

## RESEARCH ARTICLE

# Tool Wear Monitoring Based on Transfer Learning and Improved Deep Residual Network

NAN ZHANG, (Member, IEEE), JIAWEI ZHAO<sup>1</sup>, (Member, IEEE), LIN MA, (Member, IEEE),  
HAOQIANG KONG, (Member, IEEE), AND HUAQIANG LI<sup>2</sup>, (Member, IEEE)

School of Mechanical Engineering, Inner Mongolia University of Technology, Hohhot 010051, China

Corresponding author: Huaqiang Li (lihuaqiang2002@126.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 51965051, in part by the Natural Science Foundation of Inner Mongolia under Grant 2019LH05029, in part by the Inner Mongolia Science and Technology Project under Grant 2021GG0259, in part by the Inner Mongolia University Basic Scientific Research Project under Project JY20220191, and in part by the Research Fund Project of Inner Mongolia University of Technology under Project DC2200000930.

**ABSTRACT** Considering the complex structure weight of the existing tool wear state monitoring model based on deep learning, prone to over-fitting and requiring a large amount of training data, a monitoring method based on Transfer Learning and Improved Deep Residual Network is proposed. First, the data is preprocessed, one-dimensional cutting force data are transformed into two-dimensional spectrum by wavelet transform. Then, the Improved Deep Residual Network is built and the residual module structure is optimized. The Dropout layer is introduced and the global average pooling technique is used instead of the fully connected layer. Finally, the Improved Deep Residual Network is used as the pre-training network model and the tool wear state monitoring model combined with the model-based Transfer Learning method is constructed. The results show that the accuracy of the proposed monitoring method is up to 99.74%. The presented network model has the advantages of simple structure, small number of parameters, good robustness and reliability. The ideal classification effect can be achieved with fewer iterations.

**INDEX TERMS** Tool wear, condition monitoring, deep residual network, transfer learning, convolutional neural network.

## I. INTRODUCTION

Tool wear state has great influence on the dimensional accuracy, surface integrity and machining efficiency of the workpiece. In the cutting process, due to thermodynamic coupling, chatter, abnormal wear and breakage of tool, these factors directly affect the service life of cutting tools, and even lead to abnormal machine shutdown and personnel injury. Therefore, it is great significance to monitor the tool wear state, it can effectively improve tool durability, reduce production costs, increase production efficiency and processing quality, and raise the reliability and safety of the processing process.

Tool wear condition monitoring can be divided into direct measurement method and indirect measurement method [1]. The direct measurement method directly measures the state change of the tool. The wear area is directly observed by

optical measurement method, isotope method and resistance method. Direct measurement method has high accuracy. Due to environmental constraints, it is inconvenient to conduct direct observation and the cost is high in the actual processing. The indirect measurement method can monitor the tool wear state by analyzing the associated dynamic information during the cutting process. Indirect method is low cost and strong anti-interference ability. The indirect measurement method has become a hot topic in tool wear monitoring with the rapid development of signal processing and intelligent fault diagnosis. The commonly used models of indirect measurement method in tool wear state monitoring include neural network [2], fuzzy inference [3], fuzzy neural network [4], dynamic Bayesian network [5], support vector machine [5], etc. The proportion of tool condition monitoring models is shown in Fig. 1.

Neural networks have made great progress in tool wear monitoring. Deep learning network models, represented by

The associate editor coordinating the review of this manuscript and approving it for publication was Yizhang Jiang<sup>3</sup>.

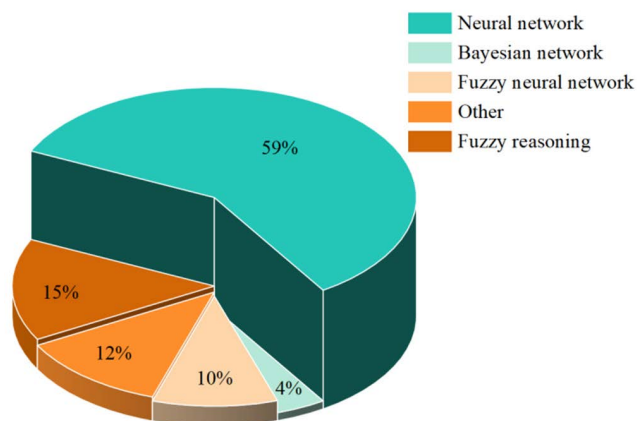


FIGURE 1. Proportion of tool condition monitoring model.

convolutional neural network and deep residual network, have been gradually applied to tool wear monitoring in recent years. Fatemeh et al. [6] enhanced the influence of tool wear in signals. The convolutional neural network (CNN) was used to estimate tool wear and different data sets were validated. It shown that the method was feasible. Convolutional neural networks can automatically identify features from multi-scale signal matrices, the method was verified by S45C steel workpiece end milling experiments with different processing parameters [7]. Wu et al. [8] built the experimental system on the machine tool. The wear image information of all blades in machining gap was obtained by matching frame frequency of industrial camera and spindle speed of machine tool. Then convolutional neural network was used to identify the tool wear state. The method has effectiveness and practicability. The cutting force generated by mechanical force model instead of experimental cutting force, which was used to predict tool wear state. Convolutional neural network was applied to confirmed the feasibility of this method by Su [9]. Convolutional neural network is also widely used in machining surface roughness estimation, bearing fault diagnosis, tool wear detection and other aspects [10].

Residual network has a good performance in tool wear state monitoring [11], intelligent diagnosis of rolling bearings [12] and fault diagnosis of gear boxes [13]. It has strong versatility and operability. Ma et al. [14] used deep residual network to achieve data-driven fault diagnosis, experimental results shown that the detection accuracy of the proposed method is significantly higher than the early faults. Peng et al. [15] fused multi-source information. Deep residual network (DRNN) was used for fault diagnosis of rotating machinery. The presented method performed well in feature learning, model training, noise resistance, fault tolerance and fault diagnosis.

Although deep learning has made great achievements in the field of fault diagnosis. But the deep learning model still has some flaws. The training of deep learning needs a lot of data, once the data is insufficient, it will cause the phenomenon of over-fitting. The maximum pooling layer of deep learning is

poor in retaining local information, it is difficult to process missing data, complicated weight of network structure and easy to ignore the correlation of attributes in data set. In addition, the parameters of the model increase exponentially when hidden layers are increase. It is huge to train a deep learning network model in the number of label samples, time and computing power.

To solve the above problems, this paper put forward a tool wear state monitoring method based on Transfer Learning and Improved Deep Residual Network. Firstly, wavelet transform is used to process cutting force signals for eliminating the influence of manual features and obtaining the required spectrum. Secondly, convolution series omits a large number of training tuning parameters. Thirdly, the residual module is improved to place the LeakyReLU activation function before the convolution layer. Fourthly, the Dropout layer is added to randomly lose network units and weights connected to them during training to reduce overfitting and improve accuracy. Fifthly, global mean pooling technique is used to reduce the number of training parameters and test time of the model. Finally, all the convolutional layers of the pre-training network are frozen to construct the required full connection layer for Transfer Learning after the construction of the Improved Deep Residual Network and the tool wear state monitoring is realized.

## II. DATA PREPROCESSING

The data collected in this paper were the cutting force signals of a certain type of aero-engine blisk. The signal was a one-dimensional time-varying unsteady signal. When the deep learning model in the field of Computer Vision (CV) is a pre-training model, one-dimensional signals should be converted into two-dimensional image data. There are two transformation methods. The first method is based on data reconstruction, which directly intercepts and splice the one-dimensional cutting force signal to reconstruct the two-dimensional image. The method is simple in operation, but the frequency domain information of the signal is ignored. The second method is based on time-frequency domain conversion, including Fourier Transform (FT) [16] and Wavelet transform (WT) [17]. Fourier Transform (FT) has poor ability to characterize non-stationary signals, which may lead to information loss when extracting deep features. Wavelet transform (WT) uses window function with fixed area and variable shape to balance time resolution and frequency resolution through multi-resolution analysis. Therefore, wavelet transform will be used in this paper to process the data.

### A. WAVELET TRANSFORM

French scholar Morlet [18] put forward wavelet transform. It has been widely used in signal processing, image processing, applied mathematics and many other fields. It is mainly used in filtering, denoising, compression and transmission in signal processing and analysis. The wavelet function is defined as follows:

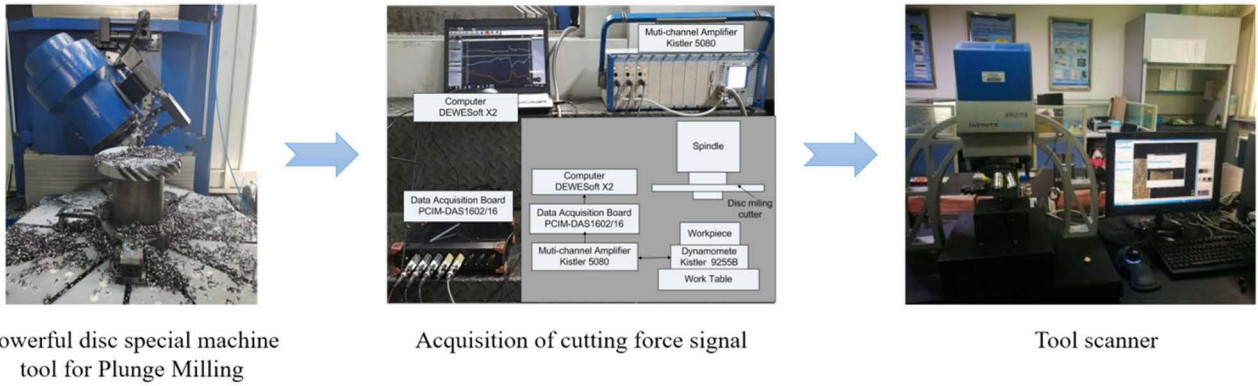


FIGURE 2. The data acquisition process.

If  $\Psi(t) \in L^2(\mathbb{R})$  satisfies the admissibility condition:

$$C_\Psi = \int_{-\infty}^{\infty} \frac{|\hat{\Psi}(\omega)|^2}{|\omega|} d\omega < \infty \quad (1)$$

Then  $\Psi(t)$  is called an admissible wavelet (integral wavelet, fundamental wavelet), Where  $\hat{\Psi}(\omega)$  is the Fourier transform of  $\Psi(t)$ . The wavelet function generated by the fundamental wavelet can be expressed as:

$$\Psi_{a,b}(t) = |a|^{-1/2} \Psi\left(\frac{t-bR}{a}\right) \quad (2)$$

where,  $a$  is the expansion factor and  $b$  is the translation factor. The area of  $\Psi_{a,b}(t)$  is fixed, the size of  $a$  only affects the length of  $\Psi_{a,b}(t)$ . The discrete wavelet function can be written as:

$$\Psi_{j,k}(t) = a_0^{-j/2} \Psi(a_0^{-j}t - kb_0), \quad j, k \in \mathbb{Z} \quad (3)$$

The discrete wavelet transform coefficient can be represented as:

$$W_{j,k}(t) = \int_{-x}^{\infty} f(t) \overline{\Psi_{j,k}(t)} dt \quad (4)$$

The reconstruction formula of discrete wavelet transform is as follows:

$$f(t) = C \sum_{-\infty}^{\infty} \sum_{-\infty}^{\infty} W_{j,k} \Psi_{j,k}(t) \quad (5)$$

where  $C$  is a constant independent of the signal.

### B. BLISK PROCESSING TEST AND SIGNAL ACQUISITION

The workpiece material was titanium alloy TC17. The disk and blades used in the test were provided by Zhuzhou Diamond Cutting Tools Co., LTD. The blade arrangement of disk milling cutter was left, middle and right. The blade model is indexable YBG212. It is made of hard alloy steel with TiAlN non-coating. The parameters of disk and blade were shown in Table 1. The powerful compound machine developed by Northwestern Polytechnical University was used. The processing method was climb milling and cooling by emulsion.

TABLE 1. Disk milling cutter and blade parameters.

The blade number	Diameter	Thickness	Cutting tool rake Angle	Sampling rate
(PCs)	(mm)	(mm)	(°)	(KHz)
39	420	15	15	10
Blade width	Blade setting Angle	After the Angle	Radius of tool tip arc	
(mm)	(°)	(°)	(mm)	
6	±2	0	0.8	

The monitoring device for cutting forces was the piezoelectric force measuring instrument Kistler 9255B. The tool wear in flank was measured by the IFM-G4 automatic tool measuring instrument. The dynamometer was fixed on the work table and the workpiece was fixed on the dynamometer with four 12mm bolts. The instantaneous cutting force signal was amplified by multi-channel amplifier Kistler 5080 and transmitted to computer by data acquisition card PCI-DAS602/16. The data was recorded by DEWESoftX2 software and the cutting force signal was collected. The data acquisition process was shown in Fig. 2.

### C. PROCESS THE DATA

According to the spindle speed with 70r/min and the feed speed with 20mm/min, 30mm/min and 40mm/min, the cutting force signals were divided into three categories. A total of 5164 sets of data were used to analyze tool wear at different feed rates. All-pass filter was used to remove interference from cutting force signal. The cutting force signal, spectrum diagram and scale diagram were shown in Fig. 3, Fig. 4, and Fig. 5, respectively. The first Type data groups (S with 70r/min and F with 20mm/min) were shown in Fig. 6, the second type (with 70r/min and F with 30mm/min)S were shown in Fig. 7 and the third type (S with 70r/min and F with 40mm/min) were shown in Fig. 8.

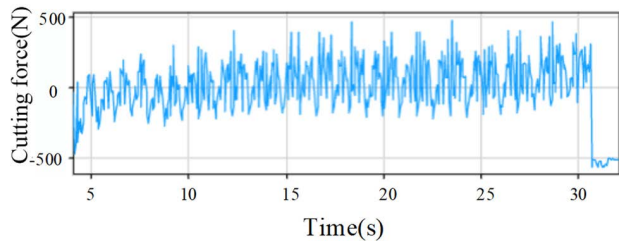


FIGURE 3. Cutting force data.

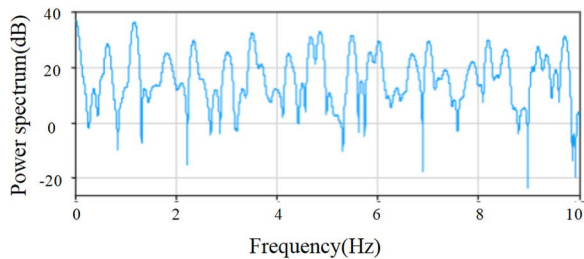


FIGURE 4. Spectrum.

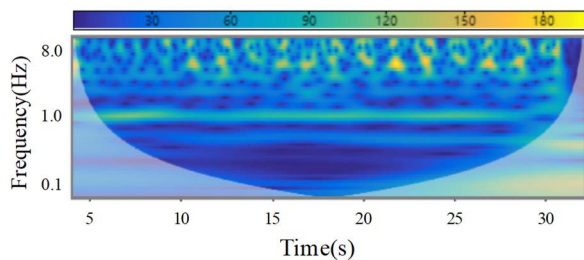


FIGURE 5. Scale figure.

The input signal was decomposed into some subband signals by wavelet transform. Some of these subband signals can be used to represent the trend of the signal, the coefficients of these subbands can be set to 0. The signal was reconstructed by discrete wavelet inverse transformation to remove the trend of cutting force signal. The reconstructed signal was shown in Fig. 9. Fig. 9 shows the first group of data reconstruction signals of the first type.

### III. DEEP LEARNING MODEL FRAMEWORK

#### A. CONVOLUTIONAL NEURAL NETWORK

Convolutional Neural Network (CNN) is a typical deep learning network architecture inspired by biological visual perception mechanism. Typical convolutional neural networks are composed of convolutional layer, pooling layer, full connection layer and Softmax classification layer, as shown in Fig. 10.

The convolutional layer is the core of the convolutional neural network, which can extract the required features according to the objective function. The convolution layer is to extract features by calculating the inner product of the

overlapping region between the 2d convolution kernel and the corresponding input image and traversing every pixel on the whole image through a nonlinear activation function. The mathematical model of the convolution layer is expressed as follows:

$$x_j^l = f \left( \sum x_i^{l-1} * k_{ij}^l + b_j^l \right) \tag{6}$$

where  $x_j^l$  is the value of the  $j$  feature map of layer  $l$ ,  $f$  is nonlinear activation function,  $x_i^{l-1}$  is the  $i$ th feature map value of layer  $l - 1$ ,  $k$  is convolution kernel,  $b_j^l$  is the  $j$ -th bias parameter of layer  $l$ . The activation function used is Rectified Linear Units(ReLU), its mathematical model is as below:

$$\text{ReLU}(x) = \max(0, x) \quad x \in (-\infty, +\infty) \tag{7}$$

The pooling layer is usually located after the convolution layer to reduce the size of the feature graph and introduce invariability. Its mathematical model is represented as follows:

$$x_j^l = f \left( \beta_j^l \text{down} \left( x_i^{l-1} + b_j^l \right) \right) \tag{8}$$

where down() is a sub-sampling function. The full connection layer is the last layer of convolutional neural network, which is used to perform classification and regression tasks. Its mathematical model is written as below:

$$\hat{y} = f(Wx + b) \tag{9}$$

where  $\hat{y}$  is the output of full connection,  $W$  is the weight matrix multiplied by the input eigencolumn vector  $x$ ,  $b$  is the offset column vector,  $f(\cdot)$  is the activation function. The full connection layer adopts SoftMax function as the activation function output. The SoftMax function is defined as:

$$S_i = \frac{e^{z_i}}{\sum_1^N e^{z_n}}, \quad i = 1, \dots, N \tag{10}$$

$$z_i^l = \sum w_{ij}^l x_i^{l-1} \tag{11}$$

where  $w_{ij}^l$  is the weight from the  $j$  neuron to the  $i$  neuron in layer  $l - 1$ .

#### B. IMPROVED DEEP RESIDUAL NETWORK

He et al. [19] presented the deep residual network to solve the problem of gradient disappearance in convolutional neural network extracting deep features. The deep residual network [20] is composed of multiple residual modules with jumping connections. The residual module can effectively solve the problems of gradient dispersion and gradient explosion after network deepening. The residual module is defined as follows:

$$y_l = F(x_l, \{W_l\}) + x_l \tag{12}$$

$$x_{l+1} = f(y_l) \tag{13}$$

where  $x_l$  and  $y_l$  represent the input and output of the residual module,  $W_l$  is the weight matrix,  $F(x_l, \{W_l\})$  is the residual mapping to be learned by the network,  $f(\cdot)$  is the Rectified

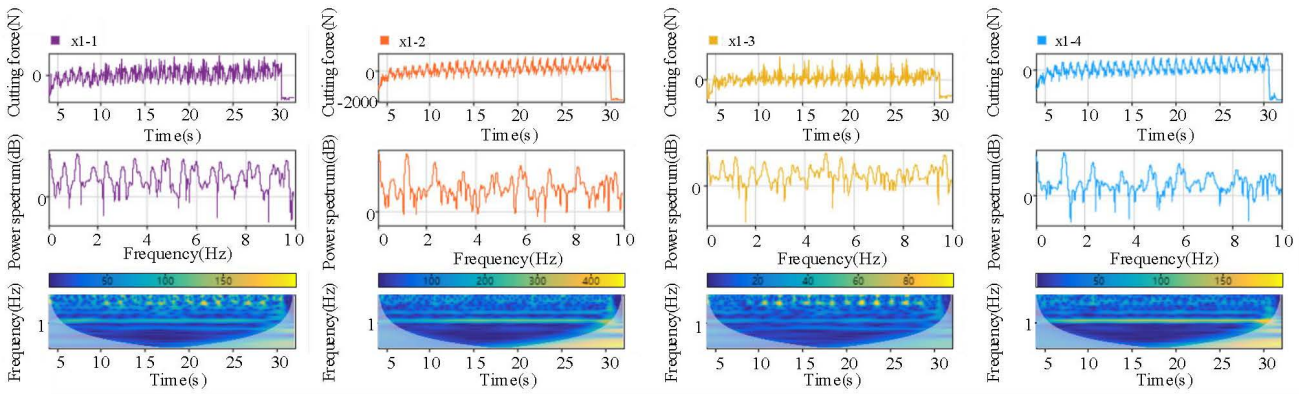


FIGURE 6. The first type of data group (S with 70r/min and F with 20mm/min).

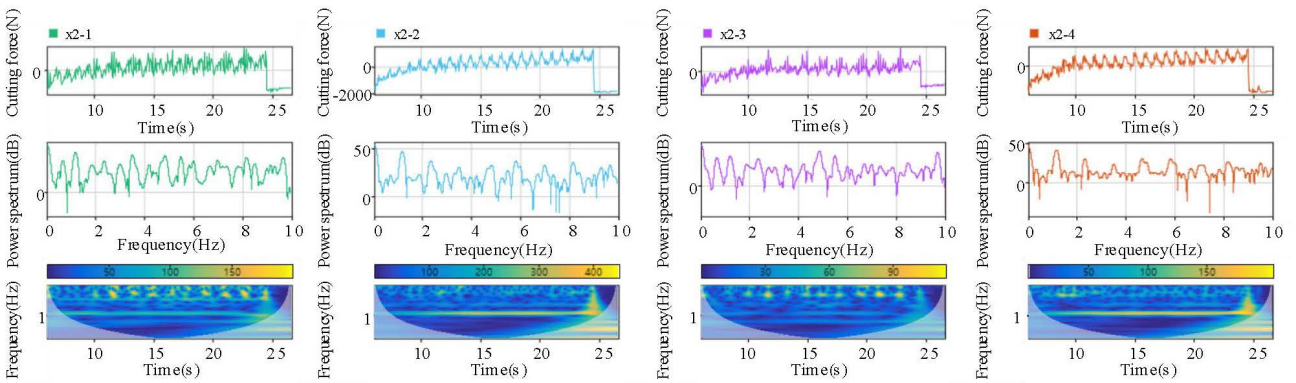


FIGURE 7. The second type of data group (S with 70r/min and F with 30mm/min).

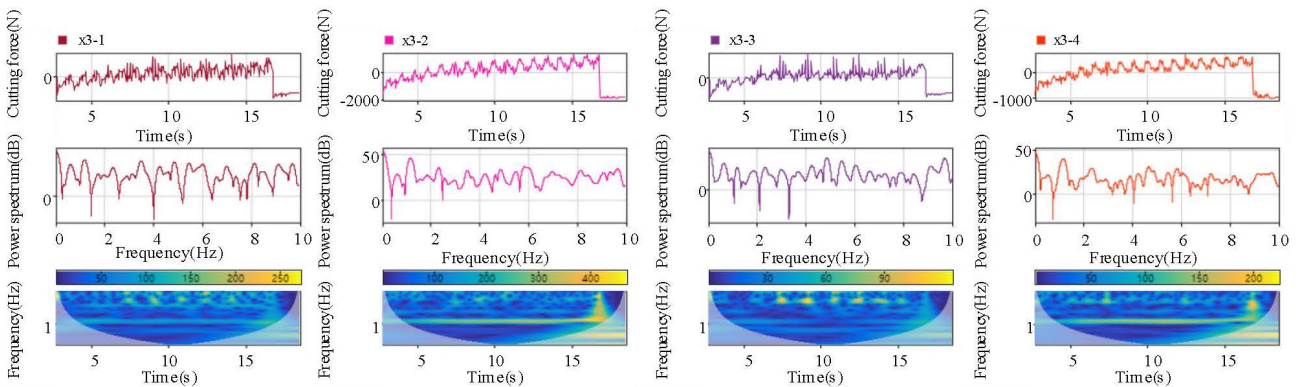


FIGURE 8. The third type of data group (S with 70r/min and F with 40mm/min).

Linear Units (ReLU). For the layer residual module, the learning characteristics from shallow to deep are written as follows:

$$x_L = x_l + \sum_{i=1}^{L-1} F(x_i, \{W_i\}) \quad (14)$$

When the corrected Linear Units (ReLU) [21] input is less than 0, the weight cannot be updated and the subsequent

training is in a silent state. Therefore, the advanced activation function LeakyReLU was used instead of ReLU function to improve the robustness of the model. The mathematical expression of the LeakyReLU function is as follows:

$$y = \max(0, x) + \varepsilon \min(0, x) \quad (15)$$

where  $\varepsilon$  is a small constant. It can retain some values of the negative axis. There is still a small gradient of non-zero output

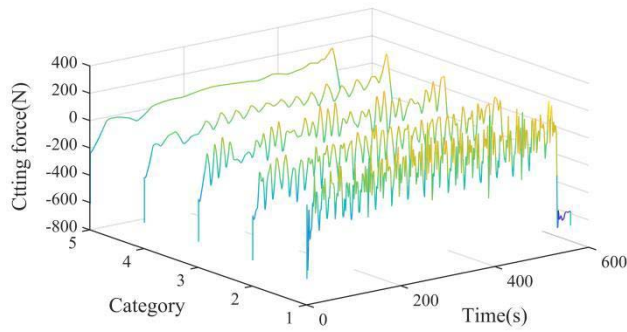


FIGURE 9. Reconstructed signal group.

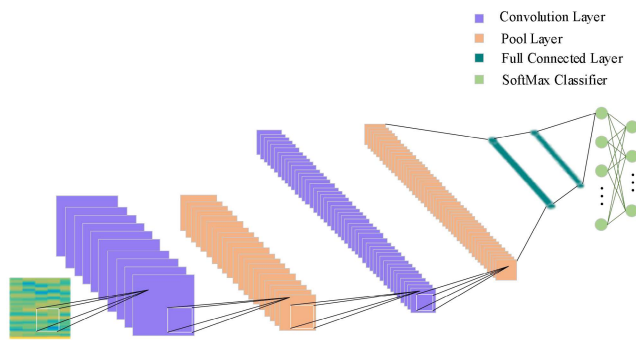


FIGURE 10. Convolutional neural network.

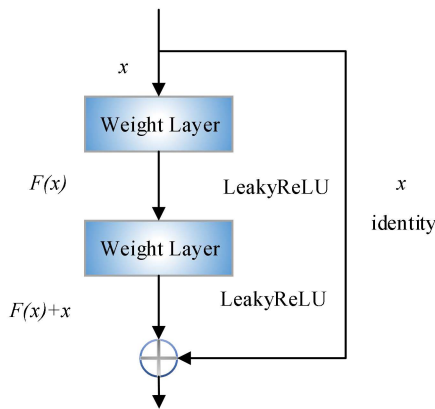


FIGURE 11. Residual module.

when neurons are not activated. It can avoid the pitfalls of the ReLU function.

The residual module is shown in Fig. 11. By learning residuals, part of the original input information is directly transmitted to the next layer through the identity mapping layer to alleviate the problem of feature loss in the information transmission of the convolution layer. This structure improves the expression ability of the model and avoids the degradation caused by the deepening of network layers.

The residual optimization structure is adopted to facilitate the construction of deep network and reduce the number of network parameters and computation, as shown in Figure.11. Firstly, a  $1 \times 1$  convolution layer is used for dimensionality

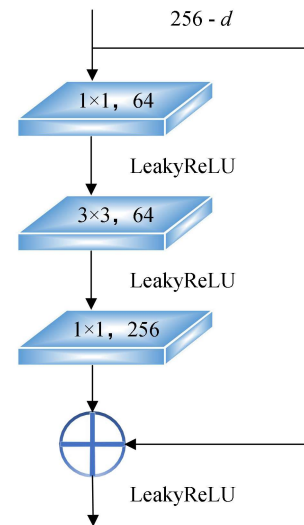


FIGURE 12. Optimized structure.

reduction. Then, another  $1 \times 1$  convolution layer is used for dimensionality reduction after the  $3 \times 3$  convolution layer. Compared with the traditional residual structure, the proposed structure not only ensures the accuracy of calculation, but also reduces the number of parameters and saves calculation force. In Fig. 12,  $d$  is the input data dimension and the parameters in the box are the convolution kernel size and output channel of each layer.

The data set collected in the actual operation often contains a certain amount of abnormal data that will induce the network to learn as a rule, lead to over-fitting problems and reduce model accuracy. Therefore, Dropout technology will introduce into the residual module in this paper. Dropout technology was proposed by Hinton et al. [22] to solve the overfitting problem. The idea of Dropout technology is to randomly discard some of the hidden layer neurons to reduce the chance of abnormal data to learn and diminish the impact of abnormal data on the network. The operation diagram of Dropout is shown in Fig. 13. Figure 13(a) shows the deep feedforward network and Figure 13(b) shows the deep feedforward network after Dropout. The number of intermediate features, redundancy and complex co-adaptive relationships between neuron are reduced by adding a Dropout layer. It can also increase the orthogonality between the features of each layer and avoid over-fitting.

Furthermore, the LeakyReLU activation function is placed before the convolution layer to achieve direct connection between the input and output, which maximizes the retention of transmitted information. Since the number of redundant parameters in the fully connected layer is not conducive to the operation of the network, the global average pooling layer is adopted in this paper to replace the full connection layer. The global average pooling directly pools the input features without neurons, which can simplify the network structure and omit a lot of network parameters. The internal structure of

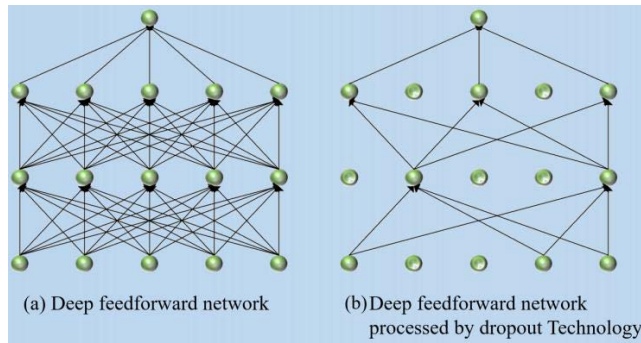


FIGURE 13. Dropout operation schematic diagram.

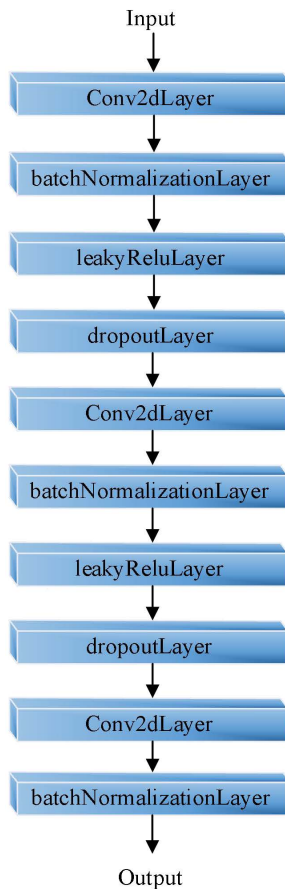


FIGURE 14. Improved residual module.

the Improved Residual Module is shown in Fig. 14. The structure of the Improved Deep Residual Network is displayed in Fig. 15.

### C. TRANSFER LEARNING

The realization methods of Transfer Learning are divided into sample transfer [23], feature transfer [24], model transfer [25] and relationship transfer [26]. The model transfer method was adopted in this paper. The Improved Deep Residual Network was used as a pre-training network model, in which

all convolutional layers were frozen, all required connection layers were built and global fine-tuning was performed. In Transfer Learning, synthetic samples are used as source domain  $D_s$  and the training set after data enhancement is used as target domain  $D_t$ . Samples in source domain and target domain follow the same label distribution. The data from  $D_s$  is pre-trained to the model and the parameters are saved. The weight matrix can be recorded as  $W_s$ .  $W_s$  was applied to initialize the parameters of the target network. The data of  $D_t$  was used to retrain the whole network and a new model was obtained through fine-tuning.

Yosinski et al. [27] evaluated the migration ability of convolution layers at different positions through a large number of experiments. It was found that the features extracted from the lower convolutional layer had strong migration ability, while the features extracted from the higher convolutional layer that related to specific tasks were not suitable for migration and needed to be retrained on new data sets.

Therefore, the proposed network is adopted freezing of the low level network parameter and the high level network parameters that migrated from the network parameters trained by cutting force data. In this paper, most of the architecture of the Improved Deep Residual Network model was retained when pre-training the network model. The fully connected layer was rebuild. The parameters of the first seven residual network modules are frozen and used as shallow feature extractors. The parameters of the 8th residual module are randomly initialized to learn the deep features. Softmax classification of layer is used to obtain the tool wear status.

### D. TRANSFER LEARNING WITH IMPROVED DEEP RESIDUAL NEURAL NETWORKS

Aiming at the deficiency of traditional deep learning model, an Improved Deep Residual Network model based on Transfer Learning and Improved Deep Residual Network was proposed for tool wear monitoring. Transfer Learning is adopted to train the target domain to establish the sample recognition and classification model, which can avoid the time and computing power wasted by repeated training on the network model. The deep residual network is used to solve the problems of gradient dispersion, gradient explosion and degradation after network deepening. The residual module is improved to enhance the robustness of the model during training. The schematic diagram of Transfer Learning and Improved Deep Residual Network framework is shown in Fig.16, and the basic process is shown in Fig.17.

Firstly, the data was preprocessed as follows:

- The cutting force signals were collected.
- The one-dimensional cutting force were converted into two-dimensional signals by wavelet transform.

Then, the deep residual network was set up as follows:

- The LeakyReLU activation function was applied to replace ReLU function in the residual module to improve the robustness of the model.

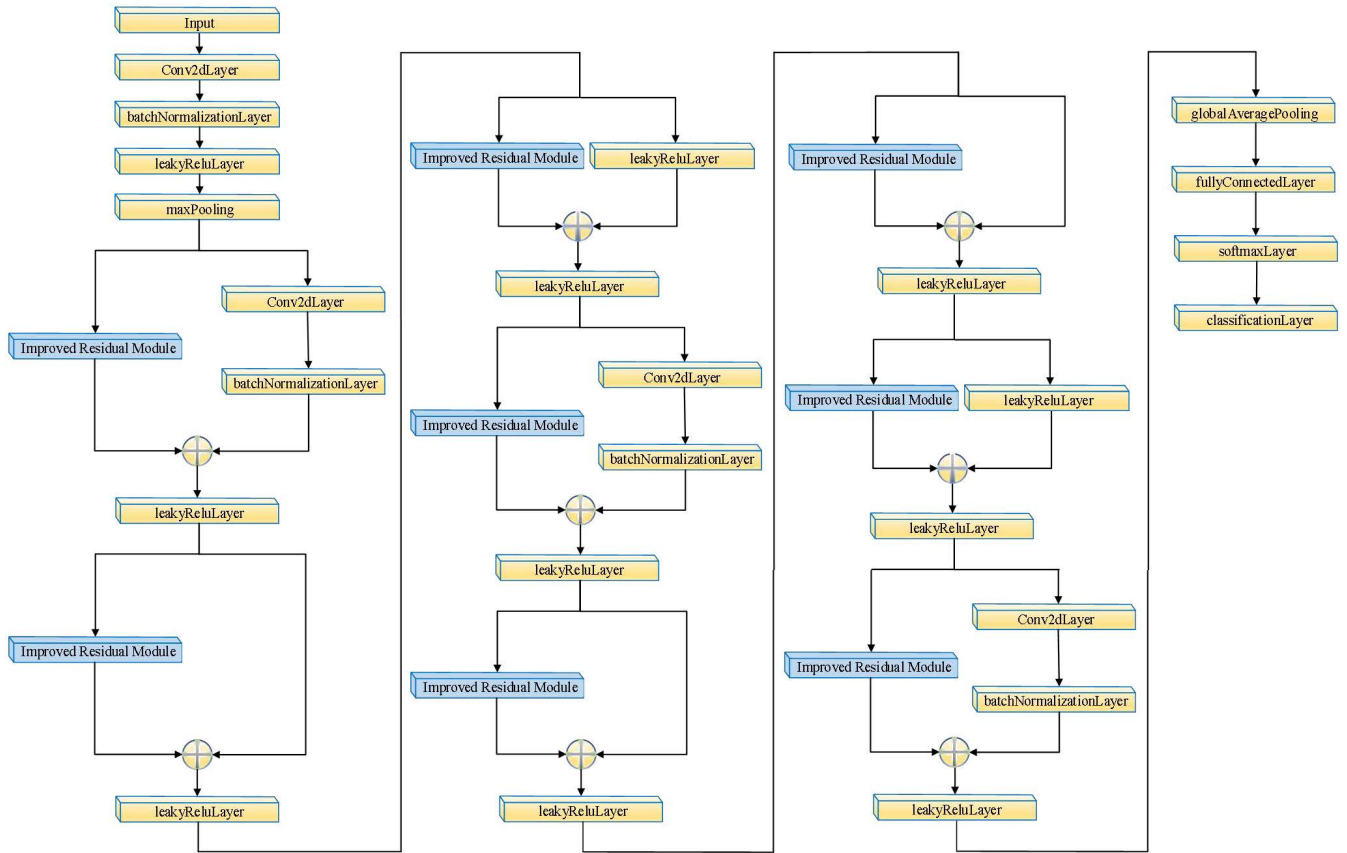


FIGURE 15. Improved deep residual network structure diagram.

- The LeakyReLU activation function was placed before the convolutional layer to maximally retain the transfer information.
- Dropout technology was introduced into the residual module to discard part of the redundant information in the network, which can improve the over-fitting phenomenon of the network.
- The global average pooling layer was used instead of the fully connected layer to simplify the network structure.

Finally, the model was migrated as follows:

- After the Improved Deep Residual Network was built, all convolutional layers were frozen and the fully connected layer was reconstructed for Transfer Learning.
- The tool wear state monitoring was completed by training network.

#### IV. TOOL WEAR MONITORING APPLICATIONS

##### A. EXPERIMENTAL ENVIRONMENT AND PARAMETER INDEX

Deep learning framework Matlab2021a was adopted to verify the feasibility of the presented network model in this paper. The experimental environment is shown in Table 2. To improve the performance of the proposed networks in this paper, ablation experiments were used to determine the

TABLE 2. Table of experimental environment.

Environment	Configuration information
Hardware	CPU: Intel(R) Core(TM) i7-8750H Memory: 8GB GPU: NVIDIA GeForce GTX 1050Ti
Software	Matlab2021a, DEWESoftX, Windows11, Origin2021.

optimal network parameters. The parameters of the ablation experiment are shown in Table 3.

##### B. EXPERIMENTAL RESULTS

In the case of wear, the cutting force data were collected in this paper. The spindle speed was 70 r/min, the feeding speed were 20 mm/min, 30 mm/min and 40 mm/min. A total of 5164 sets of data were divided into three categories. The ratio of training set, validation set and test set is 6:3:1. In the Improved Deep Residual Network model, and then Transfer Learning. The accuracy of the ablation experiment training process was shown in Fig. 18(a). The Model 1, Model 2 and Model 3 are the same network models with different parameters and the Model 4 is the Improved Deep Residual Network without Transfer Learning. The accuracy



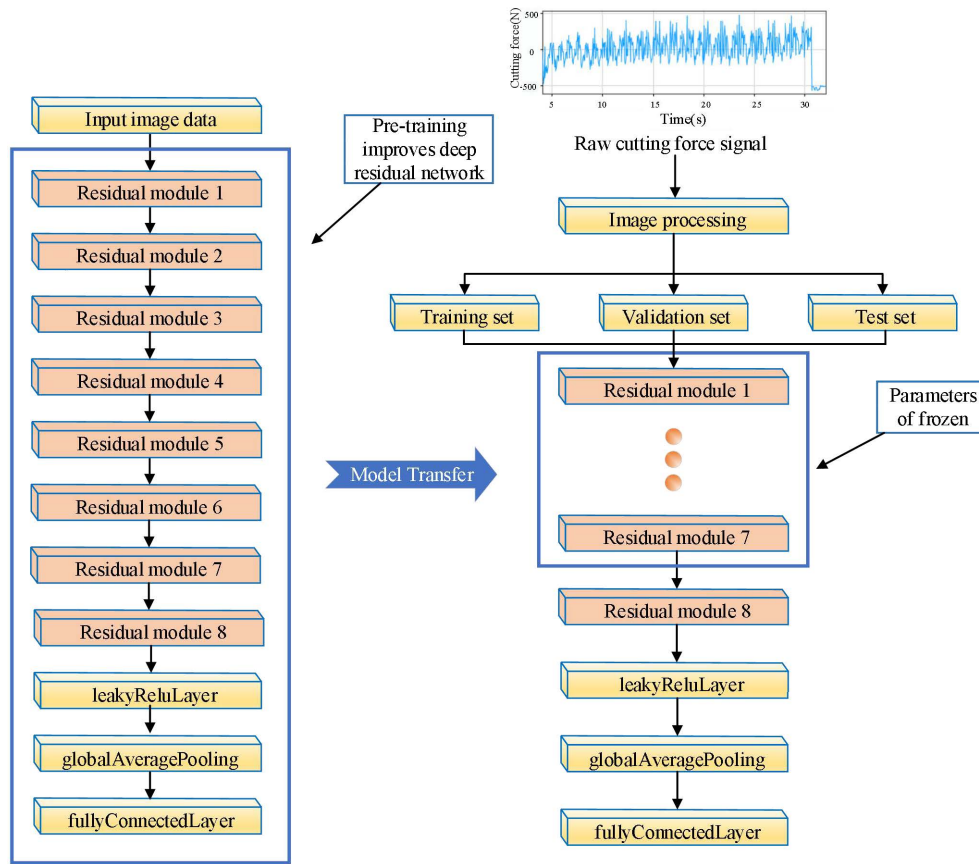


FIGURE 16. Transfer learning and improved deep residual network framework.

TABLE 3. Ablation experiment parameter index.

Parameter types		Transfer Learning and Improved Deep Residual Network1 (Model1)	Transfer Learning and Improved Deep Residual Network2 (Model2)	Transfer Learning and Improved Deep Residual Network3 (Model3)	Improved Deep Residual Network4 without Transfer Learning (Model4)	Transfer Learning and Deep Aesidual Networks (ResNet50)
Dataset Parameters	Training set	3099	3099	3099	3099	3099
	Validation set	1550	1550	1550	1550	1550
	Test set	515	515	515	515	515
	Data categories	3	3	3	3	3
Network model parameters	Iterative rounds	40	40	40	40	40
	Number of iterations per round	4	4	4	4	4
	Small batch size	128	128	128	128	128
	Random inactivation	0.5	0.1	0.5	0.5	0
	Vector	0.001	0.001	0.01	0.001	0.001

of Model 1, Model 2, Model 3, Model 4 and ResNet50 [28] were 99.74%, 94.55%, 92.33%, 84.78% and 91.67%, respectively. The accuracy of Model 1 was the highest. It was decided to use the Model 1 as the final network. Fewer training times can achieve ideal classification effect. The loss rate of ablation experimental training was shown in Fig. 18(b). The confusion matrix of Model 1 was displayed in Fig. 19,

where Fig. 19(a), Fig. 19(b), Fig. 19(c) and Fig. 19(d) are for the training set, the validation set, the test set and the overall dataset, respectively.

C. ALGORITHM CONTRAST

In order to verify the reliability of the proposed model in this paper, comparison experiments were conducted using

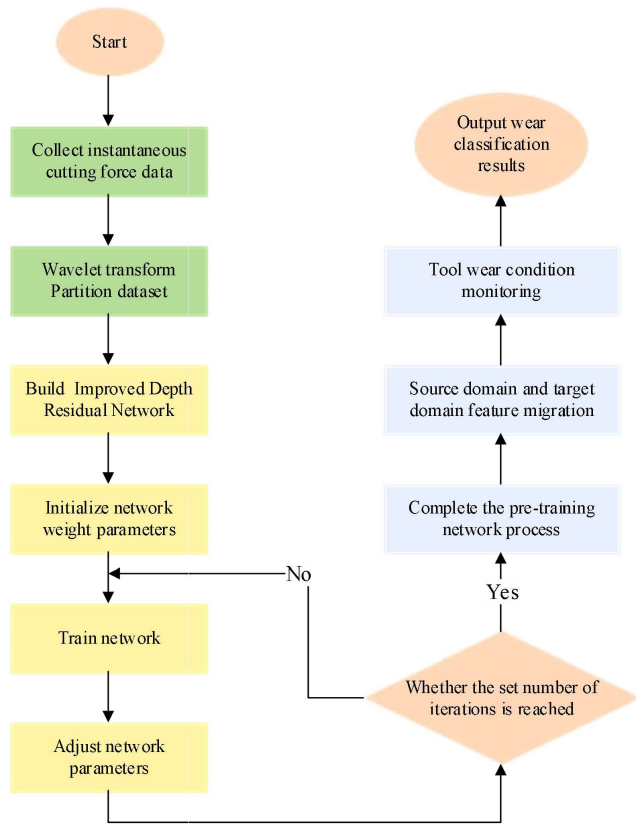


FIGURE 17. Flow chart of tool wear monitoring.

Support Vector Machine(SVM), Decision Tree Algorithm(DT), Naive Bayes(NB), K-Nearest Neighbor(KNN), Neural Network(NN) and classical Convolutional Neural Network (VGG16) [29] for comparative experiments. The VGG16 network is a Convolutional Neural Network model proposed by Simonyan et al. in “Very Deep Convolutional Networks for Large Scale Image Recognition”. The VGG16 model has better performance in classification tasks and its name comes from the Oxford University Geometry Group (Visual Geometry Group) abbreviation. The results were

shown in Fig. 20. The classification accuracy of VGG16 network, SVM, DT, NB, KNN and NN were 95.5%, 78.8%, 76.5%, 72.0%, 79.9% and 75.7%, respectively.

As shown in the figure, the model of Transfer Learning and Improved Deep Residual Network used in this paper had higher accuracy.

### V. DISCUSS

The one-dimensional convolutional neural network was used to train time signals for bearing fault diagnosis by Zhang et al. [30] Due to the initial phase difference of time samples, the learning process of convolutional neural network would be interfered. Therefore, wavelet transform was applied to convert the collected one-dimensional time-varying unsteady cutting force signals into two-dimensional data and the two-dimensional signals were used as the input of the deep learning network model in this paper. Wavelet transform has good effect on characterizing abrupt and singular signals and is suitable for processing cutting force signals. Moreover, all-pass filter is adopted to reduce the correlation between different features extracted.

Tool wear monitoring based on deep learning has been studied deeply. Zhang et al. [31] adopted wavelet packet decomposition to process the cutting force data and fused the signal characteristics. The sample error was 8.2%. However, the problem of gradient dispersion or gradient explosion caused by the deepening of network layers was ignored. Deep residual network was applied to solve this problem in the paper. The residual structure was improved to replace the traditional ReLU activation function with an advanced LeakyReLU activation function and introduce a Dropout layer. In order to make full use of effective resources and save computational power, the Improved Deep Residual Network was used as the pre-training model and the model migration method was adopted to tool wear monitoring of machining aero-engine blisk.

To verify the feasibility of the proposed network, ablation experiments were used to compare network models with different parameters. The proposed method was compared

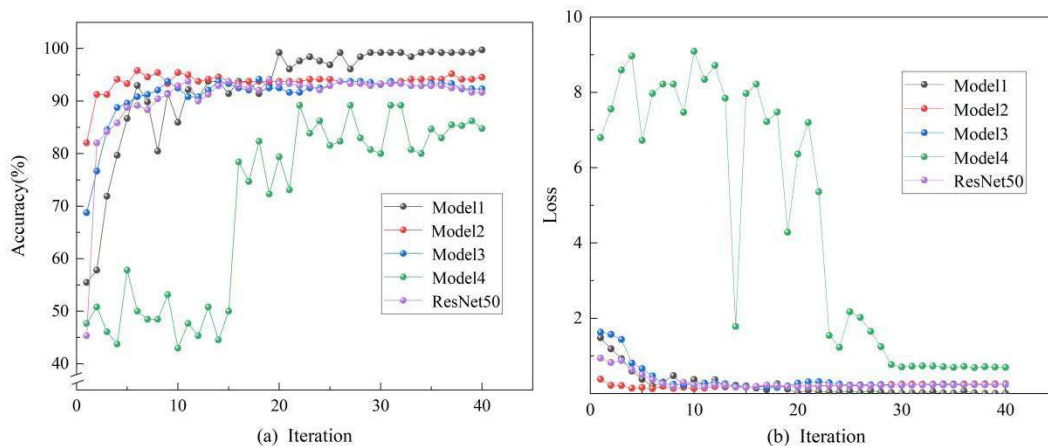


FIGURE 18. Raining accuracy and loss rate.

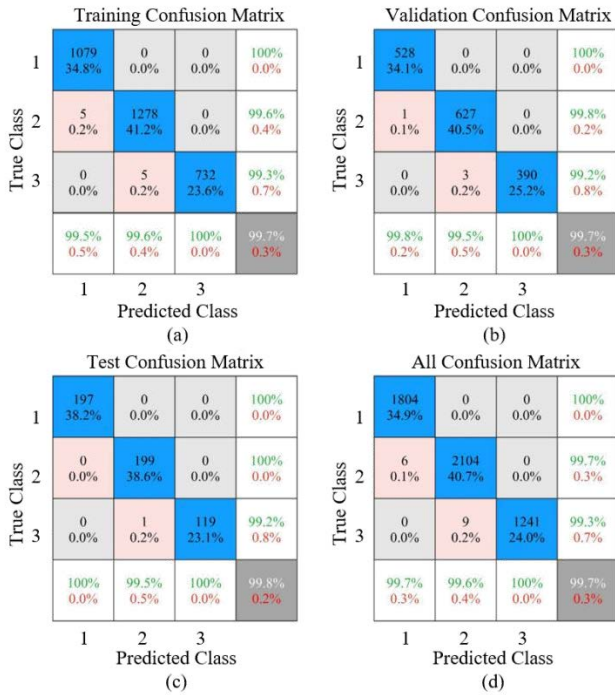


FIGURE 19. Confusion matrix.

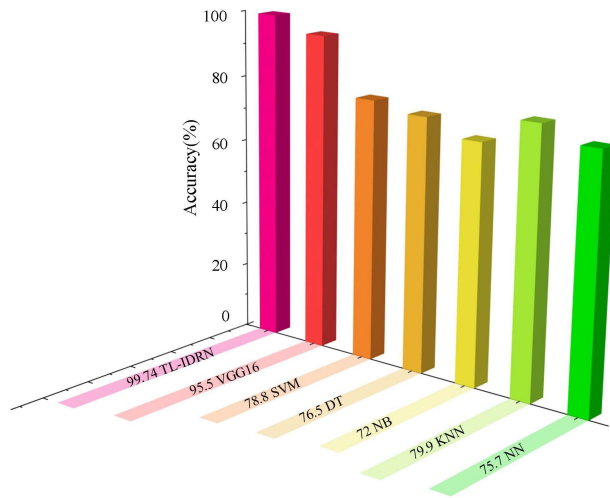


FIGURE 20. Comparative analysis.

with several commonly used machine learning methods in section 4. The traditional model K algorithm with the highest accuracy of shallow learning was 79.9% and the VGG16 network with the highest accuracy of deep learning was 95.5%. The input of shallow learning was one-dimensional time-varying unsteady cutting force signal without wavelet transform. The shallow learning was difficult to deal with missing data, prone to over-fitting problems and easy to ignore the correlation of attributes in the data set, which lead to low accuracy in generally. Data was preprocessed and wavelet transform was used in traditional deep learning.

But the problem of network degradation caused by gradient explosion and gradient disappearance had not been solved.

The spectrum graph was directly input the presented method of Transfer Learning and Improved Deep Residual Network, which can avoid the complex process of feature extraction and data reconstruction and can learn representative features directly from raw data. Moreover, it can combine feature extraction and fault diagnosis in a model (end-to-end). The structure nodes of the proposed network in this paper are 112 layers. The structure nodes of the traditional deep residual network ResNet50 are 177 layers. The accuracy rate of the proposed network was 8.07% higher than that of ResNet50. Experimental results showed that the proposed network simplified the network structure, improved the computing efficiency and enhanced the robustness.

## VI. CONCLUSION

A tool wear monitoring method based on Improved Deep Residual Network and Transfer Learning was presented in this paper. Dropout layer was introduced to prevent the occurrence of overfitting phenomenon. The LeakyReLU activation function was added to improve the robustness. The Transfer Learning model was applied to pre-trained to improve the accuracy and generalization performance of the model monitoring. It was compared with convolutional neural network, support vector machine and other machine learning. The conclusions were as follows:

- The proposed model is suitable for tool wear monitoring. The data is directly input the model without manual feature extraction of processed 2D data. The “end-to-end” model structure has better operability and versatility. The model can be changed by modifying parameters, which has strong flexibility and growth.
- Compared with traditional machine learning, the accuracy of the presented model was 99.74% and the VGG16 network with the highest accuracy in traditional machine learning was 95.5%. The proposed model has excellent fault diagnosis performance.
- The feasibility of the proposed method was verified by cutting force signals in this paper. Other types of signals can be used to verify the method in future studies. The method can also be applied to other mechanical equipment fault diagnosis.

## REFERENCES

- [1] R. A. Patil and S. L. Gombi, “Experimental study of cutting force on a cutting tool during machining using inverse problem analysis,” *J. Brazilian Soc. Mech. Sci. Eng.*, vol. 40, no. 10, pp. 1–8, Oct. 2018.
- [2] W. H. Dong, X. Q. Xiong, Y. Ma, and X. Y. Yue, “Woodworking tool wear condition monitoring during milling based on power signals and a particle swarm optimization-back propagation neural network,” *Appl. Sci.*, vol. 11, p. 9026, Sep. 2021.
- [3] S. T. Kumaran, T. J. Ko, and R. Kurniawan, “Grey fuzzy optimization of ultrasonic-assisted EDM process parameters for deburring CFRP composites,” *Measurement*, vol. 123, pp. 203–212, Jul. 2018.
- [4] C. J. Lin, J. Y. Jhang, and S. H. Chen, “Tool wear prediction using a hybrid of tool chip image and evolutionary fuzzy neural network,” *Int. J. Adv. Manuf. Technol.*, vol. 118, pp. 1–16, Jan. 2021.

- [5] M. S. Alajmi and A. M. Almeshal, "Estimation and optimization of tool wear in conventional turning of 709M40 alloy steel using support vector machine (SVM) with Bayesian optimization," *Materials*, vol. 14, no. 14, p. 3773, 2021.
- [6] A. Fatemeh, A. Tahan, and M. Thomas, "Tool condition monitoring using spectral subtraction and convolutional neural networks in milling process," *Int. J. Adv. Manuf. Technol.*, vol. 98, nos. 9–12, pp. 3217–3227, 2018.
- [7] X. C. Cao, B. Q. Chen, B. Yao, and W. P. He, "Combining translation-invariant wavelet frames and convolutional neural network for intelligent tool wear state identification," *Comput. Ind.*, vol. 106, pp. 71–84, Apr. 2019.
- [8] X. F. Wu, Y. H. Liu, X. L. Zhou, and A. Mou, "Automatic identification of tool wear based on convolutional neural network in face milling process," *Sensors*, vol. 19, no. 18, p. 3817, 2019.
- [9] S. Su, G. Zhao, W. Xiao, Y. Yang, and X. Cao, "An image-based approach to predict instantaneous cutting forces using convolutional neural networks in end milling operation," *Int. J. Adv. Manuf. Technol.*, vol. 2021, pp. 1657–1669, May 2021.
- [10] H. Y. Chen and C. H. Lee, "Deep learning approach for vibration signals applications," *Sensors*, vol. 21, no. 11, p. 3929, 2021.
- [11] B. L. Yan, L. D. Zhu, and Y. C. Dun, "Tool wear monitoring of TC4 titanium alloy milling process based on multi-channel signal and time-dependent properties by using deep learning," *J. Manuf. Syst.*, vol. 61, pp. 495–508, Oct. 2021.
- [12] X. Y. Hao, Y. Zheng, L. Lu, and H. Pan, "Research on intelligent fault diagnosis of rolling bearing based on improved deep residual network," *Appl. Sci.*, vol. 11, no. 22, p. 10889, 2021.
- [13] Z. Ye and J. B. Yu, "AKRNet: A novel convolutional neural network with attentive kernel residual learning for feature learning of gearbox vibration signals," *Neurocomputing*, vol. 447, pp. 23–37, Aug. 2021.
- [14] S. Ma, F. Chu, and Q. Han, "Deep residual learning with demodulated time-frequency features for fault diagnosis of planetary gearbox under nonstationary running conditions," *Mech. Syst. Signal Process.*, vol. 127, pp. 190–201, Jul. 2019.
- [15] B. S. Peng, H. Xia, and X. Z. Lv, "An intelligent fault diagnosis method for rotating machinery based on data fusion and deep residual neural network," *Appl. Intell.*, vol. 52, pp. 1–15, Jun. 2021.
- [16] A. H. Reza, M. Navideh, and B. Mahmoud, "A novel damage identification method based on short time Fourier transform and a new efficient index," *Structures*, vol. 33, pp. 3605–3614, Oct. 2021.
- [17] Z. Liu, X. M. Liu, and S. L. Xie, "A novel fault diagnosis method for analog circuits based on multi-input deep residual networks with an improved empirical wavelet transform," *Appl. Sci.*, vol. 12, no. 3, p. 1675, Mar. 2022.
- [18] Z. C. Cui, Y. F. Gao, and J. Hu, "LOS/NLOS identification for indoor UWB positioning based on Morlet wavelet transform and convolutional neural networks," *IEEE Commun. Lett.*, vol. 25, no. 3, pp. 879–882, Mar. 2021.
- [19] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [20] N. Zhang, "A novel hybrid model integrating residual structure and bi-directional long short-term memory network for tool wear monitoring," *Int. J. Adv. Manuf. Technol.*, vol. 120, pp. 1–16, Apr. 2022.
- [21] B. Liu, "Understanding the loss landscape of one-hidden-layer ReLU networks," *Knowl.-Based Syst.*, vol. 220, May 2021, Art. no. 106923.
- [22] N. G. Aidan, Z. Ivan, S. Kevin, G. Yarin, and E. H. Geoffrey, "Learning sparse networks using targeted dropout," 2019, *arXiv:1905.13678*.
- [23] C. Y. Pan, J. Huang, J. X. Gong, and X. H. Yuan, "Few-shot transfer learning for text classification with lightweight word embedding based models," *IEEE Access*, vol. 7, pp. 53296–53304, 2019.
- [24] W. Xu, Y. Wan, T. Y. Zuo, and X. M. Sha, "Transfer learning based data feature transfer for fault diagnosis," *IEEE Access*, vol. 8, pp. 76120–76129, 2020.
- [25] C. Wang, D. Chen, J. Chen, X. Lai, and T. He, "Deep regression adaptation networks with model-based transfer learning for dynamic load identification in the frequency domain," *Eng. Appl. Artif. Intell.*, vol. 102, Jun. 2021, Art. no. 104244.
- [26] J. Li, Y. Horiguchi, and T. Sawaragi, "Counterfactual inference to predict causal knowledge graph for relational transfer learning by assimilating expert knowledge—Relational feature transfer learning algorithm," *Adv. Eng. Informat.*, vol. 51, Jan. 2022, Art. no. 101516.
- [27] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" 2014, *arXiv:1411.1792*.
- [28] P. S. S. Sreedhar and N. Nandhagopal, "Classification similarity network model for image fusion using ResNet50 and GoogLeNet," *Intell. Automat. Soft Comput.*, vol. 31, no. 3, pp. 1331–1344, 2022.
- [29] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. on Learn. Represent.* (Computational and Biological Learning Society), 2015, doi: [10.48550/arXiv.1409.1556](https://arxiv.org/abs/1409.1556).
- [30] W. Zhang, C. Li, G. Peng, Y. Chen, and Z. Zhang, "A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load," *Mech. Syst. Signal Process.*, vol. 100, pp. 439–453, Feb. 2018.
- [31] X. D. Zhang, C. Han, M. Luo, and D. H. Zhang, "Tool wear monitoring for complex part milling based on deep learning," *Appl. Sci.*, vol. 10, no. 19, p. 6916, 2020.



**NAN ZHANG** (Member, IEEE) received the Ph.D. degree in mechatronic engineering from Northwestern Polytechnical University, in 2021. She is currently an Associate Professor with the School of Mechanical Engineering, Inner Mongolia University of Technology. Her research interests include machining mechanism, condition monitoring, residual life prediction, and system reliability.



**JIawei ZHAO** (Member, IEEE) is currently pursuing the master's degree with the School of Mechanical Engineering, Inner Mongolia University of Technology. His research interests include image processing, intelligent diagnosis, condition monitoring, remaining life prediction, and deep learning.



**LIN MA** (Member, IEEE) is currently pursuing the master's degree with the School of Mechanical Engineering, Inner Mongolia University of Technology. His mainly engaged in the research of remaining service life of milling cutter, tool wear, and machine learning.



**HAOQIANG KONG** (Member, IEEE) is currently pursuing the master's degree with the School of Mechanical Engineering, Inner Mongolia University of Technology. His mainly engaged in the research of milling cutter residual service life, tool wear, and tool reliability.



**HUAQIANG LI** (Member, IEEE) received the master's degree in mechanical engineering from Northeastern University, in 2005. He is a senior experimenter at Inner Mongolia University of Technology. His research interest includes mechanical structure design, analysis, and control.

• • •