Hindawi
Shock and Vibration
Volume 2020, Article ID 8863388, 12 pages
https://doi.org/10.1155/2020/8863388

Research Article

Multisignal VGG19 Network with Transposed Convolution for Rotating Machinery Fault Diagnosis Based on Deep Transfer Learning

Jianye Zhou,[1] Xinyu Yang,[2] Lin Zhang,[2] Siyu Shao ,[2] and Gangying Bian[2]

[1] *Graduate School of Air Force Engineering University, Xi'an 710051, China*
[2] *Air and Missile Defence College of Air Force Engineering University, Xi'an 710051, China*

Correspondence should be addressed to Siyu Shao; cathygx.sy@gmail.com

Received 23 July 2020; Revised 11 November 2020; Accepted 25 November 2020; Published 8 December 2020

Academic Editor: Changqing Shen

To realize high-precision and high-efficiency machine fault diagnosis, a novel deep learning framework that combines transfer learning and transposed convolution is proposed. Compared with existing methods, this method has faster training speed, fewer training samples per time, and higher accuracy. First, the raw data collected by multiple sensors are combined into a graph and normalized to facilitate model training. Next, the transposed convolution is utilized to expand the image resolution, and then the images are treated as the input of the transfer learning model for training and fine-tuning. The proposed method adopts 512 time series to conduct experiments on two main mechanical datasets of bearings and gears in the variable-speed gearbox, which verifies the effectiveness and versatility of the method. We have obtained advanced results on both datasets of the gearbox dataset. The dataset shows that the test accuracy is 99.99%, achieving a significant improvement from 98.07% to 99.99%.

1. Introduction

Fault diagnosis refers to the status monitoring of equipment, which has reached the prediction of its fault time and the classification of faults. With the continuous development of intelligent fault diagnosis, a series of fault diagnosis methods such as expert system [1], artificial neural network [2], fuzzy control [3], and fault tree [4] have appeared.

However, the above models are based on manual feature extraction for fault diagnosis. When feature extraction does not meet the model requirements, the diagnosis effect will no longer be excellent. Second, manual features are used for different classification tasks. This means that the features used for accurate predictions are not suitable for other scenarios in some cases. It is difficult to design a set of functions that can generate reliable predictions under all conditions.

As a new algorithm based on artificial neural networks, deep learning [5, 6] solves the problems of gradient disappearance in the BP neural network [7, 8]. The deep architecture has multiple hidden layers, which is able to learn hierarchical representations directly from the original data. Through model training, the deep architecture can automatically choose proper representations based on the training data to help make accurate predictions in subsequent classification stages. Due to the above advantages, deep learning is widely used in image processing [9, 10], medical image analysis [11], speech recognition [12, 13], target detection [14, 15], text detection [16, 17], natural language processing [18, 19], and other fields.

Since the deep belief network (DBN) was first applied to fault diagnosis in 2013 [20], the application of deep learning in the field of fault diagnosis in rotating machinery has entered a stage of rapid development. Today, in addition to the deep belief network (DBN), there are also automatic encoders (AE) and its variants [21, 22], recurrent neural networks (RNN) [23, 24], deep neural networks (DNN) [25, 26], generative adversarial networks (GAN) [27, 28], and other models that all have been applied in the field of fault diagnosis in rotating machinery. Because the above models have good feature extraction capabilities for large amounts of data, they can produce good results when they are applied in the field of redundant data and various types of fault diagnosis. After training the current model, fast and high-precision fault classification and life prediction can be achieved, thereby monitoring the health of the machine.

However, in order to achieve the desired performance, deep learning models require a large amount of data to train the weights of its nodes in the training stage of the model, which leads to too much training time, and the data requirement for model training is also strict in a complex network training. If the amount of data is not enough or the model size is too large, the required accuracy rate cannot be achieved.

Transfer learning as an algorithm model was proposed by Sinno Jialin Pan and Qiang Yang in early 2009 [29]. Its purpose is to use the trained model parameters for other new model training to reduce the training time of the new model and reduce the amount of training data required. Transfer learning is widely used in training new models having a large number of sources that are not related to the new training model, but are similar and have sparse target data related to the new model. Migrating parameters to a new model reduces training time and gets higher accuracy by using limited target data to train models. Based on these advantages, transfer learning is used to solve the above problems in deep learning.

At the same time, transposed convolution as an upsampling technique can expand the image from a smaller resolution to a higher resolution to achieve more excellent feature extraction. Transposed convolution [30] is currently widely used in the field of image recognition. In fault diagnosis, when the target data is too small, upsampling technology can be used to expand the data, and it can also specifically increase the proportion of useful features, clean up some redundant features, and rationally allocate and use data to improve model training efficiency. With the excellent feature extraction ability of convolutional neural networks, fewer target data can be used to achieve ideal fault diagnosis accuracy.

Therefore, this paper proposes a framework utilizing the transfer learning model VGG19 to reduce training time and transposed convolution feature enhancement features to reduce the sample size. Based on the idea of the sensor array, the paper proposes a method of multisensor signal detection, which processes the sensor signals under multiple channels at the same time to improve the diagnostic accuracy of the model. At the same time, on the basis of the inverse convolutional neural network, the original input signal is upsampled, and the size of the input signal is expanded so that the feature map input to the model can better match the input requirements of the model. Finally, the paper adopts the concept of transfer learning and uses VGG19 as the backbone network to realize the rapid diagnosis of rotating machinery faults by designing the corresponding input signal preprocessing module and classifier module.

The rest of the paper is organized as follows. Section 2 introduces the theoretical background of the proposed approach, including deep convolutional neural networks, transposed convolution, and transfer learning. In Section 3, the overall mechanical fault diagnosis system is illustrated in detail. In Section 4, experimental studies on the datasets are carried out to verify the effectiveness of the proposed model, together with performance comparisons to other methods. Conclusions and future work are presented in Section 5.

2. Methodology

2.1. Convolutional Neural Network. Convolutional neural networks (CNNs) have a wide range of applications in image processing. Deep convolutional neural networks can automatically learn hierarchical features from input images, where features from higher levels are more abstract than those from lower levels. Generally speaking, convolutional neural networks contain three layers: convolutional layer, pooling layer, and fully connected layer. The convolution layer and the pooling layer are combined to form a convolution block, and several such blocks are stacked to build a deep-level architecture. Usually, a fully connected layer is used as the last layer to perform classification or regression. Figure 1 shows a representative CNNs architecture. CNN is a typical supervised feed forward neural network [31, 32]. Its training goal is to learn abstract features by alternating and superimposing convolution kernels and pooling operations. The structure mainly includes three parts: convolutional layer, pooling layer, and fully connected layer, as shown in Figure 1.

The convolution layer uses multiple convolution kernels to extract multiple features. The output of each layer is to convolve multiple input features. The mathematical model shows as follows:

$$C_j^l = \phi\left(\sum_{i=1}^{M^{l-1}} C_i^{l-1} * k_{ij}^l + b_j^{l-1}\right), \tag{1}$$

where * indicates the convolution operation, which indicates the connection weights of the feature maps i and j of the layers l-1 and l. b_j is the offset value, and φ is the nonlinear activation function.

The role of the pooling layer is to reduce the size of the feature space by half through the pooling operation, thereby reducing the feature dimension while ensuring the feature invariance. Common pooling operations are maximum pooling and mean pooling.

Finally, after multiple convolution operations and pooling, the input data is passed to one or more fully connected (FC) layers, and the results will be used as the input of the top-level classifier (for example, soft-max).

2.2. Convolution and Transposed Convolution. In the convolutional neural network, convolution is the operation of the input image, the convolution kernel, and the mapping relationship corresponding to the convolution kernel to make the output image size reduced but the features are more concentrated. In the convolution operation, the size of input and output satisfies the following relationship [29]:

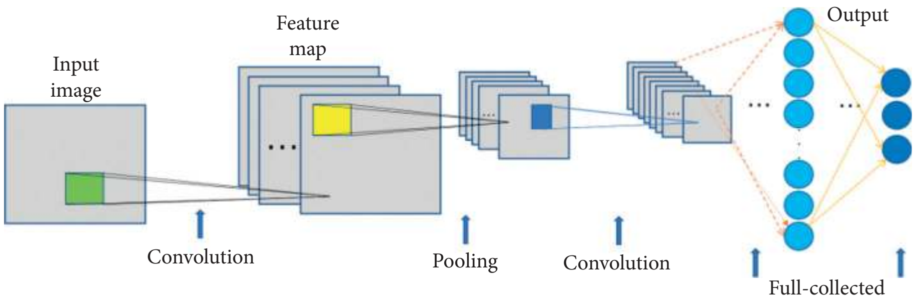

Figure 1: The representative CNN architecture.

$$o = \left[\frac{i + 2p - k}{s} \right] + 1, \tag{2}$$

where o represents the feature size of the output image, i represents the feature size of the input, k represents the size of the convolution kernel, s represents the step size of each operation, p represents padding, and $[x]$ represents rounding down.

For the convolution processing of the two-dimensional image, the paper can treat it as a matrix by multiplying it into another matrix of different scales. Let us take a square matrix as an example to illustrate the relationship between the matrix and the convolution operation.

Take the following simple convolution layer operation as an example, its parameters are $i = 4$, $k = 3$, $s = 1$, and $p = 0$. The process is shown in Figure 2. After the convolution operation, the output is 2.

For the above convolution operation, we expand the 3×3 convolution kernel shown in Figure 2 to a sparse matrix of [4, 15] as shown below, whose name is matrix C.

$$C = \begin{bmatrix} w_{0,0} & w_{1,0} & w_{0,2} & 0 & w_{1,0} & w_{1,1} & w_{1,2} & 0 & w_{2,0} & w_{2,1} & w_{2,2} & 0 & 0 & 0 & 0 & 0 \\ 0 & w_{0,0} & w_{0,1} & w_{0,2} & 0 & w_{1,0} & w_{1,2} & w_{1,2} & 0 & w_{2,0} & w_{2,1} & w_{2,2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & w_{0,0} & w_{0,1} & w_{0,2} & 0 & w_{1,0} & w_{1,1} & w_{1,2} & 0 & w_{2,0} & w_{2,1} & w_{2,2} & 0 \\ 0 & 0 & 0 & 0 & 0 & w_{0,0} & w_{0,1} & w_{0,2} & 0 & w_{1,0} & w_{1,1} & w_{1,2} & 0 & w_{2,0} & w_{2,1} & w_{2,2} \end{bmatrix}, \tag{3}$$

where the nonzero element wi, j represents the i-th row and j-th column of the convolution kernel.

Expanding the 4×4 input features into a matrix X of [1, 16], the final output matrix is $Y = CX$. Y is the output matrix whose size is $4 * 1$. Then, rearranging the 2×2 output features will get the final result. From the above analysis, we can see that the calculation of the convolutional layer can actually be converted into matrix multiplication.

It can be seen that the forward propagation process of the convolution layer is actually the reverse propagation process of the deconvolution layer, and the backward propagation process of the convolution layer is the forward propagation process of the deconvolution layer. Because the forward and backward calculations of the convolutional layer are multiplied by C and CT, respectively, and the forward and backward calculations of the deconvolution layer are multiplied by CT and (CT) T, respectively, their forward propagation and backpropagation are just swapped. Figure 3 shows a deconvolution operation corresponding to the convolution calculation of Figure 2, where their input and output relationships are exactly opposite. Among them, the parameters of transposed convolution are $i' = 2$, $k' = 3$, $s' = 1$, and $p = 2$.

Through the schematic diagram, the relationship between the input and output of the deconvolution layer when $s = s' = 1$ is

$$o' = i' + (k' - 1) - 2p'. \tag{4}$$

Therefore, the paper can think of transposed convolution as the inverse process of convolution operation, but it is worth noting that transposed convolution can only restore the size of the output matrix or graph, but not their parameters.

2.3. Transfer Learning. Transfer learning is a term in machine learning, which refers to the influence of one kind of learning on another kind of learning, or the influence of learned experience on the completion of other activities. Migration widely exists in the learning of various knowledge, skills, and social norms. Simply speaking, given a source domain named Ds and a target domain Dt, transfer learning attempts to apply the knowledge previously learned from Ds to the target domain Dt. Transfer learning methods can be divided into four categories: case-based, feature-based, parameter-based, and-relationship based. Here, transfer learning can help train the target model by initializing the target model with parameters transferred from the pre-trained model. The paper adopts the method of parameter transfer.

In short, transfer learning refers to training data as domains. This paper adopts only one source domain Ds and one target domain Dt. Each domain is composed of

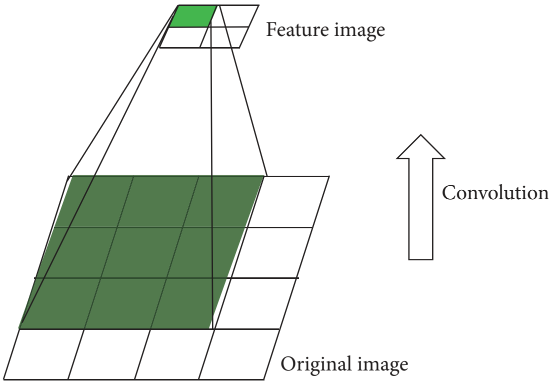

Figure 2: The process of convolution.

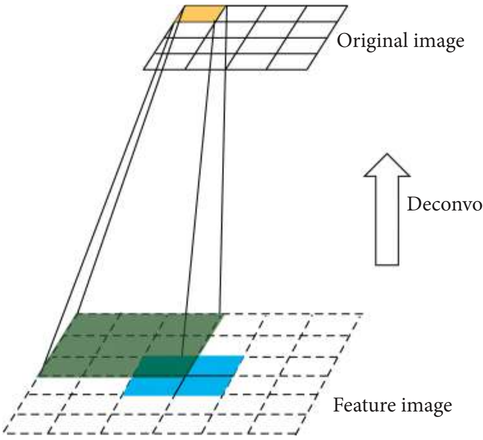

Figure 3: The process of deconvolution.

characteristic space X and marginal probability distribution $P\ (X)$, where $X = \{x_1, x_2, x_3, \ldots, x_n\}$. Based on parameter transfer learning, the deep learning model will be pretrained in the source domain, then the parameters will be transferred, and the target domain will be used for retraining and fine-tuning to achieve the goal of identifying unknown label samples in target domain. The cross domain reuse of parameters can improve the effect of model reconstruction and the efficiency of new data preprocessing. [33].

Training a deep architecture from scratch is difficult in practice. A large deep neural network contains a large number of weights, which are initialized randomly before training and iteratively updated according to the labelled data and the loss function. Iteratively updating the ownership weight is very time-consuming, and because of the limited training data, the deep architecture may present overfitting data by training.

Transfer learning provides a very effective idea to solve the above problems of deep learning. It uses a pretrained deep convolutional neural network that has been trained on another dataset. As mentioned before, CNNs can learn hierarchical representations from images, and the knowledge embedded in the pretrained model's weights can be transferred to new tasks. The low-level convolutional layer extracts low-level features such as edges and curves and is suitable for general image classification tasks, while the operation of the later layer can learn more abstract representations for different application fields. Therefore, lower-level representations can be transmitted, and only higher-level representations need to be learned from the new dataset. The process of updating the weights of higher hidden layers is called fine-tuning, and its success depends on a certain extent on the "distance" between the source dataset and the target dataset. Training a deep convolutional neural network with pretrained weights has been successfully applied to many different tasks, and most existing pretrained models are trained from enough natural image data. For example, in speech recognition, although there are large gaps between various languages, some studies have demonstrated the effectiveness of applying pretrained models to speech recognition tasks in different languages. Inspired by these achievements in other fields such as speech recognition and image recognition, we studied the knowledge transfer of multichannel mechanical datasets to natural images. [29].

3. Machine Fault Diagnosis by Deep Transfer Learning Based on Multichannel Datasets

In order to achieve high-precision mechanical system state detection, we propose a fault detection model based on a deep convolutional neural system. For multichannel fault datasets, it discards the traditional time-frequency processing of single-channel data into a graph for feature extraction, but directly uses multichannel data information to form an array graph, which is the input to the convolutional nerve after normalization. The graph trains the network and uses the excellent feature extraction capabilities of the neural network to achieve high-precision classification of faults. In order to reduce the training time, the model adopts the method of transfer learning, using similar data for training, and the dataset about fault diagnosis for fine-tuning to improve the accuracy of the experiment.

The paper proposes a framework that can learn fault features from multichannel mechanical signals and determine the type of fault through training. The flow of the proposed mechanical fault diagnosis model is shown in Figure 4, which includes data preparation, multichannel combined imaging, fine-tuning pretrained model establishment, and model application. First, sample and combine data from 8 channels. During the process, 512 data points are collected each time. Then, they are combined into an $8 * 512$ array image and converted into a $64 * 64$ format. After that, the paper uses transposed convolution to expand the image into a $256 * 256$ array image. Finally, the image is inputted into a pretrained model for training and fine-tuning to achieve high-precision classification of faults. The pretrained model used in this paper is VGG19 with a depth of 19 layers [34]. It was proposed by the Visual Geometry Group of Oxford University in 2014 and obtained accurate classification performance on the ImageNet dataset. Figure 5 illustrates the details of VGG19.

The pretrained model used in this paper is trained using the dataset on ImageNet. The target dataset is a picture formed by the array signal of the multichannel mechanical

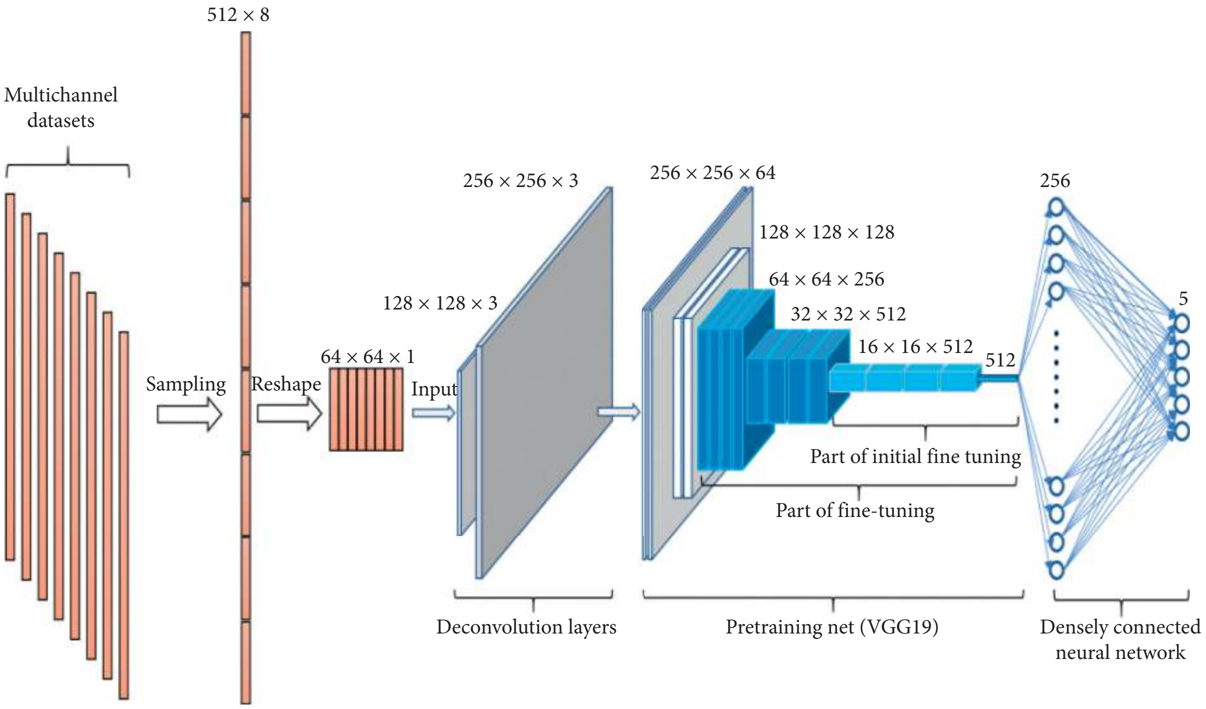

Figure 4: The flow of the proposed mechanical fault diagnosis.

dataset. The pretrained model used in this paper is trained through the dataset on ImageNet. Because the natural picture is not similar to the array signal but rather similar to an image, more convolutional blocks are needed to be fine-tuned for mechanical datasets.

3.1. Data Preparation. During operation, the vibration signals are collected by 8 sensors installed on the mechanical equipment. In order to train and fine-tune the internal structure of the pretraining, the model requires a specific size of RGB image. The purpose of this operation is to maintain the aspect ratio of the feature map input to the convolutional layer, which is a fixed 1 : 1 for operation with convolutional network. At the same time, another important factor is that VGG16 is used as the backbone network, which makes the model's input image match the final output feature map. The paper adopts the following method: select 512 signal data from each of the 8 channels for array combination and convert them into $64 * 64 * 1$ images. Since the converted image is a grayscale image with only one channel, it will need to expand the image to a 2D image with a size of $256 * 256 * 3$. The training dataset is used to train the pretraining model and adjust its weights, while the test dataset is only used to verify the performance of the deep model, not to participate in the training process.

3.2. Pretrained Model Creation and Fine-Tuning. In order to reduce the model training time and have no or low influence in the accuracy rate, this experiment adopted a phased training method: In the first step, the experiment migrates the parameters of the VGG19 model and trains the transposed convolution block so that it can better extract the features of the input image. The first step of training the transposed convolution block will cover the entire experiment. When the accuracy of the model reaches 98%, the second step of the experiment will be carried out. In order to reduce the training time, the experiment only fine-tunes the parameters of the fifth convolution block of the VGG19 model, allowing it to achieve accurate classification of fault types, thereby improving the accuracy of the experiment. In order to further improve the accuracy of the model, in the third step, the paper adjusts the parameters of the fourth and fifth convolution blocks of VGG19 so that the features extracted through the convolution operation can more accurately describe the fault and achieve a high model accuracy classification. The transfer learning training and fine-tuning process is shown in Figure 6.

3.3. Model Application. The test dataset is used to test the accuracy of the designed model in the mechanical fault diagnosis task, and the pretrained deep learning model that has been debugged is applied to the diagnosis of the machine working state.

4. Experimental Verification

In order to test the performance of the proposed fault diagnosis system and verify its effectiveness, the paper adopts gear and bearing datasets to conduct experiments and compares with traditional feature-based classification methods and deep learning-based classification methods. In addition, the model proposed in the paper is also compared with three different models, including the VGG16 fault

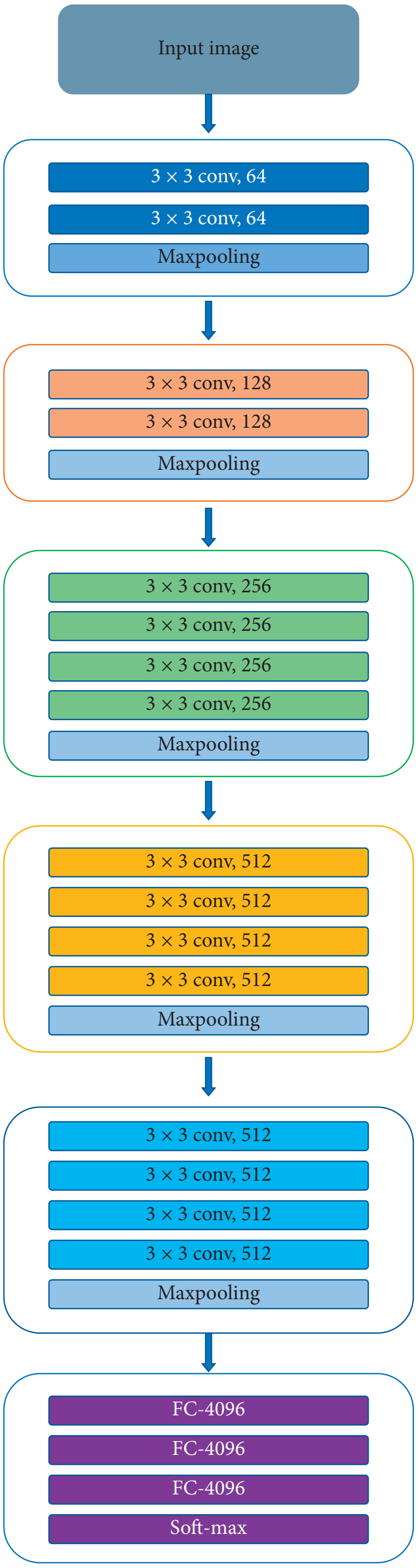

Figure 5: The architecture of VGG19.

diagnosis model trained with a single-channel signal, a convolutional neural networks model with three convolutional blocks and a three-layer symmetric autoencoder model.

4.1. Gearbox Dataset and Bearing Dataset. The gearbox dataset was collected from the drivetrain dynamic simulator (DDS) shown in Figure 7 [34]. This dataset contains 2 subdatasets, including bearing data and gear data, which are both acquired on DDS. There are two kinds of working conditions with rotating speed-load configuration set to be 20-0 and 30-2. Within each file, there are 8 rows of signals which represent 1-motor vibration, 2, 3, 4-vibration of planetary gearbox in three directions x, y, and z, 5-motor torque, and 6, 7, 8-vibration of parallel gear box in three directions x, y, and z. Signals of rows 2, 3, and 4 are all effective. [34] Different fault types of bearings and gearboxes are shown in Table 1.

The bearing dataset contains two different loads and five different working conditions, which include four types of faults and one state of health. Therefore, the fault diagnosis of the powertrain dynamic simulation system is a ten-level classification task. Each failure type contains 1800 training samples, and the entire gear and bearing training dataset is 18000 images, and there are 496 sets of data between the test set data and the training set data. So, there are a total of 247 test sets. The specific experimental data grouping situation is shown in Table 2. The entire gear and bearing test sets are 2470 images. In order to verify the effectiveness of the method in dealing with the mixed faults, the faults and working conditions are combined to form a mixed dataset containing 4 gear failures, 1 health state, and two working states. Each working state contains 1800 training samples and 247 test samples, and the training and test dataset contains 2047 samples. The entire dataset is divided into three parts: training, test, and unused data. In order to correspond to the fault types in 10, the output layer of the pretrained VGG19 model is replaced with the output layer of 10 neurons to predict the corresponding label.

The accuracy of the gear and bearing fault classification is shown in Figures 8 and 9. In addition, in order to verify the proposed generalization ability of the model and whether there is overfitting in the classification experiment, this article also conducted ten-fold cross-validation experiments and listed the loss function of the two sets of data. For the experiment, the detailed results are shown in Figures 10 and 11 and Figures 12 and 13. At the same time, in order to verify the effectiveness of the proposed model, the paper also lists the literature [34–36] for the accuracy of the failure classification experiments shown in Table 3. The paper also compares the accuracy of the VGG16 model. The comparison of the accuracy of the two models is shown in Figure 14.

The results show that the VGG19 pretrained model that uses multichannel array signals to generate images is superior to other machine learning models based on single-channel signal processing in terms of classification accuracy, and the performance is improved by about 6%, and the failure rate of the gearbox dataset reaches 99.99% accuracy.

In conclusion, the paper proposes a framework that adopts the concept of transfer learning for training. Therefore, theoretically speaking, as long as the corresponding form of one-dimensional sensing signal is satisfied, the framework proposed in this paper is adaptable.

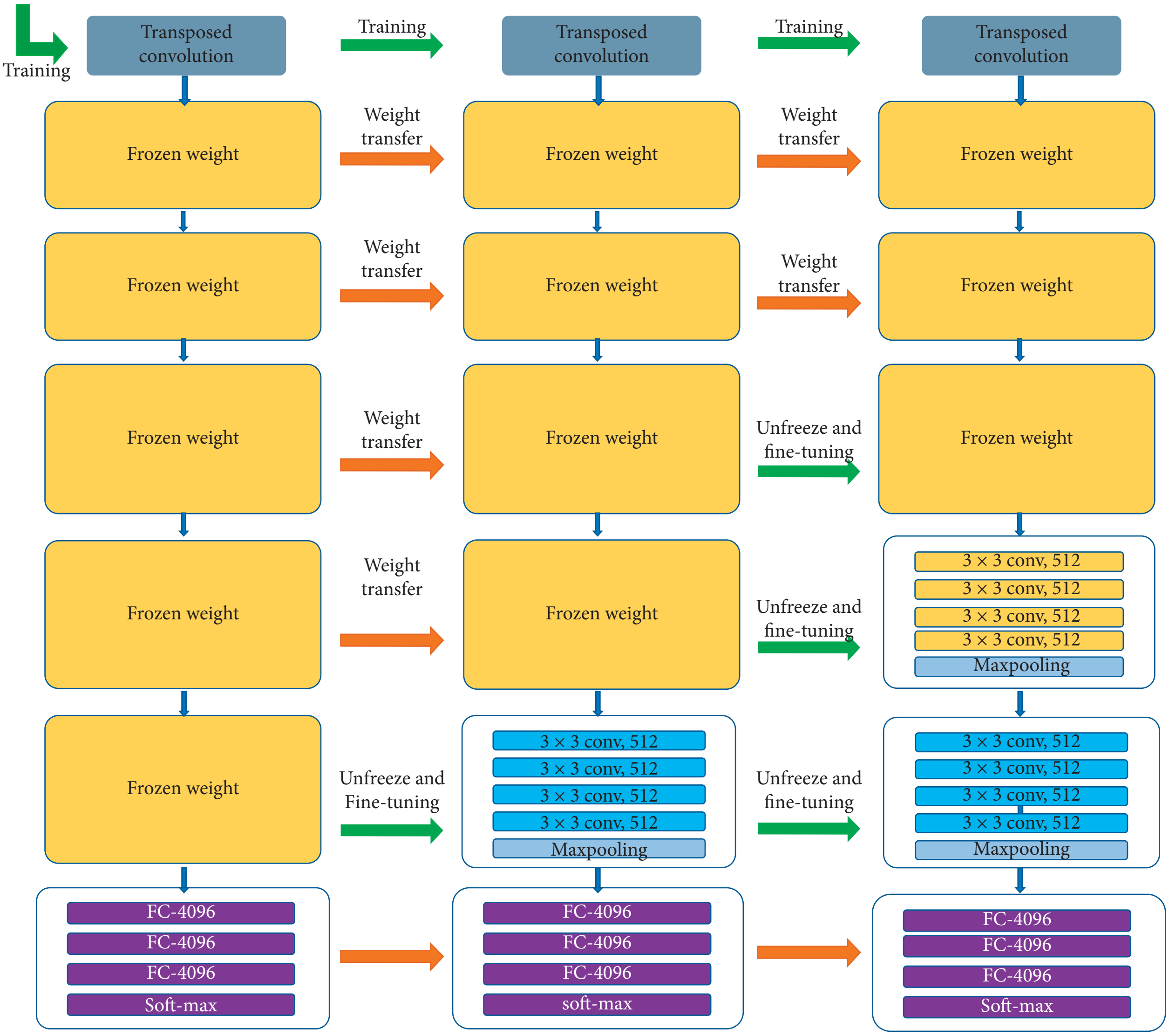

Figure 6: The process of the experiment.

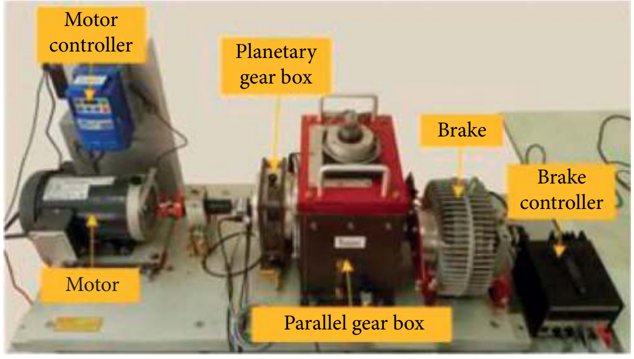

Figure 7: Experimental setup for gearbox dataset [34].

Table 1: Bearing and gearbox fault types description.

Location	Type	Description
Gearbox	Chipped	Crack occurs in the gear feet
	Miss	Missing one of feet in the gear
	Root	Crack occurs in the root of gear feet
	Surface	Wear occurs in the surface of gear
Bearing	Ball	Crack occurs in the ball
	Inner	Crack occurs in the inner ring
	Outer	Crack occurs in the outer ring
	Combination	Crack occurs in the both inner and outer rings

Table 2: The experimental data grouping situation.

Load	Fault types	Label	The number of experiment groups Training/testing
0 and 2	Health	0	1800/247
0 and 2	Chipped/ball	1	1800/247
0 and 2	Miss/inner	2	1800/247
0 and 2	Root/outer	3	1800/247
0 and 2	Surface/combination	4	1800/247

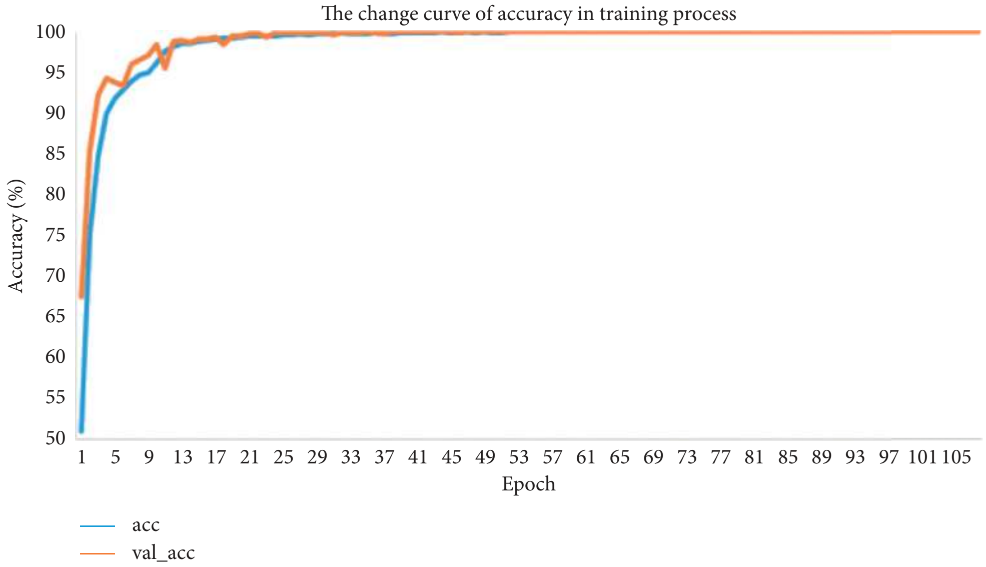

Figure 8: The accuracy of bearing dataset experiment.

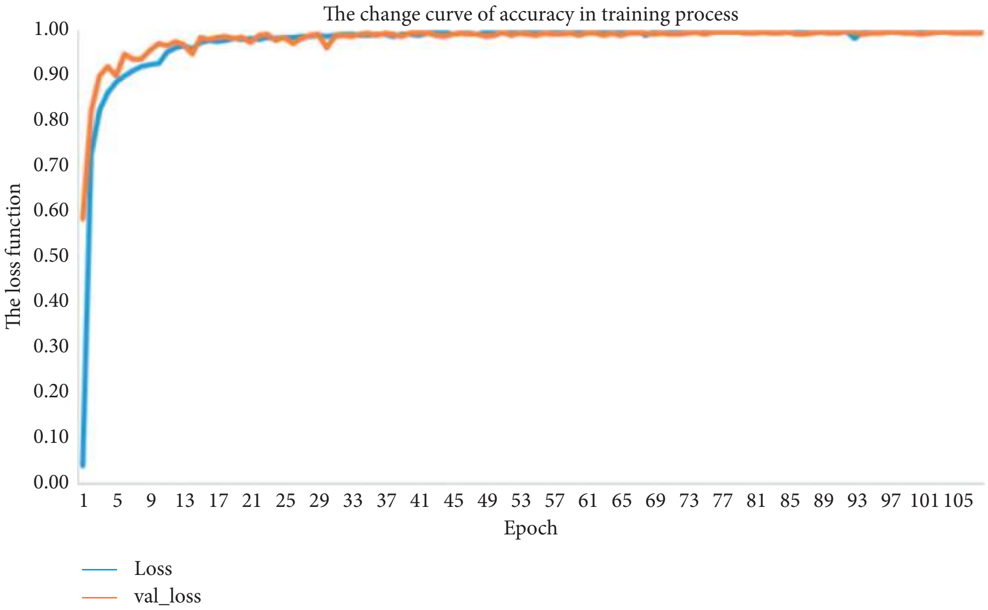

Figure 9: The accuracy of gearbox dataset experiment.

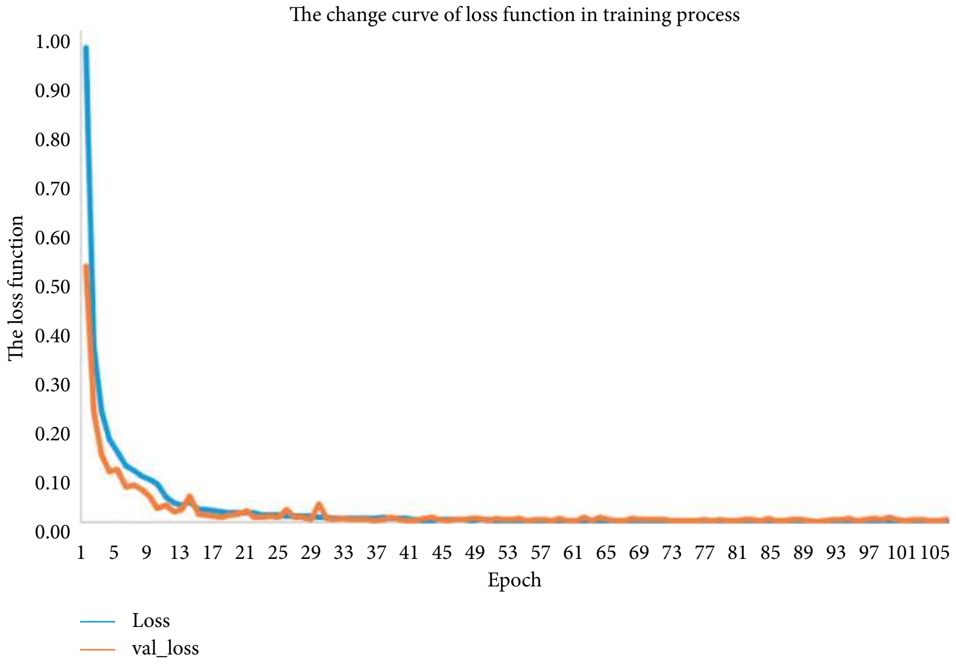

Figure 10: The loss function of bearing dataset experiment.

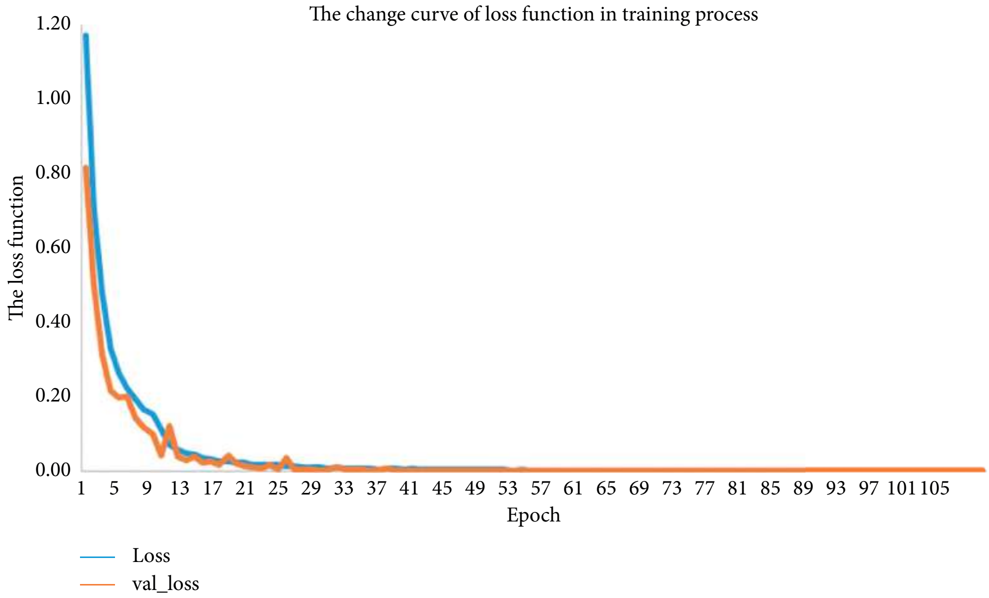

Figure 11: The loss of function gearbox dataset experiment.

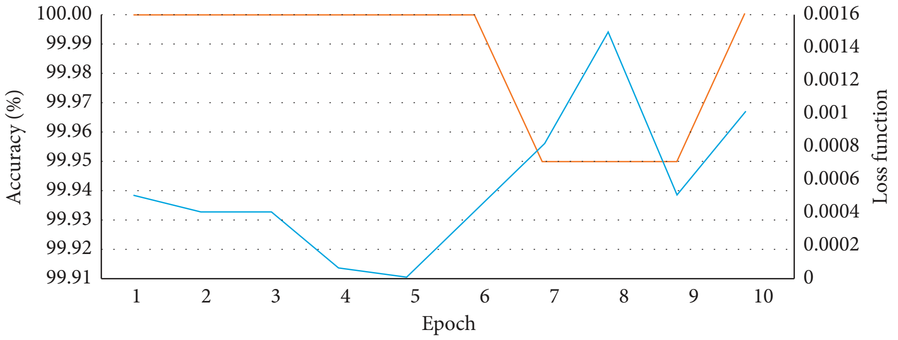

Figure 12: Ten-fold cross-validation result of gearbox dataset.

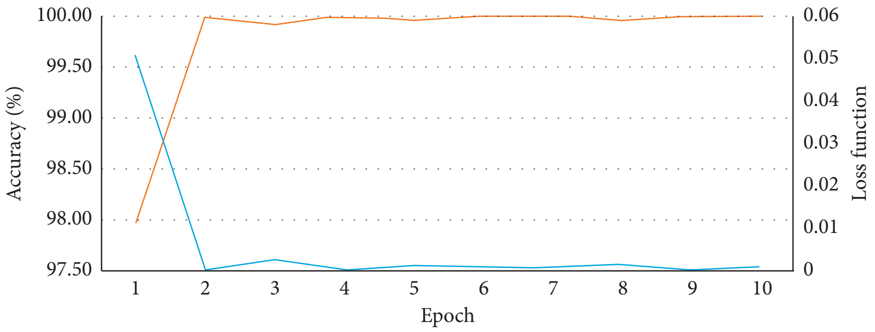

Figure 13: Ten-fold cross-validation result of bearing dataset.

Table 3: Classification results for the gearbox dataset.

	Fault diagnosis method	Time window	Channel	Bearing (20-0)	Bearing (30-2)	Gearset (20-0) (%)	Gearset (30-2) (%)
[35]	Premodel (VGG16)	1024	1	98.90%	98.84%	98.70	98.07
[36]	SAE-DNN	20000	1	87.50%	92.10%	92.70	91.90
	GRU	20000	1	91.20%	92.40%	93.80	90.70
	BiGRU	20000	1	93.00%	93.60%	93.80	90.70
	LFGRU	20000	1	93.20%	94.00%	94.80	95.80
[38]	DSR	1024	1	*	*	99.59	99.78
	SACNN	1024	1	*	*	99.86	99.88
	CNN trained from scratch	512	1	98.08%	98.08%	97.80	97.80
	DNN	512	1	97.56%	97.56%	93.68	93.68
	Premodel (VGG19)	512	8	99.78%	99.78%	99.99	99.99

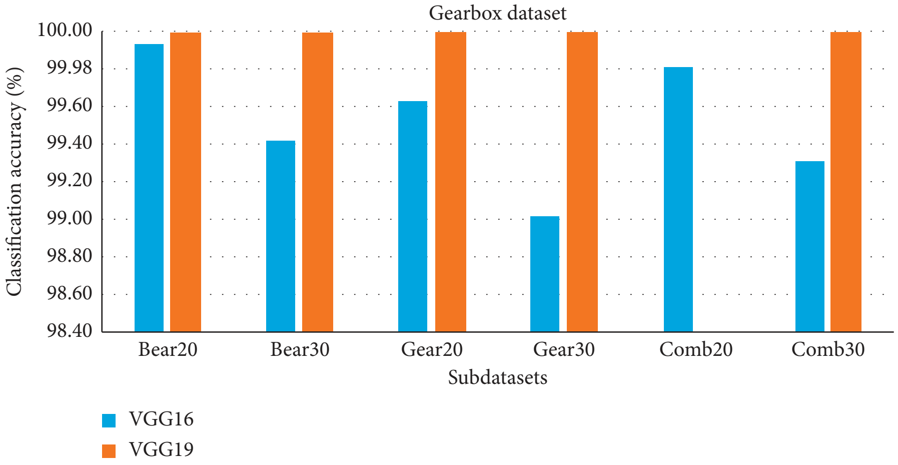

Figure 14: Classification accuracy of VGG16 and VGG19 on the gearbox dataset.

Therefore, even though we recommend using similar sensors for data and collection, the data source is not limited.

5. Conclusion

In summary, we have developed a deep transfer learning framework for machine fault diagnosis and classification. According to our comparative analysis, the new model using the transposed convolution and VGG19 framework has achieved good results in the experiments of the gearbox which include bearing and gear two datasets, and not only reduce the training time and the number of samples for a single training but also made significant progress in classification accuracy. It is foreseeable that deep learning will make great achievements in other fields, such as fault detection and classification of different types of mechanical systems and control engineering.

Data Availability

The data used to support the findings of this study are available at https://github.com/cathysiyu/Mechanical-datasets.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] Y. M. Park and G. W. Kim, "A logic based expert system (LBES) for fault diagnosis of power system," *IEEE Transactions on Power Systems*, vol. 12, no. 1, pp. 363–369, 1997.

[2] B. Li, M.-Y. Chow, Y. Tipsuwan, and J. C. Hung, "Neural-network-based motor rolling bearing fault diagnosis," *IEEE Transactions on Industrial Electronics*, vol. 47, pp. 1060–1069, 2000.

[3] H. J. Lee and D. Y. Park, "A fuzzy expert system for the integrated fault diagnosis," *IEEE Transactions on Power Delivery*, vol. 15, no. 2, pp. 833–838, 2000.

[4] N. H. Ulerich and G. J. Powers, "On-line hazard aversion and fault diagnosis in chemical processes: the digraph + fault-tree method," *IEEE Transactions on Reliability*, vol. 37, no. 2, pp. 171–177, 1988.

[5] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.

[6] Y. Lecun, B. Boser, J. S. Denker et al., "Backpropagation applied to handwritten zip code recognition," *Neural Computation*, vol. 1, no. 4, pp. 541–551, 1989.

[7] J. Schmidhuber, "Deep learning in neural networks: an overview," *Neural Net-works*, vol. 61, pp. 85–117, 2015.

[8] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 25, no. 2, http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf, 2012.

[9] B. Bayar and M. C. Stamm, "A deep learning approach to universal image manipulation detection using a new convolutional layer," in *Proceedings of the 4th ACM Workshop on Information Hiding and Multimedia Security*, Vigo, Spain, June 2016.

[10] G. Cheng, C. Yang, X. Yao, L. Guo, and J. Han, "When deep learning meets metric learning: remote sensing image scene classification via learning discriminative CNNs," *IEEE Transactions on Geoscience & Remote Sensing*, vol. 56, no. 5, pp. 1–11, 2018.

[11] A. H. Abdi, C. Luong, T. Tsang et al., "Automatic quality assessment of echocardiograms using convolutional neural networks: feasibility on the apical four-chamber view," *IEEE Transactions on Medical Imaging*, vol. 36, no. 6, pp. 1221–1230, 2017.

[12] K. Sakamoto, E. Ota, T. Ozawa, H. Nishimura, and H. Tanaka, "Feasibility study on deep learning scheme for sign language motion recognition," in *Complex, Intelligent, and Software Intensive Systems Conference on Complex,* Springer, Cham, Switzerland, 2018.

[13] S.-H. Park and C.-K. Kim, "Sign language translation system based on deep learning for speech disorders," in *Proceedings of the Symposium of the Korean Institute of Communications & Information Sciences*, JeongSeon, Korea, January 2018.

[14] X. Zhou, W. Gong, W. Fu, and F. Du, "Application of deep learning in object detection," in *Proceedings of the 2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS)*, IEEE, Wuhan, China, May 2017.

[15] S. Hoermann, P. Henzler, M. Bach, and K. Dietmayer, "Object detection on dynamic occupancy grid maps using deep learning and automatic label generation," in *Proceedings of the 2018 IEEE Intelligent Vehicles Symposium*, Changshu, China, June 2018.

[16] X. Ren, Y. Zhou, Z. Huang, J. Sun, X. Yang, and K. Chen, "A novel text structure feature extractor for Chinese scene text detection and recognition," *IEEE Access*, vol. 5, pp. 3193–3204, 2017.

[17] X. Ren, Y. Zhou, J. He, K. Chen, X. Yang, and J. Sun, "A convolutional neural network-based Chinese text detection algorithm via text structure modeling," *IEEE Transactions on Multimedia*, vol. 19, no. 3, pp. 506–518, 2017.

[18] T. Yoshimasa, "Deep learning and natural language processing," *Brain and Nerve = Shinkei Kenkyu no Shinpo*, vol. 71, no. 1, pp. 45–55, 2019.

[19] W. Stephen, K. Roberts, S. Datta et al., "Deep learning in clinical natural language processing: a methodical review," *Journal of the American Medical Informatics Association*, vol. 27, no. 3, pp. 457–470, 2020.

[20] C. Shen, Y. Qi, J. Wang, G. Cai, and Z. Zhu, "An automatic and robust features learning method for rotating machinery fault diagnosis based on contractive autoencoder," *Engineering Applications of Artificial Intelligence*, vol. 76, pp. 170–184, 2018.

[21] 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. 72-73, pp. 303–315, 2016.

[22] G. Xiaojie, S. Changqing, and C. Liang, "Deep fault recognizer: an integrated model to denoise and extract features for fault diagnosis in rotating machinery," *Applied Sciences*, vol. 7, no. 1, p. 41, 2016.

[23] H.-H. Pan, X. Tang, S. Meng, X. He, and F. Meng, "An improved bearing fault diagnosis method using one-dimensional CNN and LSTM," *Journal of Mechanical Engineering*, vol. 64, no. 7/8, pp. 443–452, 2018.

[24] X. Jiang, C. Shen, J. Shi, and Z. Zhu, "Initial center frequency-guided VMD for fault diagnosis of rotating machines," *Journal of Sound and Vibration*, vol. 435, no. 22-460X, pp. 36–55, 2018.

[25] H. Fang, H. Shi, Y. Dong, H. Fan, and R. Shuai, "Spacecraft power system fault diagnosis based on DNN," in *Proceedings of the 2017 Prognostics & System Health Management Conference IEEE*, Harbin, China, July 2017.

[26] T. N. Babu, A. Aravind, A. Rakesh, M. Jahzan, and D. Rama Prabha, "Application of EMD ANN and DNN for self-aligning bearing fault diagnosis," *Archives of Acoustics*, vol. 43, no. 2, pp. 163–175, 2018.

[27] H. Liu, J. Zhou, Y. Xu, Y. Zheng, X. Peng, and W. Jiang, "Unsupervised fault diagnosis of rolling bearings using a deep neural network based on generative adversarial networks," *Neurocomputing*, vol. 315, pp. 412–424, 2018.

[28] Z. Wang, J. Wang, and Y. Wang, "An intelligent diagnosis scheme based on generative adversarial learning deep neural networks and its application to planetary gearbox fault pattern recognition," *Neurocomputing*, vol. 310, pp. 213–222, 2018.

[29] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.

[30] S. Kim, K. Lee, S. Doo, and B. Shim, "Moving target classification in automotive radar systems using transposed convolutional networks," in *Proceedings of the 2018 52nd Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, USA, October 2018.

[31] J. Gu, Z. Wang, J. Kuen et al., "Recent advances in convolutional neural networks," *Pattern Recognition*, vol. 77, pp. 354–377, 2015.

[32] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," 2012, http://arxiv.org/abs/1207.0580.

[33] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Proceedings of the 2014 European Conference on Computer Vision*, Springer International Publishing, Zurich, Switzerland, September 2014.

[34] S. Siyu, S. McAlee, 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.

[35] R. Zhao, D. Wang, R. Yan, K. Mao, F. Shen, and J. Wang, "Machine health moni- toring using local feature-based gated recurrent unit networks," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 2, pp. 1539–1548, 2018.

[36] D. NikolaosL. Leontarisa et al., "A Deep Learning framework for simulation and defect prediction applied in microelectronics," *Simulation Modelling Practice and Theory*, vol. 100, Article ID 102063, 2020.