Gearbox Fault Diagnosis Method Based on Improved MobileNetV3 and Transfer Learning

Yanping DU, Xuemin CHENG, Yuxin LIU, Shuihai DOU, Juncheng TU*, Yanlin LIU, Xianyang SU

Abstract: Under different working conditions of gearbox, the feature extraction of fault signals is difficult, and large difference in data distribution affects the fault diagnosis results. Based on the problems, the research proposes a method based on improved MobileNetV3 network and transfer learning (TL-Pro-MobilenetV3 network). Three time-frequency analysis methods are used to obtain time-frequency distribution. Among them, short time Fourier transform (STFT) combined with Pro-MobilenetV3 network takes the shortest time and has the highest accuracy. Furthermore, transfer learning is introduced into the model, and the optimal training parameters are selected training the network. Using the dataset from Southeast University, the TL-Pro-MobilenetV3 model is compared with four classical fault diagnosis models. The experimental results show the accuracy of the method proposed can reach 100% and the training time is the shortest in two working conditions, proving the proposed model has a good performance in generalization ability, recognition accuracy and training time.

Keywords: deep learning; gearbox fault diagnosis; STFT; TL-Pro-MobileNetV3 network; transfer learning

1 INTRODUCTION

The Gearbox is the core rotating component in mechanical equipment. Its running status is directly related to whether the entire mechanical system can work efficiently [1, 2]. In actual work, the operating conditions of gearbox are complex and changeable, and the measured signals are easy to show obvious non-linearity and non-stationary, which makes it difficult to extract fault features. At the same time, gearbox often operates under different working conditions, resulting in large data distribution differences. It makes gearbox fault diagnosis more difficult. Besides, the rotating parts in the gearbox typically suffer from fatigue damage in the early stage. These minor faults are often easily ignored due to the interference of some noise. If they are allowed to develop, they will cause severe losses. Therefore, it is of great significance to choose the appropriate method to diagnose the gearbox fault signal and deal with the gearbox fault problem in time.

Some scholars use the vibration signal to detect the fault of the gearbox and obtain good results. Ma et al. [3] used the symplectic quaternion singular mode decomposition (SQSMD) method to decompose multivariate vibration signals into a series of independent significant components. The results proved the effectiveness and superiority of SQSMD in gear fault diagnosis. Wang et al. [4] used a vibration signal analysis method based on integrated empirical mode decomposition and support vector machine to diagnose diesel engine faults. The method accurately identified multiple fault information from the collected vibration signals under different states of the gearbox. Chen et al. [5] proposed a hyper parameter selection strategy based on vibration signals for long short-term memory (LSTM) identification and gearbox fault detection, and effectively utilized the vibration signal information of gearbox. Liu et al. [6] proposed a time-domain diagnosis method for gearbox faults based on measured vibration signals, which combined fast dynamic time and correlation kurtosis technology to characterize local fault signal of gear. The method achieved a good diagnosis result.

The method of gearbox fault diagnosis based on vibration signals has high accuracy and diversity. It usually uses the signal model to analyze the measured signal, extract the characteristic signal value, and evaluate whether the state of mechanical equipment is unusual. However, the method relies too much on expert experience to extract signal features. It does not dig deep into the coupling degree between high-dimensional signals, only analyzes the signal data of the object to be tested, and does not make enough use of its in-depth information. Because deep learning can adaptively extract and identify features, it has been widely used in many fields. As a typical deep learning model, Convolutional Neural Network (CNN) has been widely used in fault diagnosis field for its powerful data mining ability [7, 8].

Zhang et al. [9] used a hybrid attention improved residual network (HA-ResNet) based on deep learning for fault diagnosis of a wind turbine gearbox and the recognition accuracy of the method was significantly improved. To realize the diagnosis of a large number of typical fault samples, Han et al. [10] used the deep learning neural network combined with two-dimensional empirical mode decomposition for feature extraction and fault diagnosis. They proposed an intelligent fault diagnosis method for rotating machinery based on feature selection and deep learning. Li et al. [11] offered an enhanced selective integrated deep learning combined with beetle antenna search (BAS) algorithm to improve the performance of rolling bearing fault diagnosis. Still, this method has higher hardware requirements and a lengthy training process. Wang et al. [12] proposed a fault diagnosis method for rotating machinery under time-varying working conditions based on unbonded touchless order tracking (TOT) and deep learning. However, in the process of data collection, 14 steering gearboxes of 7 gantry cranes need to be tracked and tested, which is time-consuming and laborious.

In recent years, some scholars have introduced transfer learning into the fault diagnosis field [13-15]. Liu et al. [16] proposed a probabilistic wind power prediction method based on transfer learning and machine learning model, achieving more accurate quantile prediction results by introducing transfer learning. Li et al. [17] used transfer learning with limited labelled data for nuclear power plant fault diagnosis. They demonstrated the feasibility and superiority of transfer learning for atomic power plant fault diagnosis under limited labelled data. Wang et al. [18] used transfer learning based on improved GRU to estimate state-of-charge (SOC) for lithium-ion batteries with small target sample sets. The results show that the proposed method can

estimate the SOC of lithium-ion batteries quickly and accurately without requiring a large number of training samples and high computing costs. Therefore, transfer learning only needs a small amount of sample data to train a neural network and can achieve good classification effect. Shao et al. [19] proposed a high-accuracy machine fault diagnosis method using deep transfer learning. Transfer learning gives the target model a proper initialization and reduces the number of parameters needed to be updated. Wang et al. [20] invented a transfer learning model for detecting six classes of phonocardiogram recordings, which achieved high accuracy even in noisy backgrounds. Introducing transfer learning into the neural network can reduce the complexity of the network and improve classification accuracy.

We propose a gearbox fault diagnosis method based on MobileNetV3 network and transfer learning, aiming at the problems that the feature extraction is difficult, and large difference in data distribution under different working conditions affects the fault diagnosis results under different working conditions. Three different time-frequency analysis methods are introduced, the structure of the MobileNetV3 model is described in detail, a model training method based on transfer learning is designed, and the steps of fault diagnosis using the model are given. Finally, the accuracy of the proposed model is verified by using the gearbox fault dataset published by Southeast University.

The main contributions of this paper are as follows:

(1) We combine different time-frequency analysis methods and the proposed models, and the time-frequency analysis methods are used to obtain information on the time-frequency domain. With the powerful data processing capabilities of the multi-level deep learning model, the feature engineering automation is realized, and the information of the samples is fully utilized.

(2) We propose a model based on the MobileNetV3 network and adopt the transfer learning strategy. Compared with traditional methods, it achieves a higher accuracy rate and requires less training time. Experiments have verified that the TL-Pro-MobileNetV3 network performs better in feature extraction, training time, and generalization ability under two working conditions. It effectively solves the problem that under different working conditions, the gearbox feature extraction is difficult and the large difference of data distribution affects the fault diagnosis accuracy.

2 RELATED THEORY
2.1 Time-Frequency Analysis Method

In the raw signals of the gearbox, the collected fault signal is usually composed of 1D non-stationary signals. To make use of the advantages of the deep learning model in the 2D image classification, the signals must be converted into 2D image data. First, we can implement this process by equidistant interception and splicing of signals. The method is simple and fast but lacks frequency-domain information of signals. Secondly, the 1D vibration signals can also be converted by the time-frequency analysis method. The methods mainly include STFT, continuous wavelet transform (CWT) and Wigner-Ville distribution (WVD), Hilbert-Huang transform, Choi-Williams distribution, Gabor transform, constant Q transform etc. According to the structure of the model and the characteristics of above methods, STFT, CWT, and WVD

are selected as the signal conversion methods and we will choose the best time-frequency analysis method from them. The three time-frequency analysis methods are described in detail below.

The principle of STFT is to add a sliding window to the signal, and then perform Fourier transform to obtain the spectrum of the whole time-domain signal. As shown in Eq. (1):

$$S_{\text{STFT},x}(\omega,\tau) = \int_{-\infty}^{\infty} x(t)g(t-\tau)\mathrm{e}^{-j\omega t}\mathrm{d}t \tag{1}$$

where, $g(t-\tau)$ is the window function, using Hanning window.

WVD is a bilinear time-frequency distribution function connecting time and frequency, which can clearly show the distribution law of the signal in the time-frequency domain. The expression is shown in Eq. (2):

$$WVD_x(t,\omega) = \int_{-\infty}^{\infty} x(t+\frac{\tau}{2})x^*(t-\frac{\tau}{2})\mathrm{e}^{-j\omega\tau}\mathrm{d}\tau \tag{2}$$

where, $x(t+\frac{\tau}{2})x^*(t-\frac{\tau}{2})$ is instantaneous autocorrelation function of the signal.

Wavelet transform has a good performance of partial analysis ability in both time-domain and frequency-domain, providing excellent resolution results for a periodic signal [21]. We use a continuous wavelet transform for time-frequency analysis, and the expression is shown in Eq. (3):

$$WT(a,b) = \frac{1}{\sqrt{a}}\int_{-\infty}^{\infty} x(t)\psi(\frac{t-b}{a})\mathrm{d}t \tag{3}$$

where, a is a stretching factor, b is a translation factor, $\psi(t)$ represents the mother wavelet.
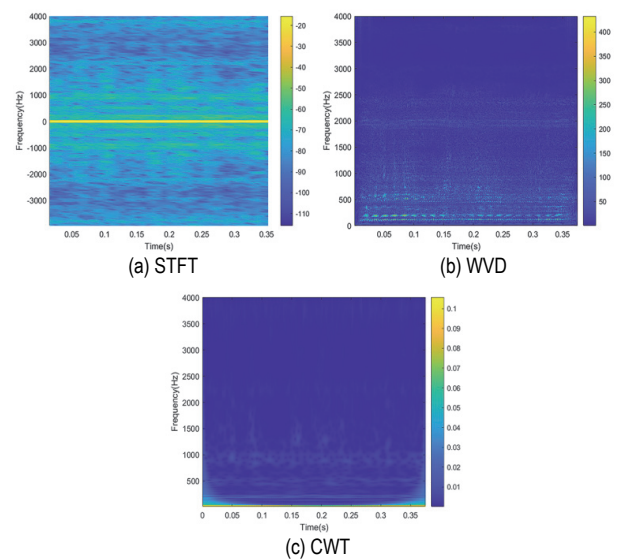

Figure 1 The time-frequency images of three methods

We use the three time-frequency analysis methods introduced above to process the raw signals. The raw

signals are mapped to 2D space through the ways above. In this way, 1D are converted to time-frequency images.

The time-frequency images of three methods are shown in Fig. 1.

2.2 MobileNetV3 Network

MobileNet is a lightweight network developed by the Google team for embedded devices. Because of its few parameters and low latency, it is widely applied to mobile terminal devices. MobileNetV3 is inherited and developed from MobileNetV1 and MobileNetV2 [22]. Based on the two models, Squeeze-and-Excitation (SE) module [23] is introduced to suppress obscure information and strengthen prominent features, which could further improve the classification accuracy of the model.

2.2.1 Depthwise Separable Convolution (DSC)

DSC is composed of depthwise convolution (DW) and pointwise convolution (PW), as shown in Fig. 2. Unlike standard convolution (SC) operation, DW carries out convolution calculation for each channel of the feature maps and applies a single convolution kernel, which significantly reduces the model calculation amount on the premise of preserving essential features. PW is used to carry out 1×1 convolution operation on the feature maps to achieve information fusion between channels.

Figure 2 Operation process of DW and PW
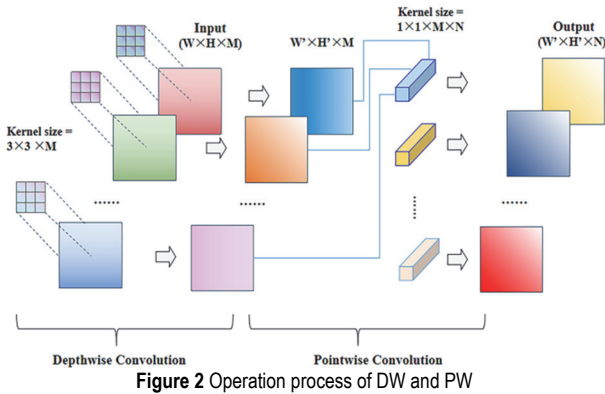
Suppose the input feature map size is $W \times H \times M$, convolution kernel size is $k \times k$, and the output feature map size is $W' \times H' \times N$. The comparison of parameter amount and calculation amount between DSC and SC is shown in Tab. 1.

Table 1 The comparison of parameter amount and the calculation amount.

Type	Parameters number	Calculation
SC	$k \times k \times M \times N$	$k \times k \times M \times N \times W' \times N'$
DSC	$k \times k \times M$	$k \times k \times M \times W' \times N'$
	$M \times N$	$M \times N \times W' \times N'$
TC/DSC	$1/N + 1(k \times k)$	$1/N + 1(k \times k)$

2.2.2 SE Module

SE mainly includes two operations: squeeze and excitation. The primary implementation process of the SE module is described in detail below. For the input feature matrix, each channel is globally pooled to obtain a vector (C is equal to the number of channels of the input feature matrix). After two "fully-connected layer-activation"

operations, a vector is obtained, and multiply the vector with input feature matrix by bit. This module aims to realize feature recalibration by dynamically adjusting the weight of each channel, and improving the network performance at a small computing cost. The expression of a squeeze is shown in Eq. (4):

$$z = F_{sq}(i,j) = \frac{1}{h \times w} \sum_{i=1}^{h} \sum_{j=1}^{w} f(i,j) \tag{4}$$

The expression of the excitation is shown in Eq. (5):

$$s = F_{ex}(z,W) = \sigma(g(z,W)) = \phi(W_2 \delta(W_1 z)) \tag{5}$$

where, σ and δ represent hard-Sigmoid and ReLU function, respectively.

2.2.3 Activation Function

MobileNetV3 uses a new activation function $h-switch$ to replace $switch$, which solves the problem of complex derivation and not conducive to quantization. The expression is shown in Eq. (6):

$$h-switch[x] = x \frac{\mathrm{Re}\,Lu6(x+3)}{6} \tag{6}$$

2.2.4 Bottleneck Structure

The bottleneck adopts an inverted residuals structure, as shown in Fig. 3. First, PW (see 2.2.1) is used for dimension reduction. Second, the features are further extracted through the DSC (see 2.2.1), and then the weights of each channel are calibrated through SE (see 2.2.2). The expanded feature is compressed into the original dimension by 1×1 convolution. Through the operations above, the computational load of the model is significantly reduced.

Figure 3 The detailed bottleneck structure
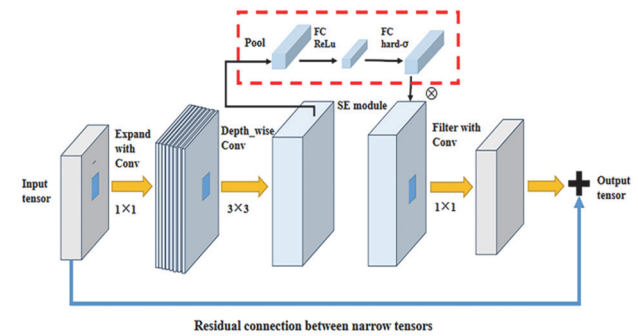
2.3 Transfer Learning

Transfer learning means given the labelled source domain $D_s = \{x_s, P(x_s)\}$, learning tasks $T_s = \{y_s, f_s(\cdot)\}$, target domain $D_t = \{x_t, Q(x_t)\}$ and its learning tasks $T_t = \{y_t, f_t(\cdot)\}$ $(D_s \neq D_t$ or $T_s \neq T_t)$, the knowledge learned from D_s and T_s is transferred to D_t to help improve the learning of $f_t(\cdot)$ in D_t. Compared with traditional machine learning algorithms, transfer learning does not

require too many sample data in the model training process. It has great advantages in the case of few samples in D_t and sufficient samples in D_s and its schematic diagram is shown in Fig. 4.
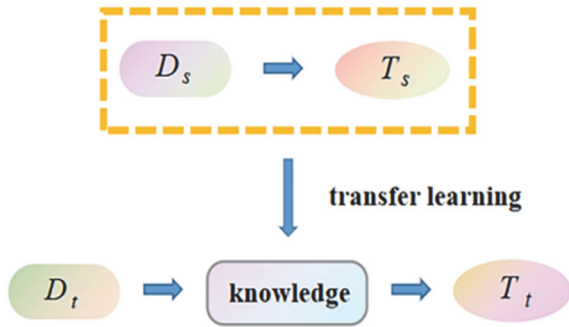

Figure 4 The schematic diagram of transfer learning

Transfer learning includes sample-based transfer learning, feature representation-based transfer learning, relational knowledge-based transfer learning, and parameter-based transfer learning. We adopt the last method. In the task, the similarity between two domains above is the key to success. If the target domain dataset distribution is similar to the source domain dataset, only the fully-connected layer plays the role of classification, and needs to be trained. If the difference between two datasets is large, more high-level convolutional layers or even the whole network need to be trained and updated.

3 TRANSFER LEARNING FOR IMPROVED MOBILENETV3 NETWORK

The model proposed, Pro-MobileNetV3, retains most of the structure of the MobileNetV3-Large network. To retain more features of the time-frequency images, the stride of convolution kernel is set to 1 in the last three bottleneck layers. We improve the last bottleneck structure of the MobileNetV3-Large network, and then apply concat operation to integrate features and reduce the number of parameters.
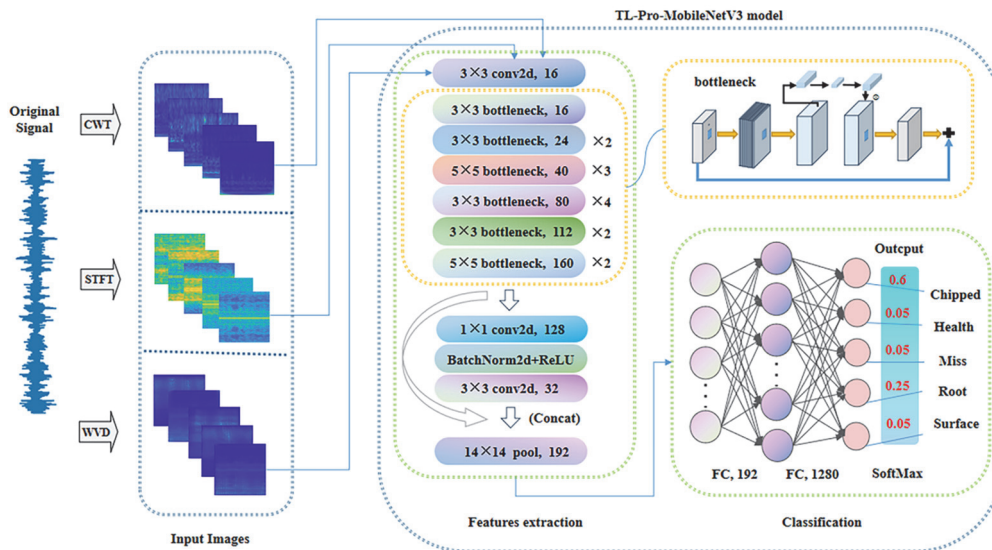

Figure 5 Flow chart of the TL-Pro-MobileNetV3 identification model

The last fully-connected layer is removed, then a fully connected layer with 5 neurons is added according to the number of states for the gear sample dataset. The Softmax is used for classification. The detailed structure is shown in Fig. 5. Due to a small number of labelled samples in fault diagnosis, it is far from the 10 million sample images of the extensive dataset ImageNet. To achieve a wonderful classification effect, sufficient amount of data is an essential prerequisite. Because the MobileNetV3-Large network has completed good classification performance on the ImageNet dataset, some parameters trained on the ImageNet dataset are transferred to the Pro-MobileNetV3 network to get the TL-Pro-MobileNetV3 model. Detailed information about the TL-Pro-MobileNetV3 network is shown in Tab. 2. The time-frequency images of the signals are taken as the target dataset, and the gearbox fault features are automatically extracted from the time-frequency image to diagnose the gearbox fault. Because natural images and time-frequency maps are not similar, more convolutional blocks need to be fine-tuned during training the gearbox dataset compared with pictures more similar to natural images [19].

Table 2 Detailed architecture of the TL-Pro-MobileNetV3 model

Input	Operator	exp size	#out	SE	NL	stride
$224^2 \times 3$	conv2d, 3×3	16	16	-	RE	1
$112^2 \times 16$	bottleneck, 3×3	16	16	-	RE	1
$112^2 \times 16$	bottleneck, 3×3	64	24	-	RE	2
$56^2 \times 24$	bottleneck, 3×3	72	24	-	RE	1
$56^2 \times 24$	bottleneck, 5×5	72	40	√	RE	2
$28^2 \times 40$	bottleneck, 5×5	120	40	√	RE	1
$28^2 \times 40$	bottleneck, 5×5	120	40	√	RE	1
$28^2 \times 40$	bottleneck, 3×3	240	80	-	HS	2
$14^2 \times 80$	bottleneck, 3×3	200	80	-	HS	1
$14^2 \times 80$	bottleneck, 3×3	184	80	-	HS	1
$14^2 \times 80$	bottleneck, 3×3	184	80	-	HS	1
$14^2 \times 80$	bottleneck, 3×3	480	112	√	HS	1
$14^2 \times 112$	bottleneck, 3×3	672	112	√	HS	1
$14^2 \times 112$	bottleneck, 5×5	672	160	√	HS	1
$14^2 \times 160$	bottleneck, 5×5	960	160	√	HS	1
$14^2 \times 160$	conv2d, 1×1	-	128	-	RE	1
$14^2 \times 128$	Batchnorm2d	-	128	-	-	-
$14^2 \times 128$	conv2d, 3×3	-	32	-	-	1
$14^2 \times 192$	pool, 14×14	-	192	-	-	1
$1^2 \times 192$	Fully-connected layer1	-	1280	-	HS	1
$1^2 \times 1280$	Fully-connected layer2	-	5	-	-	1

The detailed fault diagnosis process is shown in Fig. 5, including creating a sample dataset, model training, and model application.

(1) Building a sample dataset. In this paper, STFT, WVD, and CWT methods are used to convert time-domain signals into three-channel image data. We need to remove the surrounding coordinates and characters of the images above. Then the images are compressed and converted into a RGB time-frequency image. The corresponding time-frequency image is labeled according to the normal or fault states of the gearbox. The generated image is divided into the training set and the validation set, which are used to train and verify the Pro-MobileNetV3 network model.

(2) Model training. The parameters obtained from the MobileNet-Large network are used as the initialization parameters of the proposed model. The parameters of the changed structure are re-initialized to a random value close to 0. The TL-Pro-MobileNetV3 model can be obtained by iterating through certain rounds of training, and the parameters of TL-Pro-MobileNetV3 are persisted for the application of the model.

(3) Model application. The validation set separated from the gearbox dataset is used to verify the fault diagnosis performance of the TL-Pro-MobileNetV3 model after persistence, the accuracy of classification prediction for fault samples on the validation set is taken as the fault diagnostic accuracy.

4 EXPERIMENTAL VERIFICATION
4.1 Introduction to Experimental Data

The experimental data used in this paper comes from the public gearbox dataset of Southeast University [23], which is the parallel gearbox data collected from the Drivetrain Dynamic Simulator (DDS) shown in Fig. 6. The dataset contains four fault types and one healthy type. The detailed fault types are shown in Tab. 3.
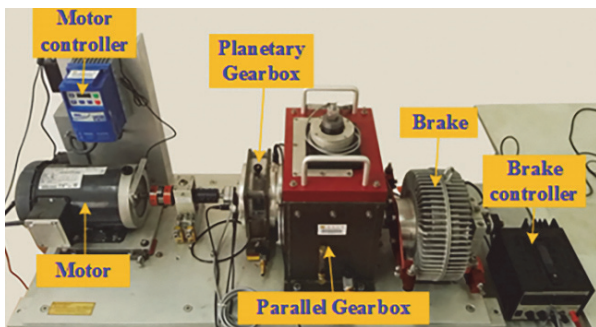

Figure 6 Experimental setup for gearbox dataset

Table 3 Gearbox different types description

Variables	Variable definitions	Values
Normal	Health	20 HZ - 0 V and 30 HZ - 2 V
Chipped	Crack occurs in the gear feet	20 HZ - 0 V and 30 HZ - 2 V
Miss	Missing one of feet in the gear	20 HZ - 0 V and 30 HZ - 2 V
Root	Crack occurs in the root of gear feet	20 HZ - 0 V and 30 HZ - 2 V
Surface	Wear occurs in the surface of gear	20 HZ - 0 V and 30 HZ - 2 V

Four fault types and one health type under two working conditions are selected as the gearbox fault diagnostic dataset. We use the three time-frequency analysis methods

introduced above to process the time-domain signals for obtaining time-frequency images. Each fault type consists of 800 training samples. The two gearbox datasets have a total of 8000 samples, which are divided into training sets and validation sets according to the proportion of 4:1. The detailed division is shown in Tab. 4.

Table 4 Dataset partitioning under different working conditions cited

	Type	Normal	Chipped	Miss	Root	Surface	Total	Working conditions
Dataset A	training set	640	640	640	640	640	4000	20 HZ - 0 V
	validation set	160	160	160	160	160		
Dataset B	training set	640	640	640	640	640	4000	30 HZ - 2 V
	validation set	160	160	160	160	160		

4.2 Test Platform and Parameter Setting

All experiments are carried out on the same platform. The experimental running environment is the Ubuntu18.04 operating system, the CPU is 12 core Intel (R) Xeon (R) Gold 5320, and the GPU is RTX A4000 (16 GB). The model proposed in this paper is implemented using python3.8 and pytorch1.8 combining the model with transfer learning to establish a gearbox fault diagnosis model.

During the experiment, the adjustment and use of parameters are shown in Tab. 5. In terms of learning rate, the strategy of automatically adjusting the learning rate is used, and the loss value of the validation set is used as an indicator. We set the initial learning rate to 0.1, 0.01, 0.001 and 0.0001, and the patience value of the loss function to 3, 5, 7, 10, the decreased amplitude to 0.1, 0.01, and 0.001 respectively to verify the recognition accuracy of the proposed model. When the initial learning rate is 0.1, the patience value is 3, and the decreased amplitude is 0.1 of the initial learning rate, the proposed model performs best. When the loss value exceeds three times without any change, the learning rate decay strategy is implemented to realize automatic adjustment of the learning rate. The Dropout is set to 0.2. In the experiment, all models are trained using Adaptive Moment Estimation (Adam). The cross-entropy function is used to calculate the loss value.

Table 5 Experiment setting variables parameters

Type	Description	Working conditions
Batch_size	Number of samples processed per batch	32
Basic learning rate	The initial learning rate	0.01
Learning rate decay	The learning rate is automatically adjusted according to the index, and the decrease is 0.1 of the initial learning rate	0.1
Dropout	Delete neurons randomly	0.2
Adam	An optimizer that updates parameters based on training data	-

4.3 Experiments and Result Analysis

In this paper, we verify the comprehensive performance of three time-frequency analysis methods

combined with deep learning. To reduce the impact of random initial values, each time-frequency analysis method is trained five times and its average value is taken. The fault diagnosis results are shown in Tab. 6.

Table 6 Diagnosis results of different time-frequency analysis methods

	Type	STFT	CWT	WVD
Accuracy	Dataset A	99.875%	98.875%	98.375%
	Dataset B	99.75%	98.625%	98.32%
Time	Dataset A	13.67	13.82	13.92
	Dataset B	13.69	13.85	13.94

In this table, the accuracy of the three methods is above 98% in the two working conditions, among which the accuracy of STFT is the highest, reaching 99.75% and 99.875%, respectively. From the perspective of training efficiency, WVD takes the longest training time, with an average iteration time of about 13.92 s per round, while STFT takes the least training time, approximately 13.67 s. Through comprehensive comparison, STFT takes less time and has the highest accuracy for 50 iterations.

To further evaluate the advantages of the STFT method, Fig. 7 shows the training and validation accuracy curves of the Pro-MobileNetV3 network for 50 iterations.
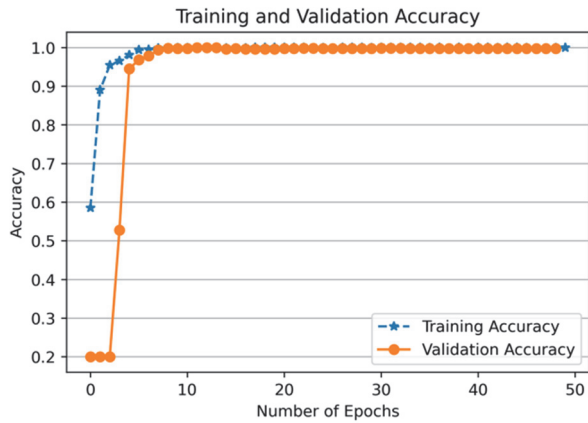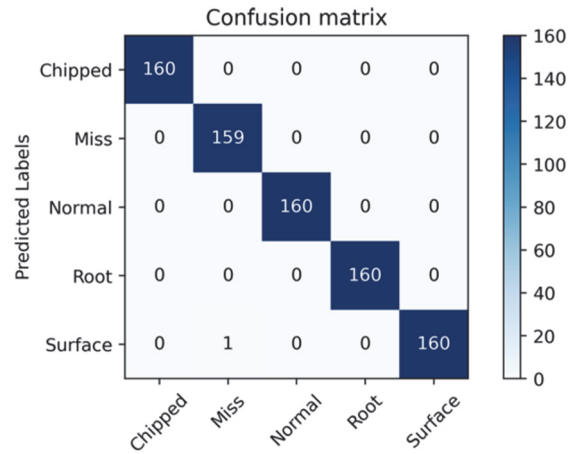

(a) Dataset A
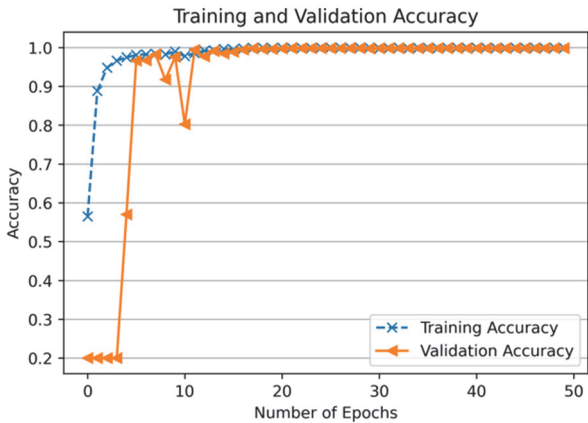
(b) Dataset B
**Figure 7** Accuracy curves of STFT

The abscissa represents the number of iterations and the ordinate represents the recognition accuracy that is used to quantify the difference between the real value and the model predicted value. Fig. 7a shows the accuracy has reached about 99.875% after 12 iterations on Dataset A. Although there are small fluctuations, it gradually becomes stable afterward. Fig. 7b shows that the accuracy converges to 99.75% after about 20 iterations on Dataset B. These

results show that the proposed model can complete the fault diagnosis of gearbox well. Fig. 8 shows the confusion matrix of the STFT method. The abscissa represents the true label and the ordinate represents the predicted label. The recognition accuracy of the validation set on Chipped, Normal, Root, and Surface all reach 100% in Fig. 8, indicating that the gearbox fault diagnosis based on the STFT method and deep learning has good resolution accuracy. Therefore, further research is to use the STFT method combining with deep learning to establish the gearbox diagnosis model.
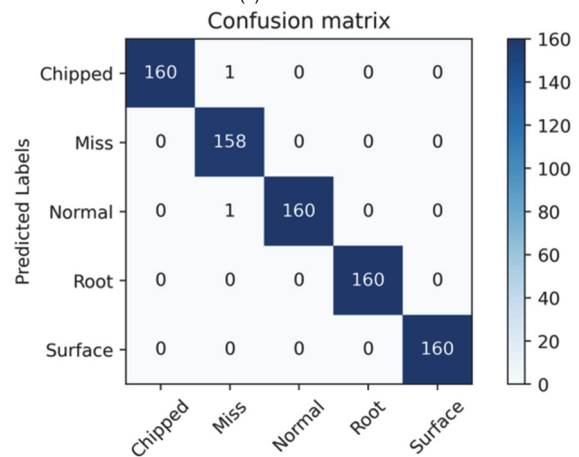
(a) Dataset A

(b) Dataset B
Figure 8 Confusion matrix of STFT

To verify the impact on network performance after introducing transfer learning, the parameters of the MobileNetV3-Large network on ImageNet dataset are used as the initial parameters of the proposed model for training. To ensure the fairness of the experiments and the reliability of the results, five experiments are conducted in the same environment.

Under two working conditions, the recognition accuracy results of the validation set are shown in Fig. 9. Compared with the Pro-MobilenetV3, the TL-Pro-MobilenetV3 network has better classification accuracy and achieves 100% recognition accuracy in each type. These results show that the accuracy after introducing transfer learning is better than that without transfer learning on the gearbox dataset. TL-Pro-MobilenetV3 network can achieve accurate prediction of fault types.
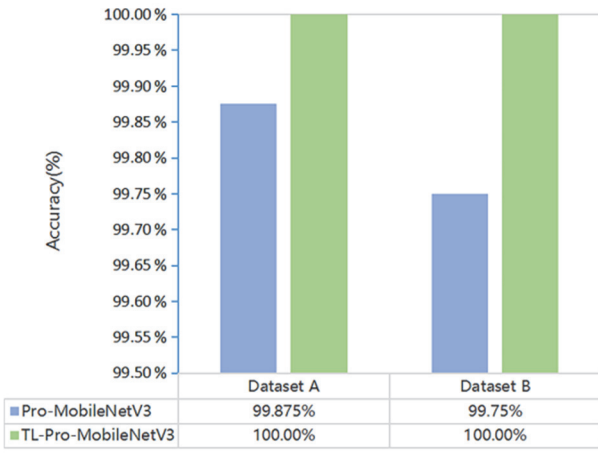
Figure 9 Validation accuracy before and after introducing transfer learning
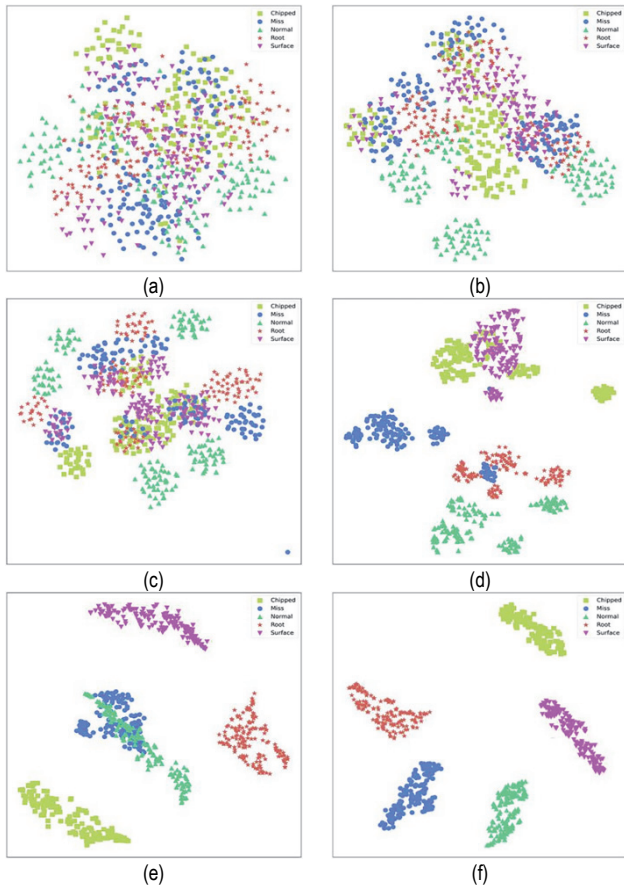

Figure 10 T-SNE for feature visualization (feature representations for signals extracted from the raw signal, four bottleneck structures and the last fully - connected layer, respectively)

To further analyze the diagnosis and classification effect, the data trained on Dataset A is subjected to feature dimension reduction using T-SNE technology in some structures of the TL-Pro-MobileNetV3 network. As Fig. 10 shows, the 2D feature vector was visualized. The visualization of dimensionality reduction shows that, without classification, all kinds of fault features are mixed together and difficult to distinguish. After model training, there are apparent five types of distribution in the fully-connected layer.

To further verify the advantages of the model after introducing transfer learning method, four models the RexNet50, ResNext50, MobilenetV2 and MobilenetV3-Large are compared with the proposed model. The detailed
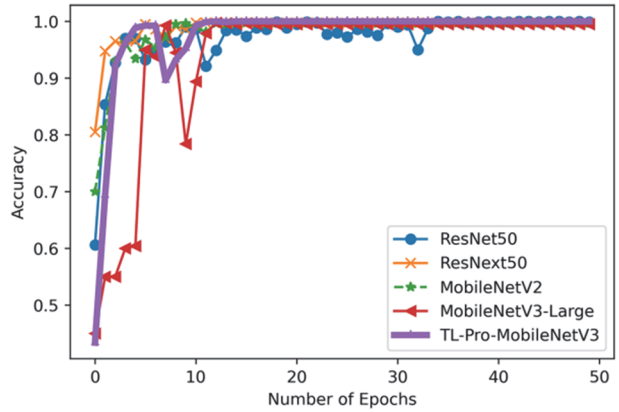
comparative results are shown in Tab. 7. The TL-Pro-MobileNetV3 network proposed has the best diagnostic accuracy on both Dataset A and Dataset B. In Fig. 11 and Fig. 12, the accuracy is stable to 100% after 13 iterations on Dataset A and converges to 100% after 15 iterations on Dataset B. Moreover, the time required to train from scratch is the shortest in all deep learning models.

Table 7 Diagnosis results of the different deep learning models.

Method	Dataset A	Dataset B
ResNet50	99.63%	99.52%
ResNext50	99.73%	99.55%
MobileNetV2	99.71%	99.65%
MobileNetV3	99.75%	99.68%
TL-Pro-MobiileNetV3	100%	100%
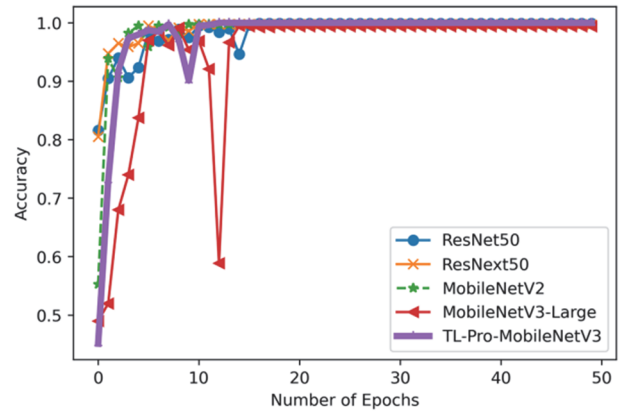
(a) Dataset A

(b) Dataset B

Figure 11 Accuracy curves of different deep learning models

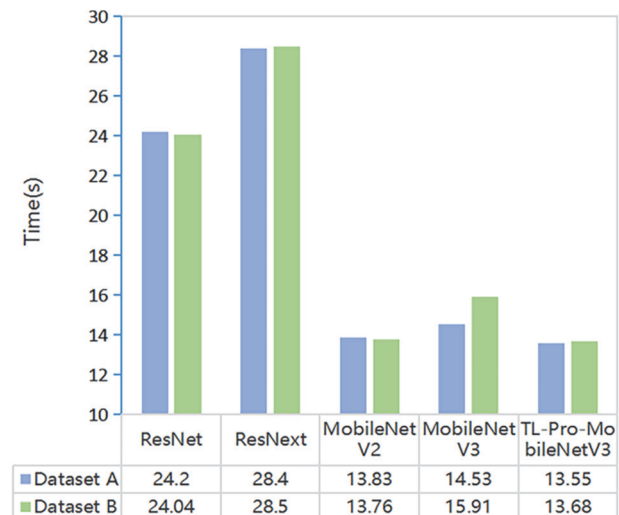|	ResNet	ResNext	MobileNet V2	MobileNet V3	TL-Pro-Mo bileNetV3
Dataset A	24.2	28.4	13.83	14.53	13.55
Dataset B	24.04	28.5	13.76	15.91	13.68

Figure 12 Training time for different deep learning models

5 CONCLUSION

In the paper, a gearbox fault diagnosis method (TL-Pro-MobileNetV3) based on MobileNetV3 network and transfer learning is proposed for the problem that under different working conditions, the gearbox feature extraction is difficult and the large difference in data distribution causes the low diagnostic accuracy. The parameters trained on the ImageNet dataset are successfully transferred to the gearbox fault diagnosis field. The TL-Pro-MobilenetV3 model is applied to gearbox fault diagnosis under different working conditions.

We combine three different time-frequency methods with the Pro-MobileNetV3 model. The model shows good classification ability in the three methods. Among them, the accuracy based on STFT method reached 99.75% and 99.875% respectively in the two working conditions. The experimental results show that the proposed model achieves high accuracy in gearbox fault diagnosis under different working conditions. It can adaptively extract the fault features of the time-frequency image for gearbox, and shows good classification ability combined with STFT method.

Based on the STFT method, the TL-Pro-MobileNetV3 is validated experimentally. The results show each iteration takes 13.55 s on average. In the 13-th iteration, the fault diagnosis accuracy rate reaches 100% and tends to be stable. In another working condition, the network can also achieve 100% fault diagnosis accuracy. Therefore, the proposed model has good performance in generalization ability, recognition accuracy, and training time under different working conditions. It can realize accurate and efficient classification of gearbox faults, which is of great significance for difficult to extract gearbox fault features and have large difference in data distribution under different working conditions.

Since this paper only explores two working conditions of gearbox and the training time needs to be reduced, the aspects will be further explored in the future:

(1) Research a gearbox fault diagnosis model suitable for more diverse working conditions to further enhance model generalization ability.

(2) Explore deep learning algorithms with shorter fault diagnosis time while maintaining high accuracy.

6 REFERENCES

[1] Yu, X., Li, Z., He, Q., Y, Y., Du, M., & Peng, Z. (2021). Gearbox fault diagnosis based on bearing dynamic force identification. *Journal of Sound and Vibration*, *511*, 253-116360. https://doi.org/10.1016/j.jsv.2021.116360

[2] Zhu, X., Wang, R., Fan, Z., Xia, D., Liu, Z., & Li, Z. T. (2022). Gearbox fault identification based on lightweight multivariate multi directional induction network. *Measurement*, *193*,110977. https://doi.org/10.1016/j.measurement.2022.110977

[3] Ma, Y., Cheng, J., Hu, N., Cheng, Z., & Yang, Y. (2021). Symplectic quaternion singular mode decomposition with application in gear fault diagnosis. *Mechanism and Machine Theory*, *160*, 104266. https://doi.org/10.1016/j.mechmachtheory.2021.104266

[4] Wang, A., Riehle, Li, Y., Du, X., & Zhong, C. (2021). Diesel Engine Gearbox Fault Diagnosis Based on Multi-features Extracted from Vibration Signals. *IFAC-Papers on Line*, *54*, 33-38. https://doi.org/10.1016/j.ifacol.2021.10.137

[5] Chen, Y., Rao, M., Feng, Ke., & Zuo, M. (2022). Physics-Informed LSTM hyperparameters selection for gearbox fault detection. *Mechanical Systems and Signal Processing*, *171*, 108907. https://doi.org/10.1016/j.ymssp.2022.108907

[6] Liu, H. & Dhupia, J. S. (2014). A time domain approach to diagnose gearbox fault based on measured vibration signals. *Journal of Sound and Vibration*, *333*, 2164-2180. https://doi.org/10.1016/j.jsv.2013.11.033

[7] Ye, A., Zhou, X., & Miao, F. (2022). Innovative Hyperspectral Image Classification Approach Using Optimized CNN and ELM. *Journal of Psychophysiology*, *11*(5), 775-775. https://doi.org/10.3390/electronics11050775

[8] Gong, W., Chen, H., Zhang, Z., Zhang, M., Wang, R., Guan, C., & Wang, Q. (2019). A Novel Deep Learning Method for Intelligent Fault Diagnosis of Rotating Machinery Based on Improved CNN-SVM and Multichannel Data Fusion. *Sensors (Basel, Switzerland)*, *19*(7), 1693. https://doi.org/10.3390/s19071693

[9] Zhang, K., Tang, B., Deng, L., & Liu, X. (2021). A hybrid attention improved ResNet based fault diagnosis method of wind turbines gearbox. *Measurement*, *179*, 109491. https://doi.org/10.1016/j.measurement.2021.109491

[10] Han, D., Liang, K., & Shi, P. (2020). Intelligent fault diagnosis of rotating machinery based on deep learning with feature selection. *Journal of Low Frequency Noise, Vibration and Active Control*, *39*(4), 939-953. https://doi.org/10.1177/1461348419849279

[11] Li, X., Jiang, H., Niu, M., & Wang, R. (2020). An enhanced selective ensemble deep learning method for rolling bearing fault diagnosis with beetle antennae search algorithm. *Mechanical Systems and Signal Processing*, *142*, 106752. https://doi.org/10.1016/j.ymssp.2020.106752

[12] Wang, T., Riehle, H. J., & Elbert, T. (2020). Fault diagnosis of rotating machinery under time-varying speed based on order tracking and deep learning. *Journal of Vibroengineering*, *22*(2), 366-382. https://doi.org/10.21595/jve.2019.20784

[13] He, Z., Shao, H., Lin, J., Cheng, J., & Yang, Y. (2020). Transfer fault diagnosis of bearing installed in different machines using enhanced deep auto-encoder. *Measurement*, *152*, 107393. https://doi.org/10.1016/j.measurement.2019.107393

[14] He, C., Jiang, Y., Sun, L., & Wu, B. (2021). Gearbox fault diagnosis based on transfer learning with RseNet50 model. *Journal of Physics: Conference Series*, *1986*(1), 012093. https://doi.org/10.1088/1742-6596/1986/1/012093

[15] He, Z., Shao, H., Wang, P., Lin, J., Cheng, J., & Yang, Y. (2020). Deep transfer multi-wavelet auto-encoder for intelligent fault diagnosis of gearbox with few target training samples. *Knowledge-Based Systems*, *191*, 105313. https://doi.org/10.1016/j.knosys.2019.105313

[16] Liu, Y. & Wang, J. (2022). Transfer learning based multi-layer extreme learning machine for probabilistic wind power forecasting. *Applied Energy*, *312*, 118729. https://doi.org/10.1016/j.apenergy.2022.118729

[17] Li, J., Lin, M., Li, Y., & Wang, X. (2022). Transfer learning with limited labeled data for fault diagnosis in nuclear power plants. *Nuclear Engineering and Design*, *390*, 111690.

https://doi.org/10.1016/j.nucengdes.2022.111690

[18] Wang, Y., Chen, Z., & Zhang, W. (2022). Lithium-ion battery state-of-charge estimation for small target sample sets using the improved GRU-based transfer learning. *Energy*, *244*, 123178. https://doi.org/10.1016/j.energy.2022.123178

[19] Shao, S., McAleer, S., Yan, R., & Baldi, P. (2018). Highly-Accurate Machine Fault Diagnosis Using Deep Transfer Learning. *IEEE Transactions on Industrial Informatics, 15*(4), 2446-2455. https://doi.org/10.1109/TII.2018.2864759

[20] Wang, M., Guo, B., Hu, Y., Zhao, Z., Liu, C., & Tang, H. (2022). Transfer Learning Models for Detecting Six Categories of Phonocardiogram Recordings. *Journal of Cardiovascular Development and Disease*, *9*(3), 86. https://doi.org/10.3390/jcdd9030086

[21] Chen, R., Huang, X., Yang, L., Xu, X., Zhang, X., & Zhang, Y. (2019). Intelligent fault diagnosis method of planetary gearboxes based on convolution neural network and discrete wavelet transform. *Computers in Industry*, *106*, 48-59. https://doi.org/10.1016/j.compind.2018.11.003

[22] Deng, T. & Wu, Y. (2022). Simultaneous vehicle and lane detection via MobileNetV3 in car following scene. *PloS one*, *17*(3), e0264551. https://doi.org/10.1371/journal.pone.0264551

[23] Li, X., Shen, X., Zhou, Y., Zhou, Y., Wang, X., & Li, T. (2020). Classification of breast cancer histopathological images using interleaved DenseNet with SENet (IDSNet). *PloS one*, *15*(5), e0232127. https://doi.org/10.1371/journal.pone.0232127

Contact information:

Yanping DU
Department of Mechanical and Electrical Engineering,
Beijing Institute of Graphic Communication,
Beijing 102600, China
E-mail: duyanping@bigc.edu.cn

Xuemin CHENG
Department of Mechanical and Electrical Engineering,
Beijing Institute of Graphic Communication,
Beijing 102600, China
E-mail: 1147818561@qq.com

Yunxin LIU
Department of Mechanical and Electrical Engineering,
Beijing Institute of Graphic Communication,
Beijing 102600, China
E-mail: 1325683817@qq.com

Shuihai DOU
Department of Mechanical and Electrical Engineering,
Beijing Institute of Graphic Communication,
Beijing 102600, China
E-mail: doushuihai@bigc.edu.cn

Juncheng TU
(Corresponding author)
Department of Mechanical and Electrical Engineering,
Beijing Institute of Graphic Communication,
Beijing 102600, China
E-mail: tutuworking@163.com

Yanlin LIU
Department of Mechanical and Electrical Engineering,
Beijing Institute of Graphic Communication,
Beijing 102600, China
E-mail: wn25816482021@163.com

Xianyang SU
Department of Mechanical and Electrical Engineering,
Beijing Institute of Graphic Communication,
Beijing 102600, China
E-mail: suxianyang323@163.com