and each solar cell sub image was reconstructed as the combination of the basis images in the inspection stage. Dong and Luo[7] reduced the dimension of sensitive features extracted from vibration signals with principal component analysis, following by the LS-support vector machine (LS-SVM) model for bearing degradation prediction.

Non-statistical based methods, such as neural network (NN), are used for their nonlinear approximation for various applications. Sparse filtering studied by Lei et al.[3] was used for feature extraction directly from vibration signals, followed by a softmax function for classification. Jia et al.[8] demonstrated the effectiveness of stacked autoencoder which was trained by frequency spectra on five fault datasets. Gan and Wang[9] studied a new hierarchical diagnosis network structure, with one part of two successive layers followed by one deep belief network (DBN) for fault detection and the other part composed of several DBNs for the further fault classification.

Chen et al.[10] designed a CNN structure using statistical measures and spectral information as input for the fault diagnosis of gearboxes. Wang et al.[11] studied a CNN with its parameters selected by particle swarm optimization. Shao et al.[12] fine-tuned VGG16 to achieve a faster convergence, and obtained significant promotion from 94.8% to 99.64%. However, the performance of CNN-based fault diagnosis methods relies on the volume of labeled data greatly. This research investigates the TCGAN method, which can generate data with the similar distribution of the input labeled data through an adversarial training, to handle with the situation with limited labeled data. In TCGAN, the labeled dataset can be enlarged, promoting the performance of fault classification.

3 Introduction of GAN

GAN has been applied in image generation, video generation, and some other interesting researches[13]. The structure of GAN consists of the generator and discriminator parts. As presented in Fig. 1, the generator

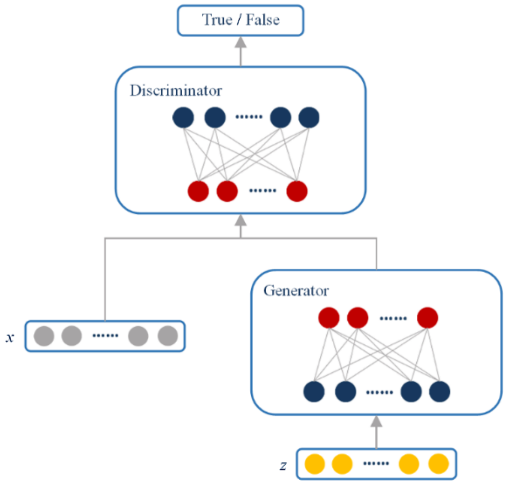

Fig. 1 Structure of GAN.

part aims to learn the data distribution from the real labeled data, while the discriminative part is to classify that whether the samples belong to the real labeled data or the pseudo samples. Both generator and discriminator parts should be trained in the training of GAN.

The training process of GAN can be presented as Formula (1):

$$E_{x\sim p_{\text{data}}(x)}\left[\log D\left(x\right)\right]+E_{z\sim p_z(z)}\left[\log\left(1-D\left(G\left(z\right)\right)\right)\right] \tag{1}$$

where E denotes the expected value. x denotes the samples from real data and z represents the noise. D and G mean the discriminative and generative parts of GAN, respectively. p_{data} and p_z are the data distribution of real data and a random noise, respectively.

The reason to choose GAN for fault diagnosis is its possibility for classification with limited labeled data by generation pseudo data during the adversarial training of discriminator and generator. In GAN, the generator will learn and imitate the distribution of real data, and thus the training dataset can be enlarged. Finally, the performance of classification can be improved in the limited labeled data environment.

However, there are still some obstacles in the application of GAN on fault diagnosis. Firstly, the selection of network structures for discriminative part and generative part is essential. It is obviously not powerful enough to deal with data of high complexity when these two parts are designed as multilayer perceptron (MLP). Secondly, GAN is an unsupervised method and it cannot be used directly for classification. Thirdly, the balance between discriminator and generator can be easily broken, which may result in instability and even collapse[14].

In this research, TCGAN is developed based on the following aspects: (1) both structures of TCGAN are designed as CNN for its powerful and representative ability of feature extraction; (2) a softmax layer is added after the discriminator of TCGAN for the fault classification; and (3) a threshold-control method is proposed in TCGAN to adjust the synchronization between discriminator and generator dynamically and automatically.

4 Proposed TCGAN for Fault Diagnosis

This section contains three parts: (1) data pre-processing process; (2) the training method of TCGAN; and (3) threshold-control dynamic adjustment of TCGAN.

4.1 Data pre-processing process

To make full use of the image processing ability of CNN, the vibration signals in 1D format will be transformed into 2D images using conversion process[15]. The conversion process is shown in Fig. 2. Firstly, M sequencing pieces of signals will be selected, and the length for each piece is also M. Secondly, these pieces will be connected to form a sequence with the length of M^2. Finally, the sequence of signals will be converted into images by Eq. (2):

$$I\left(j,k\right)=\text{round}\left\{\frac{L\left(j\times M+k\right)-\text{Min}\left(L\right)}{\text{Max}\left(L\right)-\text{Min}\left(L\right)}\times 255\right\} \tag{2}$$

where $L(i)$, $i=1,\ldots,M^2$ denotes the signals. The round( ) function means it will return an integer and $I(j,k)$, $j=1,\ldots,M, k=1,\ldots,M$, represents the pixel values varying from 0 to 255 after the conversion, and it can be seen that the size of the converted image is $M\times M$.

4.2 Training method of TCGAN

The TCGAN will be trained with the transformed 2D images. The generator part in TCGAN is used for data generation, and the softmax layer is used in discriminator part for fault classification. Both discriminator and generator are designed as CNN structures.

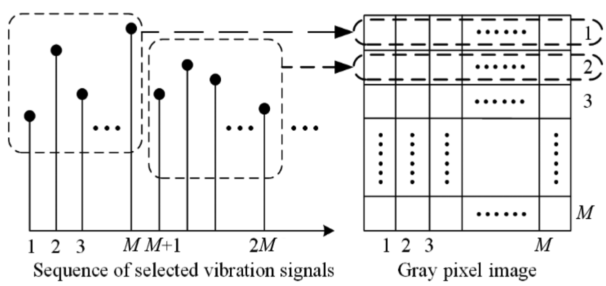

Fig. 2 Signal-to-image conversion.

4.2.1 Network structures for discriminator and generator

In this research, the structures of discriminator and generator part in TCGAN are designed as CNN network which are presented in Figs. 3 and 4, respectively. The discriminative part is designed as the classical LeNet-5 network. As presented in Fig. 3, the discriminator part has four convolution layers followed with two full-connected (FC) layers. It should be noted that the pooling layer is followed after each convolution layer and it is not presented in Fig. 3. The ReLU activation function is chosen for every layer for the nonlinear mapping. L2 regularization is also adopted to protect the network from overfitting. The input of discriminative structure is 2D images of real data or generated samples which are transformed from 1D vibration signals, and the output of discriminative part is the prediction of fault classification.

The generator part has four deconvolution layers. The deconvolution layer is the inverse process of convolution, and it will restore the size and information of images back into the condition where they have been convolved before. The activation function is ReLU, except that the output layer uses Tanh function. For every deconvolution layer, batch normalization method is used to avoid overfitting and to speed up the convergence. The input of the generative structure is random noises and the output of generator is images with the same size of converted 2D real data, which will be fed as the input of discriminator.

4.2.2 Softmax for classification in TCGAN

GAN is an unsupervised method with unlabeled data and the output of discriminator part is true or false. To

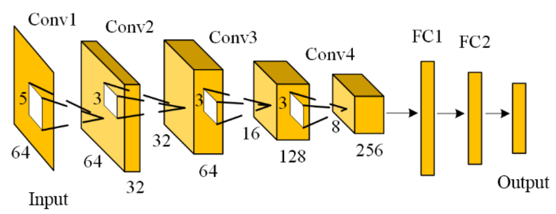

Fig. 3 Discriminator of TCGAN method.

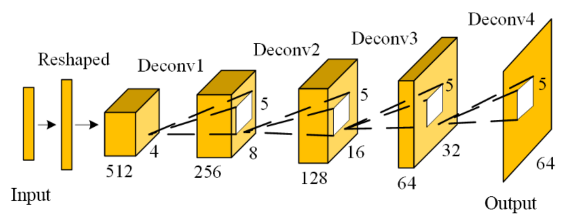

Fig. 4 Generator of TCGAN method.

make it possible for fault classification with labeled data, a softmax layer is added in discriminative structure in TCGAN. The biggest probability value will be chosen as the fault prediction label for the input data.

4.2.3 Performance evaluation

For the performance evaluation, the accuracy is selected as the main index to evaluate the classification precision for fault diagnosis. The accuracy will be calculated for the percentage of correctly classified data in one batch as Eq. (3):

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{P + N} \times 100\% \tag{3}$$

where TP is short for true positive, representing the number of normal data correctly classified as the normal ones. TN is short for true negative, representing the number of faults correctly classified. P and N are the number of normal data and fault data, respectively. During the training process, the accuracy of testing dataset will be calculated to evaluate the performance of TCGAN.

4.3 Dynamic adjustment of TCGAN with threshold-control method

As pointed in Ref. [14], the relationship between the discriminator and generator parts may lead the network to an unstable situation, and the balance of these two parts becomes unstable when the number of given labeled data is less. Generally, the relationship between the discriminator and generator parts is set unchanged during the training process. In this research, a new threshold-control method is proposed, named TCGAN, in which this relationship can be adjusted through the dynamic change of synchronization between discriminator and generator.

In TCGAN, at the beginning of the training process, the discriminator will be trained a little bit more than the generator to accelerate the convergence of TCGAN. When the network is nearly well trained, the threshold will slow down the training speed of discriminative part to protect the network from overtraining.

The flowchart of the threshold-control on TCGAN is explained in Fig. 5, where D stands for discriminator part and G is generator part. At the initialization process, a threshold value is set. When the network is initialized, the initial value for the relationship between D and G is set as 2 (TrainRelaSteps=2), which means that discriminator will be trained twice for one training of generator. For every 200 training steps, a testing accuracy (TestAcc) will be calculated to evaluate

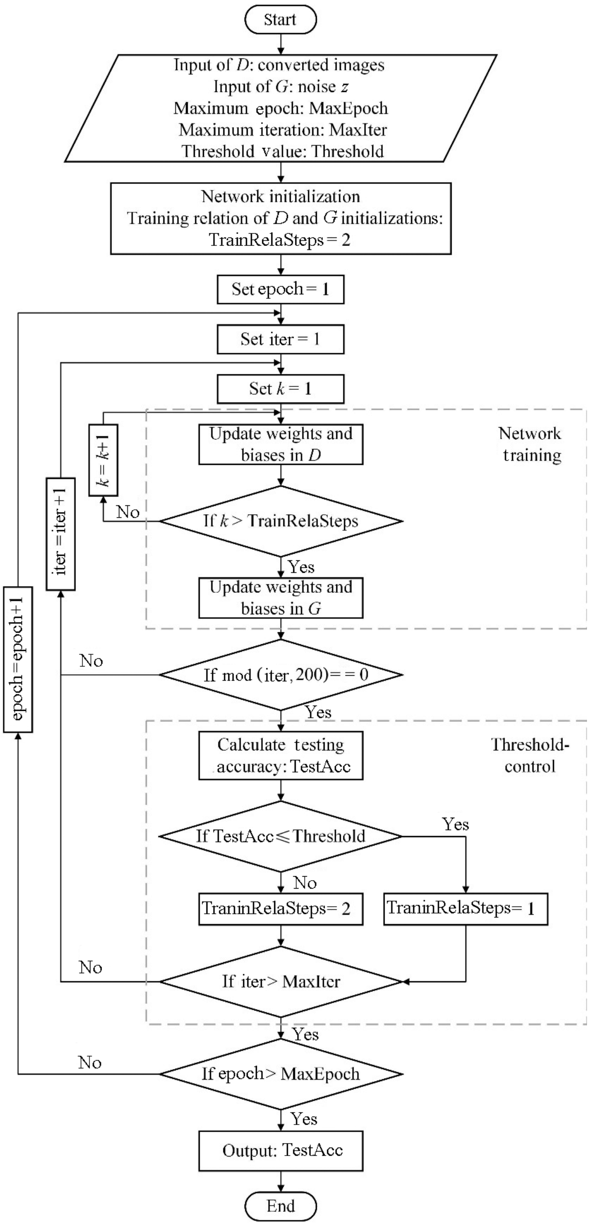

Fig. 5 Flowchart of the threshold-control method.

whether the training strategy between D and G should be adjusted. When the testing accuracy has not reached the threshold value, the network will keep the training strategy as it is. And the parameter TrainRelaSteps will be adjusted to 1 when the testing accuracy gets higher than the threshold value, meaning the training strategy will be changed into one discriminative training for one generative training. When the testing accuracy goes down, the threshold will force TCGAN back to the strategy before. Through this way, the relationship between D and G can be adjusted automatically and dynamically.

As shown in Fig. 5, the threshold-control method is validated after each assigned step (in this research, it is set as 200) of the network training instead of every training step. In this way, the network can get enough time to adjust itself to get more stable values of weights and biases during the updating.

5 Experiment Result

This section shows the experimental results of TCGAN on CWRU and SPCP datasets. The results of each experiment are the average value for ten repeats. For each case study, there are two kinds of experiments: (1) The comparisons under different capacities of limited labeled datasets, in which TCGAN with limited labeled data is tested and the accuracies reach as high as 99.96% and 99.898% on CWRU and SPCP dataset, respectively. The effectiveness of threshold-control method is also validated. (2) The comparisons with other methods, in which TCGAN is tested with the whole labeled data for a fair comparison to validate it further.

TCGAN is conducted under Ubuntu with Titian XP GPU, and the hyper-parameters are as follows: batch size is 100; learning rate is 0.0001; the dimension of the random noise z fed into generator is 100; both discriminator and generator are optimized with AdamOptimizer; L2 regularization is adopted with the penalty set as 1×10^{-5} to avoid overfitting for discriminative training; and the threshold value is set as 0.99.

5.1 Case 1: CWRU bearing dataset

5.1.1 Data description

In this section, TCGAN will be tested on the CWRU bearing dataset. There are four conditions in CWRU dataset, and they are the normal condition, roller fault, inner race fault, and outer race fault. Each fault type contains three damage sizes, which are 0.18 mm, 0.36 mm, and 0.54 mm, respectively. So there will be ten classifications. The proportion of train and test data is set as five to one. The whole labeled data are separated into five subsets for comparison, which are named as follows for convenience: 2000 samples, 1000 samples, 750 samples, 500 samples, and 250 samples. In this way, experiments can be conducted to validate the potential of TCGAN in limited labeled data.

5.1.2 Comparisons under different capacities of limited labeled datasets

To validate the ability of TCGAN with limited labeled data, different experiments are conducted. TCGAN with and without threshold is also tested. The experiments are conducted on 1000 samples, 750 samples, 500 samples, and 250 samples. The test accuracies are shown in Fig. 6, and the curves are very similar, which indicate that the number of data samples has no significant effects on the convergence of accuracy.

Accuracy comparisons with/without threshold are presented in Table 1, the best accuracy is obtained under 1000 samples, with 99.96%, showing the potential of TCGAN. The accuracies under 250, 500, and 750 samples with threshold method are 98.67%, 99.69%, and 99.86%, respectively, showing that the performance will grow better with support of more labeled data.

Table 1 also demonstrates the effectiveness of the threshold-control method. Higher accuracies can be achieved when TCGAN using threshold-control method in all subsets. When TCGAN is conducted under 1000 samples, the accuracy of threshold-control method is 99.96%, while that in non-threshold-control method is 99.91%. This promotion grows up to 3.80% when TCGAN is conducted under 250 samples. It can be inferring that, along with the decrease of data, the performances of TCGAN become worse. But the TCGAN without threshold-control method can be affected more. This result indicates that the threshold-control method can help improve the performance of TCGAN. Additional, threshold-control method can help TCGAN achieve a faster and more stable convergence, as illustrated in Fig. 7.

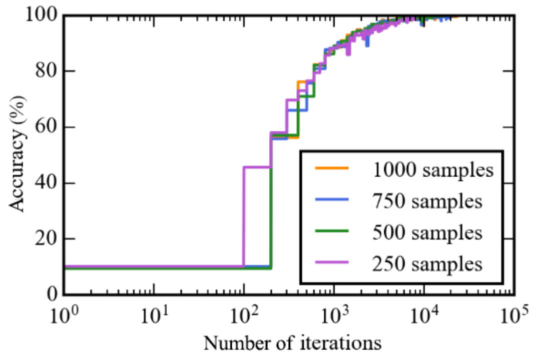

Fig. 6 Test accuracies under different capacities of datasets in Case 1.

Table 1 Accuracy comparisons with/without threshold in different samples of limited labeled datasets in Case 1.

Dataset	Accuracy (%)	
	With threshold	Without threshold
1000 samples	99.96	99.91
750 samples	99.86	99.80
500 samples	99.69	99.18
250 samples	98.67	94.87

5.1.3 Comparisons with other methods with the whole labeled data

In the subsection, TCGAN is compared with other

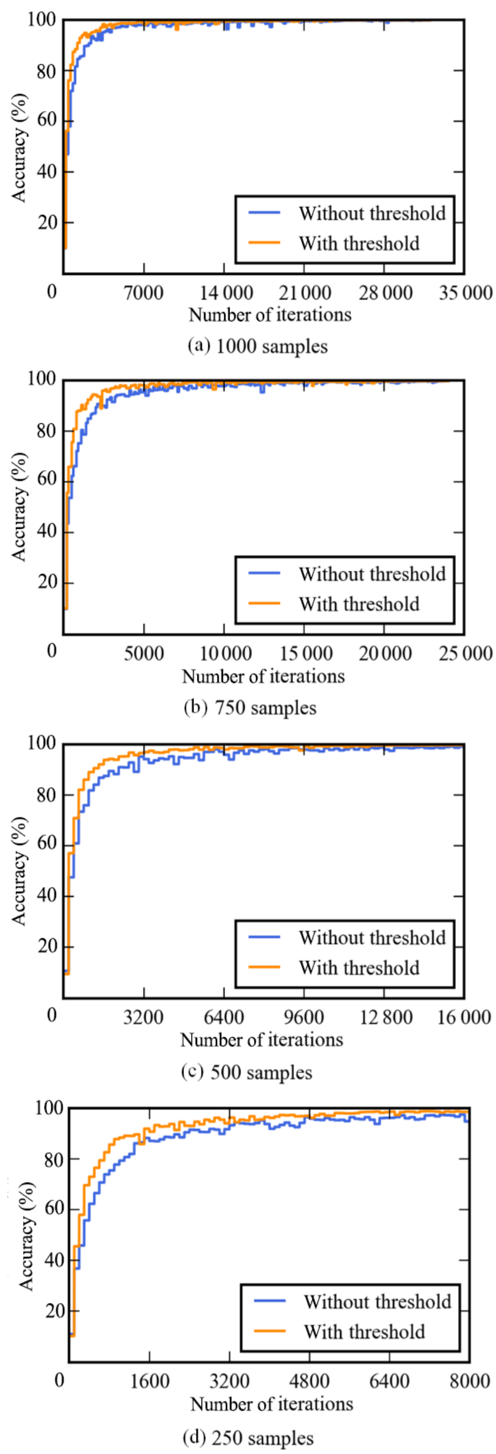

Fig. 7 Accuracy comparisons with/without threshold under different capacities of limited labeled datasets in Case 1.

methods, including statistic-based methods like EEMD+AR+SVM[2] and SVM with EEMD[16], and non-statistical methods such as DCNN[15], sparse filter (SF)[3], HDBN[9], sparse auto-encoder (SAE)[2], and ADCNN[17]. For a fair comparison, TCGAN is tested under the whole data and the results are presented in Table 2.

In Table 2, TCGAN reaches the mean accuracy of 99.99%, outperforming all the other methods. It should be noticed that, the mean accuracy of 99.96% can be obtained with 1000 samples in Table 1, which is very close to 99.99% conducted under the whole data, meaning a nearly same result can still be reached by cutting half of the labeled data. This result validates the effectiveness of TCGAN within limited data.

Figure 8 shows the logarithmic coordinate diagram of loss, where the loss decreases quickly at the beginning of training, then slows down to its convergence. Figure 9 illustrates the train and test accuracies, and the test accuracy grows up in stairs during the first hundreds of iterations because it was measured intermittently for every 200 iterations.

Table 2 Accuracy comparison results with other methods in Case 1.

Method	Accuracy (%)	Method	Accuracy (%)
TCGAN	99.99	SF	99.66
EEMD+AR+SVM	98.65	HDBN	99.03
SVM with EEMD	97.91	SAE	92.20
DCNN	99.79	ADCNN	98.10

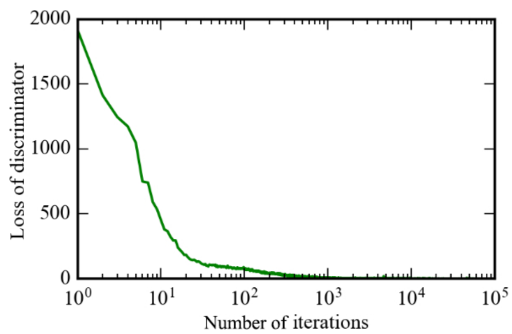

Fig. 8 Loss of discriminator with 2000 samples in Case 1.

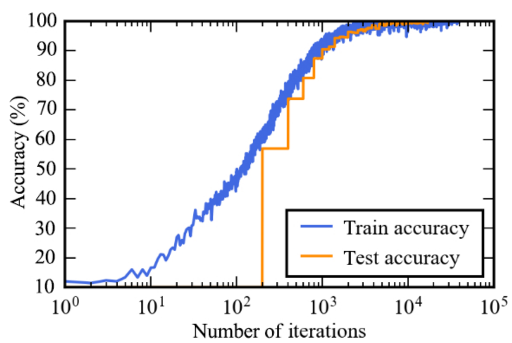

Fig. 9 Train and test accuracy with 2000 samples in Case 1.

5.2 Case 2: SPCP bearing dataset

5.2.1 Data description

In this case, TCGAN is validated on SPCP dataset, which is composed of five health conditions, which include the normal condition, impeller wearing, inner race wearing, outer race wearing, and bearing roller wearing. The proportion of train and test data is set as five versus one. SPCP dataset is a bit smaller than CWRU dataset, and we depart it into 2000 samples, 1000 samples, and 500 samples.

5.2.2 Comparisons with different capacities of limited labeled datasets

This part presents the experiment results under 1000 samples and 500 samples, and the mean accuracies with threshold are 99.898% and 99.536%, respectively, as shown in Table 3.

This result indicates that TCGAN can obtain good performances with limited labeled data in this dataset. The effectiveness of threshold-control method is also demonstrated by comparisons without threshold-control method. When tested with 1000 samples, the accuracy of threshold-added method is 99.898%, and the accuracy without threshold is 99.797%. The difference between them is 0.101%, which is extremely small. When tested under 500 samples, the difference grows up to 0.644%, where the accuracy of threshold-control method is 99.536% and the accuracy without threshold-control method is 98.892%. The result demonstrates that the threshold-control method can improve the performance, and this superiority grows apparent with the decrease of data. The convergence curves given in Fig. 10 show a more stable and a faster convergence with threshold.

5.2.3 Comparisons with other methods with the whole labeled data

When compared TCGAN with other methods, the competitors selected are SBTDA[18], SURF-based PNN[19], and CNN[15]. Their results are presented in

Table 3 Accuracy comparisons with/without threshold in different samples of limited labeled datasets in Case 2.

Dataset	Accuracy (%)	
	With threshold	Without threshold
1000 samples	99.898	99.797
500 samples	99.536	98.892

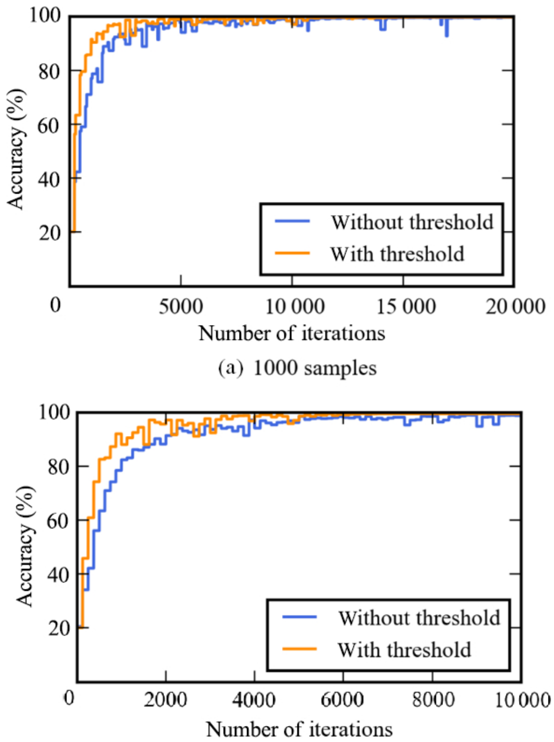

Fig. 10 Comparisons with/without threshold under different capacities of limited labeled datasets in Case 2.

Table 4. The accuracy of proposed TCGAN reaches as high as 99.985%. The curves of loss function and accuracy are presented in Figs. 11 and 12, respectively. The loss converges with a fast speed at the beginning, and then it becomes gentle, and finally a stable and high accuracy is reached.

Compared Table 3 with Table 4, the accuracy under 1000 samples is 99.898%, which is very close to the result under the whole labeled data, which is 99.985%. And the result under 1000 samples can still get a superior accuracy than all the other methods in Table 4, which are conducted under the whole data. This indicates that TCGAN has the obvious advantage for dealing with limited labeled data by comparing with other methods.

Table 4 Accuracy comparison results with other methods in Case 2.

Method	Accuracy (%)	Method	Accuracy (%)
TCGAN	99.985	SURF-based PNN	98.330
SBTDA	99.600	CNN	97.481

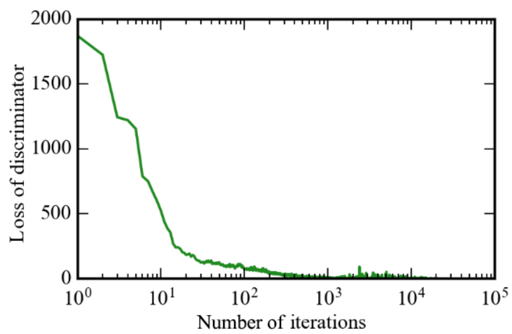

Fig. 11 Loss of discriminator with 2000 samples in Case 2.

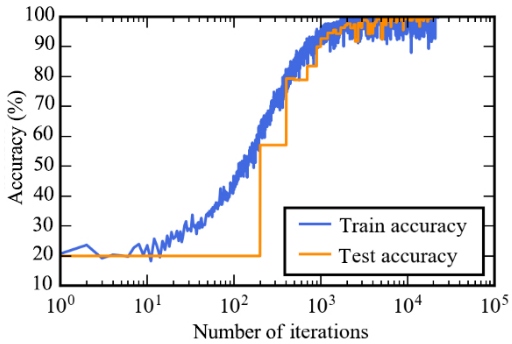

Fig. 12 Train and test accuracy with 2000 samples in Case 2.

6 Conclusion

In this research, a new TCGAN is proposed for the fault diagnosis with limited labeled data. In TCGAN, the generator is designed and it can imitate the distribution of the input labeled data, and generate pseudo data to enlarge the whole training dataset. The discriminator is added with a softmax layer for the fault classification. A threshold-control method is proposed in TCGAN to adjust the synchronization between discriminator and generator dynamically and automatically. Several experiments are conducted under CWRU and SPCP datasets. The accuracies under limited labeled data achieve 99.96% and 99.898% for the two datasets, and these results validate the effectiveness of threshold-control method with limited labeled data.

The limitations of this research for the real applications are the following aspects. Firstly, the synchronization between the generator and discriminator can be further investigated for the general purpose. Secondly, the designed CNN-based discriminator and generator structures can be combined with the newly developed CNN models. The future direction can be extended in the following ways: (1) A new adaptive synchronization mechanism can be developed; and (2) more powerful CNN models can be tested in the GAN for the fault diagnosis.

Acknowledgment

This work was supported in part by the National Key R&D Program of China (No. 2018AAA0101700), the National Natural Science Foundation of China (No. 51805192), and the State Key Laboratory of Digital Manufacturing Equipment and Technology of Huazhong University of Science and Technology (No. DMETKF2020029).

References

[1] L. Wen, X. Li, and L. Gao, A new reinforcement learning based learning rate scheduler for convolutional neural network in fault classification, *IEEE Transactions on Industrial Electronics*, doi: 10.1109/TIE.2020.3044808.

[2] Y. Qi, C. Shen, D. Wang, J. J. Shi, X. X. Jiang, and Z. K. Zhu, Stacked sparse autoencoder-based deep network for fault diagnosis of rotating machinery, *IEEE Access*, vol. 5, pp. 15 066–15 079, 2017.

[3] Y. Lei, F. Jia, J. Li, S. B. Xing, and S. X. Ding, An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data, *IEEE Transactions on Industrial Electronics*, vol. 63, no. 5, pp. 3137–3147, 2016.

[4] Y. LeCun, Y. Bengio, and G. Hinton, Deep learning, *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[5] L. Wen, L. Gao, X. Li, and B. Zeng, Convolutional neural network with automatic learning rate scheduler for fault classification, *IEEE Transactions on Instrumentation and Measurement*, vol. 70 pp. 1–12, 2021.

[6] D. M. Tsai, S. C. Wu, and W. Y. Chiu, Defect detection in solar modules using ICA basis images, *IEEE Transactions on Industrial Informatics*, vol. 9, no. 1, pp. 122–131, 2013.

[7] S. Dong and T. Luo, Bearing degradation process prediction based on the PCA and optimized LS-SVM model, *Measurement*, vol. 46, no. 9, pp. 3134–3152, 2013.

[8] F. Jia, Y. Lei, J. Lin, X. Zhou, and N. Lu, Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data, *Mechanical Systems and Signal Processing*, vol. 72, pp. 303–315, 2016.

[9] M. Gan and C. Wang, Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of rolling element bearings, *Mechanical Systems and Signal Processing*, vol. 72, pp. 92–104, 2016.

[10] Z. Q. Chen, C. Li, and R. V. Sanchez, Gearbox fault identification and classification with convolutional neural networks, *Shock and Vibration,* vol. 2015, p. 390134, 2015.

[11] F. Wang, H. K. Jiang, H. D. Shao, W. J. Duan, and S. P. Wu, An adaptive deep convolutional neural network for rolling bearing fault diagnosis, *Measurement Science and Technology*, vol. 28, p. 095005, 2017.

[12] S. Shao, S. McAleer, R. Yan, and P. Baldi, Highly-accurate machine fault diagnosis using deep transfer learning, *IEEE Transactions on Industrial Informatics*, vol. 15, no. 4, pp. 2446–2455, 2018.

[13] A. Dosovitskiy and T. Brox, Generating images with perceptual similarity metrics based on deep networks, presented at Advances in Neural Information Processing Systems (NIPS), Barcelona, Spain, 2016, pp. 658–666.

[14] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, Generative adversarial nets, presented at Advances in Neural Information Processing Systems (NIPS), Montreal, Canada, 2014, pp. 2672–2680.

[15] L. Wen, X. Y. Li, L. Gao, and Y. Y. Zhang, A new convolutional neural network based data-driven fault diagnosis method, *IEEE Transactions on Industrial Electronics*, vol. 65, no. 7, pp. 5990–5998, 2018.

[16] X. Zhang, Y. Liang, and J. Zhou, A novel bearing fault diagnosis model integrated permutation entropy, ensemble empirical mode decomposition and optimized SVM, *Measurement*, vol. 69, pp. 164–179, 2015.

[17] X. Guo, L. Chen, and C. Shen, Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis, *Measurement*, vol. 93, pp. 490–502, 2016.

[18] Y. Zhang, X. Li, L. Gao, and P. G. Li, A new subset based deep feature learning method for intelligent fault diagnosis of bearing, *Expert Systems with Applications*, vol. 110, pp. 125–142, 2018.

[19] C. Lu, Y. Wang, M. Ragulskis, and Y. Cheng, Fault diagnosis for rotating machinery: A method based on image processing, *PLoS ONE*, vol. 11, no. 10, p. e0164111, 2016.

Liang Gao received the PhD degree in mechatronic engineering from Huazhong University of Science and Technology, China in 2002.

He is a professor at the Department of Industrial & Manufacturing System Engineering, State Key Laboratory of Digital Manufacturing Equipment & Technology, School of Mechanical Science & Engineering, Huazhong University of Science and Technology. He has published more than 170 refereed papers. His research interests include operations research and optimization, big data, machine learning, etc.

Xinyu Li received the PhD degree in industrial engineering from Huazhong University of Science and Technology, China in 2009.

He is a professor at the Department of Industrial & Manufacturing System Engineering, State Key Laboratory of Digital Manufacturing Equipment & Technology, School of Mechanical Science & Engineering, Huazhong University of Science and Technology. He has published more than 80 refereed papers. His research interests include intelligent algorithm, big data, machine learning, etc.

Long Wen received the PhD degree in industrial engineering from Huazhong University of Science and Technology, China in 2014.

He is a professor at the School of Mechanical Engineering and Electronic Information, China University of Geosciences. He had published more than 20 refereed papers, and his research interests include deep learning, automatic machine learning, fault diagnosis, intelligent algorithm, etc.

Sican Cao received the BS degree in mechatronic engineering from Xidian University, China in 2016, and the master degree from Huazhong University of Science and Technology in 2019. Now she is working at China Carrier RTAC, HUAWEI Technical Service Co., LTD.