2021 Global Reliability and Prognostics and Health Management (PHM-Nanjing) | 978-1-6654-0131-9/21/$31.00 ©2021 IEEE | DOI: 10.1109/PHM-Nanjing52125.2021.9612760

Online Milling Chatter Detection Using Deep Residual Network and Transfer Learning

Junyu Cong
School of Mechanical Engineering
Tianjin University
Tianjin, China
congjunyu@tju.edu.cn

Guofeng Wang
School of Mechanical Engineering
Tianjin University
Tianjin, China
gfwangmail@tju.edu.cn

Fei Wang
School of Mechanical Engineering
Tianjin University
Tianjin, China
fwang@tju.edu.cn

Jianming Che
School of Mechanical Engineering
Tianjin University
Tianjin, China
tcjm@tju.edu.cn

Xingchen Yu
School of Mechanical Engineering
Tianjin University
Tianjin, China
13752177679@163.com

He Geng
School of Mechanical Engineering
Tianjin University
Tianjin, China
genghe@tju.edu.cn

Wenhua Han
School of Mechanical Engineering
Tianjin University
Tianjin, China
HWH20201520@163.com

***Abstract*—In this paper, a novel approach of the online chatter detection in the milling process is presented based on the deep residual network (Resnet) and transfer learning. Considering the non-stationary characteristic of the chatter signal, the short-time Fourier transform (STFT) is used to transform the vibration acceleration signal into a two-dimensional spectrogram as input to identify the stable, transition and chatter states. A model for variable depth-of-cut conditions at 6000 rpm was trained by fusing the process information with the features down-dimensioned by the deep residual network, and the classification accuracy reached 99.2%. Since the spectrum structure of the signal changes at high speed and the data no longer satisfy the assumption of identically distribution, a model applicable to variable depth-of-cut conditions at 15000 rpm is obtained with 95.6% classification accuracy by using small samples and transfer learning method. The experimental results demonstrate that the method is suitable for early chatter monitoring under variable path and variable parameter conditions in cavity milling.**

Keywords- online chatter detection; deep residual network; transfer learning; variable path and variable parameter conditions

I. INTRODUCTION

In recent years, highly intelligent cutting process monitoring technology has become the key to improve productivity and product quality. Undoubtedly, the detection of machine chatter is one of the most important topics in the milling process. Milling chatter occurs more frequently when higher material removal rates are used. Milling chatter will lead to poor surface finish of the workpiece, tool breakage and noise. Therefore, online monitoring technology for early chatter can effectively avoid the hazards caused by chatter and save cost while improving efficiency. Online chatter detection can be divided into three main parts: acquisition of signal, feature extraction and identification model.

Benefiting from the development of sensing technology, a large number of signals reflecting the cutting status of the machine can be collected and transmitted to the host computer in real time for further processing. In this paper, 29 papers on cutting chatter detection from the SCI database for the year 2020 were retrieved. 16 of them used vibration acceleration signals, 10 used cutting force signals, 6 used sound signals, 2 used current signals, 2 used images, and 1 used eddy current displacement signals. Four of them used more than one type of signal. It can be seen that accelerometers have been widely used because of its advantages of low price, convenient installation and disassembly, and not easy to be disturbed by environmental factors.

At present, scholars mainly extract the features reflecting cutting chatter from three aspects: time domain, frequency domain and time-frequency domain. In the time domain, the amplitude, energy and waveform of the signal have received

This work is supported by the National Key Research and Development Program of China (2019YFB1704802).

more attention because chatter often leads to violent vibration between the tool and the workpiece. Shi et al. proposed a chatter identification method based on feature reduction and reinforced k-nearest neighbors method, using the proportion of data points falling into different amplitude intervals as detection features [1]. Liu et al. obtained the residual signal by subtracting the cutting force signal from the simulated dynamic cutting force, and then applied the standard deviation and root mean square metrics to the residual cutting force [2]. Considering the energy shift phenomenon brought by chatter, many scholars use energy [3-4] and energy entropy [5-7] as detection indexes. In the frequency domain, many features around the chatter frequency were proposed. Li et al. used an adaptive filter to filter the harmonics of the spindle rotation frequency and the identified colored noise components. Then, the difference in power spectral entropy (ΔPSE) between the unfiltered and filtered signals was used as an indicator [8]. Similarly, frequency domain features such as wavelet packet entropy [9], multiscale power spectrum entropy [10], Rényi entropy [11], and vibration fundamental frequency [12] were used to detect chatter. In the time-frequency domain, considering that the chatter signal is a non-stationary time series, scholars have proposed a series of features around the time-frequency diagram, such as the logarithmic spectrum distance [13] and the spectrum density of each frequency band [14].

The extracted features will be used as the input of the identification model to obtain the desired detection results. The current mainstream identification models are divided into threshold models [2,4,5,6,8,9,11-13], machine learning models (artificial features) [1,7,10,14] and deep learning models (automatically extracted features) [15-20]. The threshold model delineates a warning line between stable and chatter states, which is simple and practical, but the threshold value is often set artificially and has been criticized. Traditional machine learning models can automatically obtain the current cutting state according to the classifier, but the accuracy still relies heavily on the artificial features. In recent years, deep learning models have started to be applied in the field of mechanical fault diagnosis and monitoring. Although it increases the time cost of model training, its features are automatically extracted by deep neural networks and fed into the end classifier, which gets rid of the subjectivity of artificial features and can obtain a high-precision identification model.

In this paper, a chatter detection method based on deep learning and transfer learning is proposed for the cavity milling process of mirror-milling machine. The monitoring model is obtained by training multiple sets of variable depth-of-cut data at low speed, and the cutting state is divided into stable, transition and chatter states. Then, based on the transfer learning theory, a monitoring model applicable to variable depth-of-cut milling at high speed is obtained using a small sample data set.

II. RESEARCH METHODS

A. Short-time Fourier transform

The Fourier transform only reflects the characteristics of the signal in the frequency domain and does not allow the signal to be analyzed in the time domain. The short-time Fourier transform (STFT) adds a time variable to the traditional Fourier spectrum, thus enabling it to study the time-varying characteristics of a signal, which is essentially a Fourier transform with a window [21]. Assuming that the non-stationary signal is stationary at short time intervals of the time window, a set of local spectrum of the signal is obtained by shifting the window on the time axis and analyzing the signal segment by segment.

The short-time Fourier transform is defined as in (1). For a signal $x(t)$, assume that $h(\tau - t)$ is a window function centered at t (τ as a time variable), $x(\tau)h(\tau - t)$ is equivalent to taking a slice of the signal near the analysis time point t. Sliding the window function and applying the Fourier transform to each slice gives the short-time Fourier transform result. For a given time t, $STFT(t, f)$ can be considered as the Fourier spectrum at that moment.

$$STFT(t, f) = \int_{-\infty}^{\infty} x(\tau)h(\tau - t)e^{-j2\pi f\tau} d\tau \tag{1}$$

B. Deep residual network

In 2015, He et al. [22] proposed the deep residual network (Resnet), which is a form of neural network using shortcut connection. The deep residual network itself does not have a fixed structure and parameters, which makes the deep residual network very flexible and can be effectively inserted into other models to improve the model performance.

The residual block is the most important component of the deep residual network, as shown in Fig. 1. The shortcut connection in the residual block is directly added to the output of the convolutional layer after crossing the convolutional layers.

By repeatedly stacking such residual blocks, a deep residual network is obtained. The residual block can be expressed in the form of (2):

$$y_l = h(x_l) + F(x_l, W_l), x_{l+1} = f(y_l) \tag{2}$$

where x_l and x_{l+1} denote the input and output of the lth residual block, respectively, and W_l denotes the parameter of the lth residual block; $h(x_l)$ is a constant mapping, $f(y_l)$ represents the RELU activation function; F is the residual function, representing the learned residuals. If both $h(x)$ and $f(y)$ are constant mappings, then (2) can be expressed as (3):

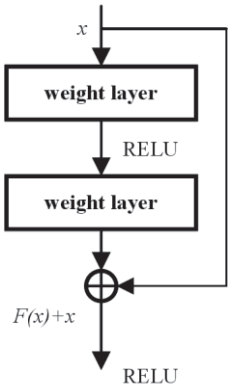

Figure 1. Residual block

$$x_{l+1} = x_l + F(x_l, W_l) \tag{3}$$

By recursion, the learning feature of any deep block L can be obtained as (4):

$$x_L = x_l + \sum_{i=l}^{L-1} F(x_i, W_i) \tag{4}$$

Equation (4) shows a very good back propagation property, assuming that the loss function is ε , (5) can be obtained according to the chain rule of back propagation：

$$\frac{\partial \varepsilon}{\partial x_l} = \frac{\partial \varepsilon}{\partial x_L}\frac{\partial x_L}{\partial x_l} = \frac{\partial \varepsilon}{\partial x_L}\left(1 + \frac{\partial}{\partial x_l}\sum_{i=l}^{L-1} F(x_i, W_i)\right) \tag{5}$$

The above equation shows that the gradient consists of two parts, $\frac{\partial \varepsilon}{\partial x_L}$ is the flow of information without any weighting, and the other part is $\frac{\partial \varepsilon}{\partial x_L}\left(\frac{\partial}{\partial x_l}\sum_{i=l}^{L-1} F(x_i, W_i)\right)$ through the weighted layer, and the linear nature of the connection between the two parts ensures that the information can be directly back-propagated to the shallow layer. The presence of "1" also ensures that the gradient does not vanish. By means of short splicing paths in the residual network, the gradient can be propagated coherently in a very deep network without being superimposed by too many convolutional gradients. This essentially avoids the gradient vanishing problem.

C. Transfer learning

Traditional machine learning assumes that the training data and the validation data obey the same distribution. However, in many cases, this identical distribution assumption is not satisfied. Training new models for data with different distributions often requires us to re-label a large amount of training data to meet our needs, but labeling new data is very expensive and requires a lot of human and material resources. Transfer learning can transfer knowledge from existing data and use it to help future learning. The goal of transfer learning is to take the knowledge learned from one environment and use it to help the learning task in the new environment. Therefore, transfer learning does not make the same distribution assumption as traditional machine learning.

Unlike sample transfer and feature transfer, model transfer can be performed by means of fine tune, specifically, freezing some of the convolutional layers of the pre-trained model (usually most of the convolutional layers near the input) and training the remaining convolutional layers (usually some of the convolutional layers near the output) and the fully connected layers. By this approach, we can use a small data sample and the pre-trained model to obtain a new model with appreciable accuracy for a new data distribution.

D. Our method

A procedure of our method is shown in Fig. 2. The main work of this paper is as follows. (1) The vibration acceleration signal is collected by triaxial accelerometer and transmitted to the host computer. (2) Segmentation of the signal under the full path and short-time Fourier transform are performed to obtain the corresponding spectrogram. (3) The images are resized and normalized to reduce the training time and to speed up the gradient descent solution, i.e., to improve the convergence speed of the model. The pre-processed data are divided into training and validation sets. (4) A deep residual network is trained for the dataset at low speeds, where the process information of each sample is fused with the reduced-dimensional feature tensor after the last residual block to increase the robustness of the model under variable conditions. (5) A new model is trained using fine tune and a small dataset at high speed, which is transferred from the model in (4) and optimized for hyperparameters using the Tree-structured Parzen Estimator (TPE). (6) The performance of the model is evaluated using the Accuracy and t-SNE Visualization algorithm. (7) Application of the model to an online monitoring task.

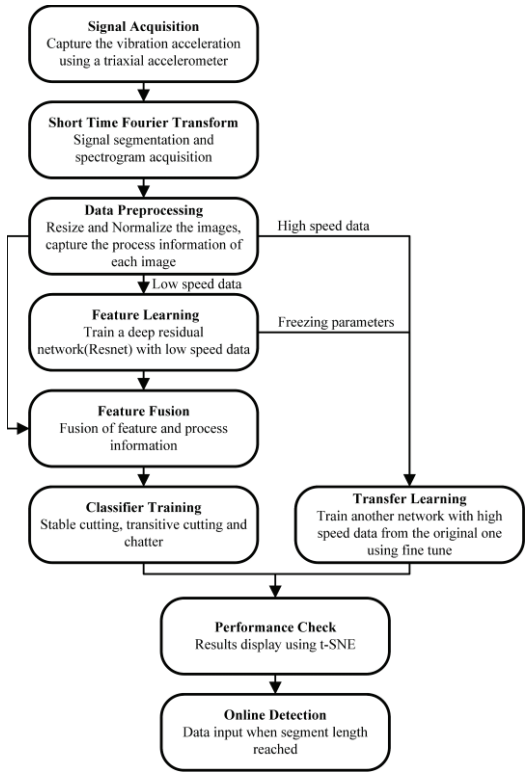

Figure 2. Procedure of method

III. EXPERIMENTS AND RESULTS

A. Experimental setup and measurement

In this paper, a series of experiments were conducted on a mirror-milling machine. The sensors are installed as shown in Fig. 3. A total of two triaxial accelerometers are installed, one of which (PCB356A26) is mounted on the spindle housing and the other (PCB356A03) is mounted at the unmachined position on the workpiece. The vibration acceleration signal is collected using NI9234 acquisition card with the sampling frequency for 10240Hz, and displayed and saved by the software written in Labview. Cutting experiments were conducted by milling an aluminum alloy thin-walled workpiece with a two-teeth end mill cutter. Unlike the rectilinear milling, cavity milling in the experiment includes feeding in both x and y directions, accounting for variations in the milling path.

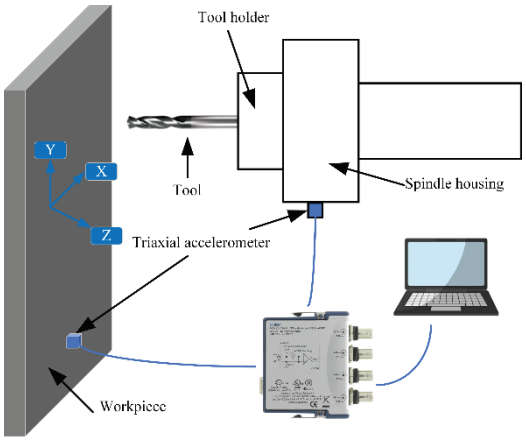

Figure 3. Experimental setup

The cutting parameters for each group of experiments are shown in TABLE I. Six groups of experiments at 6000 rpm were used to train a deep residual network, and three groups of experiments at 15000 rpm were used to train a new model transferred from the low-speed model.

B. Short time Fourier transform and Data preprocessing

Hopf bifurcation, period doubling bifurcation and saddle-node bifurcation are typical destabilization phenomena during milling. When chatter occurs, a new fundamental frequency of motion is generated, which is referred to here as the vibration fundamental frequency [12]. The appearance of the vibration fundamental frequency means the system is destabilized, and this fundamental frequency will be superimposed on the tooth passing frequency to become the chatter frequency. Due to the multi-order nature of the tooth passing frequency, the chatter frequency is also a multi-order frequency.

As the vibration signal on the spindle contains many electromagnetic interference frequencies, and the thin-walled workpiece has the largest vibration in the vertical direction, the z-direction signal on the workpiece is chosen in this paper. The length of the time window used to cut the signal is set as 10 spindle rotation periods, and the overlap of two consecutive segments is set as 25%. In order to obtain the trend of frequency change with time, the short-time Fourier transform is used to obtain the spectrogram of each signal segment. The frequency components in the spectrogram and the surface roughness of the workpiece are used as the judgment criteria to assign labels to each segment of the signal. Fig. 4 shows the signal time domain amplitude, spectrogram and workpiece surface for the stable, transition and chatter stages, respectively.

As can be seen from Fig. 4, when the machining is in the stable stage, the signal time domain amplitude is small, and the spectrogram shows regular spectral lines at the spindle rotation frequency (100Hz), tooth passing frequency (200Hz) and its multiples, and the surface of the workpiece is smooth without chatter mark. When the machining is in the transition stage, the signal time domain amplitude does not change much compared with the stable stage, and new frequency components appear in the spectrogram near the above-mentioned spectral line, at this time the chatter frequencies are in the pregnant stage, and there is still no obvious chatter mark on the surface of the workpiece. When the machining is in the chatter stage, the signal time domain amplitude increases significantly, the chatter frequencies dominate, and the energy is transferred from the cutting frequencies to the main chatter frequency.

TABLE I. THE CUTTING PARAMETERS

No.	Spindle speed (rpm)	Feed rate (mm/min)	Depth of cut(mm)	Width of cut(mm)
1	6000	100	0.1	6
2	6000	100	0.2	6
3	6000	200	0.4	6
4	6000	200	0.8	6
5	6000	200	1.2	6
6	6000	300	0.2	6
7	15000	100	0.1	6
8	15000	100	0.2	6
9	15000	200	0.8	6

According to the comprehensive judgment criteria given in this paper, the dataset is produced and the training and validation sets are randomly generated according to the ratio of 80%-20%. For the low-speed model, the training set has a total of 1824 samples and the validation set has a total of 480 samples, covering group 1-6, with an identical number of samples in each category. For the high-speed model, the total number of samples is about 1/3 of that used in the low-speed model, with a total of 624 samples in the training set and 160 samples in the validation set, covering group 7-9, with an identical number of samples in each category. The pixel size of each sample is resized to 128*128 to reduce the training time, and the samples are normalized.

C. Deep residual network and transfer learning

In this paper, we construct a training model based on an 18-layer deep residual network (Resnet18) [22], and the network structure is shown in Fig. 5, where the dimensions and sizes of the output results of each layer are labeled in the figure according to the format of dimension (number of channels) * feature map width * feature map height. The main part of the model consists of four residual blocks with different number of channels. From the lobe diagram indicating the stability of milling machining, it can be learned that spindle speed and depth of cut are the two main factors affecting cutting chatter. For the purpose of increasing the applicability and robustness of the model, the output feature tensor of the last residual block is combined with the feature tensor which is composed of the spindle speed and depth of cut of the sample, and then the 512*4*8 features are fed into the average pooling layer and the fully connected layer. The model for 6000 rpm uses the Xavier weight initialization method, selects the Adam optimizer with a fixed-step learning rate decay and adopts a cross-entropy loss function, named Model 1. The cross-entropy can be described as:

$$H(p,q) = -\sum_{i=1}^{n} p(x_i)\log(q(x_i)) \tag{6}$$

where $p(x_i)$ represents the real probability distribution of the samples while $q(x_i)$ represents the predicted probability

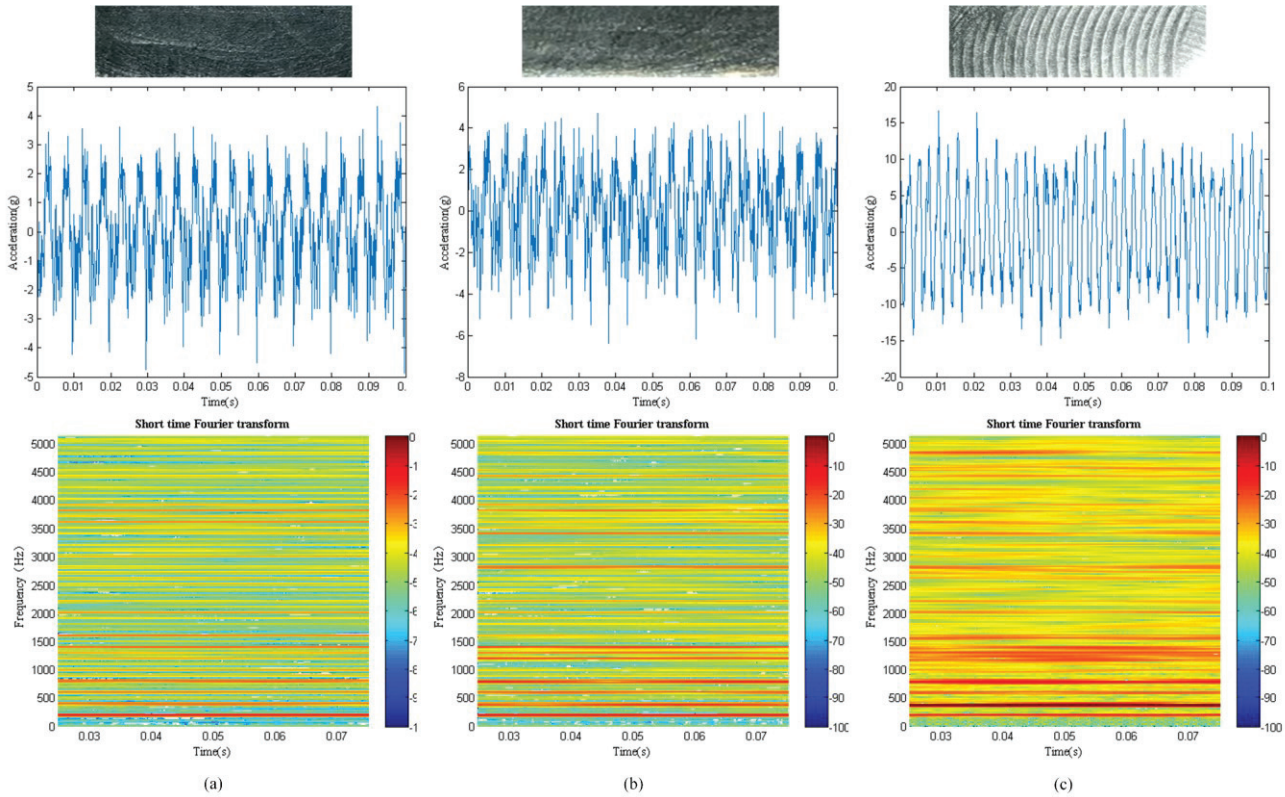

Figure 4. Time domain amplitude, spectrogram and workpiece surface (a) stable stage (b) transition stage (c) chatter stage

distribution. The purpose of neural network is to make the predicted distribution approach the real distribution by training.

Since the input features reflect the time-frequency characteristics of the signal, and the spectrum structure of milling processing is greatly influenced by the spindle speed, when the monitoring domain comes from the low-speed region to the high-speed region, the input features will no longer satisfy the assumption of identical distribution, therefore, fine tune is used to perform transfer learning on the trained low-speed model. The parameters of the first three residual blocks in the model were frozen, and the parameters of the last residual block, the average pooling and the fully connected layer were opened for training. A transfer model applicable in 15000 rpm was obtained using small samples. Two hyperparameters, the learning rate and the learning rate decay coefficient, were optimized using the TPE algorithm, named Model 2.

D. Classification results and discussions

The loss and classification accuracy curves of Model 1 are shown in Fig. 6 and Fig. 7, respectively. The loss function of the training set converges after about 60 epochs and the loss function of the validation set converges after about 80 epochs as seen in Fig. 6, and the accuracy of the validation set reaches 99.2%. The confusion matrix is shown in TABLE II.

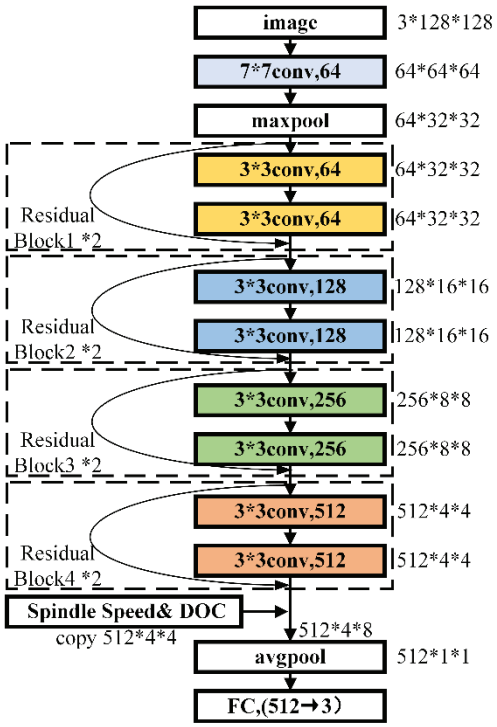

Figure 5. Network structure

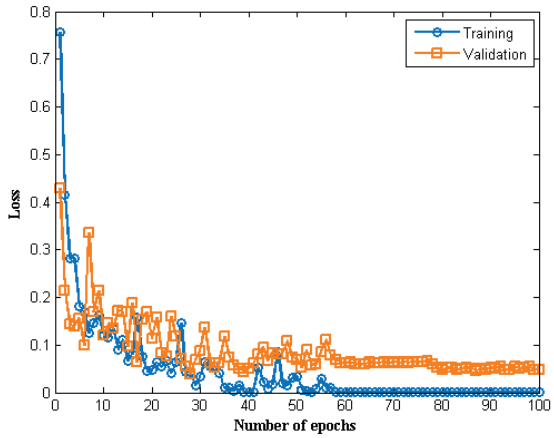

Figure 6. Loss curve of Model 1

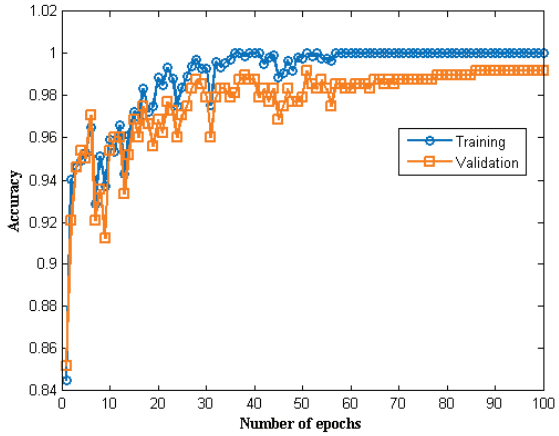

Figure 7. Accuracy curve of Model 1

TABLE II. CONFUSION MATRIX OF MODEL 1

Predicted labels	Target labels		
	Stable	Transition	Chatter
Stable	159	1	0
Transition	1	157	0
Chatter	0	2	160
Accuracy	99.38%	98.13%	100%

For a real chatter detection task, the advantages of our model are illustrated by Fig. 8. The signal in Fig. 8 is spliced with two signals. These two signals are acquired when milling two different edges (not in the same circle) in the cavity milling. The cyan line is the artificially extracted Root Mean Square (RMS) curve. The RMS value reflects the energy of the signal and is a time-domain indicator commonly used to detect chatter, since the signals on both sides of the black dashed line belong to different circles, it leads to a significant difference in the RMS value of respective stable cutting stage. According to the 3σ criterion, if the RMS value of the left signal (stable stage) is used to calculate the threshold, the threshold is 2.0897 (red solid line), and chatter can be detected at the green solid line. The model in this paper can detect chatter at the solid orange line according to the classification result, which is 0.5s earlier than the artificial feature which is very valuable for the early detection of chatter. Also, we note that by the time the machining proceeds to the right side, all stable samples will be misidentified as chatter samples, which is caused by the feature's inability to adapt to the milling path variation. However, if the RMS value of the right signal (stable stage) is used to calculate the threshold, the threshold is 2.7532 (black solid line), and although there will be no further misclassification, chatter will be detected much later due to the high threshold. As for the frequency domain features, unlike rectilinear milling, the main chatter frequency changes under such a complex milling path, leading to the failure of the artificial features extracted around the frequency band. In summary, our model has better robustness under variable conditions.

Model 2 is trained using fine tune, and the model accuracy is more sensitive to the values of hyperparameters. The two important hyperparameters, learning rate and learning rate decay coefficient, are optimized using the Tree-structured Parzen Estimator Approach (TPE) algorithm. With loss as the objective function, the optimal combination of hyperparameters is obtained after 30 iterations and used for the training of Model 2, and the optimization process is shown in Fig. 9.

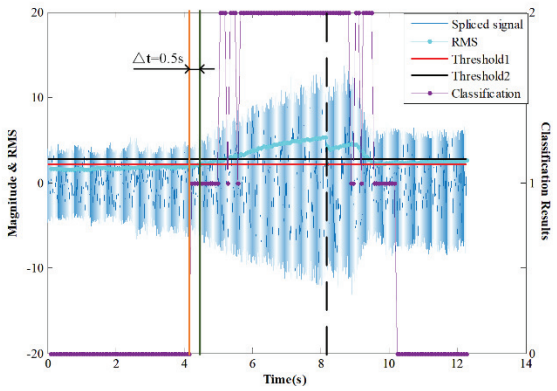

Figure 8. Real chatter detection task

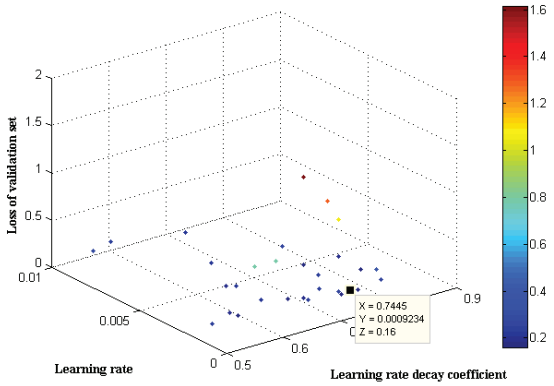

Figure 9. Optimization process with TPE algorithm

The loss and classification accuracy curves of Model 2 are shown in Fig. 10 and Fig. 11, respectively. The loss function of the training and validation sets converge after about 20 epochs as seen in Fig. 10, and the accuracy of the validation set reaches 95.6%.

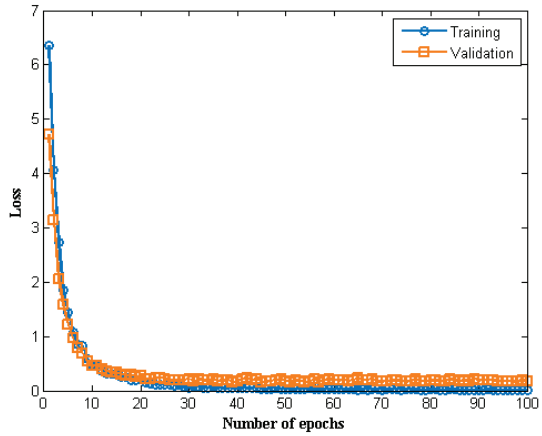

Figure 10. Loss curve of Model 2

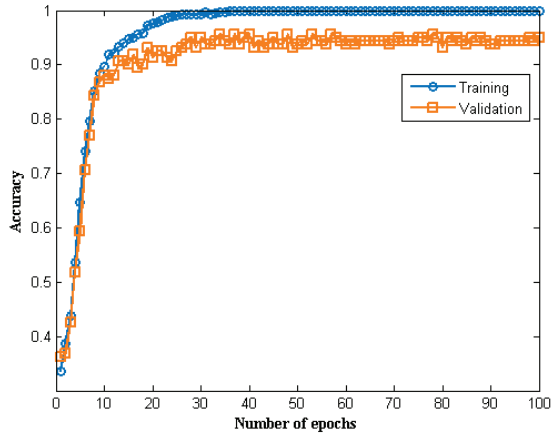

Figure 11. Accuracy curve of Model 2

The confusion matrix of Model 2 is shown in Table 3. We can clearly see that the accuracy of model 2 is greatly improved with fusing the process information.

Feature dimension reduction can be performed on the input images of model 2 using the t-SNE visualization algorithm, and then its two-dimensional distribution is plotted as shown in Fig. 12, where "0,1,2" represents the true labels of the samples. Similarly, feature dimension reduction can be performed on the 1*3 dimensional features after the fully connected layer, and then its two-dimensional distribution is plotted as shown in Fig. 13. The comparison shows that the distribution state of the input image changes from chaotic to clear after the classification of the proposed model, which proves the effectiveness of the method in this paper.

TABLE III. CONFUSION MATRIX OF MODEL 2

With fusing the process information			
Predicted labels	**Target labels**		
	Stable	**Transition**	**Chatter**
Stable	52	0	0
Transition	1	50	3
Chatter	0	3	51
Accuracy	98.11%	94.34%	94.44%
Without fusing the process information			
Predicted labels	**Target labels**		
	Stable	**Transition**	**Chatter**
Stable	31	0	0
Transition	11	42	1
Chatter	11	11	53
Accuracy	58.49%	79.25%	98.15%

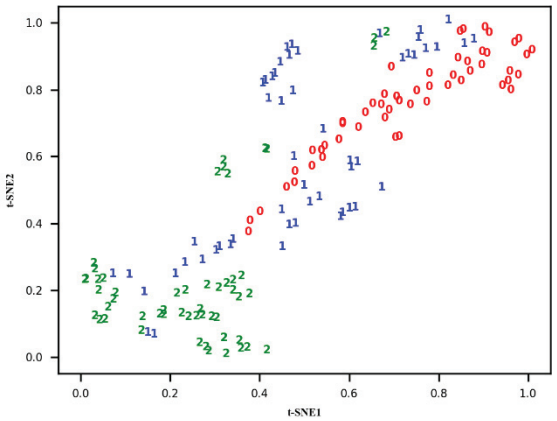

Figure 12. Two-dimensional distribution of input samples

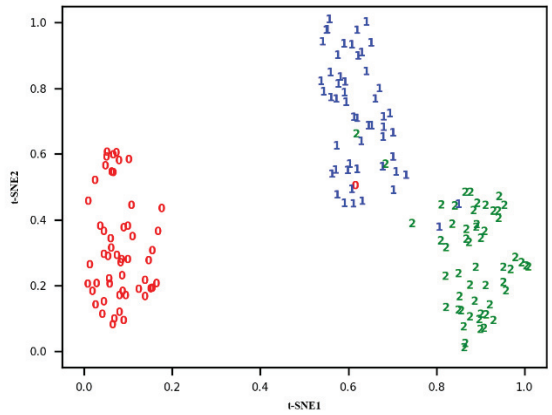

Figure 13. Two-dimensional distribution of output samples

IV. CONCLUSIONS

In this paper, an online milling chatter detection method integrating deep residual network and process information is proposed. A model applicable to variable depth-of-cut conditions at low speed was first developed. Then due to the change in data distribution at high speed, the model was extended to the high-speed milling region using a transfer learning approach. Both models perform well and some conclusions can be drawn that:

(a) The features automatically extracted by the method proposed in this paper can accurately reveal the relationship between the vibration signal and the machining state, and perform well not only in rectilinear milling, but also under the complex path of cavity milling.

(b) Using the transfer learning method, a new model for high speed with considerable precision can be obtained by using small sample data and the pretrained weight, saving the extra overhead associated with extensive experimentation.

ACKNOWLEDGMENT

This work is supported by the National Key Research and Development Program of China (2019YFB1704802).

REFERENCES

[1] F. Shi, H. Cao, X. Zhang, and X. Chen, "A reinforced k-Nearest Neighbors method with application to chatter identification in high-speed milling," IEEE Transactions on Industrial Electronics, vol. 67, no. 12, pp. 10844-10855, Dec 2020.

[2] M. K. Liu, M. Q. Tran, C. H. Chung, and Y. W. Qui, "Hybrid model- and signal-based chatter detection in the milling process," Journal of Mechanical Science and Technology, vol. 34, no. 1, pp. 1-10, Jan 2020.

[3] E. Wang, P. Yan, and J. Liu, "A hybrid chatter detection method based on WPD, SSA, and SVM-PSO," Shock and Vibration, vol. 2020, Jun 2020.

[4] R. Wang, J. Niu, and Y. Sun, "Chatter identification in thin-wall milling using an adaptive variational mode decomposition method combined with the decision tree model," Proceedings of the Institution of Mechanical Engineers Part B-Journal of Engineering Manufacture, Jul 2020.

[5] Q. Zhang, X. Tu, F. Li, and Y. Hu, "An effective chatter detection method in milling process using Morphological Empirical Wavelet Transform," IEEE Transactions on Instrumentation and Measurement, vol. 69, no. 8, pp. 5546-5555, Aug 2020.

[6] J. Tao et al., "Timely chatter identification for robotic drilling using a local maximum synchrosqueezing-based method," Journal of Intelligent Manufacturing, vol. 31, no. 5, pp. 1243-1255, Jun 2020.

[7] L. Zhu, C. Liu, C. Ju, and M. Guo, "Vibration recognition for peripheral milling thin-walled workpieces using sample entropy and energy entropy," International Journal of Advanced Manufacturing Technology, vol. 108, no. 9-10, pp. 3251-3266, Jun 2020.

[8] X. Li, S. Wan, X. Huang, and J. Hong, "Milling chatter detection based on VMD and difference of power spectral entropy," International Journal of Advanced Manufacturing Technology, vol. 111, no. 7-8, pp. 2051-2063, Dec 2020.

[9] L. Ding, Y. Sun, and Z. Xiong, "Adaptive removal of time-varying harmonics for chatter detection in thin-walled turning," International Journal of Advanced Manufacturing Technology, vol. 106, no. 1-2, pp. 519-531, Jan 2020.

[10] K. Li, S. He, B. Li, H. Liu, X. Mao, and C. Shi, "A novel online chatter detection method in milling process based on multiscale entropy and gradient tree boosting," Mechanical Systems and Signal Processing, vol. 135, Jan 2020.

[11] Z. Chen, Z. Li, J. Niu, and L. Zhu, "Chatter detection in milling processes using frequency-domain Renyi entropy," International Journal of Advanced Manufacturing Technology, vol. 106, no. 3-4, pp. 877-890, Jan 2020.

[12] L. Chang, W. Xu, and G. Lei, "Identification of milling chatter based on a novel frequency-domain search algorithm," International Journal of Advanced Manufacturing Technology, vol. 109, no. 9-12, pp. 2393-2407, Aug 2020.

[13] Y. Liu, X. Wang, J. Lin, and X. Kong, "An adaptive grinding chatter detection method considering the chatter frequency shift characteristic," Mechanical Systems and Signal Processing, vol. 142, Aug 2020.

[14] I. Kvinevskiy, S. Bedi, and S. Mann, "Detecting machine chatter using audio data and machine learning," International Journal of Advanced Manufacturing Technology, vol. 108, no. 11-12, pp. 3707-3716, Jun 2020.

[15] W. Zhu, J. Zhuang, B. Guo, W. Teng, and F. Wu, "An optimized convolutional neural network for chatter detection in the milling of thin-walled parts," International Journal of Advanced Manufacturing Technology, vol. 106, no. 9-10, pp. 3881-3895, Feb 2020.

[16] M.-Q. Tran, M.-K. Liu, and Q.-V. Tran, "Milling chatter detection using scalogram and deep convolutional neural network," International Journal of Advanced Manufacturing Technology, vol. 107, no. 3-4, pp. 1505-1516, Mar 2020.

[17] F. Shi, H. Cao, Y. Wang, B. Feng, and Y. Ding, "Chatter detection in high-speed milling processes based on ON-LSTM and PBT," International Journal of Advanced Manufacturing Technology, vol. 111, no. 11-12, pp. 3361-3378, Dec 2020.

[18] Y. Fu et al., "Automatic feature constructing from vibration signals for machining state monitoring," Journal of Intelligent Manufacturing, vol. 30, no. 3, pp. 995-1008, Mar 2019.

[19] H. N. Gao, D. H. Shen, L. Yu, and W. C. Zhang, "Identification of cutting chatter through deep learning and classification," International Journal of Simulation Modelling, vol. 19, no. 4, pp. 667-677, Dec 2020.

[20] R. K. Vashisht and Q. Peng, "Online chatter detection for milling operations using LSTM Neural Networks assisted by motor current signals of ball screw drives," Journal of Manufacturing Science and Engineering-Transactions of the Asme, vol. 143, no. 1, Jan 2021.

[21] F. Hlawatsch and G. F. Boudreaux-Bartels, "Linear and quadratic time-frequency signal representations," IEEE Signal Processing Magazine, vol. 9, no. 2, pp. 21-67, Apr 1992.

[22] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770-778, Jun 2016.