Expert Systems With Applications 190 (2022) 116197

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Zero-shot learning for compound fault diagnosis of bearings

Juan Xu [a,b], Long Zhou [a,b], Weihua Zhao [c], Yuqi Fan [b], Xu Ding [b], Xiaohui Yuan [d,*]

[a] Key Laboratory of Knowledge Engineering with Big Data, Hefei University of Technology, Ministry of Education, China
[b] Hefei University of Technology, China
[c] Luoyang Bearing Research Institute Co. Ltd, China
[d] University of North Texas, Denton, TX, 76203, USA

ARTICLE INFO

Keywords:
Fault diagnosis
Zero-shot Learning
Compound fault diagnosis
Semantics

ABSTRACT

Due to the concurrency and coupling of various types of faults, and the number of possible fault modes grows exponentially, thereby compound fault diagnosis is a difficult problem in bearing fault diagnosis. The existing deep learning models can extract fault features when there are a large number of labeled compound fault samples. In industrial scenarios, collecting and labeling sufficient compound fault samples are unpractical. Using the model trained on single fault samples to identify unknown compound faults is challenging and innovative. To address this problem, we propose a Zero-shot Learning Compound Fault Diagnosis Model of bearings (ZLCFDM). We design an encoding method to express the semantics of single faults and compound faults according to the fault characteristics. A convolutional neural network is developed to extract the time–frequency features of the compound fault signal. Then we embed the semantic feature of the fault into the visual space of the fault data. The Euclidean distance is used to measure the distance between the signal features and the semantic features of the compound faults to identify the categories of unknown compound faults. To validate the proposed method, we conduct experiments on a self-built testbed. The results demonstrate that the accuracy of identifying compound fault reached 77.73% when the model was trained without any compound fault samples.

1. Introduction

Bearing is an indispensable basic component in industrial equipment. According to statistics, about 30% of the rotating machinery failures are caused by the damage of bearings (Sonal, Reddy, & Kumar, 2019; Zheng, 2021). Bearing fault diagnosis is important to ensure the safety of equipment and personnel (Deng, Wang, Tang, Huang, & Zhu, 2021; Zhao, Zhang, Zhan, & Pang, 2020).

In complex industrial equipment, compound faults refer to the occurrence of two or more interrelated and cross-influencing faults of mechanical equipment at the same time. The compound faults of bearings are due to the concurrency and coupling of various types of faults, and the number of possible fault modes grows exponentially (Xia, Mao, Zhang, Jiang, & Wei, 2020; Zhang, Li, Xin, & Ma, 2020). Therefore, compound faults of the bearing are common and difficult to extract and identify (Tang, Hu, Chen, Lin, & Chen, 2021).

Traditional compound fault diagnosis methods mainly include analytical model-based (Mhamdi, Dhouibi, Liouane, & Simeu-Abazi, 2013; Piacentino & Talamo, 2013), qualitative experience-based (Chatti, Merzouki, Ould-Bouamama, & Gehin, 2014; Ubar, Raik, Kostin, & Kousaar, 2012), and signal analysis-based (Chen, Peng, Wang, & Yu, 2020; Shao, Lin, Zhang, & Wei, 2020; Zhao, Cheng, Gao, Yan, & Wang, 2020) compound fault diagnosis. The analytical model establishes an accurate physical or mathematical model for some specific faults. The qualitative experience method uses incomplete prior knowledge to establish qualitative model reasoning for diagnosis. The signal analysis method identifies the feature of compound faults by analyzing the compound fault signal. These methods require prior knowledge, which makes it difficult to apply them in real-world industrial scenarios.

A deep neural network provides an end-to-end learning approach to map the original data to the expected output without prior knowledge (Fan, Shen, Xu, Xu, & Yuan, 2021; Fan, Shen, Yuan, & Xu, 2020; Liu et al., 2021). In recent years, deep neural network models have been well developed to identify faults and assist diagnosis (Hoang, Tran, Van, & Kang, 2021; Ma, Sun, & Chen, 2018; Shao, Jiang, Zhao, & Wang, 2017; Sun, Yan, & Wen, 2018). Most deep learning-based methods focus on single fault diagnosis, whereas the ones developed for compound fault diagnosis mainly attracted attention on designing different structures of deep neural networks (Huang, Li, & Cui, 2019;

* Corresponding author.
E-mail addresses: xujuan@hfut.edu.cn (J. Xu), zhou_long321@163.com (L. Zhou), zys_zhao@163.com (W. Zhao), yuqi.fan@hfut.edu.cn (Y. Fan), dingxu@hfut.edu.cn (X. Ding), xiaohui.yuan@unt.edu (X. Yuan).

https://doi.org/10.1016/j.eswa.2021.116197
Received 21 June 2021; Received in revised form 22 September 2021; Accepted 4 November 2021
Available online 17 November 2021

Lin, Han, Fan, & Li, 2020; Sun, Wang, Sun, & Jin, 2019). These models are trained from a large number of labeled examples to achieve satisfactory performance. The compound faults of bearings are due to the concurrency and coupling of various types of faults, and the number of possible fault modes grows exponentially. However, in actual working conditions, the number of fault modes of compound faults increases exponentially, and the characteristics of faults are diversified (Xia et al., 2020; Zhang, Li, et al., 2020). In addition, the compound fault samples are difficult to collect and labeled, which limits the application of the existing deep learning-based compound fault diagnosis methods.

A more practical strategy is to obtain labeled examples of single fault bearing that is easy to implement. The problem becomes using the examples of this fault data to recognize the unknown compound faults of bearings. Gao, Gao, Li, and Zheng (2020) applied zero-shot learning for fault diagnosis under unknown working loads, in which the fault classes of the training and test sets are the same, but the data distribution differs. Combined with the capsule networks and ensemble learning technique, Huang et al. propose a deep ensemble capsule network (DECN) for intelligent compound fault diagnosis, which effectively decouples the compound fault into two individual faults (Huang, Li, Li, & Cui, 2020). Xing, Lei, Wang, Lu, and Li (2022) proposed a label description space (LDS) embedded model for zero-shot compound fault diagnosis. The proposed method extracts feature of single fault examples using a locally connected restricted Boltzmann machine (LCRBM). When LDS is established, the dimension equals the number of single fault types. Generally, the LDS built with such a method results in a relatively low dimension, which limits its performance.

To address the problem of learning from single-fault data for classifying the unknown compound faults, this paper introduces a shared middle-layer embedding between the features and semantics. The proposed model achieves compound fault classification using a model that learns from single-fault examples to minimize the difference between fault features and signal-derived semantics. Unlike DECN what requires a number of sensors to collect vibration signals and a set of pre-training capsule network models, our method takes the signals acquired by one sensor and extracts distance features for model development, which greatly reduces the complexity of the model.

The main contributions of this paper can be summarized as follows:

1. We propose a Zero-shot Learning Compound Fault Diagnosis Model (ZLCFDM) that is trained with the vibration signals of single-fault to identify the unknown compound faults. It enables the compound fault diagnosis with no or extremely scarce examples.
2. We devise fault semantics to express the prior knowledge of the single-fault and compound fault. The vibration signals of the single fault are used to construct fault semantics, whereas the fault semantics of the compound faults are derived from the fault semantics of the single fault. This enables learning from single-fault examples and classifying the compound faults.
3. Our zero-shot learning method maps the fault semantics such that they have a matched dimensionality with the signal features. The training process of our method matches the signal features of the single-faults with their fault semantics. The trained model identifies the compound faults by computing the Euclidean distance between the signal features and the fault semantics.

The rest of the paper is organized as follows. Section 2 provides a review of related methods in two aspects: zero-shot learning based on the embedded model and attribute definition method of zero-shot learning, Section 3 presents the details of our proposed method. Section 4 discusses the experimental results. Section 5 concludes this paper with a summary.

2. Related work

2.1. Zero-shot learning based on embedded model

Zero-shot learning is a target classification technique to solve the problem of missing category labels, which can be classified when the training set and the testing set are disjoint (Verma, Arora, Mishra, & Rai, 2018; Zhang, Wang, Liu, et al., 2020; Zhang, Xiang, & Gong, 2017). The basic idea of zero-shot learning is to use some data of visible categories, supplemented by relevant common knowledge information or prior knowledge as attribute labels, to train a certain learning model and finally identify the data of unknown categories (Li, Wang, Hu, Lin, & Zhuang, 2017; Rahman, Khan, & Porikli, 2020; Zhang, Liu, Long, Zhang, & Shao, 2020).

The prevalent implementation approach is the embedded model-based method, which is used in our paper. The model embeds all the features and semantic attributes of category labels into a certain space and then classifies the samples according to the similarity measure. The existing published works can be divided into three directions: semantic space embedded model, public space embedded model, and visual space embedded model.

The semantic space-based method (Fu, Xiang, Kodirov, & Gong, 2015; Xu, Hospedales, & Gong, 2015) directly maps the features of the sample to the semantic space, finds the semantics closest to the features in the semantic space using similarity measurement, and then labels their corresponding category. Xu et al. (2015) used the semantic word vector space as a certain space to embed video and category labels, and established a mapping between the features of each category and the human-explainable semantic description, thus realizing the zero-shot classification of actions in the video. Fu et al. (2015) took into account the maintenance of manifold structure within semantic space, proposed to use a graph model to model the semantic manifold structure contained in semantic embedding space, and used an absorbing Markov chain to measure the distance in the manifold structure.

The public space-based method (Guo, Ding, Jin, & Wang, 2016; Yu, Ji, Li, et al., 2018) maps the features and semantics of samples into a subspace. In this subspace, the model can save the semantic and visual feature information of the original space as much as possible, and then classified according to the matching of semantics and features. Yu et al. (2018) proposed a direct push zero-sample image classification method, which used a self-training strategy to integrate testing samples into the framework of model learning to further improve the model performance. Guo et al. (2016) proposed a method to establish a direct connection between images and category labels. Based on learning category sharing model space, test categories were also included in the training stage for training models through the joint learning framework.

The visual space-based method (Changpinyo, Chao, Gong, & Fei, 2016) maps the semantics of the sample to visual space, which can solve the Hubness problem (Marco, Angeliki, & Georgiana, 2015) caused by semantic space. Shojaee and Baghshah (2016) proposed a semi-supervised learning method to map the semantics of the category into the visual feature space and make the sum of the features of the category in the visual space close to the semantics of the category using clustering. Changpinyo et al. (2016) proposed the method of using the synthetic classifier to classify zero samples. By introducing virtual class, the visible class and invisible class are connected and the invisible class is finally identified.

2.2. Attribute definition method of zero-shot learning

In the embedded model-based method, the definition of semantic attributes of category labels is the core problem. Semantic attributes are a kind of prior knowledge to represent the specific categories of objects. There are two general strategies for the definition of semantic attributes: attribute learning (Mikolov, Sutskever, Chen, Corrado,

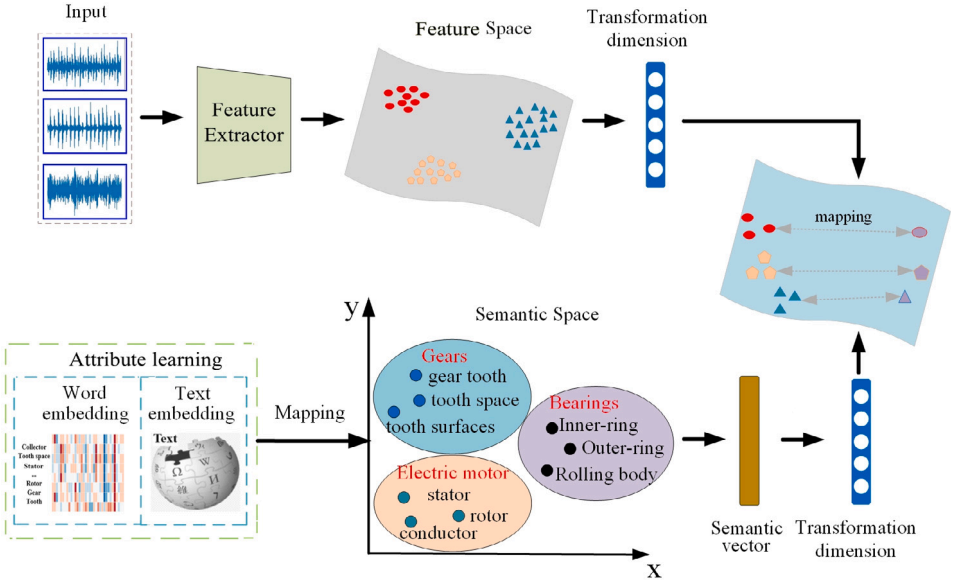


Fig. 1. The exiting attribute definition method of the fault semantics is given in (a),(b), which correspond to the Word embedding and the Text embedding. Due to the words learned from attribute learning, similar words have a close distance in semantic space, while dissimilar words have a far distance in semantic space, it is difficult to identify these words.

& Dean, 2013; Pennington, Socher, & Manning, 2014) and manual definition (Lampert, Nickisch, & Harmeling, 2014).

The attribute learning approaches mainly include word embedding and text embedding. Word embedding methods use the natural language processing models, e.g., Word2Vec and GloVe to embed category label names into a value space to obtain their vector metric representation. The text embedding methods describe the text about the categories and then convert the text description into the fault semantics of the categories. The manual definition approaches encode some attributes of the categories through the features of the data to obtain fault semantics. Thereby, they are only suitable for image data because the features of the image data, e.g., color, position, and habit, can be visually seen.

Although the two existing attribute definition methods have achieved favorable results in image data, however, they cannot be well adapted to the fault category attributes of the vibration data in the field of fault diagnosis. The main reasons are as follows:

1. If we define fault semantics using manually defined methods, we generally use fault frequency, peak value, peak–peak value, etc. to describe fault features, but the types of the features are too few to describe the fault category attributes, so the manually defined method cannot accurately represent the semantic attribute of different faults.
2. If we define fault semantics using attribute learning methods, since similar words are so close in semantic space, as shown in Fig. 1, fault semantics are hard to distinguish, especially in the fault field, the words of inner ring, outer ring, and rolling body faults, are very adjacent in semantic space. It is difficult for the model to find the fault semantics closest to the visual feature, thus the classification ability of the model will be seriously degraded.

In summary, the aforementioned methods have obvious deficiencies in the category labels applied to compound fault diagnosis. Thereby it is urgent to explore an effective fault semantics definition method in fault diagnosis.

3. Zero-shot learning compound fault diagnosis model

3.1. Problem statement

Without loss of generality, given a labeled training dataset of vibration signal of single faults, the dataset contains K classes and N labeled samples, denoted as $D_s = \{x_i, y_i, S_i\}_{i=1}^{N}$, where x_i, y_i is the ith training sample and the corresponding category label, y_i is Y_s in the category label and Y_s is the label space for the category, S_i is the fault semantics of the ith seen category sample with dimension $R \times 1$.

There is also an unlabeled testing dataset of vibration signal of compound faults, including L classes and M unlabeled samples, denoted as $D_u = \{\hat{x}_i, \hat{S}_i\}_{i=1}^{M}$, where $\hat{x}_i$ is the ith testing sample, $\hat{y}_i$ is the label of the testing set, $\hat{S}_i$ is the fault semantics of the ith category sample with dimension $R \times 1$, $\hat{y}_i \in Y^u$, and Y_u is the label space for the category. The dataset and category sets satisfy the following conditions:

$$\begin{cases} I(p(D_s); p(D_u)) = 0 \\ Y_s \cap Y_u = \emptyset \\ \varphi(S) = \hat{S} \end{cases} \tag{1}$$

where $I(\cdot)$ computes the mutual information of two distributions $p(D_s)$ and $p(D_u)$. That is, the data distributions of D_s and D_u are different, the category sets Y_s and Y_u are disjoint, and compound semantics $\hat{S}$ is obtained from single semantics S by function $\varphi(\cdot)$.

By training sample x_i and the corresponding fault semantics S_i, our model $F(x, S; \theta)$ learns the mapping relationship between x_i and S_i.

$$F(x, S; \theta) : (x_i, S_i) \rightarrow y_i, \tag{2}$$

where x, $x_i \in D_s$ and θ denotes model parameters that minimize the difference between the signal and its fault semantics:

$$\theta = \arg\min_{w} \sum_{i=1}^{N} D(x_i, S_i),$$

where $D(\cdot)$ is the distance between x_i and S_i.

Finally, given a testing sample $\hat{x}_i$ and fault semantics $\hat{S}_i$ of unknown category, the model predicts it class label $\hat{y}_i$.

$$F(x, S; \theta) : (\hat{x}_i, \hat{S}_i) \rightarrow \hat{y}_i, \tag{3}$$

where x, $\hat{x}_i \in D_u$.

3.2. Data pre-processing

Wavelet transform is used to transform one-dimensional time-domain signals into two-dimensional time–frequency domain images.

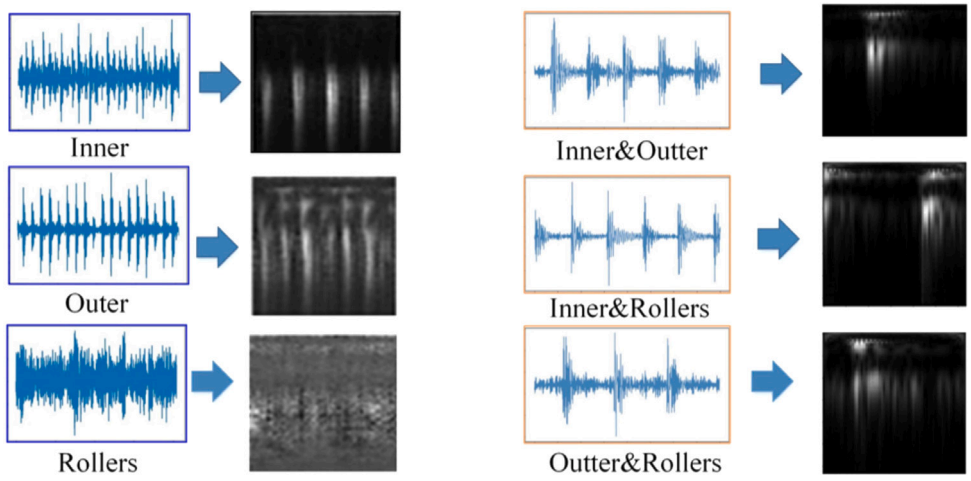


Fig. 2. Transformation the one-dimensional time-domain signals to two-dimensional images.

It can be used for multi-scale analysis of vibration signals. Wavelet transform is computed as follows:

$$WT(a,t) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f(t) * \psi(\frac{t-\tau}{a})dt, \tag{4}$$

where a is the scaling factor, which is used for scaling the wavelet function, τ is the translation factor, which is used to control the translation of wavelet function, $f(t)$ is the vibration signal of fault, and $\psi(t)$ is wavelet function given as follows:

$$\psi(t) = e^{(-t^2/2)} cos(5t), \tag{5}$$

Fig. 2 illustrates six examples of the one-dimensional time-domain vibration signals, each of which contains 256 data points and the wavelet transform results. The center rate is 0.8125. The wavelet coefficients in the time–frequency domain are reformatted as a 64×64 matrix.

3.3. Model structure

To overcome the problem that compound fault samples are difficult to be collected and labeled, we propose a Zero-shot Learning Compound Fault Diagnosis Model of bearings (ZLCFDM) to identify compound fault by training a model from single fault data of bearings. As shown in Fig. 3, the model consists of three parts: a feature extraction model, a semantic processing module, and a semantic embedding module.

Feature extraction module: Since the fault features extracted by CNN have favorable discrimination, we choose the CNN as the feature extractor. The network consists of normalized convolutional layers, pooling layers, flatten layers, and normalized fully connected layers. For convenience, we use C, P, F, and FC to denote normalized convolutional layers, pooling layers, flatten layers, and normalized fully connected layers, respectively. The input of the feature extractor is the 2D time–frequency matrix. The output of the last fully connected layer is the feature vector. The architecture of our CNN is shown in Table 1.

Semantic processing module: The existing semantic definition cannot be directly applied to fault diagnosis. Hence, we propose a novel semantic descriptor for fault categories and use the vibration feature of the original signals as the semantic of the fault attribute. We select the R data points of the vibration signal of the single fault, e.g., the vibration signal of the inner ring fault:

$$f = (v_1, v_2, \dots, v_k, \dots, v_R), \tag{6}$$

where the R is much greater than one period of the vibration signal. We assume that any single fault semantic, e.g., the semantic of inner ring fault is:

$$S_i = (a_1, a_2, \dots, a_k, \dots, a_R). \tag{7}$$

Table 1
Architecture of CNN.

Notation	Description	Kernel Size	Stride	Kernel Number
Input	Input signal	64×64	/	/
C1	Convolution	5×5	1×1	32
P1	Pooling	2×2	2×2	32
C2	Convolution	5×5	1×1	64
P2	Pooling	2×2	2×2	64
F	Flatten	16384×1	/	1
FC1	Fully-connected	2048×1	/	1
FC2	Fully-connected	2048×1	/	1

The threshold λ_i represents the threshold value of the vibration signal of single fault f. If the dimension v_k of signal f is greater than λ_i, the dimension a_k is set to 1, otherwise a_k is set to 0, then we obtain the R dimension semantic S_i for a single fault:

$$a_k = \begin{cases} 1, & v_k \geq \lambda_i \\ 0, & v_k < \lambda_i \end{cases} \tag{8}$$

where λ_i is the threshold for the vibration signal of a fault category and is computed as follows:

$$\lambda_i = \frac{1}{a} \max f, \tag{9}$$

where a is the empirically determined hyperparameter and f is the vibration signal. The choice of a should retain the characteristics of the vibration signal of the single-fault category i.

We obtain the semantic of single fault $S = \{S_1, S_2, \dots, S_N\}$, where $S_i \in S$, and N is the number of the single fault samples. The fault semantics of the compound fault $\hat{S}_i$ is obtained by logical or operation of the semantic of the single fault that constitutes the compound fault:

$$\hat{S}_i = S_1 \parallel S_2 \parallel \dots \parallel S_i, \tag{10}$$

where $\hat{S}_i \in \hat{S}$, $S_1, S_2, \dots, S_i \in S$, and $S_1, S_2, \dots, S_i$ are the single fault semantics that make up the compound fault semantics $\hat{S}_n$. The compound fault semantics is $\hat{S} = \{\hat{S}_1, \hat{S}_2, \dots, \hat{S}_M\}$, where M is the number of the compound fault samples.

Semantic embedding module: Semantic embedding module embeds fault semantics into features to match the fault semantics with features, which is achieved with the full connection of the two layers. The input of the embedding layer is fault semantics.

Specifically, our model establishes a mapping relationship $F(f(x), S; \theta)$, where $f(x)$ is the visual feature vector extracted by CNN, and the optimization goal of the model is to minimize the parameter θ to match the fault semantics and visual feature. After the model

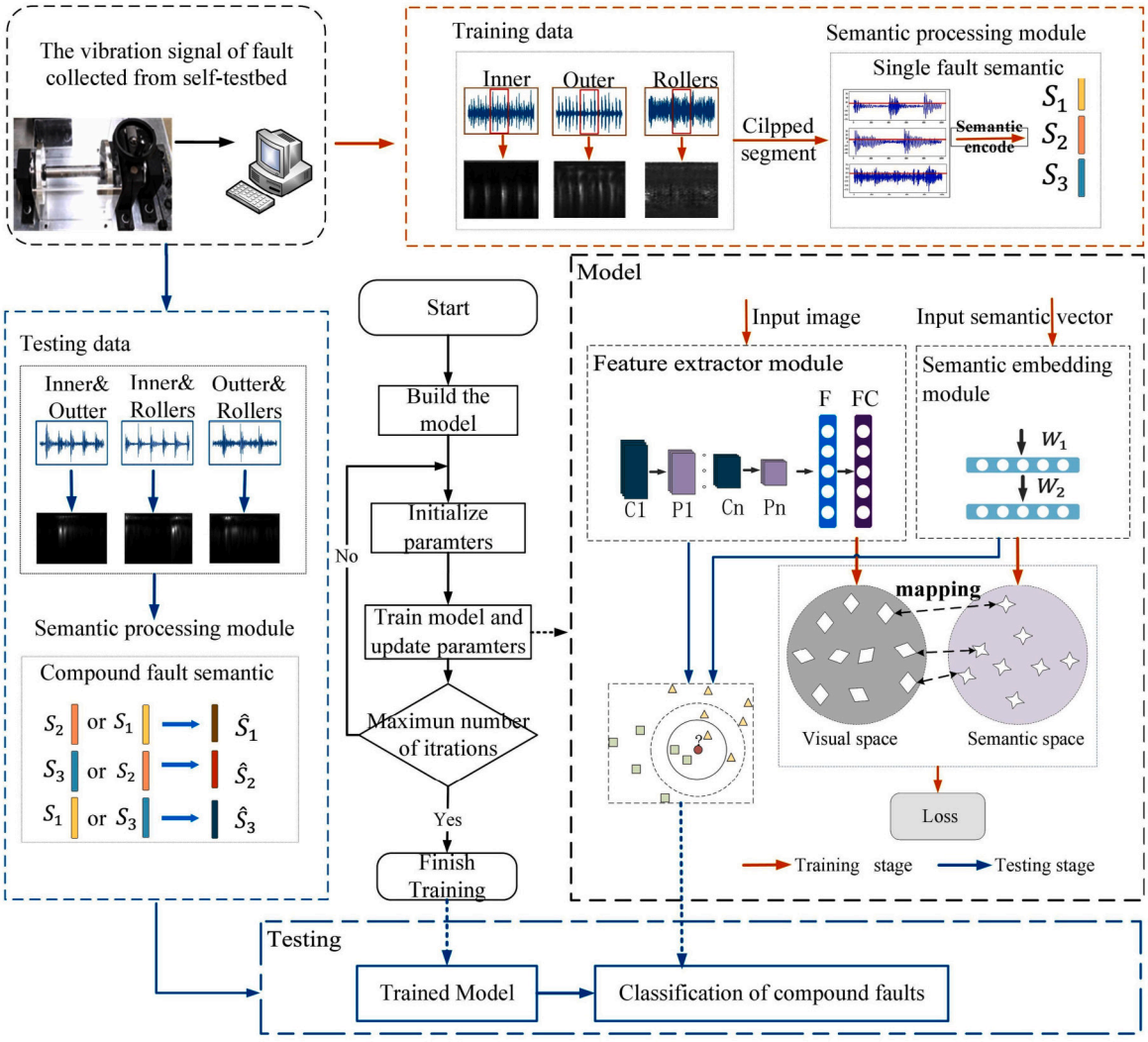


Fig. 3. Flow chart of the proposed method.

establishes the matching relationship, we input the testing sample and fault semantics into the model to find the nearest matching relationship as follows:

$$i = \arg \min_{W_1, W_2 \in \theta} D(f(\hat{x}_i), W_2(W_1(\hat{S}))), \tag{11}$$

where $\hat{S}_i$ is the fault semantic of the ith compound fault sample, and W_1 and W_2 are the two-layer weight vectors of the embedded layer, and i is the ith of compound fault category. We employ Euclidean distance to measure the nearest matching relationship as follows:

$$D(f(\hat{x}_i), W_2(W_1(\hat{S}_i))) = \sqrt{\sum_{i=1}^{M} (f(\hat{x}_i) - W_2(W_1(\hat{S}_i)))^2}, \tag{12}$$

where $f(\hat{x}_i)$ is the feature vector of compound fault $\hat{x}_i$, $\hat{S}_i$ is a semantic of compound fault, and $W_2(W_1(\hat{S}_i))$ denotes the projection of $\hat{S}_i$ via two fully connected layers of the network to match the dimension with $f(\hat{x}_i)$. M is the number of the compound fault samples. A KNN based on Euclidean distance is used to measure the correlation between features and fault semantics.

3.4. Optimized objective function

The loss function of the model is composed of two parts: the classification loss L_1 and the embedding loss L_2. The classification loss L_1 can be expressed as:

$$L_1 = -\sum_{i=1}^{N} (y_i \log p_i + (1 - y_i) \log(1 - p_i)), \tag{13}$$

where y_i is the real label of the sample and p_i is the predicted value of the model.

We embed the semantic feature of fault categories into the visual space in the two full-connection layers. The embedding loss function L_2 is designed to constantly optimize the parameters of the model to make the fault semantics better express the features:

$$L_2(W_1, W_2) = \frac{1}{N} \sum_{i=1}^{N} \|f(x_i) - W_2(W_1(S_i))\|^2 + \alpha(\|W_1\|^2 + \|W_2\|^2), \tag{14}$$

where W_1 and W_2 are the parameters of the two-layer full connection layer, $f(x_i)$ is the feature vector of single fault extracted by the feature extractor, S_i is the fault semantic of single fault corresponding to the ith sample, and α is a hyperparameter.

3.5. Model training and application

The algorithm of our proposed method is shown in Table 2. In the training stage, the examples of the single fault data are transformed using wavelet transform to obtain 2D time–frequency images, from which features are extracted. A number of consecutive data points are

Table 2
ZLCFDM Algorithm.

Require: The training samples and labels; Number of fault samples N; Number of iterations I_1, I_2 , μ, β_1, β_2, $\epsilon = 10^{-8}$, α, λ;
Trained convolutional neural network
Randomly initialize φ
for I_1 epochs **do**
Draw training samples $\{(x^{(1)}, y^{(1)}) \cdots (x^{(n)}, y^{(n)})\}$
Compute adapted parameters with Adam
Update:
$\nabla_\varphi L_1 \leftarrow \nabla_\varphi \sum_{i=1}^{N} \left(y_i \log p_i + (1 - y_i) \log (1 - p_i)\right)$
$\varphi \leftarrow \text{Adam}\left(\nabla_\varphi L_1, \mu, \beta_1, \beta_2, \epsilon = 10^{-8}\right)$
end for

fault semantics processing
Single fault semantics processing:
Select the original data fragment in R dimension, $S_i = (a_1, a_2, \ldots, a_k, \ldots, a_R)$
for i to N:
for k to R:
$v_k > \lambda_i ? a_k = 1 : a_k = 0$ #λ_i is threshold of ith single fault
end for
end for
Compound fault semantics processing:
Single fault semantics logic or processing: $\hat{S}_i = S_1 \| S_2 \| \ldots \| S_i$

Training the Embedding
randomly initialize W_1, W_2
for I_2 epochs **do**
extracted features $f(x_i)$
Update:
$\nabla_{W_1,W_2} L_2 \leftarrow \nabla_{W_1,W_2} \sum_{i=1}^{N} (\left\| f(x_i) - W_2(W_1(S_i)) \right\|^2 + \alpha(\|W_1\|^2 + \|W_2\|^2))$
$W_1, W_2 \leftarrow \text{Adam}\left(\nabla_{W_1,W_2} L_2, \alpha, \mu, \beta_1, \beta_2, \epsilon = 10^{-8}\right)$
end for

Identify compound fault categories
Input compound data $\hat{x}_i$ and fault semantics $\hat{S}_i$
$i = \arg\min_{W_1,W_2} D(f(\hat{x}_i), W_2(W_1(\hat{S})))$
End

selected and coded through the semantic processing module to obtain a vector. The semantic embedding module embeds this fault semantics into the visual space. Finally, the trained model makes the features match the fault semantics.

In the testing stage, the compound fault data of the testing samples are transformed into time–frequency images using wavelet transform. Then the feature extractor extracted the features of the time–frequency images. By logical OR operation of the semantic of the single fault, we obtain compound fault semantic. The fault categories of the sample are identified by calculating the Euclidean distance between the visual feature and the fault semantics.

4. Experiment results and discussion

4.1. Experiment description

We use the self-built testbed to collect vibration signals of the bearing faults to evaluate the effectiveness of our proposed method and conduct comparison studies. The testbed is shown in Fig. 4. The speed of the bearing is controlled by the three-phase motor through flexible coupling. The acceleration sensor is installed on the bearing seat to collect vibration signals, and the sampling frequency is 51,200 Hz.

Our dataset contains three types of single faults: rolling body fault (BF), inner ring fault (IF), and outer ring fault (OF), and four types of compound faults: inner ring & outer ring compound fault (IF&OF), inner ring & rolling body fault (IF&BF), outer ring & rolling body fault (OF&BF), and inner ring & outer ring & rolling body compound fault (IF&OF&BF). Examples of the vibration signals of these seven fault cases are shown in Fig. 5. The signals of different single-fault bearings are significantly different, whereas the compound faults, which are the integration of difference sing-faults, are much complex. If we train a neural network with the vibration signals of single-fault and apply the trained model to identify the compound faults, the samples of which are not used in the training phase, the permeance of the model will extremely aggravate or complete fail.

To evaluate our method, we design two sets of compound fault diagnosis tasks. Table 3 presents the diagnosis tasks with different training and testing set sizes. The training sets of Task A and Task B are composed of single-fault examples of faults IF, OF, and BF. There is not compound fault examples in the training sets. The testing sets of Task A includes three compound faults: IF&OF, IF&BF, OF&BF and the testing sets of Task B include four compound faults: IF&OF, IF&BF, OF&BF, and IF&OF&BF. The training examples are randomly selected from the pool of vibration signals collected from our testbed. The size of the testing dataset is 2,000 for all cases that include signals of compound faults.

In our experiments, we use different values for a in the computation of the threshold, Eq. (9), for generating fault semantics. To understand the impact of a and decide the best option, we conduct experiments using different values for a in the range of $[3, 7]$. Fig. 6 shows the overall accuracy with respect to the choice of a. Without loss of generality, we conducted experiments with Task A_4 and Task B_4 that include the largest number of training examples of single-fault. By changing the values of a, the overall accuracy varies in a small range (about 7%) for both tasks. The overall accuracy reaches the maximum when a equals 5. Hence, in the rest of our experiments, we set the value of a to five, i.e., we have $\lambda_i = 1/5 \max f$ as the threshold for vibration signal f.

4.2. Evaluation of fault detection

Fig. 7 illustrates the accuracy of fault detection of Tasks A and B. As the number of training samples increases, the compound fault diagnosis accuracy of Task A and Task B increases. It reaches a maximum of 77.73% on Task A and 54.59% on Task B when the training samples of each fault category is 2000. The accuracy of image classification in zero samples is also about 50% to 60%, so the accuracy of task B is also relatively ideal. Due to the testing set of task B contains the compound fault data of the inner ring, outer ring and rolling body, the coupling of the three kinds of faults is more complex, so the accuracy of Task A is higher than that of Task B.

To better illustrate the effectiveness of the proposed method, we visualize the results of Task A_4 in Fig. 8. Fig. 8(a) shows the dimension reduction results of the original training set and testing set using the Principal Component Analysis. The different colors represent the different categories of fault data. The signals of faults are heavily overlapping and can hardly be classified. We train the feature extractor using the training samples and extract fault features of training data and testing data. Using t-distributed Stochastic Neighbor Embedding (t-SNE), we get the dimension reduction results of different fault features, as shown in Fig. 8(b). The single fault features on the training set are quite distinct and are distributed far apart in the space. Most of the compound fault features on the testing set are also separated.

Fig. 8(c) shows that the final classification results of the testing data using t-SNE. We can see that the compound fault samples are mostly separated from each other, which indicates that the features extracted by our method provide greater discriminant ability of the compound faults.

Fig. 9 illustrates the average detection accuracy of our method for compound faults at different operating conditions by varying loads and operating speeds. The training data consists of vibration signals of single fault including BF, IF, and OF on the load 0HP (Horsepower) and speed at 1500rpm. The trained model is applied to classify vibration signals of compound faults including IF&OF, IF&BF, and OF&BF at four different operating conditions: load 0HP and speed at 1200rpm, the load 0HP and speed at 1350rpm, the load 10HP and speed at 1500rpm,

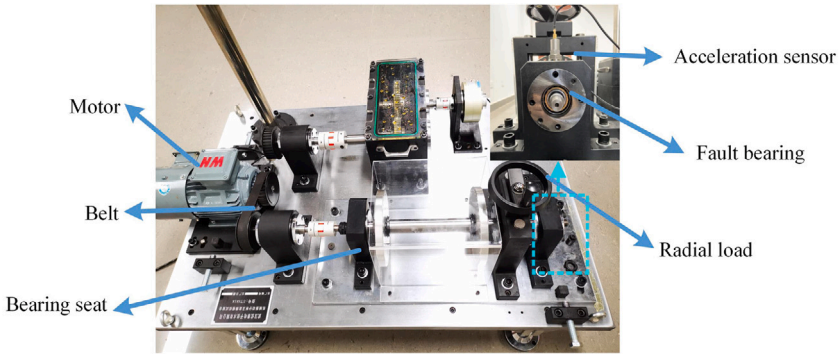


Fig. 4. Our bearing testbed.

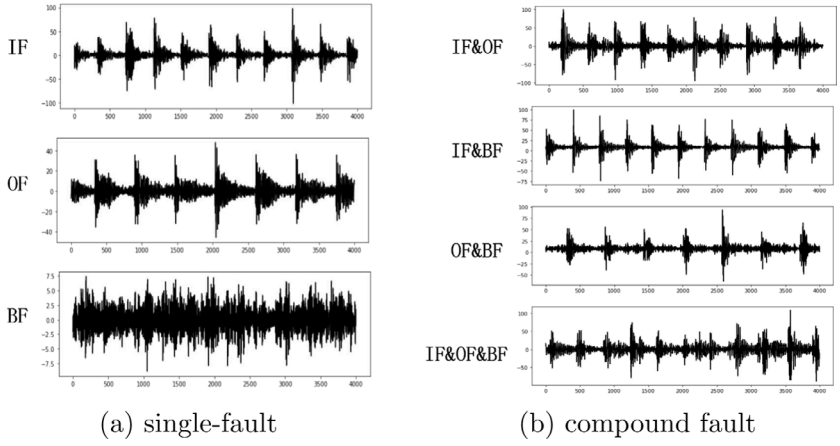


Fig. 5. Examples of vibration signals of different faults. The signals of compound fault are mixtures of two or more single-fault signals.

Table 3
Diagnosis tasks and the properties of training and testing sets.

	Fault categories (Training set)	Size	Fault categories (Testing set)	Size
Task A_1	IF/OF/BF	500	IF&OF / IF&BF / OF&BF	2,000
Task A_2	IF/OF/BF	1,000	IF&OF / IF&BF / OF&BF	2,000
Task A_3	IF/OF/BF	1,500	IF&OF / IF&BF / OF&BF	2,000
Task A_4	IF/OF/BF	2,000	IF&OF / IF&BF / OF&BF	2,000
Task B_1	IF/OF/BF	500	IF&OF / IF&BF / OF&BF / IF&OF&BF	2,000
Task B_2	IF/OF/BF	1,000	IF&OF / IF&BF / OF&BF / IF&OF&BF	2,000
Task B_3	IF/OF/BF	1,500	IF&OF / IF&BF / OF&BF / IF&OF&BF	2,000
Task B_4	IF/OF/BF	2,000	IF&OF / IF&BF / OF&BF / IF&OF&BF	2,000

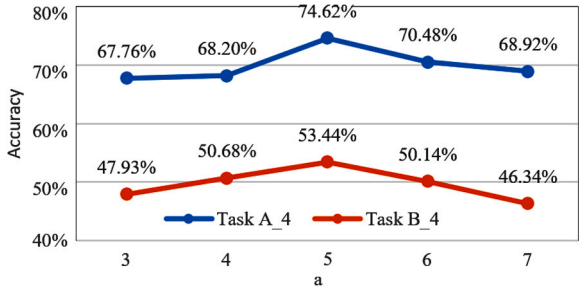
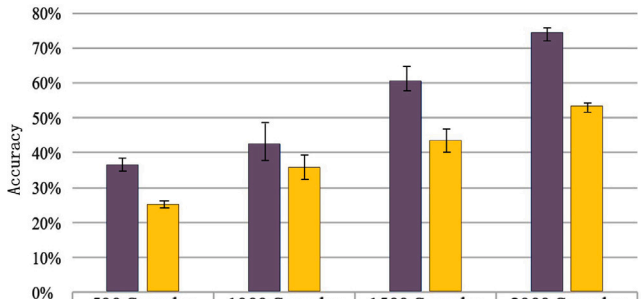


Fig. 6. Accuracy with respect to different choices of *a*.



	500 Samples	1000 Samples	1500 Samples	2000 Samples
Task A	36.35± 3.02	42.5± 9.08	60.6± 6.6	74.62± 3.11
Task B	25.24± 0.28	35.78± 9.45	43.36± 8.26	53.44± 1.15

Fig. 7. Results of Task A and Task B.

the load 20HP and speed at 1500rpm. For each case, we repeat the experiment five times and report the average performance.

The accuracy is the greatest (74.62%) when our method is applied to the data collected under the same operating conditions as that of the

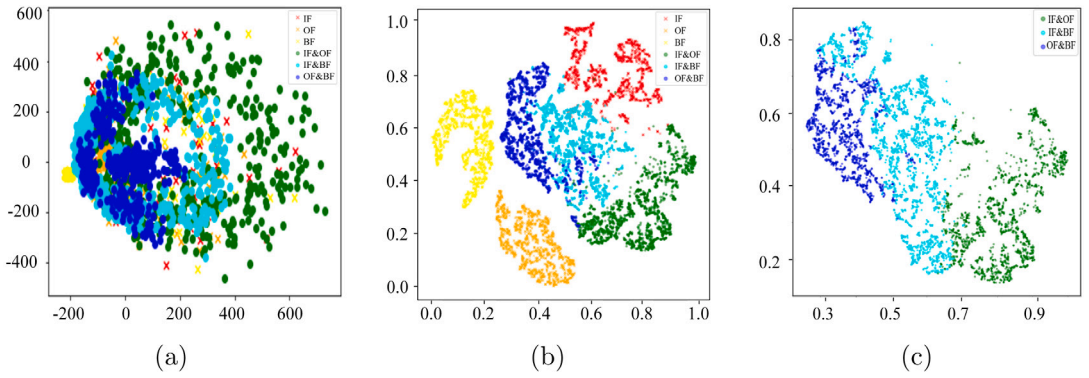


Fig. 8. Visualization of the results of Task A. (a) original distribution of the data; (b) the distribution of high-dimensional fault features; (c) the results of our method.

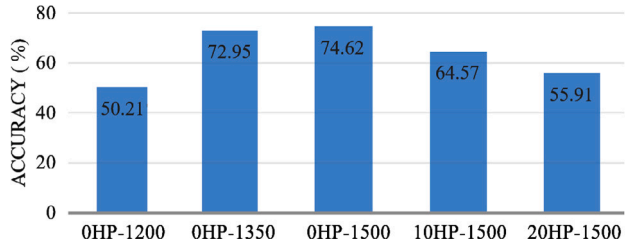


Fig. 9. Average accuracy of our method for faults at different operating conditions.

training data, i.e., 0HP and 1500rpm. When the operating condition is changed by either reducing the operating speed or increasing the load, the performance drops. This demonstrates that the different operating conditions change the similarity distribution of the dataset, which degrades the generalization ability of the model.

4.3. Analysis of semantics strategies

The first experiment is called as Word Embedding-based method. We use the word vector model, i.e., Word2Vec to train the fault corpus and then use the label of the fault as the fault semantics to embed into the visual space. The second experiment is called as Text Embedding-based method. The description sentences of fault features are constructed and the corresponding word vectors of the faults are generated through the skip-gram model, then the fault vectors are embedded into visual space. The third experiment did not use fault semantics. The trained CNN extract fault features of the training set and testing set, and we measure the distance between the feature of the testing sample and the category center of the features of the training set. The two closest category features are used as the prediction labels of the testing sample. We refer to it as CNN_D in our discussions.

For each method, we perform four experiments with different training samples, the settings of the training set and testing set are the same as Table 3. For each series of experiments, we repeat the experiment five times and report the average performance. The experimental results are shown in Fig. 10. For the Word Embedding and Text Embedding methods, although the number of training samples for each category increased from 500 to 2000, the classification accuracy of compound fault is still around 33.33%. Thereby the two methods could not distinguish compound fault categories. This is because the fault semantics generated by Word2Vec or text messages are very close in high dimensional space. When using KNN to measure the distance between the compound fault samples and the fault semantics, due to the fault semantics are too similar, the model is unable to identify the compound fault category.

Meanwhile, with the increase of the number of training samples, the classification accuracy of the CNN_D method increases from 46.87% to 59.52%. The CNN is trained by the single fault samples and extracts the features of compound fault samples. The results showed that the compound fault features have a certain distinction. Through measure the distance between the compound fault features and the center of single fault features, the model can get an improved classification accuracy.

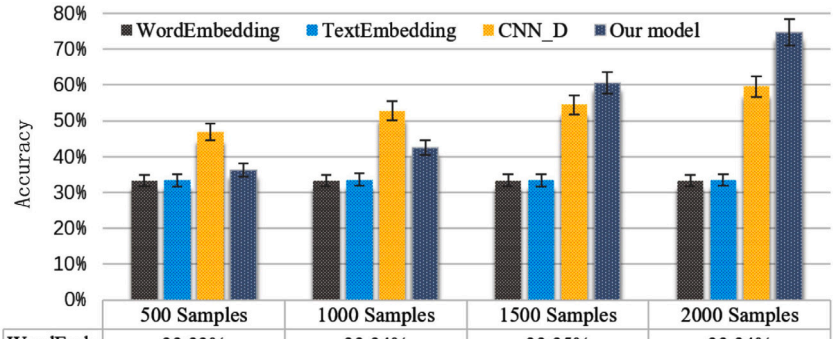
Besides the classification accuracy of our method increases from 36.35% to 74.62%. When the training sample of each category is 2,000, The classification accuracy of our method is 15% higher than the CNN_D method. It demonstrates that our semantic definition method can solve the similarity problem of different fault semantics. As the fault semantics of each compound fault distinguishes obviously, the fault semantics can be felicitously embedded in features for compound fault classification.

Fig. 11 illustrates the loss and accuracy of the compared methods. Fig. 11(a) depicts the embedding loss in the training phase. Since the third group of the experiment does not involve fault semantics, there is no embedding loss. As the number of iterations increases, the loss function reduces and eventually converges. All methods converges around 4,000 iterations. Fig. 11(b) shows the fault classification accuracy of Task A_4. In most cases, the accuracy improves as the number of iterations increases. The accuracy of our method reaches a plateau about 3,000 iterations and the accuracy is improved by more than 40% compared to the second best. The fault semantics constructed using our method is more accurate than the compared methods.

Fig. 12 shows the confusion matrices for Tasks A_4 and B_4 using CNN_D and our method. The labels for the rows are the ground truth and the labels for the columns are the predicted classes. The cells without a number indicate zero. Fig. 12(a) and (c) depict the results of CNN_D for Task A_4 and B_4, respectively. In the classification of three compound faults, CNN_D exhibited poor performance in differentiating fault types IF&BF and OF&BF with a precision of 0.44 and 0.48, respectively. In the classification of four compound faults, i.e., Task B_4, CNN_D failed to detect fault IF&OF&BF, which is mostly confused with fault IF&OF. In contrast, in the classification of three compound faults shown in Fig. 12(b), our method achieved a much-improved performance with a minimum single category precision at 0.72. The improvement rates of our method for classifying IF&BF and OF&BF are 50% and 79.5%, respectively. A similar pattern can be observed in the classification of four compound faults. This demonstrates that our semantic definition method is effective in compound fault diagnosis.

4.4. Results of different models

We compare the performance of our proposed ZLCFDM method with other two zero-shot learning methods FGN (Chen, Pan, Zhou, & He, 2019) and CADA-VAE (Edgar, Sayna, Samarth, Trevor, & Zeynep, 2019). Note that these two methods were developed for image recognition. Hence, the semantics of images are irrelevant to vibration signals



	500 Samples	1000 Samples	1500 Samples	2000 Samples
WordEmb	33.33%	33.34%	33.35%	33.34%
TextEmb	33.39%	33.57%	33.40%	33.54%
CNN_D	**46.87%**	**52.77%**	54.43%	59.52%
Our model	36.35%	42.50%	**60.60%**	**74.62%**

Fig. 10. Classification accuracy using different semantics for Task A_4.

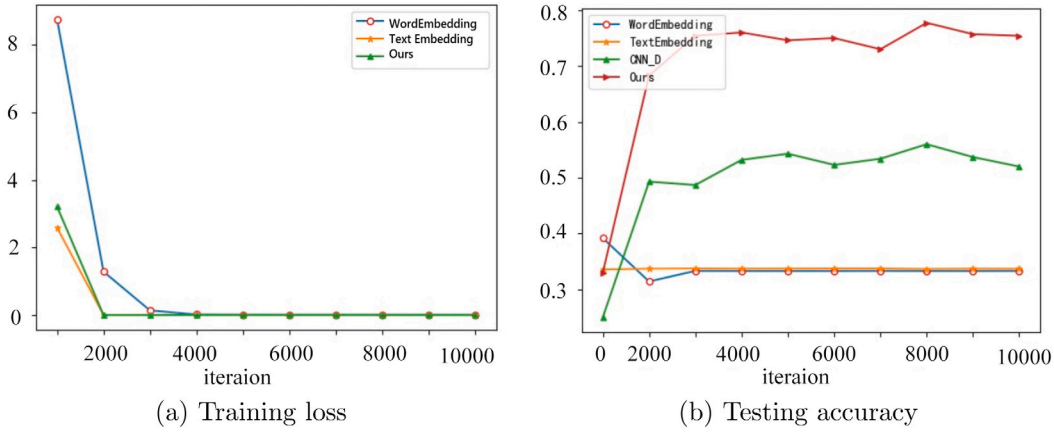


Fig. 11. Training loss and testing accuracy on Task A_4.

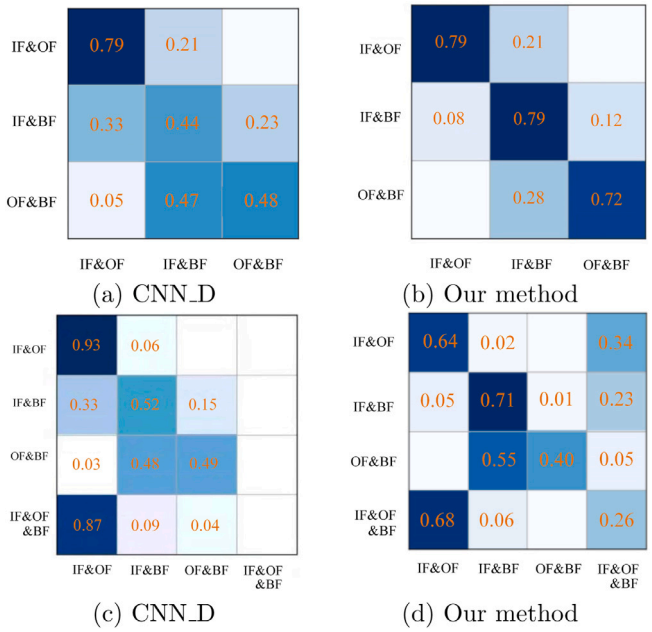


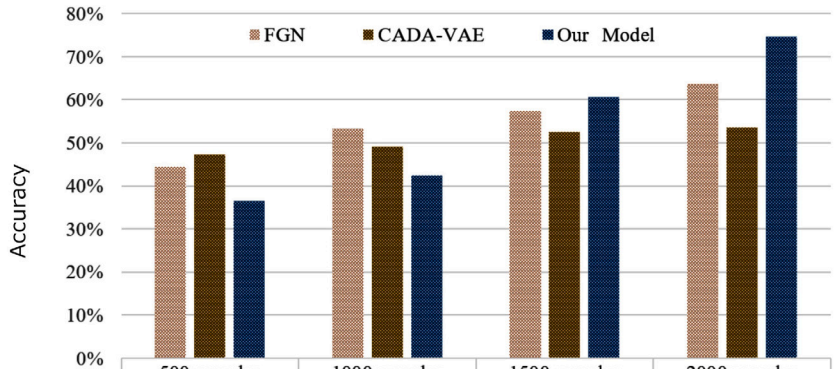
Fig. 12. Confusion matrix of CNN_D and our method on (a) and (b): Task A_4 and (c) and (d): Task B_4.

processed in this work. To realize the task of fault classification using these two methods, we replace the generation of image semantics with the fault semantics used in our proposed method and keep the rest unchanged.

Fig. 13 illustrates the average classification accuracy of the three methods using different number of training examples. With a small number of training examples, e.g., 500 or 1,000, all methods demonstrate poor performance, mostly below 50%. Both FGN and CADA-VAE are contrastive models, which designs the generation networks to derive fault features for unknown classes using semantic attributes. These two methods are advantageous in the circumstances of small training set. As we increase the size of the training set, the performance of all methods improves. The increase of accuracy of our method is the greatest. When 1,500 and 2,000 training examples are used, our method demonstrated superior performance in contrast to FGN and CADA-VAE. The best average accuracy achieved by our method is 74.62%, which is 14.5% improvement from the second best case by FGN.

5. Conclusion

To solve the problem that compound fault data are difficult to collect and label in practical industrial scenarios, we proposed a zero-shot learning compound fault diagnosis model, which uses single fault samples to predict compound fault categories. The high-dimensional features of the fault vibration data are extracted by two-layer CNN, and the fault semantics feature is designed to express the fault category information. We introduce a shared middle-layer semantic embedding space between the feature space and the fault semantics space. The



	500 samples	1000 samples	1500 samples	2000 samples
FGN	44.38%	**53.30%**	57.35%	63.75%
CADA-VAE	**47.33%**	49.02%	52.57%	53.55%
Ours Model	36.50%	42.50%	**60.60%**	**74.62%**

Fig. 13. Classification accuracy of different methods in Task A_4.

proposed model classifies the compound fault samples according to the similarity measure between the compound fault features and compound fault semantics. The experimental results show that even if there are no labeled compound fault samples, the model has a satisfactory classification accuracy. In the case of 2000 single fault samples of each category, the classification accuracy of our model is 15% higher than the CNN_D method, 41% higher than that of Word Embedding and Text Embedding methods, 11% higher than that of the FGN model, and 21% higher than that of CADA-VAE model.

However, in Task B, the testing set contains the compound fault of the inner ring, outer ring, and rolling body (IF&OF&BF). It is coupled with three kinds of single faults, and its fault characteristics are more difficult to identify. Hence the accuracy of task A is about 20% higher than that of task B, it suggests that our model can hardly distinguish the combination of the three faults. In our future work, we will explore different network structures of CNNs for improved performance and recognizing complicated compound faults in various scenarios.

CRediT authorship contribution statement

Juan Xu: Conceptualization, Methodology, Writing – original draft, Writing – review & editing. **Long Zhou:** Software, Writing – original draft, Writing – review & editing. **Weihua Zhao:** Software, Writing – original draft, Formal analysis. **Yuqi Fan:** Project administration, Resources. **Xu Ding:** Validation, Resources. **Xiaohui Yuan:** Conceptualization, Formal analysis, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported in part by the Key Research and Development Plan of Anhui Province under Grant 202104a04020003, in part by the National Nature Science Foundation of China under Grant 61806067, and in part by the Fundamental Research Funds for the Central Universities under Grant PA2021KCPY0045.

References

Changpinyo, S., Chao, W., Gong, B. Q., & Fei, S. (2016). Synthesized classifiers for zero-shot learning. In *Proceedings of IEEE conference on computer vision and pattern recognition*. (pp. 5327–5336). Las Vegas, NV.

Chatti, N., Merzouki, R., Ould-Bouamama, B., & Gehin, A. L. (2014). Signed bond graph for multiple faults diagnosis. *Engineering Applications of Artificial Intelligence, 36*, 134–147.

Chen, J. L., Pan, T. Y., Zhou, Z. T., & He, S. L. (2019). An adversarial learning framework for zero-shot fault recognition of mechanical systems. In *IEEE international conference on industrial informatics. vol. 2019*. pp. 1275–1278.

Chen, B., Peng, F., Wang, H., & Yu, Y. (2020). Compound fault identification of rolling element bearing based on adaptive resonant frequency band extraction. *Mechanism and Machine Theory, 154*.

Deng, Z. W., Wang, Z. Y., Tang, Z. H., Huang, K. K., & Zhu, H. Q. (2021). A deep transfer learning method based on stacked autoencoder for cross-domain fault diagnosis. *Applied Mathematics and Computation, 408*.

Edgar, S., Sayna, E., Samarth, S., Trevor, D., & Zeynep, A. (2019). Generalized zero-and few-shot learning via aligned variational autoencoders. In *Proceedings of the IEEE computer society conference on computer vision and pattern recognition. vol. 2019*.

Fan, Y., Shen, G., Xu, X., Xu, J., & Yuan, X. (2021). Flight track pattern recognition based on few labeled data with outliers. *Journal of Electronic Imaging, 30*(3), 031204(1-18).

Fan, Y., Shen, G., Yuan, X., & Xu, J. (2020). Target track recognition from few-labeled radar data with outliers. In *International conference on urban intelligence and applications*. (pp. 206–214).

Fu, Z., Xiang, T., Kodirov, E., & Gong, S. (2015). Zero-shot object recognition by semantic manifold distance. In *Proceedings of the IEEE computer society conference on computer vision and pattern recognition*. (pp. 2635–2644).

Gao, Y. P., Gao, L., Li, X. Y., & Zheng, Y. W. (2020). A zero-shot learning method for fault diagnosis under unknown working loads. *Journal of Intelligent Manufacturing, 31*(4), 899–909.

Guo, Y., Ding, G., Jin, X., & Wang, J. (2016). Transductive zero-shot recognition via shared model space learning. In *Proceedings of AAAI conference on artificial intelligence*. (pp. 3–8). Phoenix, AZ.

Hoang, D. T., Tran, X. T., Van, M., & Kang, H. J. (2021). A deep neural network-based feature fusion for bearing fault diagnosis . *Sensors, 21*(1), 1–13.

Huang, R., Li, W., & Cui, L. (2019). An intelligent compound fault diagnosis method using one-dimensional deep convolutional neural network with multi-label classifier. In *IEEE international instrumentation and measurement technology conference*.

Huang, R., Li, J., Li, W., & Cui, L. (2020). Deep ensemble capsule network for intelligent compound fault diagnosis using multisensory data. *IEEE Transactions on Instrumentation and Measurement, 69*(5), 2304–2314.

Lampert, C., Nickisch, H., & Harmeling, S. (2014). Attribute-based classification for zero-shot visual object categorization. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 36*(3), 453–465.

Li, Y., Wang, D., Hu, H., Lin, Y., & Zhuang, Y. (2017). Zero-Shot recognition using dual visual-semantic mapping paths. In *Proceedings of the 2017 IEEE conference on computer vision and pattern recognition*. (pp. 5207–5215).

Lin, M., Han, P., Fan, Y., & Li, C. (2020). Development of compound fault diagnosis system for gearbox based on convolutional neural network. *Sensors, 20*(21), 1–14.

Liu, Y., Ye, T., Zeng, Z., Zhang, Y., Wang, G., Chen, N., et al. (2021). Generative adversarial network-enabled learning scheme for power grid vulnerability analysis. *International Journal of Web and Grid Services, 17*(2), 138–151.

Ma, M., Sun, C., & Chen, X. (2018). Deep coupling autoencoder for fault diagnosis with multimodal sensory data. *IEEE Transactions on Industrial Informatics*, *14*(3), 1137–1145.

Marco, B., Angeliki, L., & Georgiana, D. (2015). Hubness and pollution: Delving into cross-space mapping for zero-shot learning. In *Proceedings of the 53rd annual meeting of the association for computational linguistics and the 7th international joint conference on natural language processing*. (pp. 270–280). Beijing, China.

Mhamdi, L., Dhouibi, H., Liouane, N., & Simeu-Abazi, Z. (2013). Multiple fault diagnosis using mathematical models. In *Proceedings of the 9th asian control conference*.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Proceedings of advances in neural information processing systems*. (pp. 3111–3119). Lake Tahoe.

Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing*. (pp. 1532–1543). Doha.

Piacentino, A., & Talamo, M. (2013). Innovative thermoeconomic diagnosis of multiple faults in air conditioning units: Methodological improvements and increased reliability of results. *International Journal of Refrigeration*, *63*(8), 2343–2365.

Rahman, S., Khan, S. H., & Porikli, F. (2020). Zero-shot object detection: Joint recognition and localization of novel concepts. *International Journal of Computer Vision*, *128*(12), 2979–2999.

Shao, H., Jiang, H., Zhao, H., & Wang, F. (2017). A novel deep autoencoder feature learning method for rotating machinery fault diagnosis. *Mechanical Systems and Signal Processing*, *95*, 187–204.

Shao, H., Lin, J., Zhang, L., & Wei, M. (2020). Compound fault diagnosis for a rolling bearing using adaptive dtcwpt with higher order spectra. *Quality Engineering*, *32*(3), 342–353.

Shojaee, S. M., & Baghshah, M. (2016). Semi-supervised zero-shot learning by a clustering-based approach. arXiv:1605.09016.

Sonal, G., Reddy, S., & Kumar, D. (2019). Review of smart health monitoring approaches with survey analysis and proposed framework. *IEEE Internet of Things Journal*, *6*, 2116–2127.

Sun, G., Wang, R., Sun, F., & Jin, Q. (2019). Intelligent detection of a planetary gearbox composite fault based on adaptive separation and deep learning. *Sensors*, *19*(23).

Sun, J., Yan, C., & Wen, J. (2018). Intelligent bearing fault diagnosis method combining compressed data acquisition and deep learning. *IEEE Transactions on Instrumentation and Measurement*, *67*(1), 185–195.

Tang, T., Hu, T. H., Chen, M., Lin, R. L., & Chen, G. R. (2021). A deep convolutional neural network approach with information fusion for bearing fault diagnosis under different working conditions. *Journal of Mechanical Engineering Science*, *235*(8), 1389–1400.

Ubar, R., Raik, J., Kostin, S., & Kousaar, J. (2012). Multiple fault diagnosis with BDD based boolean differential equations. In *Proceedings of the 13th biennial baltic electronics conference*.

Verma, V., Arora, G., Mishra, A., & Rai, P. (2018)Generalized zero-shot learning via synthesized examples. In *Proceedings of IEEE conference on computer vision and pattern recognition*. (pp. 4281–4289). Salt Lake City.

Xia, M. X., Mao, Z. H., Zhang, R., Jiang, B., & Wei, M. H. (2020). A new compound fault diagnosis method for gearbox based on convolutional neural network. In *2020 IEEE 9th data driven control and learning systems conference*. (pp. 1077–1083).

Xing, S. B., Lei, Y. G., Wang, S. H., Lu, N., & Li, N. P. (2022). A label description space embedded model for zero-shot intelligent diagnosis of mechanical compound faults. *Mechanical Systems and Signal Processing*, *162*, Article 108036.

Xu, X., Hospedales, T., & Gong, S. (2015). Semantic embedding space for zero-shot action recognition. In *Proceedings of IEEE international conference on image processing*. (pp. 63–67). Quebec City.

Yu, Y. L., Ji, Z., Li, X., et al. (2018). Transductive zero-shot learning with a self-training dictionary approach. *IEEE Transactions on Cybernetics*, *48*(10), 2908–2919.

Zhang, Z. Z., Li, S. M., Xin, Y., & Ma, H. (2020). A novel compound fault diagnosis method using intrinsic component filtering. *Measurement Science & Technology*, *31*(5), Article 055103.

Zhang, H., Liu, L., Long, Y., Zhang, Z., & Shao, L. (2020). Deep transductive network for generalized zero shot learning. *Pattern Recognition*, *105*.

Zhang, L., Wang, P., Liu, L., et al. (2020). Towards effective deep embedding for zero-shot learning. *IEEE Transactions on Circuits and Systems for Video Technology*, *30*, 2843–2852.

Zhang, L., Xiang, T., & Gong, S. (2017). Learning a deep embedding model for zero-shot learning. In *Proceedings of IEEE conference on computer vision and pattern recognition honolulu*. (pp. 3010–3019).

Zhao, D., Cheng, W., Gao, R. X., Yan, R., & Wang, P. (2020). Generalized vold-kalman filtering for nonstationary compound faults feature extraction of bearing and gear. *IEEE Transactions on Instrumentation and Measurement*, *69*(2), 401–410.

Zhao, B., Zhang, X. M., Zhan, Z. H., & Pang, S. Q. (2020). Deep multi-scale convolutional transfer learning network: A novel method for intelligent fault diagnosis of rolling bearings under variable working conditions and domains. *IEEE Transactions on Industrial Electronics*, *407*, 24–38.

Zheng, J. (2021). Automatic fault diagnosis method for hydrodynamic control system. *International Journal of Fluid Power*, *22*(3), 357–372.