DCANet: Dense channel attention network based on SimCLR for fault diagnosis of aviation transmission magazine

Xie Yining[1*], Shi Jiangtao[1], Guo Jinrun[1], Yang Kaihua[2], Guan Guohui[2], Zhao Zhichao[2], Ma Shantao[3], Zhao Jing[1]

[1]Northeast Forestry University, Harbin, China; [2]China Aerospace Development Harbin Dongan Engine Co, Harbin, China; [3]Harbin University of Science and Technology, Harbin, China;

Abstract—Aero engines and their accessory subsystems operate under complex operating conditions in a production environment, and usually, the equipment will experience variable acceleration and deceleration. The existing models and methods are limited by many problems, such as network structure and poor feature extraction ability, resulting in low diagnostic accuracy under variable working condition datasets. A Dense channel attention network (DCANet) based on SimCLR is proposed to address the above problems. This method combines the advantages of DenseNet and SENet. It can enhance feature reuse, integrate the feature information of high, middle, and low layers, and realize recalibration in the feature extraction process. It effectively enhances the feature extraction ability of single-channel time-frequency images, recognizes the attention weighting of features, and extracts effective features. At the same time, SimCLR's twin network small-sample learning method is introduced to solve the problem of low accuracy due to the small amount of data. Many experiments show that the recognition rate of the proposed method is better than the existing methods under the variable working condition dataset. The highest accuracy rate is 97.12%, which has been obtained on the Dongan dataset.

Index Terms—Fault diagnosis; SimCLR; DCANet; SENet

1. Introduction

As the heart of an aircraft, an aero-engine is known as the jewel in the crown of industry. Its magazine is an important supporting and load-bearing component, which has worked under high temperature, high pressure, high strain, and strong corrosive environment for a long time. It is prone to failures, such as corrosion and impact, which will bring certain downtime failures and even destructive damage to the aero-engine [1-3]. Therefore, fault diagnosis of the aero-engine transmission magazine can effectively improve its efficiency and reduce the possibility of accidents.

The methods based on signal processing are mostly applicable to constant speed conditions. At the same time, in actual industrial production, mechanical equipment in the environment is mostly variable speed and variable conditions. The frequency components of vibration signals are complex and time-varying in these conditions, making signal feature extraction difficult. Hongmei Liu et al. [4] proposed acoustic signal rotating machinery fault diagnosis based on STFT and deep learning. In this method, STFT is first used for the time-frequency analysis of signals. Then, this method extracts time-frequency image features using a stacked sparse autoencoder(SAE), and finally, the softmax function is used for classification. The momentum contrast learning method (MoCo) [5] is outstanding in the field of self-supervised learning. This method can construct positive and negative samples from unsupervised training data and build positive and negative encoders with a deep network. Positive and negative samples are compared in the feature space by introducing the momentum parameter so that this method can effectively use the supervised information in positive and negative samples to improve the feature representation of the model. It provides a possible way to solve mechanical systems' unsupervised fault diagnosis problem.

For the problems of existing methods in practical applications, this paper proposes

Dense channel attention network based on SimCLR for fault diagnosis, we collected different speed fault data

* Corresponding author: Xie Yining Zhao Jing E-mail: yiningxie@nefu.edu.cn jingzhaohlg@163.com

of aero-engine propulsion reducer under variable working conditions and started fault diagnosis research. This method achieved the highest accuracy of 97.12% on the propulsion reducer. First, the short-time smooth property of STFT is used to perform time-frequency analysis on the propulsion reducer variable speed dataset to generate time-frequency images of relevant faults in this method. Then, we combine the design ideas of DenseNet [6] and SENet [7] to propose an image classification network model DCANet. DCANet not only enhances feature reuse and fuses feature information from high and low layers but also enables rescaling in the feature extraction process. DCANet effectively enhances the feature extraction capability of single-channel grayscale images and achieves attention weighting of features, extracting more practical features and improving model classification accuracy. Finally, a self-supervised contrast learning method is added to the model to pre-train the DCANet encoder through the SimCLR [8] framework, and the trained DCANet is migrated to the downstream fault classification task. The downstream task then fine-tunes the encoder through MLP (Multilayer Perceptron) to solve the problem of small data volume and the complex workload caused by data labeling, i.e., the model training is completed using a small amount of labeled data.

2. Related work

2.1Short-time Fourier transform

The concept of Short Time Fourier Transformation (STFT) was first proposed by Gabor in 1946 and has been widely used in speech recognition tasks to extract the time-frequency features of speech signals [9], and its formula is:

$$F(\omega,\tau)=\int_{-\infty}^{\infty} f(t)\psi^{*}(t-\tau)e^{-j\omega\tau}dt \tag{1}$$

Where $F\ (\omega,\tau)$ is the two-dimensional function of time τ and frequency ω $\psi^{*}(t-\tau)$ is the window function, the slip-time window is used to intercept the signal in frames. Short-time Fourier transform considers that the signal is smooth in a short time interval, so the nonlinear signal is divided into finite segments of smooth signal. Each small segment of a smooth signal is processed by Fourier transform. The basic implementation is splitting the nonlinear signal into frames and adding windows. Fourier transform and spectrum analysis are performed on each small segment of the signal. The spectrum analysis of all time segments is stacked, and the final overall spectrum is the spectrum analysis of the signal in the whole time segment.

2.2 Channel attention mechanism

Jie Hu's team from Momenta Unmanned proposed SENet based on the channel attention mechanism in 2017 [10]. Squeeze-and-Excitation (SE) module is proposed in SENet. This model first obtains the global features at the channel level through squeeze squeezing operation. Then the Excitation operation is used to learn the relationship between each channel and calculate the weights of each channel. Finally, the original feature map is multiplied to obtain the final features of the attention mechanism [11]. The structure of the SE module is shown in Figure 1: The SE module mainly consists of squeeze and excitation, which can be applied to any mapping, taking convolution as an example, the convolution kernel is $V=[v_1,v_2,...v_c]$, where v_c denotes the c th convolution kernel, then the output: $U=[u_1,u_2,...u_c]$

$$u_c = v_c * X = \sum_{s=1}^{c'} v_c^s * x^s \tag{2}$$

Equation 2 $*$ denotes the convolution operation, s denotes the two-dimensional convolution kernel of the channel, and the spatial features on the input channel, it will automatically learn the feature spatial relations, but since the convolution results of each channel are summed, therefore, the channel feature relations are mixed with the spatial relations learned by the convolution kernel. The purpose of the SE module is to extract the channel features from these complex relations, thus enabling the model to learn the channel feature relations directly [12].

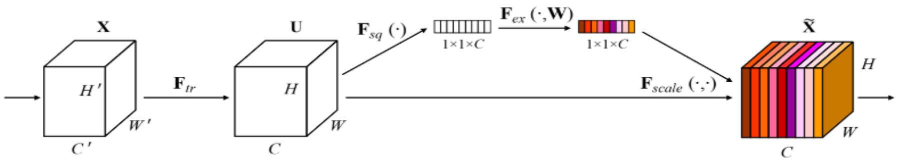


Fig 1 SENet Structure Diagram

2.3 SimCLR

The SimCLR self-supervised contrast learning framework is shown in Fig.2. A complete set of self-supervised contrast learning frameworks is proposed in SimCLR. In this model, the input sample is a positive sample through data enhancement, and other input samples are in the same batch as a negative sample. SimCLR maximizes the feature distance between the input sample and the positive sample by comparing the loss function while minimizing the feature distance between the input sample and the negative sample. In this model, the input data x is transformed into $\tilde{x}_i$ and $\tilde{x}_j$ after two different enhancements, $\tilde{x}_i$ and $\tilde{x}_j$ share the network structure and the weights of $f(\bullet)$, and after $f(\bullet)$, the respective feature representation h is obtained. $f(\bullet)$ is the encoder, which can be ResNet or other deep learning models, h is the feature representation after encoder processing. ResNet is used in the original SimCLR paper, and h represents the 2048 feature vectors of the final output of ResNet; $g(\bullet)$ is the MLP, and the feature representation h becomes the feature enhancement vector z_k after multiple MLPs. The encoder extracts the features, and the features are extracted again to meet the needs of downstream tasks

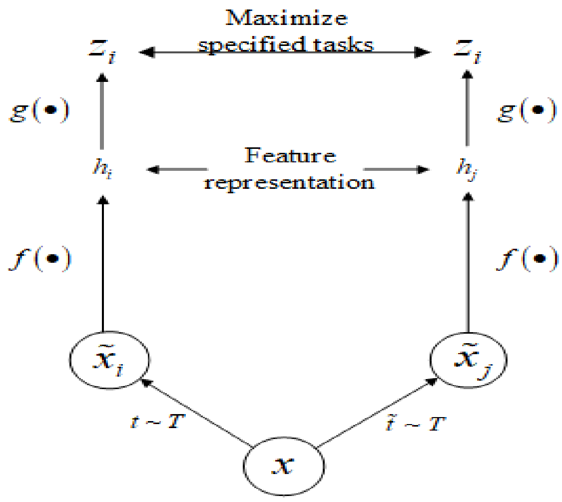


Fig. 2 Self-supervised contrastive learning framework for SimCLR

3. Proposed method

3.1 Overall Framework

An aero-engine fault diagnosis method based on DCANet and SimCLR is proposed in this paper. The structure of the model is shown in Fig. 3. The acquired raw signal is firstly processed into a single-channel grayscale time-frequency image by STFT, and then the amount of data is increased by using four methods as the input of the pre-trained network. The data enhancement methods include (a) y original signal with random noise, (b) time-frequency image with Gaussian noise, (c) 0-value padding of the head and tail of the original signal, and (d) mixing multiple signals (MixUp). This model uses DCANet as an encoder and pre-trains DCANet by using a self-supervised approach while fine-tuning DCANet using a supervised learning training network through a migration learning approach and performs similarity calculation and categorization through SimCLR.

3.2 DCANet

DCANet is a combination of SENet and DenseNet, and the overall model structure is shown in Figure 4. The network is an SE module embedded in DenseNet-BC, mainly composed of four Block modules and four SENet modules. The four Block modules have (n=3, 3, 6, 4) DenseLayers, respectively, and a Transition layer is connected between adjacent Block modules, which contains a 1x1 convolution and an average pooling to reduce feature dimension, so that realizes parameter reduction and achieves the purpose of reducing the amount of calculation and time. Each DenseLayer contains a 1x1 convolution and a 3x3 convolution, which effectively enhances feature reuse and deep information transfer and realizes the rescaling of feature channels. SENet is formed based on the channel attention mechanism and mainly consists of Squeeze and Excitation. The Squeeze operation mainly reduces the input image dimension to reduce the parameter calculation and increase the perceptual field. Excitation operation predicts the importance of each channel through a fully connected layer so that the information between channels can interact. The weight of each feature channel is determined to enhance the useful channel features and suppress the channel features which are not useful for the current task by self-learning. This operation can better achieve feature extraction for time-frequency images and further improve the accuracy of the network.

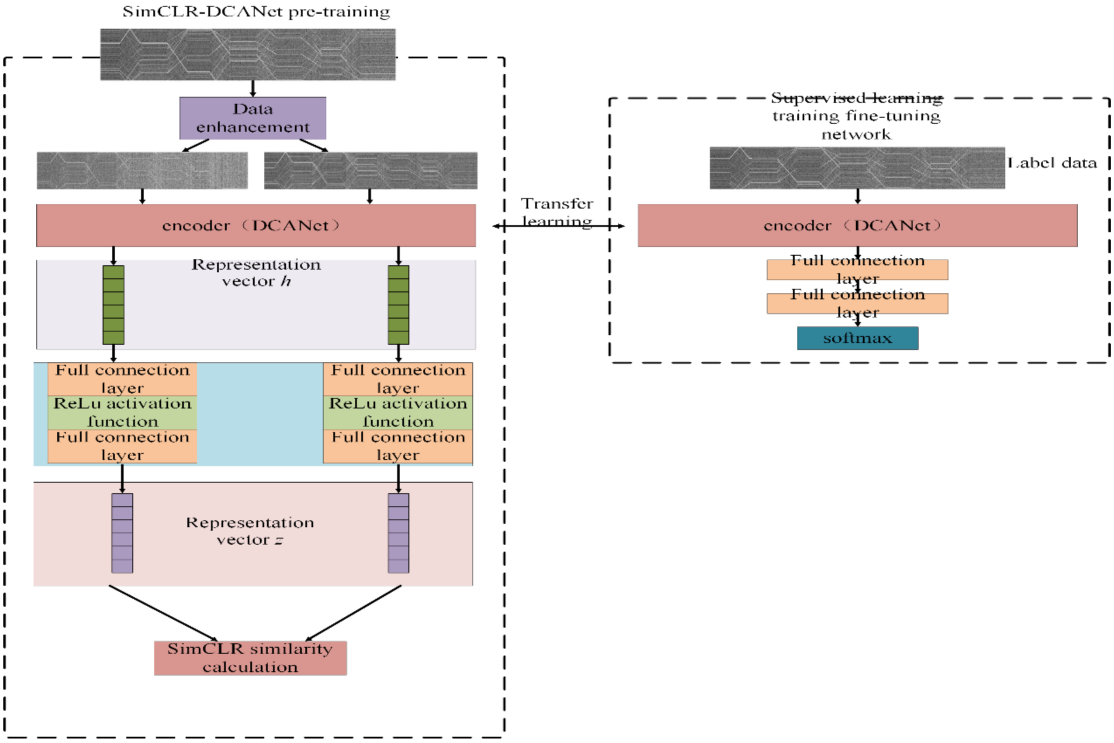


Fig3 Network structure of DCANet based on SimCLR

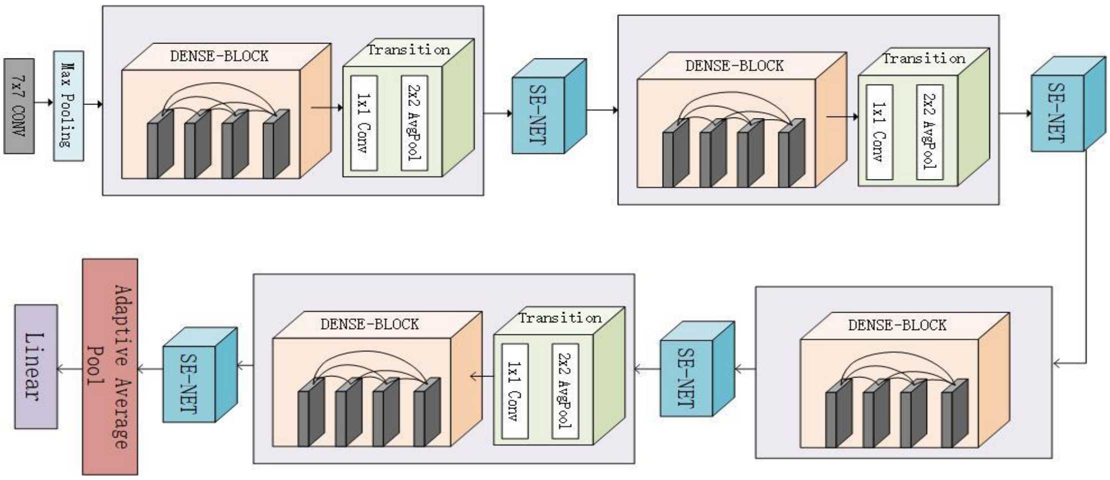


Fig 4 Model structure of DCANet

3.3 SimCLR-based twin networks

The fault diagnosis techniques based on deep learning generally require much training data. However, in actual industrial production, mechanical equipment failures occur slowly, and the relevant datasets are difficult to collect, while the workload of labeling data is large and tedious. This paper adopts a twin network small-sample learning method to address the above problems for aero-engine fault diagnosis under small samples.

The model structure is shown in Figure 5, where two DCANet have the same structure and share the weights. The input data are sample pairs of STFT time-frequency maps of the same class or different classes, and the same DCANet is used to extract features from the input STFT time-frequency maps. The SimCLR framework performs the similarity calculation to maximize the feature distance between the input and positive samples while minimizing the feature distance between the input and negative samples. The result is used as a proxy task to learn the sample features.

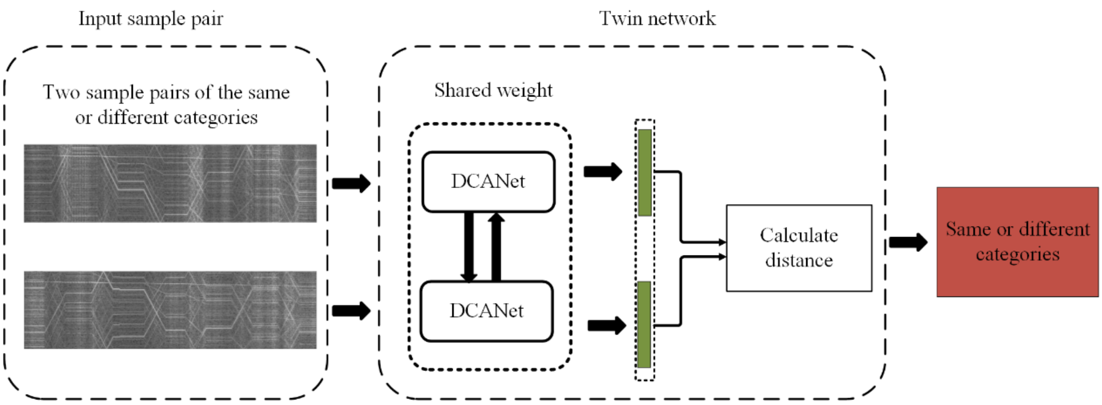


Fig 5 Few-shot learning model based on DCANet

The Adam optimizer optimizes the network, which calculates a separate adaptive learning rate for each parameter. The parameters are updated using Equation 3.

$$
\begin{aligned}
& m_w^{(T+1)} = \beta_1 m_w^T + (1-\beta_1)\nabla_W L^T \\
& v_w^{(T+1)} = \beta_2 v_w^T + (1-\beta_2)(\nabla_W L^T)^2 \\
& \hat{m}_w = \frac{m_w^{(T+1)}}{1-(\beta_1)^{T+1}} \\
& \hat{v}_w = \frac{v_w^{(T+1)}}{1-(\beta_2)^{T+1}} \\
& w^{(T+1)} = w^T - \eta \frac{\hat{m}_w}{\sqrt{\hat{v}_m} + \varepsilon}
\end{aligned}
\tag{3}
$$

The SimCLR framework defines the loss function:

$$
l_{i,j} = -\log \frac{\exp(sim(z_i, z_j)/\tau)}{\sum_{K=1}^{2N} 1_{k \neq i} \exp(sim(z_i, z_j)/\tau)} \tag{4}
$$

This loss function makes the similarity of the same category samples close to 1 and the similarity of different category samples close to 0, avoiding the above situation. Where $1_{k \neq i} \in \{0,1\}$ is the indicator function, which takes 1 when $k \neq i$ and 0 otherwise; τ is the temperature parameter. It is a constant calculated based on all positive sample pairs (i, j) or (j, i) negative sample pairs under the same batch. $sim(z_i, z_j)$ is the vector product, it represents the similarity between the feature vectors, with the following formula.

$$
sim(z_i, z_j) = \frac{z_i \cdot z_j^T}{\| z_i \| \times \| z_j \|} \tag{5}
$$

4 Experiment and results

4.1 Dataset

The experiment in this paper was validated on a total of six datasets, which include the Case Western Reserve

University (CWRU) bearing dataset, The Politecnico Di Torino Rolling Bearing Test rig bearing dataset [13], and the 4-channel variable speed dataset.

Table 1 CWRU Data Setting under Each Load

type	damage /mm	Fault label	Training set	Test set
	0.1778	0	70	30
Inner ring failure	0.3556	1	70	30
	0.5334	2	70	30
	0.1778	3	70	30
Rolling element failure	0.5334	5	70	30
	0.1778	6	70	30
Center orientation @6:00	0.3556	7	70	30
	0.5334	8	70	30
Orthogonal direction @3:00	0.1778	9	70	30
	0.5334	10	70	30
Relative direction @12:00	0.1778	11	70	30
	0.5334	12	70	30

The Case Western Reserve University Bearings (CWRU) dataset is recognized worldwide as the standard dataset for single-speed bearing troubleshooting. The experimental platform of the CWRU dataset includes fan end bearings, motors, encoders, and dynamometers. Data is collected using an acceleration sensor, and EDM causes fault damage. The load is generated by the torque sensor and encoder. The drive end bearing is SKF6205, with sampling frequencies is 12Khz and 48Khz. The fan end bearing is SKF6203, and the sampling frequency is 12Khz. As shown in Table 1, the drive end bearing includes a healthy state and three types of failure: rolling body failure, outer ring failure, and inner ring failure. The size damage diameter is 0.007inch, 0.014inch, and 0.021inch, a total of 9 damage states. This experiment selected a 48kHz sampling frequency of the drive end bearing data, each time using 48000 (per second) data points for diagnosis. Since each fault sample has approximately 5s of data, random sampling is used to expand the dataset, and each fault is randomized 100 times to generate a total of 1000 fault samples.

The Politecnico di Torino rolling bearing test rig is an experimental rig built by the DIRG laboratory of the Department of Mechanical and Aerospace Engineering of the Politecnico di Torino for the acquisition of high-speed aeronautical bearings. The rig collects various fault data under different speeds and loads. The acceleration sensor is IEPE type. The collection is the bearing 6 channels of vibration data, which includes different radial damage, a total of 6 kinds, together with normal data to form 7 classified fault data. The sampling frequency is 51,200hz, and each state lasts 10s.

The dataset is divided as shown in Table 2, and the same random sampling is used to select 51,200 points (per second) as the dataset each time, and 100 of each fault is selected as the dataset.

Table 2 Politecnico di Torino Data Setting under Each Load

name	type	size/μm	Training set	Test set
0A	Trouble-free	-----	70	30
1A	Inner ring damage	450	70	30
2A	Inner ring damage	250	70	30

3A	Inner ring damage	150	70	30
4A	Rolling ball injury	450	70	30
5A	Rolling ball injury	250	70	30
6A	Rolling ball injury	150	70	30

The variable speed dataset is from the propulsion reducer. The propulsion reducer product is installed on the test bench. The rotation of the reducer is completed by the motor driving the shaft. The product has a total of four vibration measurement points. The magnet will be vibration sensors adsorbed on the metal surface of the product to collect vibration signals, respectively fixed on the left, right, above, and below of the product.

The experimental acquisition system adopts Donghua hardware acquisition equipment and the vibration measurement software developed by our laboratory. The sensor is a voltage acceleration sensor of Wisestar Technology, model CA7002A, with a sampling frequency of 25600Hz and a sensor sensitivity of 1mv/g. The experimental acquisition system is shown in Figure 6 and consists of an industrial control computer, vibration measurement software, vibration signal collector, vibration acceleration sensor, network cable, and magnet. The industrial control computer is equipped with vibration measurement software, which will monitor the vibration signal of the equipment in real-time. After the signal is collected on the one hand, the real-time spectrum analysis (short-time Fourier transform + Hamming window) will be performed, and the spectrum analysis graph of four channels will be obtained in the computer interface, as shown in Figure 7. On the other hand, the time-domain signal data will be dropped to the disk in real time. As shown in Table 3, the acceleration and deceleration processes of the reducer in 7 different speed and load states are described, and the duration of each state is 10s, 10 samples are collected for each fault, and the duration of each sample is 100s. Because the signal volume of 100s is huge, the way of downsampling is used to reduce the data volume and, at the same time, to expand the dataset. The dataset of each failure is expanded to 100 by using the way of downsampling and randomly initializing of a sampling starting point. Each sample is 10s in this method.

As shown in Table 4: five types of failure data were collected in this experiment, namely, gear shaft bearing failure, dimension shaft outer bearing failure, bearing wear, loose bolt failure on the equipment surface, and equipment internal wear failure due to insufficient oil pressure. The five types of fault data and the normal data make up the propulsion reducer dataset with six classifications.

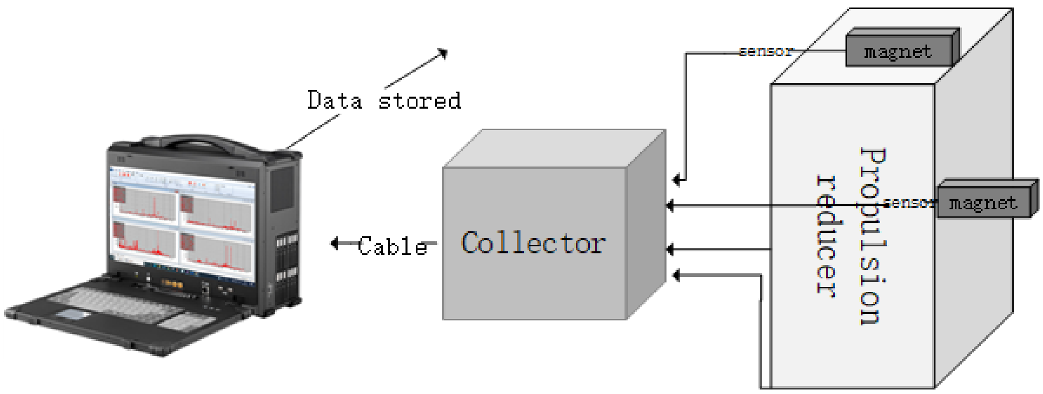


Fig 6 East An test bench diagram

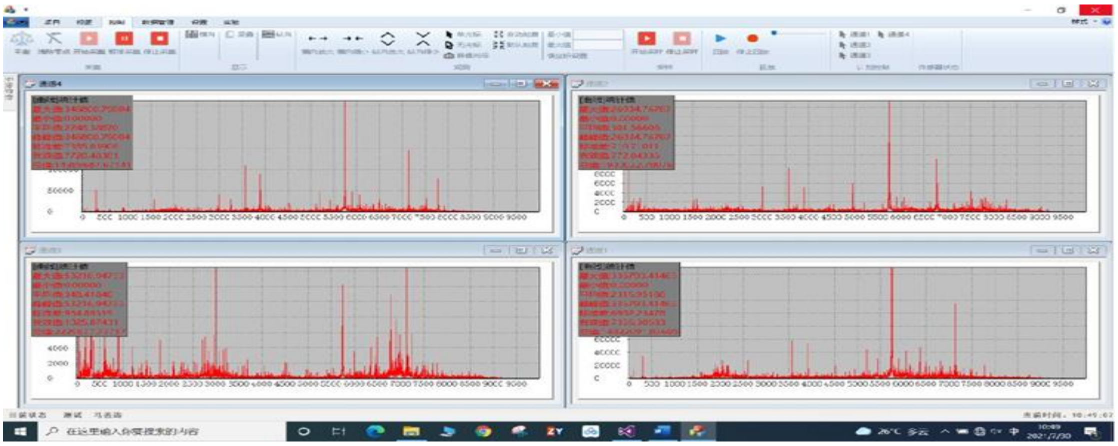


Fig 7 Time-frequency analysis diagram

Table 3 Test Bed Experimental Data Set

serial number	rotate speed /rpm	power /kw
1	5300	320
2	6744	1230
3	6744	3000
4	7678	1230
5	7678	7120
6	8300	7350
7	7885	7200

Table 4 Fault description of test bench under variable working conditions

serial number	type
1	Gear shaft bearing failure
2	Failure of the outer bearing of the dimensional shaft
3	Bearing wear
4	The fixing bolts are loose
5	Insufficient oil pressure

4.2 Evaluation Criteria

We calculate the accuracy five times and then calculate the average to prevent errors and chances. The Accuracy and Average Accuracy are calculated as follows:

$$Acc_i = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

$$AA = \frac{1}{N} \sum_{i=0}^{N} Acc_i \tag{9}$$

Where Acc_i is the accuracy of the experiment i. TP is True Positive. TN is True Negative. FN is False Negative. FP is False Positive.

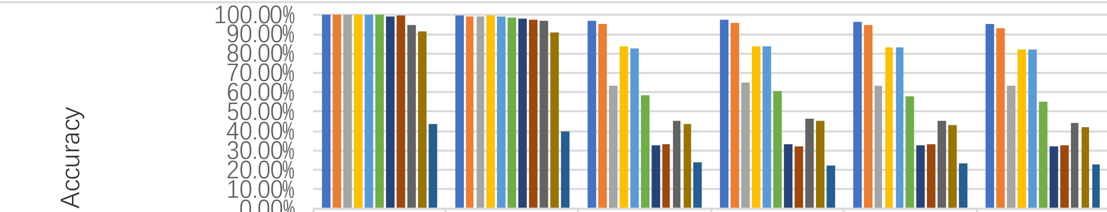
4.3 Experimental procedure

The comparison methods selected for the experiments include:

ALSTM-FCN [14] currently works best on time series classification datasets. The model achieves end-to-end classification by combining ALSTM and a Fully Convolutional Network (FCN). WDCNN is a shallow network with good results on CWRU, where the first layer extracts low-frequency features through wide convolutional kernels. Further, WDCNN extracts high-frequency features using small convolutional kernels and finally performs fault diagnosis using a multi-classification function. MSCNN is a multiscale convolutional neural network. The model first uses the average pooling of different sizes to pool the original signal. Then it uses multiscale convolution to make the feature extraction and fusion signal. Finally, it uses the full connection layer and multi-class function for realizing gearbox fault diagnosis. ResNet34, a 34-layer network structure using residual structure, is a common feature extraction method based on time-frequency image fault diagnosis. AlexNet, a convolutional neural network with an 8-layer network structure, was the champion method of ImageNet in 2012 and is also widely used in rotating machinery fault diagnosis. MLP is Multilayer Perceptron (MLP) with four fully connected layers, 4000, 1000, 100, and many fault categories, incorporating feature information for effectively classifying faults. The cross-entropy loss function (CE loss) is used.

The experiments are discussed separately from both the original signal and the time-frequency image, and two time-frequency transformation methods are used, namely STFT and HHT.

The proposed method is compared with the selected classical methods, and the time-frequency analysis methods STFT and HHT are compared, respectively. The accuracy is shown in Figure 8.



	CWRU	Politecnico di Torino	Variable speed measurement point 1	Variable speed measurement point 2	Variable speed measurement point 3	Variable speed measurement point 4
This article method STFT	100.00%	99.71%	96.61%	97.32%	96.37%	95.49%
ResNet34-STFT	100.00%	99.19%	95.23%	95.79%	94.69%	93.13%
AlexNet-STFT	100.00%	98.79%	63.37%	65.33%	63.47%	63.36%
This article method HHT	100.00%	99.34%	83.96%	83.86%	82.99%	82.23%
ResNet34-HHT	100.00%	98.89%	82.34%	83.56%	82.94%	82.16%
AlexNet-HHT	100.00%	98.39%	58.64%	60.89%	57.74%	55.25%
MSCNN	98.82%	97.98%	32.64%	33.45%	32.67%	32.17%
WDCNN	99.62%	97.49%	33.13%	32.33%	33.00%	32.93%
ALSTM-FCN	94.67%	96.62%	45.32%	46.50%	45.32%	44.12%
LSTM-FCN	91.62%	90.97%	43.46%	45.50%	43.16%	41.96%
MLP	43.62%	39.59%	23.95%	22.37%	23.13%	22.93%

Fig 8 Accuracy of each model

On the CWRU and TU datasets, all models except MLP achieve good results, and the time-frequency analysis method based on STFT and HHT achieves 100% accuracy. Therefore, the time-frequency image-based fault diagnosis method is superior to the original signal-based method, so the proposed method in this paper is based on time-frequency image analysis. Since most current methods achieve almost 100% accuracy on these two datasets, this method is only a fraction of a percentage point higher than the other methods, but the accuracy is significantly higher on the variable speed dataset.

The method's accuracy based on original signal feature extraction and classification is less than 50% on the variable speed dataset. The accuracy of the wide-convolutional-shallow convolutional neural network model, which has an impact on the CWRU dataset, is only 35%, and the accuracy of MSCNN is only 33.24%. Meanwhile, ALSTM-FCN, which performed best in the time series classification task, had slightly higher accuracy than the first two, reaching 45% accuracy. The accuracy is, on average 1.8 percentage points higher than LSTM-FCN due to the introduction of attention mechanism weights. However, the method in this paper is significantly better than other time-frequency image-based fault diagnosis methods on the variable speed dataset. Moreover, it achieves an average accuracy of 96.4% on several variable speed datasets.

4.4 SimCLR experiment

Four methods of data enhancement are used: (a) adding random noise to the original signal, (b) adding Gaussian noise to the time-frequency image, (c) filling the head and tail of the original signal with 0 values, and (d) mixing multiple signals (MixUp).

The results after fine-tuning DCANet using 100 labeled data are shown in Figure 9. The average accuracy is improved by 1.01% after adding the SimCLR framework to the DCANet network. Meanwhile, the experiments were carried out on the first channel dataset, and 100, 200, 300, 500, and 600 supervised samples were used to fine-tune DCANet, and the experimental results are shown in Figure 10. With the increased labeled sample size, the accuracy of SimCLR-DenseNet increased continuously, and the final accuracy reached 98.42%, which verifies the effectiveness of the SimCLR framework.

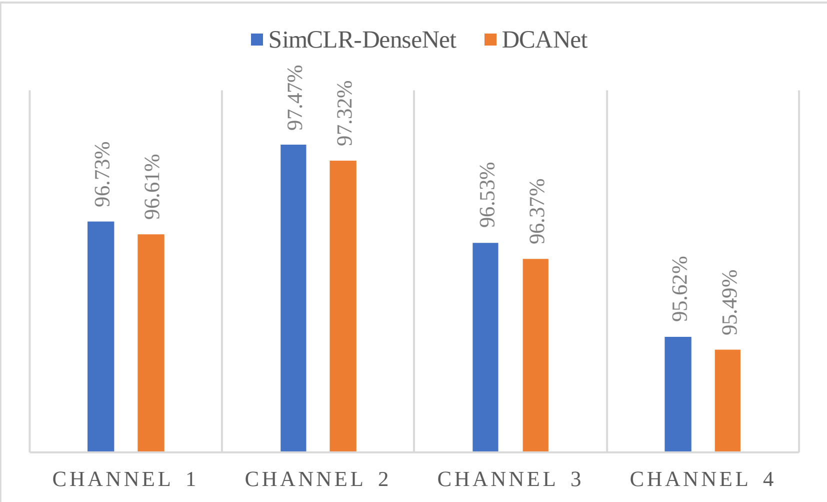


Fig 9 Self-supervised experimental results of multi-state dataset

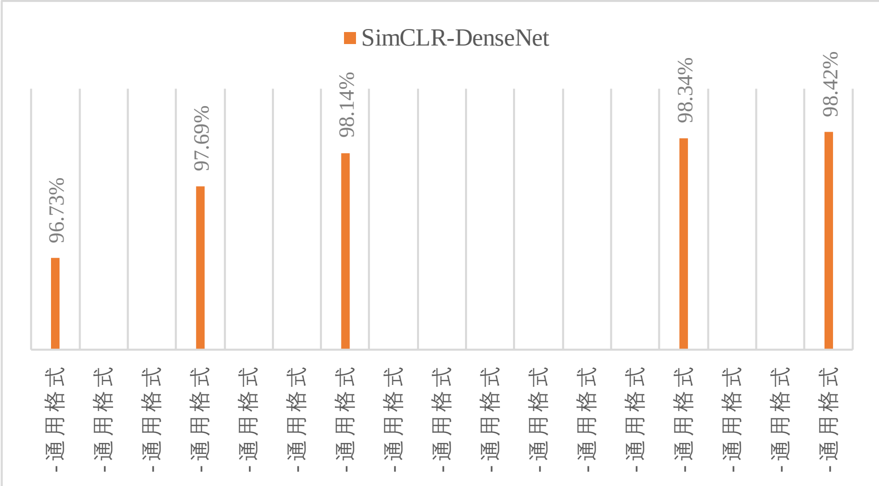


Fig 10 Fault classification accuracy under different numbers of label samples

4.5 Model ablation experiments

In this paper, four experiments are designed, which are two Dense Blocks + two SENets (Combined Experiment 1), three Dense Blocks + three SENets (Combined Experiment 2), and four Dense Blocks (Combined Experiment 3). They are all validated on the dataset of variable speed measurement point 1. We calculate the average of the five accuracies to prevent errors and chances, and the calculation formulae are detailed in 3.2.

The experimental results are shown in Table 5. The method in this paper is 33.18% higher than the combined experiment 1; it is 16.94% higher than the combined experiment 2. It shows that the network depth of using two Dense Blocks + two SENets or three Dense Blocks + three SENets in the model is insufficient to achieve effective feature extraction of fault images. The method in this paper is 15.58% higher than the combined experiment 3, which indicates that the SENet in the model significantly improves the accuracy. As seen from the analysis of the ablation experiments, the DCANet method benefits from four Dense Blocks + four SENets to achieve effective feature extraction and classification of faulty images.

Table 5 Experimental accuracy of different modules of the variable speed measuring point

	Dongan variable speed measurement point 1
This article method	96.61
Module Lab 1	63.43
Module Lab 2	79.67
Module Lab 3	81.03

5. Conclusion

A Dense channel attention network based on SimCLR for fault diagnosis is proposed to address the problems of complex actual variable working conditions of aero engines, the inapplicability of existing models, and the difficulty of fault feature extraction. This method achieves a higher accuracy than ResNet34 and is 1.74 percentage points higher than ResNet34 in the variable speed dataset while reducing the number of parameters. DCANet combines the strengths of SENet and DenseNet. Enhanced feature reuse is realized, the feature information of high, medium, and low layers is fused, and recalibration in the feature extraction process is also realized. The method effectively enhances the feature extraction capability of single-channel time-frequency images, achieves attention weighting, and extracts more effective features. After repeated tests on several variable speed datasets and comparisons with existing methods, the experiments show that the method performs well in terms of fault recognition rate under variable working condition datasets.

References

[1] Lin Jiaqi, Xu Jianguo, Liu Xingyi. Research on Fault Diagnosis Technology of Aeroengine Based on Deep Belief Network [D]. Machine Building & Automation, 2019, 048(005):179-182.

[2] Ding Yiran. Aeroengine Fault Diagnosis Based on Neural Network[J]. PRACTICAL ELECTRONICS (8):4.

[3] Hong Jiyu, Wang Huawei, Che Changchang, etc. Aeroengine Gas Circuit Fault Diagnosis Based on improved noise reduction self coding[J]. Journal of Vibration, Measurement & Diagnosis, 2019(3):603-610.

[4] Liu H, Li L, Ma J. Rolling bearing fault diagnosis based on STFT-deep learning and sound signals[J]. Shock and Vibration, 2016, 2016.

[5] He K, Fan H, Wu Y, et al. Momentum contrast for unsupervised visual representation learning[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 9729-9738.

[6] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 4700-4708.

[7] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7132-7141.

[8] Nguyen K, Nguyen Y, Le B. Semi-Supervising Learning, Transfer Learning, and Knowledge Distillation with SimCLR[J]. arXiv preprint arXiv:2108.00587, 2021.

[9] Hanifa R M, Isa K, Mohamad S. A review on speaker recognition: Technology and challenges[J]. Computers & Electrical Engineering, 2021, 90: 107005.

[10] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7132-7141.

[11] Zhang X, Ding G, Li J, et al. Deep Learning Empowered MAC Protocol Identification with Squeeze-and-Excitation Networks[J]. IEEE Transactions on Cognitive Communications and Networking, 2021.

[12] CHEN Sheng. Research on zero-sample learning method based on cluster guidance and semantic extension[D].Beijing University of Technology,2020.

[13] Daga A P, Fasana A, Marchesiello S, et al. The Politecnico di Torino rolling bearing test rig: Description and analysis of open access data[J]. Mechanical Systems and Signal Processing, 2019, 120: 252-273.

[14] Karim F, Majumdar S, Darabi H, et al. LSTM Fully Convolutional Networks for Time Series Classification[J]. IEEE Access, 2017, PP(99).