Sound & Vibration 2025, 59(1), 2242.
https://doi.org/10.59400/sv2242

Article

Engineering vibration recognition using CWT-ResNet

Wei Huang[1,*], Jian Xu[2]

[1] SINOMACH Academy of Science and Technology Co. Ltd, SINOMACH Research Center of Engineering Vibration Control Technology, Beijing 100080, China
[2] China National Machinery Industry Corporation Ltd (SINOMACH), Beijing 100080, China
*** Corresponding author:** Wei Huang, huangweiac@126.com

CITATION

Huang W, Xu J. Engineering vibration recognition using CWT-ResNet. Sound & Vibration. 2025; 59(1): 2242. https://doi.org/10.59400/sv2242

ARTICLE INFO

Received: 9 December 2024
Accepted: 27 December 2024
Available online: 7 January 2025

Abstract: Multi-source signal recognition is a common problem in engineering vibration control. Given that traditional methods often primarily rely on prior knowledge and expertise, which can limit efficiency and accuracy, this study proposed a vibration recognition model based on ResNet, utilizing continuous wavelet transform to combine signal processing with deep learning techniques. The continuous wavelet transform converts the original one-dimensional vibration signals into two-dimensional time-frequency representations with richer feature information, which are then input into the convolutional layers for automatic feature extraction, culminating in vibration recognition through the Softmax layer. To evaluate the model's performance, 20 sets of measured vibration data were tested. The results show that the proposed model achieves a recognition accuracy of 99%, excelling in both component recognition and the separation of vibration signals. Therefore, this study is of great significance for engineering vibration diagnosis, the front-end design of vibration control, and the analysis and optimization of control effectiveness.

Keywords: continuous wavelet transform; ResNet; engineering vibration recognition

1. Introduction

In the industrial realm, the concurrent operation of a multitude of apparatuses gives rise to diverse vibration sources. The precise identification of these vibration sources empowers the realization of accurate fault diagnosis. For instance, within a factory's production line, different machines exhibit distinctive vibration traits. By means of multi-source vibration recognition, faults like bearing abrasion and imbalance in a particular machine can be promptly detected, averting production halts due to equipment breakdowns and curtailing maintenance expenditures. In the transportation domain, during the course of vehicle travel, it is subject to various vibration sources including the road surface and the engine. The identification of these vibrations is conducive to optimizing the vehicle's suspension system and noise reduction design, thereby augmenting ride comfort and driving safety. In the construction arena, multi-source vibration recognition can differentiate among disparate vibration sources such as earthquakes, nearby construction activities, and vehicular traffic. This is of utmost importance for the preservation of ancient edifices and the vibration prevention of buildings housing precision instruments, facilitating the more effective implementation of corresponding protective measures.

In practical engineering, objects under vibration control often face multi-source vibrations. Precision engineering, in particular, is susceptible to various influences, including ground pulsation, personnel movement, power equipment, internal (and external) pipelines within building structures, ground traffic, rail transit, construction

activities, and wind. To ensure the accurate and effective design of vibration control systems and development of vibration control devices, it is crucial to systematically study the multi-source vibration environment affecting the control object, which includes analyzing and identifying vibration components and frequency bands. Specifically, for active and semi-active control systems employing feedback design, accurately identifying multi-source environmental vibration components is essential to ensure the control system's effective functioning, avoiding debugging and potential failures. Hou et al. [1] reviewed vibration damage identification methods used in civil engineering projects from 2010 to 2019. The review covered both classical approaches and advanced intelligent methods, including modal parameter-based methods, signal processing methods, machine learning (ML) algorithms, and Bayesian methods. Wang et al. [2] proposed a deep learning-based cable vibration recognition system, which consisted of a composite model based on Resnet-34 (Residual Neural Network of 34 layers) and Swin-B (the base model in Swin Transformer), a linear rigid body motion recognizer based on Hough linear detection, and data processing. Liu et al. [3] presented a comprehensive review of deep learning-based planetary gearbox health state recognition. Kounta et al. [4] presented an approach first based on mechanical skills first to identify the optimal signal processing, then based on deep learning to automatically detect the phenomenon of chatter in machining. Łuczak [5] highlighted the utilisation of short-time Fourier transform (STFT) and continuous wavelet transform (CWT) for extracting time–frequency components from the signal, and by extracting the features using CWT, a convolutional neural network (CNN) for fault diagnosis were carried out.

A convolutional neural network (CNN) is a type of feedforward neural network inspired by the biological visual perception mechanism and is considered one of the classical algorithms in deep learning. In recent years, cross-fusion approaches integrating CNNs with other intelligent methods have been widely adopted for vibration classification and recognition tasks. For example, Pinedo-Sanchez et al. [6] proposed a CNN model based on the AlexNet architecture to classify and diagnose wear levels in rotating systems. Nguyen et al. [7] proposed a deep learning method combining CNN and long short-term memory (CNN-LSTM) [8], which serves as the backbone for computer vision-based vibration testing technology. Liu et al. [9] applied deep recurrent neural networks (RNNs) and CNNs for vibration-based ground recognition in working faces, training and testing two deep CNN architectures, GoogLeNet and ResNet, using time-frequency scalogram data derived from continuous wavelet transformations.

ResNet is a deep residual network proposed by He et al. [10] of Microsoft Asia Research Institute. This network successfully tackles two problems that typically emerge when traditional convolutional neural networks are overly stacked. The first problem pertains to the vanishing or exploding gradients, while the second one is the degradation issue. In response to the first problem, it is put forward that through data preprocessing and the utilization of Batch Normalization (BN) layers within the network, the issue of vanishing or exploding gradients can be effectively dealt with. Regarding the second problem, the residual structure is introduced with the aim of alleviating the degradation problem. The short-circuit mechanism is the core idea of ResNet, which can effectively alleviate the network degradation problem caused by

the increase of network depth in the traditional convolutional layer. Four aspects of the widespread use of ResNet in medical image processing are discussed in Ref. [11], i.e. lung tumor, diagnosis of skin diseases, diagnosis of breast diseases, and diagnosis of diseases of the brain. Wen et al. [12] proposed a new TCNN(ResNet-50) with the depth of 51 convolutional layers for fault diagnosis. Zhang et al. [13] proposed a transfer residual neural network based on ResNet-50 for detection of steel surface defects.

In this study, a multi-source vibration signal recognition method utilizing continuous wavelet transform and ResNet is presented, where the original vibration signals are converted into two-dimensional images using continuous wavelet transform. ResNet is then employed to automatically extract features from the transformed 2D images, minimizing the influence of prior knowledge and expert experience. Finally, the vibration signals are classified using the SoftMax layer within ResNet.

2. Continuous wavelet transform

The continuous wavelet transform is a time-frequency analysis method designed for time-varying and non-stationary signals [5,14]. It converts one-dimensional vibration signals into two-dimensional time-frequency representations, containing both time and frequency domain information. Unlike the short-time Fourier transform, which uses a fixed window function, the continuous wavelet transform provides an adjustable window function that balances time and frequency resolution when analyzing non-stationary signals.

The continuous wavelet transform can conduct diverse analyses on signals within different frequency ranges and at different times. Through the multi-resolution analysis, high-frequency signals can achieve a good time resolution but a relatively poor frequency resolution. Conversely, low-frequency signals can obtain a better frequency resolution and a higher time resolution. This clearly overcomes the drawbacks of applying the Fourier transform to non-stationary signals. The wavelet transform offers a time-frequency mixed representation of signals and has highly efficient applications in numerous fields. For engineering vibration recognition, the continuous wavelet transform can be utilized to convert one-dimensional vibration time-domain signals into two-dimensional time-frequency diagrams for the training and testing of deep learning models.

For any signal x (t), its continuous wavelet transform is defined as:

$$X_{\omega}(u,v) = \int_{-\infty}^{+\infty} x(t)\varphi_{u,v}(t)dt = \frac{1}{\sqrt{v}} \cdot \int_{-\infty}^{+\infty} x(t)\varphi\left(\frac{t-u}{v}\right)dt \tag{1}$$

where u is the translation factor, determining the wavelet window's position in the time domain, v is the scaling factor, adjusting the wavelet window's size and its position in the frequency domain, and $\varphi_{u,v}(t)$ is the wavelet basis function (also referred to as the parent wavelet), expressed as follows:

$$\varphi_{u,v}(t)dt = \frac{1}{\sqrt{|u|}}\varphi\left(\frac{t-v}{u}\right) \quad u > 0 \tag{2}$$

In this study, the complex Morlet wavelet is selected as the basis function for continuous wavelet transform.

3. ResNet

The ResNet is proposed to solve the problem of gradient disappearance or gradient explosion [15,16] and degradation [17,18] caused by the deepening of the layers of convolutional neural networks. The core part of ResNet is the residual structure [19], and the schematic diagram of which is shown in **Figure 1**. The features learned by the stacked network are denoted as H(x) when the input is x, and then H(x) = F(x) + x, and the problem to be learned can be transformed to the learning residual feature F(x) = H (x) − x.

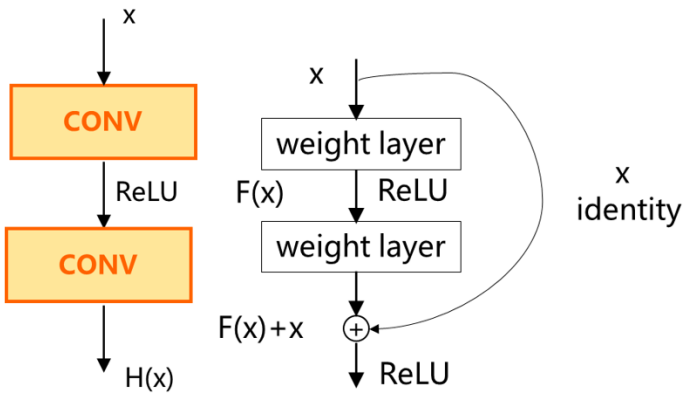


Figure 1. Residual structure.

The residual block structure can be expressed as:

$$y_l = h(x_l) + F(x_l, W_l) \tag{3}$$

$$x_{l+1} = f(y_l) \tag{4}$$

where x_l and x_{l+1} are respectively the input and output of the l-th residual block, F is the learned residual feature, $h(x_l) = x_l$ represents the identity mapping, f is the ReLU activation function. Based on these, the following can be obtained:

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

x_L is the learning feature from shallow layer l to deep layer L. Based on this, the process gradient of network backpropagation can be obtained as follows:

$$\frac{\partial loss}{\partial x_l} = \frac{\partial loss}{\partial x_L} \times \frac{\partial x_L}{\partial x_l} = \frac{\partial loss}{\partial x_L} \times \left(1 + \frac{\partial}{\partial x_L} \sum_{i=l}^{L-1} F(x_i, W_i)\right) \tag{6}$$

where $\frac{\partial loss}{\partial x_L}$ is the gradient of the loss function in layer L, and 1 in the second part shows the short-circuit mechanism, which indicates that the gradient can be propagated losslessly. Meanwhile, the other term needs to pass through layers with weight W_i to transmit the gradient. The transfer of gradient will not disappear due to

the existence of 1 in Equation (6), so that the residual block can learn new features based on the input features and have better learning performance.

For networks with a relatively small number of layers, such as ResNet18 and ResNet34, the residual structure is composed of two 3 × 3 convolutional layers with a stride of 1 on the main branch, along with a shortcut branch. In the case of networks having a larger number of layers, like ResNet50, ResNet101, and ResNet152, the main branch of the residual structure comprises three convolutional layers, specifically two 1 × 1 convolutional layers and one 3 × 3 convolutional layers. The two 1 × 1 convolutional layers serve different purposes, one is utilized for compressing the channel dimension, while the other is employed for restoring the channel dimension. Moreover, when the input channel and the output channel are not in agreement, in order to achieve consistency, the convolution on the main branch is changed to a convolution with a stride of 2, and a 1 × 1 convolution kernel with a stride of 2 must be added to the identity mapping. ResNet18 and ResNet50 are selected for comparison here, and their network structures are presented in **Table 1** [20].

Table 1. The network structures of ResNet18 and ResNet50.

Network layer	Output size	ResNet18	ResNet50
Conv1	112×112	7×7, 64, stride=2	
		3×3 max pool, stride=2	
Conv2_x	56×56	Basicblock (channel = 64) × 2	Bottleneck (channel = 64) × 3
Conv3_x	28×28	Basicblock (channel = 128) × 2	Bottleneck (channel = 128) × 4
Conv4_x	14×14	Basicblock (channel = 256) × 2	Bottleneck (channel = 256) × 6
Conv5_x	7×7	Basicblock (channel = 512) × 2	Bottleneck (channel = 512) × 3
	1×1	Average pool, 1000-d fc, softmax	

Several network structures with different depths have been designed for ResNet, and the network structure is shown in **Figure 2** [21]. The main part is composed of stacked residuals, and the number and structure of residuals are different for networks with different depths. Taking Conv3_x as an example, for the shallow network ResNet18 and ResNet34, the two residual structures on the left are stacked, where the dashed structure is in the first layer, and the solid structure is set as different stacks according to the number of network layers. For the deep network ResNet50, ResNet101, and ResNet152, the residual structure is stacked based on the right 3 layers. Similarly, the dashed structure is adopted in the first layer, and the others are stacked by the solid structure, by which the parameters and computation can be reduced and the network's training is accelerated.

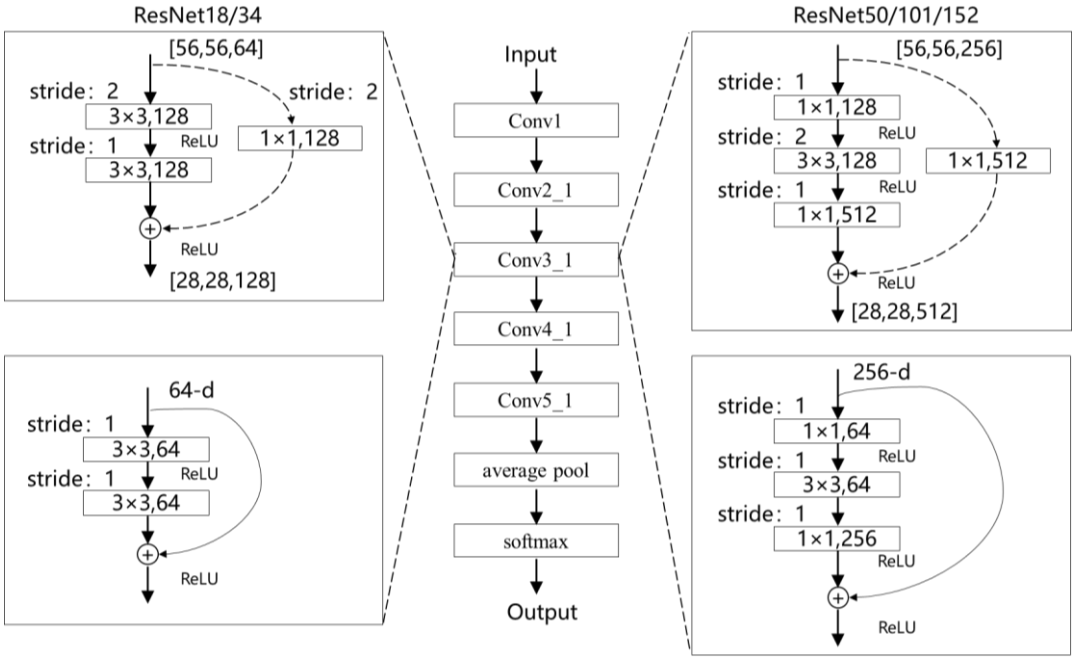


Figure 2. Shallow and deep ResNet.

ResNet50, being a significant member within the ResNet series, encompasses 3, 4, 6, and 3 Bottleneck modules in layers 2 to 5 respectively. The structure of the Bottleneck module is illustrated in **Figure 3**. Each Bottleneck module consists of two distinct types of Blocks. The first type is the Conv Block, as depicted in **Figure 3a**, and the second type is the Identity Block, as presented in **Figure 3b**.

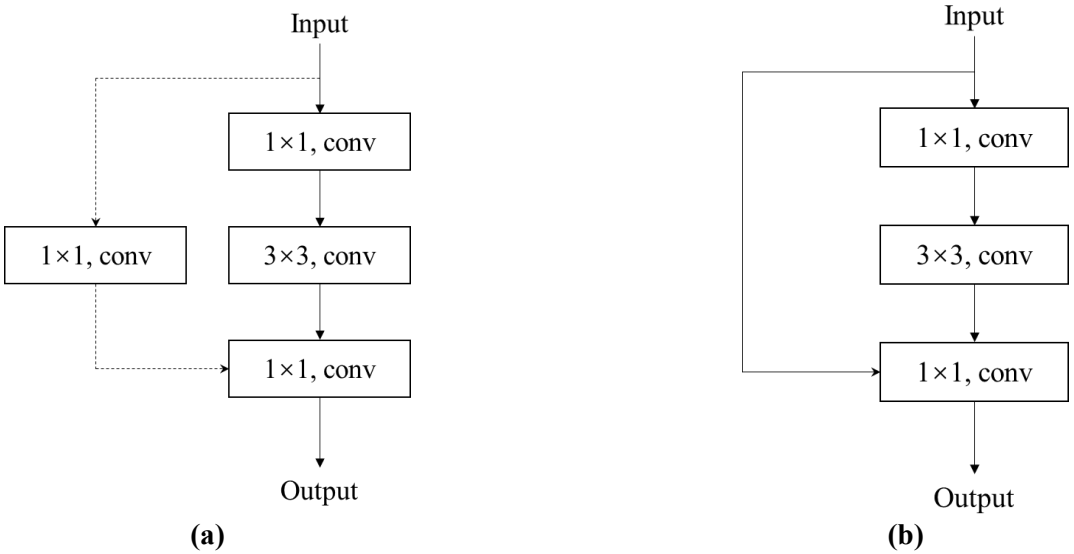


Figure 3. The structure of bottleneck. **(a)** conv block; **(b)** identity block.

ResNet50 is situated between shallow and deep networks. For training small-sample data, the model has good network expressiveness while keeping the overall number of parameters moderate, thus reducing the phenomenon of overfitting. Taking all factors into consideration, the ResNet50 is adopted in this study for the further research.

The parameter settings of ResNet50 are as follows. The optimization algorithm is *Adam*. The maximum number of epochs is 100. The size of the mini-batch is 64. The gradient threshold is 1, and the initial learning rate is 0.001.

4. Sources of multi-source vibration data

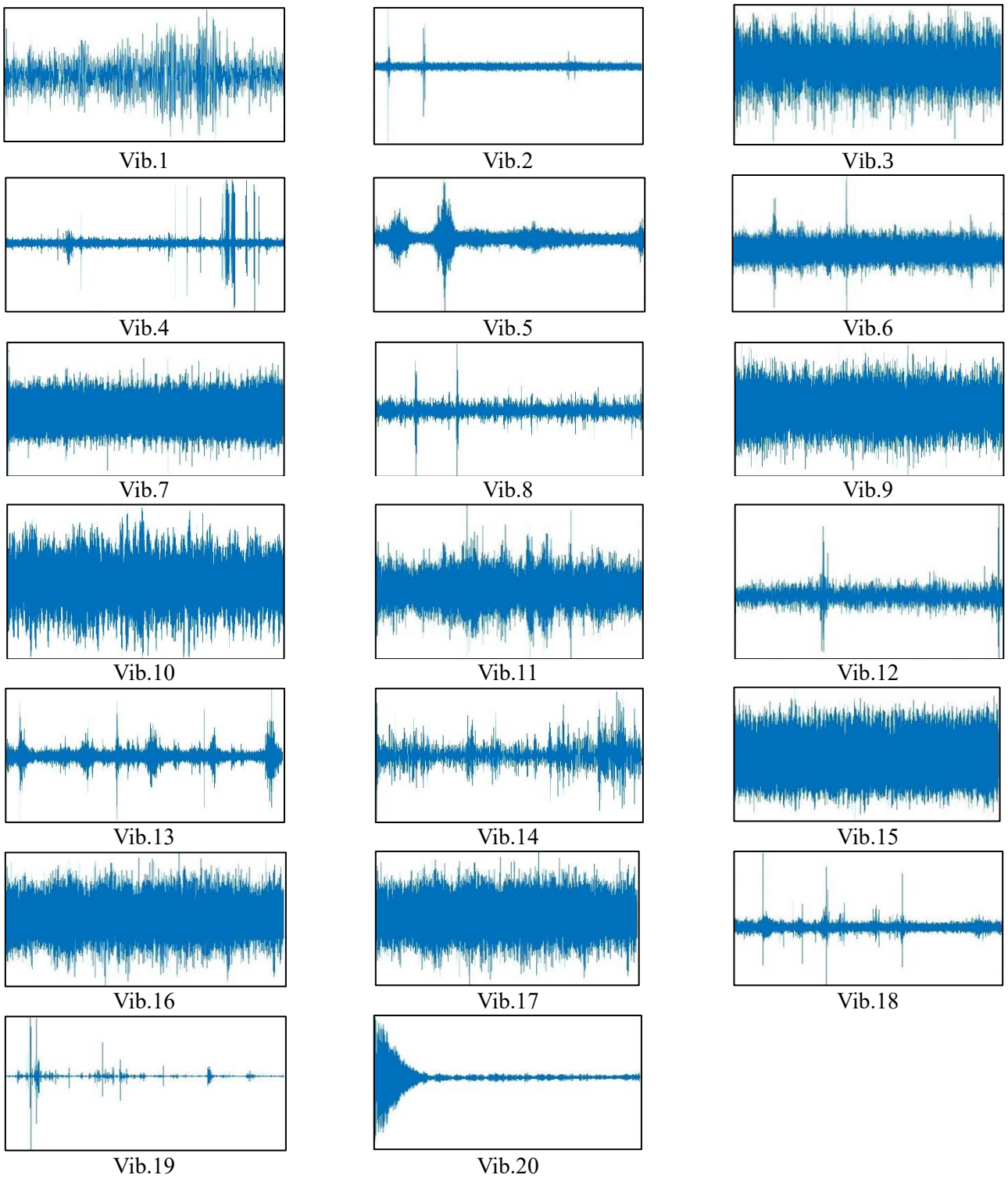


Figure 4. 20 sets of original vibration data.

This study analyzed 20 sets of original vibration data collected from actual engineering tests at a sampling frequency of 256 Hz, all representing acceleration signals, as shown in **Figure 4**. Vib.1 is data from a newly constructed building site affected by ground traffic, while Vib.2 comes from the driving ramp of an underground parking lot in an office building. Vib.3 is from the converter foundation of a steel plant, and Vib.4 is from the deep foundation pit of a newly constructed university campus. Vib.5 reflects data from the floor of a subway building structure, and Vib.6 is from the anti-vibration foundation surface of a research institution. Vib.7 represents data from an open field, and Vib.8 is from a residential floor. Vib.9 is from a site selected for a quantum testing building of a university, and Vib.10 is from an electron microscope room of a research institution. Vib.11 captures data from a gravitational wave laboratory, while Vib.12 comes from a new Internet of Things laboratory of a university. Vib.13 is from a precision optics laboratory, Vib.14 is from a chip production factory, and Vib.15 is from a rooftop fan on an office building. Vib.16 covers data from lampblack units and air conditioning power equipment situated on a university rooftop, Vib.17 is from the foundation of a research institute's mechanics laboratory, Vib.18 is from an independent large-volume foundation of an electronic industrial plant, Vib.19 is from a site selected for an electronic industrial plant, and Vib.20 is from the grid structure of a wind tunnel laboratory.

5. Identification of multi-source vibration

This study proposed a multi-source vibration recognition method, CWT-ResNet, which converts one-dimensional vibration signals into two-dimensional time-frequency representations using continuous wavelet transform, thereby enabling the representation of vibration feature information. These two-dimensional time-frequency representations are then input into the ResNet convolutional neural network, which automatically extracts relevant features. Finally, the vibration type is classified by the Softmax layer. The detailed steps are as follows:

(1) The original vibration data collected is randomly divided according to the specified sample length.

(2) Continuous wavelet transform is applied to convert one-dimensional vibration signals into two-dimensional time-frequency representations.

(3) The resulting time-frequency representations are proportionally divided into training and test sets.

(4) ResNet50 models are built (**Figure 5**), with its parameters initialized.

(5) The training set's time-frequency representations are input into the convolutional layers for model training, and the optimal model parameters are saved.

(6) The test set is input into the model for classification, and the classification results and accuracy are evaluated.

CWT-ResNet vibration identification process is summarized in **Figure 6**.

From 20 vibration datasets, 60 samples are extracted for each type of vibration, resulting in a total of 1200 samples. Each sample consists of 1024 sampling points, which are converted into 64 × 64 sample graphs to serve as model inputs. The dataset is then divided into training and test sets at a 3:1 ratio.

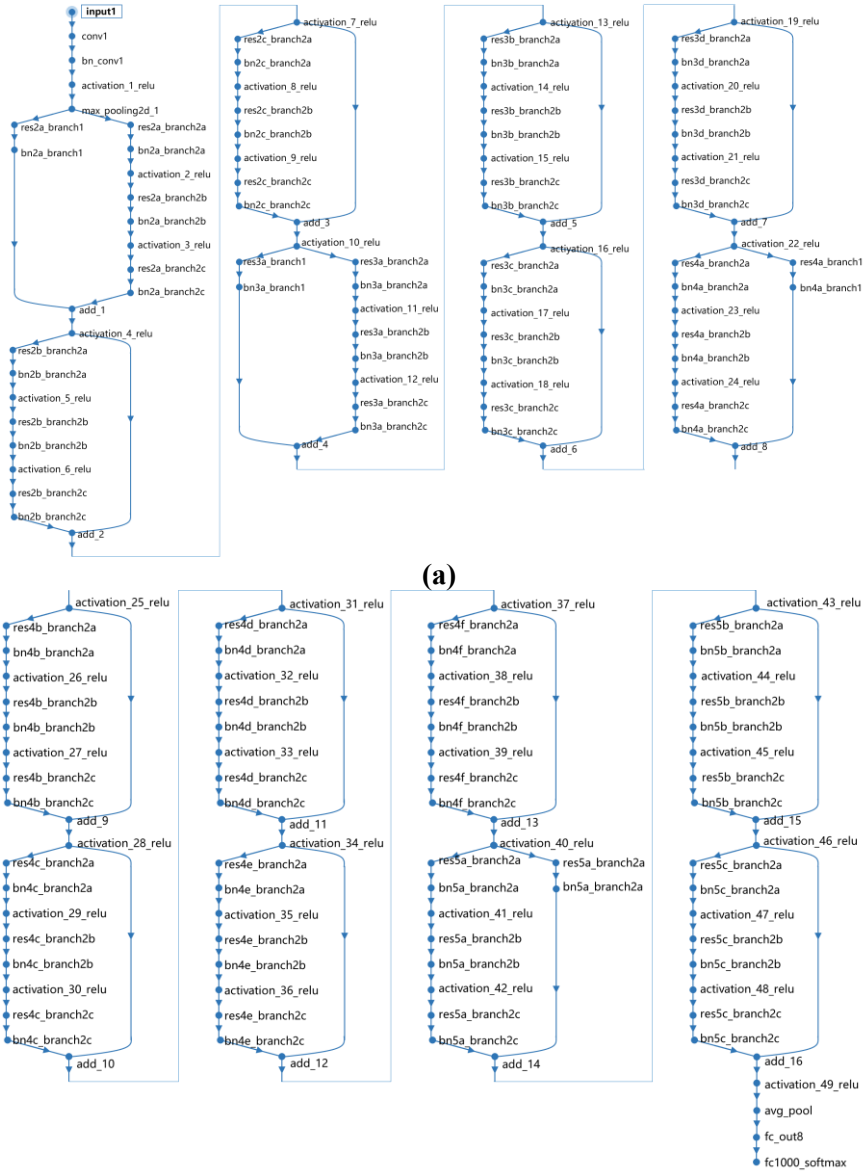


Figure 5. The net structure of adopted ResNet50. **(a)** the first half of the net structure of ResNet50; **(b)** the second half of the net structure of ResNet50.

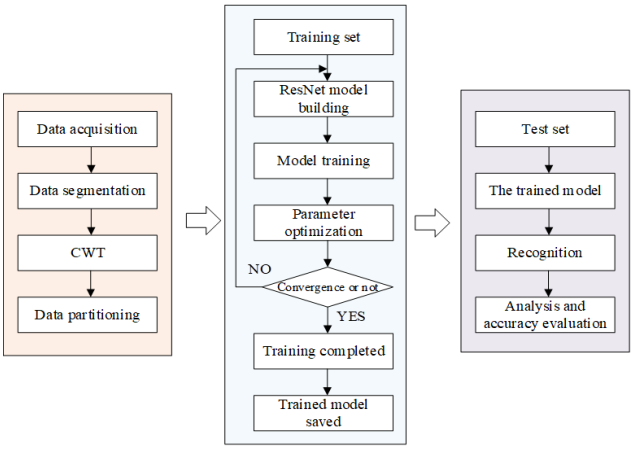


Figure 6. CWT-ResNet vibration identification process.

The continuous wavelet transform can convert one-dimensional vibration data into two-dimensional time-frequency diagrams (as shown in **Figure 7**). The total number of images is 60 × 20 = 1200, where 60 indicates the total number of samples, and 20 indicates the total categories of vibration signals. Then, these 1200 images are divided into a training set and a testing set by the ratio 900:300.

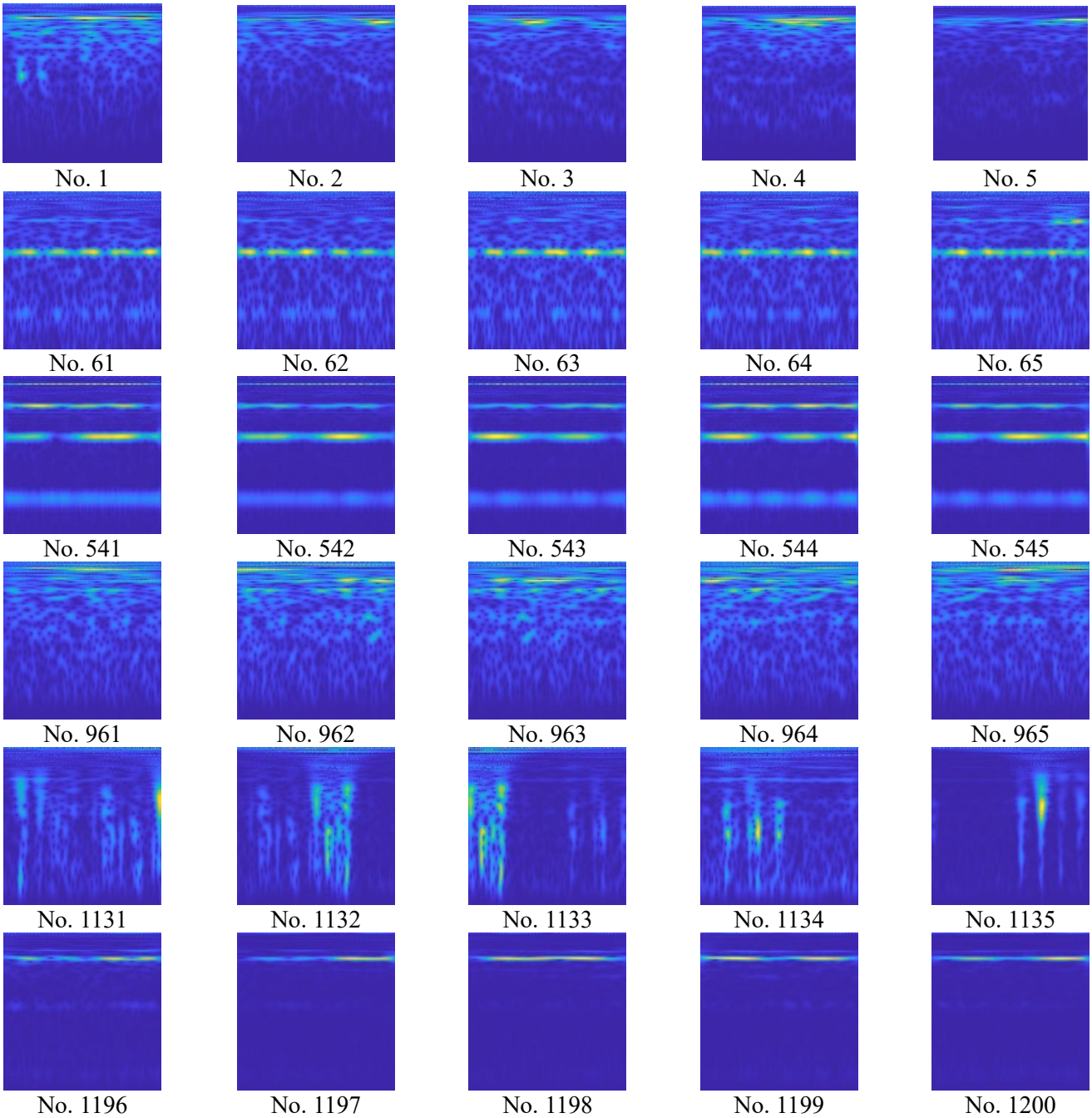


Figure 7. Two-dimensional time-frequency images of vibration signals obtained using the continuous wavelet transform.

t-SNE (t-Distributed Stochastic Neighbor Embedding) is an effective nonlinear dimensionality reduction method used to embed high-dimensional data into low-dimensional space for visualization [22]. In the high-dimensional data space, the relationships between data points are intricate and hard to comprehend intuitively. The

objective of t-SNE is to map high-dimensional data into a low-dimensional space (typically two-dimensional or three-dimensional), while preserving the local and global structural relationships among data points to the greatest extent possible. It computes the probability distributions among high-dimensional data points and constructs a joint probability distribution based on the similarity of data points. Subsequently, it searches for the corresponding probability distribution in the low-dimensional space and optimizes the positions of points in the low-dimensional space by minimizing the difference between the two probability distributions (usually using the Kullback-Leibler divergence). For instance, when handling the feature vectors of image data or text data that originally exist in a high-dimensional space, t-SNE can project these high-dimensional data onto a two-dimensional plane, bringing similar data points closer in the two-dimensional space and relatively separating data points of different categories. Thus, it helps people intuitively observe the clustering situation of data, discover patterns in the data, and assist in tasks such as data analysis and model evaluation in machine learning and deep learning.

To visually evaluate the performance of the proposed model, t-SNE is applied for visual analysis. The visualization results of the original training set data are shown in **Figure 8**, where 20 distinct colors represent the 20 vibration inputs in the dataset. The results reveal that the original data distribution is relatively disordered, with significant overlap among vibration signals from the 20 categories, making it impossible to classify and identify multi-source vibrations based on the original data alone. **Figure 9** presents the visualization results of the data processed by CWT-ResNet. Compared to the original data, the processed data exhibits a reduced degree of disorder, indicating a trend toward greater organization, with features from different categories becoming more distinguishable.

The model's classification performance is represented using a confusion matrix. As can be seen from **Figure 10**, the model achieves an accuracy of 99% on the test dataset, indicating its ability to accurately distinguish between different vibration types.

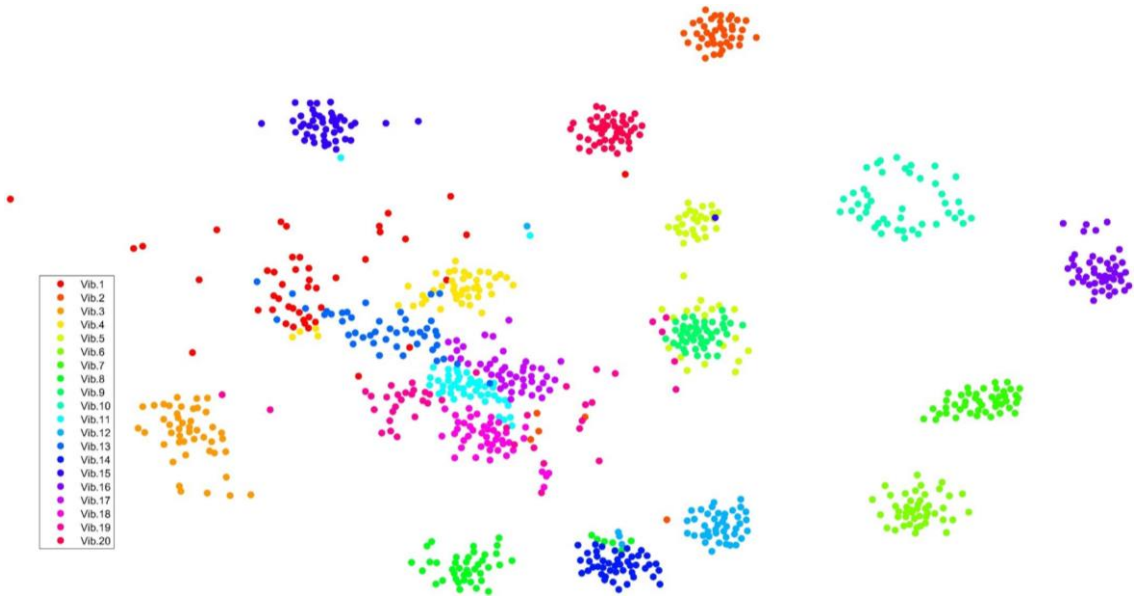


Figure 8. Original sample distribution.

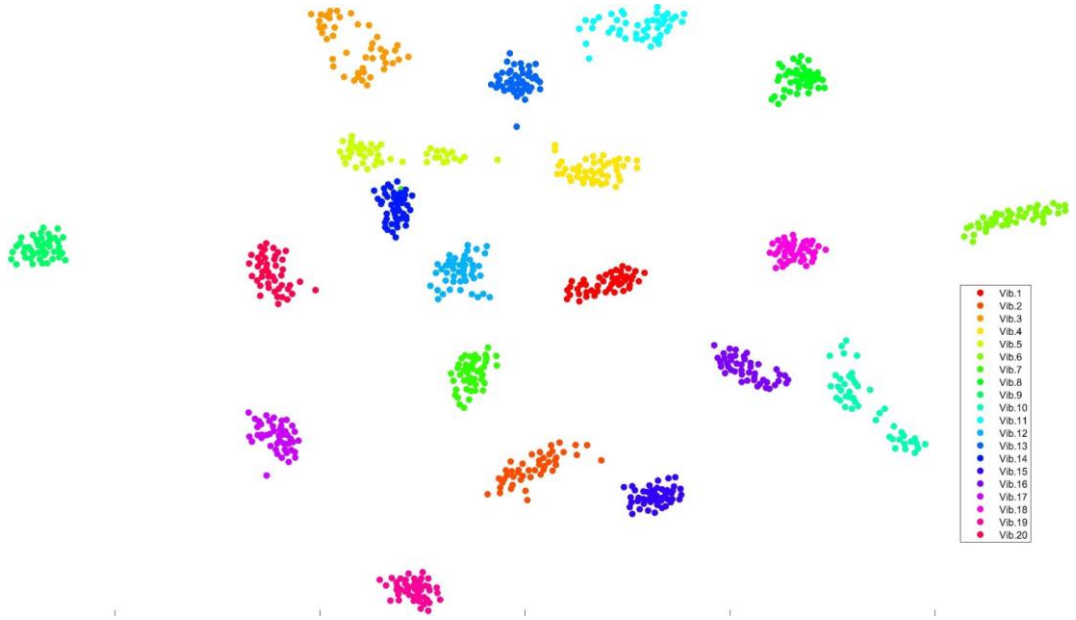


Figure 9. Sample distribution after model identification.

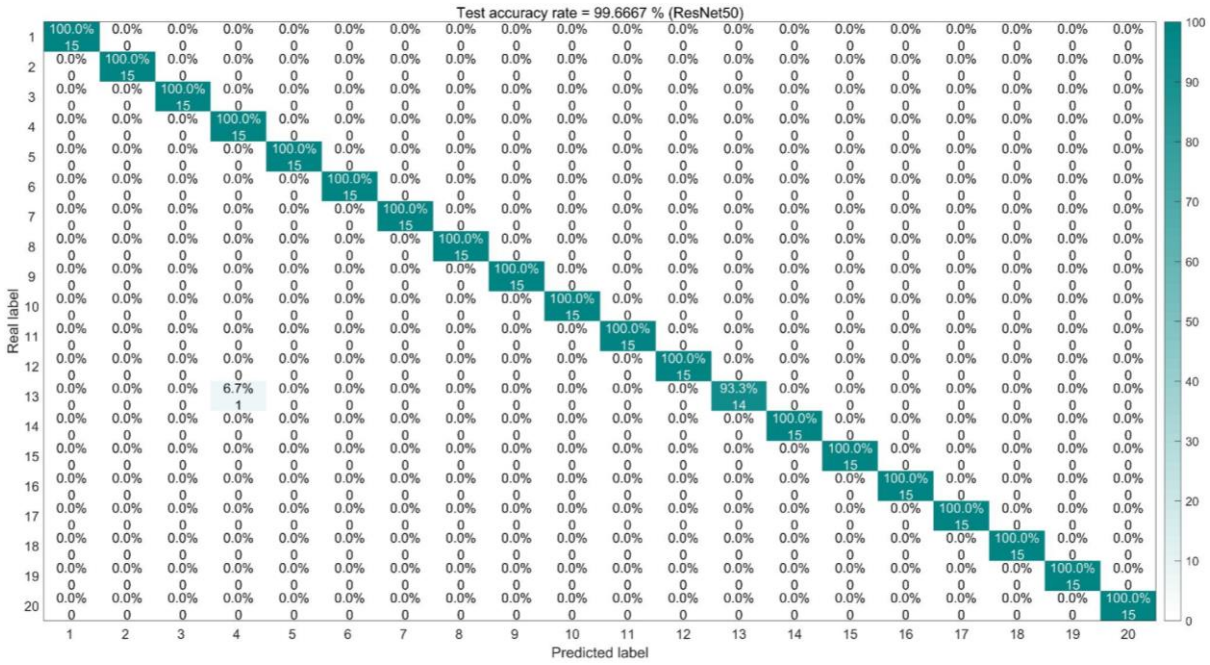


Figure 10. Confusion matrix for classification results.

Building on this research, the model is applied to vibration identification, with Vib. test selected as the subject of study (**Figure 11**). The main computational steps are as follows: the trained model is loaded, vibration data are processed through wavelet transformation, the transformed data are input into the trained model for vibration signal recognition, and the identified vibration signals are extracted in real time. The recognition results are summarized in **Table 2**.

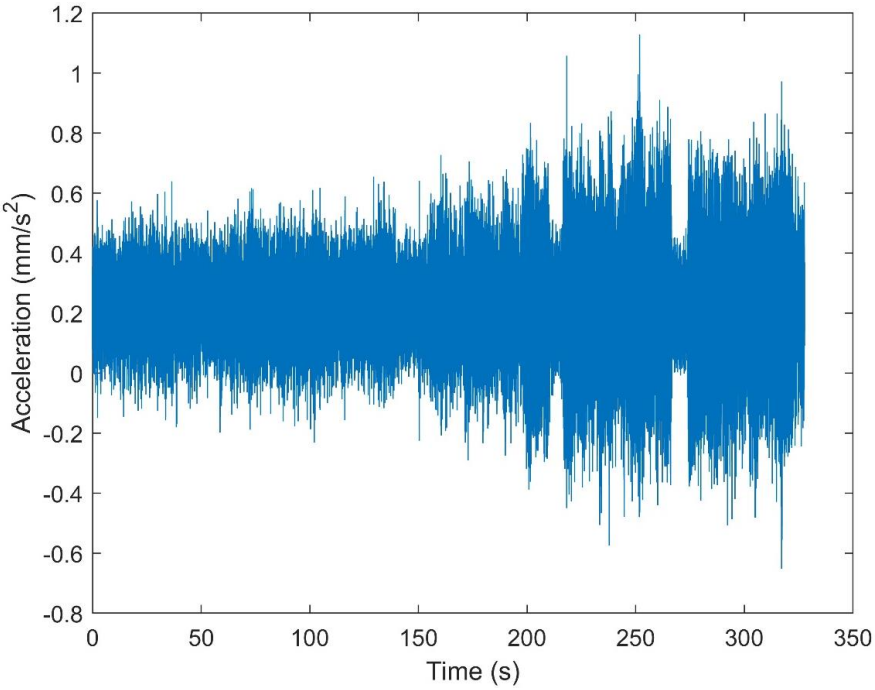


Figure 11. A selected vibration signal for identification test.

Table 2. Vibration identification results.

Time interval	Identified vibration type	Time interval	Identified vibration type
1 s~4 s	Vib.19	129 s~132 s	Vib.19
5 s~8 s	Vib.19	133 s~136 s	Vib.19
9 s~12 s	Vib.19	137 s~140 s	Vib.13
13 s~16 s	Vib.19	141 s~144 s	Vib.19
17 s~20 s	Vib.19	145 s~148 s	Vib.19
21 s~24 s	Vib.19	149 s~152 s	Vib.19
25 s~28 s	Vib.19	153 s~156 s	Vib.19
29 s~32 s	Vib.19	157 s~160 s	Vib.19
33 s~36 s	Vib.19	161 s~164 s	Vib.13
37 s~40 s	Vib.19	165 s~168 s	Vib.19
41 s~44 s	Vib.19	169 s~172 s	Vib.19
45 s~48 s	Vib.19	173 s~176 s	Vib.19
49 s~52 s	Vib.19	177 s~180 s	Vib.19
53 s~56 s	Vib.19	181 s~184 s	Vib.19
57 s~60 s	Vib.19	185 s~188 s	Vib.19
61 s~64 s	Vib.19	189 s~192 s	Vib.19
65 s~68 s	Vib.19	193 s~196 s	Vib.19
69 s~72 s	Vib.19	197 s~200 s	Vib.13
73 s~76 s	Vib.19	201 s~204 s	Vib.19
77 s~80 s	Vib.19	205 s~208 s	Vib.19
81 s~84 s	Vib.19	209 s~212 s	Vib.19
85 s~88 s	Vib.19	213 s~216 s	Vib.19
89 s~92 s	Vib.19	217 s~220 s	Vib.19

Table 2. (*Continued*).

Time interval	Identified vibration type	Time interval	Identified vibration type
93 s~96 s	Vib.19	221 s~224 s	Vib.19
97 s~100 s	Vib.19	225 s~228 s	Vib.19
101 s~104 s	Vib.19	229 s~232 s	Vib.19
105 s~108 s	Vib.19	233 s~236 s	Vib.19
109 s~112 s	Vib.19	237 s~240 s	Vib.19
113 s~116 s	Vib.19	241 s~244 s	Vib.18
117 s~120 s	Vib.19	245 s~248 s	Vib.19
121 s~124 s	Vib.19	249 s~252 s	Vib.19
125 s~128 s	Vib.19	253 s~256 s	Vib.19

As can be seen from **Table 2**, the primary component identified is Vib.19, with additional components including Vib.13 (3 groups) and Vib.18 (1 group). These components are grouped for FFT analysis, with the results presented in **Figures 12** and **13**. For Vib.13, the FFT analysis results in **Figure 12** indicate that the prominent frequency bands are consistently distributed within the ranges of 0 to 40 Hz and 60 to 120Hz, and the two ranges of components are basically equivalent. Similarly, the results for Vib.18, shown in **Figure 13**, reveal a comparable distribution of prominent frequency bands, spanning 0 to 40 Hz and 60 to 120Hz, but it is clear that the components of the latter are higher than the former.

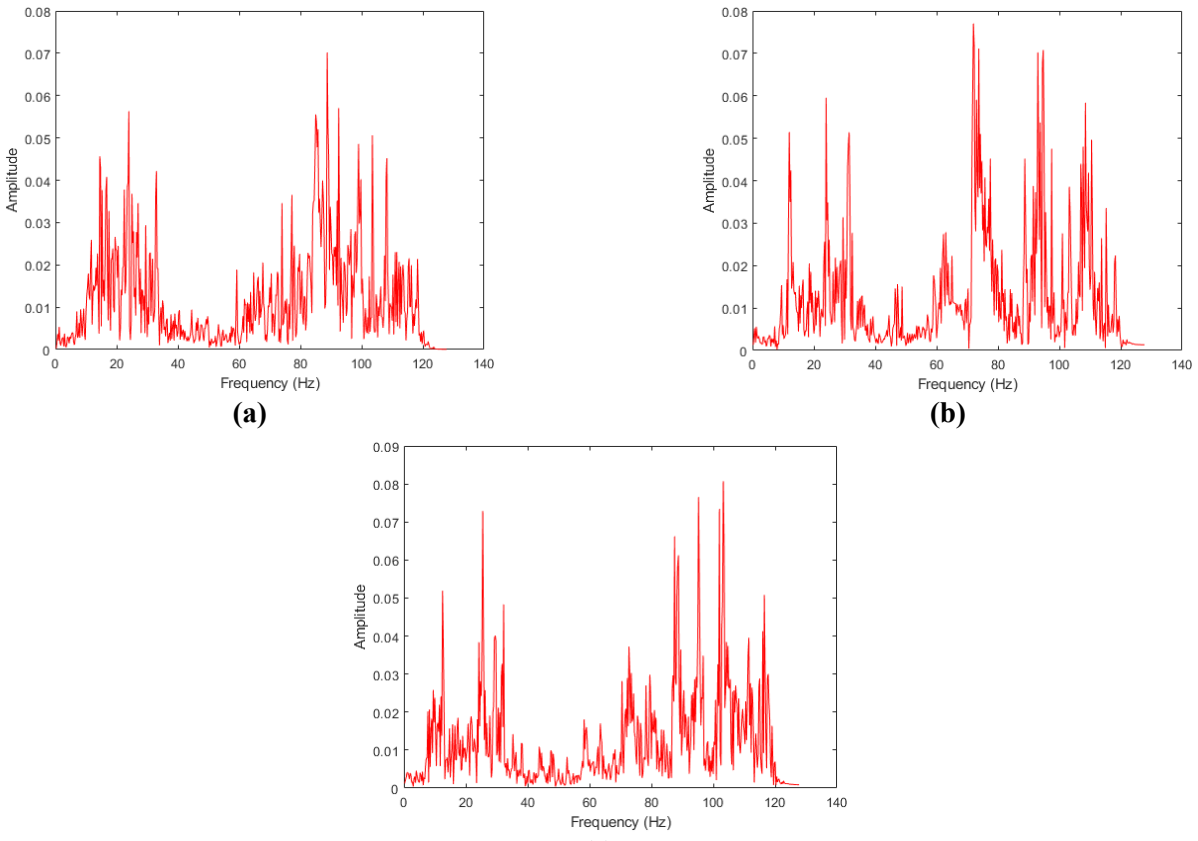
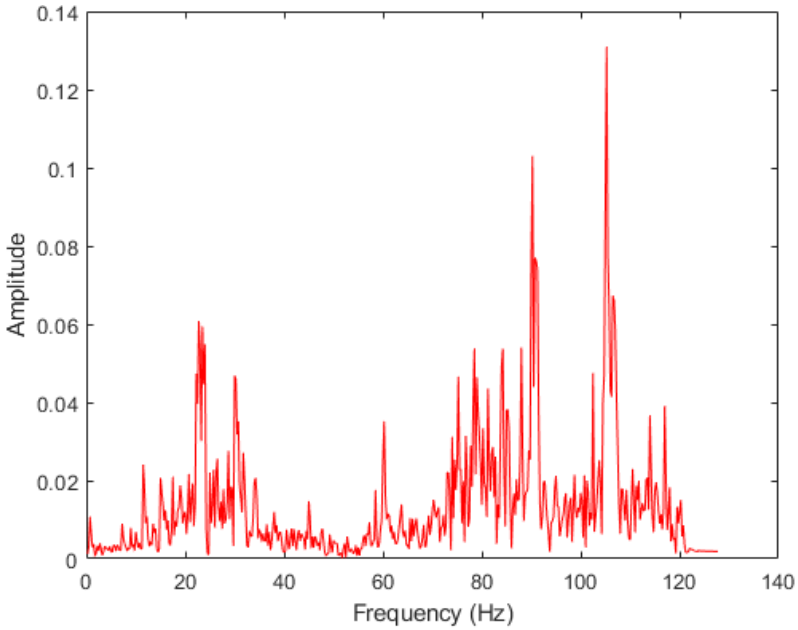


Figure 12. Spectral analysis results for identified Vib.13 components. **(a)** 137 s~140 s; **(b)** 161 s~164 s; **(c)** 197 s~200 s.



241 s~244 s

Figure 13. Spectral analysis results for identified Vib.18 components.

An FFT analysis is performed on the principal components from 1s to 136 s, as shown in **Figure 14**. The spectral analysis results indicate that the dominant frequency bands are mainly concentrated in the ranges of 0 to 50 Hz and 50 to 120 Hz, and the spectral shape and frequency band distribution differ from those observed in Vib.13 and Vib.18.

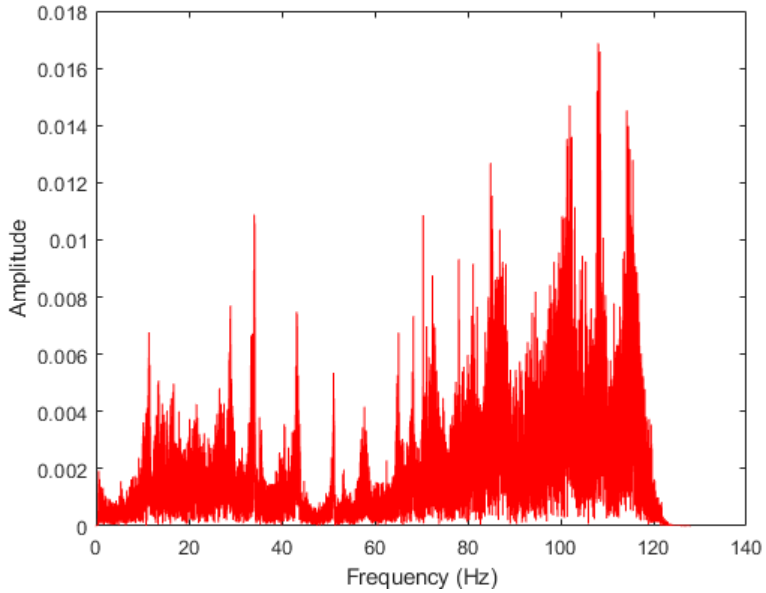


Figure 14. Spectrum analysis results for the principal components.

6. Summary

This study investigated the challenge of recognizing multi-source vibration signals, a common issue in engineering vibration control. This involves converting the vibration signals using continuous wavelet transform, followed by inputting the

transformed signals into the ResNet model. After configuring the relevant parameters, the model is trained, and the trained model is subsequently employed to recognize the vibration signals.

Given that traditional methods, which heavily depend on prior knowledge, are beset with limitations, this research introduced a novel and innovative approach. By leveraging the continuous wavelet transform, it converted one-dimensional vibration signals into two-dimensional time-frequency representations that are replete with rich and valuable feature information. Subsequently, these representations were fed into the ResNet convolutional neural network. ResNet, with its ingenious residual structure, effectively tackled problems such as gradient disappearance and network degradation, thereby facilitating efficient and effective feature extraction.

In the course of the study, 20 sets of actual engineering vibration data were meticulously analyzed. After undergoing comprehensive data preprocessing and rigorous model training, the proposed CWT-ResNet method attained an impressively high recognition accuracy of 99% on the test dataset. Through visualization using t-SNE, it was evident that the processed data exhibited a significantly higher degree of organization in contrast to the originally disordered data.

The model was further extended and applied to a specific vibration identification test, where it successfully identified the main vibration component as well as additional components. The FFT analysis of these components unearthed their distinct and characteristic frequency band properties. The research results demonstrate that the proposed method effectively identifies multi-source vibration data, with test results confirming its high accuracy. It is important to note that the results and accuracy of the multi-source vibration component identification and stripping process depend on the diversity of the training samples provided to the model.

In future work, the author plans to further expand and enrich the sample database developed in this study to enable more accurate and comprehensive vibration identification. This will not only enhance the understanding of vibration phenomena but also contribute to the development of more reliable and efficient engineering vibration control strategies. In addition, research on vibration time series prediction will be undertaken, building on the foundation established by the current vibration identification method.

Author contributions: Conceptualization, JX; methodology, WH; software, WH; validation, WH; formal analysis, WH; investigation, WH; resources, WH and JX; data curation, WH; writing—original draft preparation, WH; writing—review and editing, WH and JX; visualization, WH; supervision, JX; project administration, JX; funding acquisition, WH and JX. All authors have read and agreed to the published version of the manuscript.

Funding: The authors wish to acknowledge the Key Youth Fund of SINOMACH (Grant no. QNJJZD-2022-04), “Research and application of key technologies for micro-nano environmental vibration control of Chinese major science and technology infrastructures”. The authors also wish to express gratitude to the National Key Research and Development Program (Grant no. 2024YFF0508104), “Research and

Application of Vibration and Structural Safety Control Methods for Complicated Underground Rail Transit".

Conflict of interest: The authors declare no conflict of interest.

References

1. Hou R, Xia Y. Review on the new development of vibration-based damage identification for civil engineering structures: 2010–2019. Journal of Sound and Vibration. 2021; 491: 115741. doi: 10.1016/j.jsv.2020.115741
2. Wang W, Cui D, Ai C, et al. Target-free recognition of cable vibration in complex backgrounds based on computer vision. Mechanical Systems and Signal Processing. 2023; 197: 110392. doi: 10.1016/j.ymssp.2023.110392
3. Liu D, Cui L, Cheng W. A review on deep learning in planetary gearbox health state recognition: methods, applications, and dataset publication. Measurement Science and Technology. 2023; 35(1): 012002. doi: 10.1088/1361-6501/acf390
4. Kounta CAKA, Arnaud L, Kamsu-Foguem B, et al. Deep learning for the detection of machining vibration chatter. Advances in Engineering Software. 2023; 180: 103445. doi: 10.1016/j.advengsoft.2023.103445
5. Łuczak D. Machine Fault Diagnosis through Vibration Analysis: Continuous Wavelet Transform with Complex Morlet Wavelet and Time–Frequency RGB Image Recognition via Convolutional Neural Network. Electronics. 2024; 13(2): 452. doi: 10.3390/electronics13020452
6. Pinedo-Sánchez LA, Mercado-Ravell DA, Carballo-Monsivais CA. Vibration analysis in bearings for failure prevention using CNN. Journal of the Brazilian Society of Mechanical Sciences and Engineering. 2020; 42(12): 628. doi: 10.1007/s40430-020-02711-w
7. Nguyen DC, Salamak M, Katunin A, et al. Vibration-based SHM of railway steel arch bridge with orbit-shaped image and wavelet-integrated CNN classification. Engineering Structures. 2024; 315: 118431. doi: 10.1016/j.engstruct.2024.118431
8. Aldhyani THH, Alkahtani H. Attacks to Automatous Vehicles: A Deep Learning Algorithm for Cybersecurity. Sensors. 2022; 22(1): 360. doi: 10.3390/s22010360
9. Liu M, Liao S, Yang Y, et al. Tunnel boring machine vibration-based deep learning for the ground identification of working faces. Journal of Rock Mechanics and Geotechnical Engineering. 2021; 13(6): 1340–1357. doi: 10.1016/j.jrmge.2021.09.004
10. He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016: 770–778. doi: 10.1109/cvpr.2016.90
11. Xu W, Fu YL, Zhu D. ResNet and its application to medical image processing: Research progress and challenges. Computer Methods and Programs in Biomedicine. 2023; 240: 107660. doi: 10.1016/j.cmpb.2023.107660
12. Wen L, Li X, Gao L. A transfer convolutional neural network for fault diagnosis based on ResNet-50. Neural Computing and Applications. 2019; 32(10): 6111–6124. doi: 10.1007/s00521-019-04097-w
13. Zhang L, Bian Y, Jiang P, et al. A Transfer Residual Neural Network Based on ResNet-50 for Detection of Steel Surface Defects. Applied Sciences. 2023; 13(9): 5260. doi: 10.3390/app13095260
14. Zhao H, Liu J, Chen H, et al. Intelligent Diagnosis Using Continuous Wavelet Transform and Gauss Convolutional Deep Belief Network. IEEE Transactions on Reliability. 2023; 72(2): 692–702. doi: 10.1109/tr.2022.3180273
15. Ribeiro AH, Tiels K, Aguirre LA, et al. Beyond exploding and vanishing gradients: analysing RNN training using attractors and smoothness[C]//International conference on artificial intelligence and statistics. PMLR; 2020. pp. 2370–2380.
16. Liu M, Chen L, Du X, et al. Activated Gradients for Deep Neural Networks. IEEE Transactions on Neural Networks and Learning Systems. 2023; 34(4): 2156–2168. doi: 10.1109/tnnls.2021.3106044
17. Chen YM, Huang WT, Ho WH, et al. Classification of age-related macular degeneration using convolutional-neural-network-based transfer learning. BMC Bioinformatics. 2021; 22(S5): 1–16. doi: 10.1186/s12859-021-04001-1
18. Setiawan F, Liu AB, Lin CW. Development of Neuro-Degenerative Diseases' Gait Classification Algorithm Using Convolutional Neural Network and Wavelet Coherence Spectrogram of Gait Synchronization. IEEE Access. 2022; 10: 38137–38153. doi: 10.1109/access.2022.3158961
19. Kumar GSC, Kumar RK, Kumar KPV, et al. Deep residual convolutional neural network: an efficient technique for intrusion detection system. Expert Systems with Applications. 2024; 238: 121912. doi: 10.1016/j.eswa.2023.121912
20. Wang HM, Zhu L. Research on forest pest identification method based on face image quality assessment. Journal of Changchun University of Technology. 2024; 45(1): 53–58.

21. Zhang W, Zhang M, Jin J Q, et al. Research on status identification of disconnector based on ResNet deep learning network. High Voltage Apparatus. 2024; 60(6): 100–106.
22. Cai TT, MaR. Theoretical foundations of t-sne for visualizing high-dimensional clustered data. Journal of Machine Learning Research. 2022; 23(301): 1–54.