Measurement 186 (2021) 110084

Contents lists available at ScienceDirect

Measurement

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

Semi-supervised graph convolutional network and its application in intelligent fault diagnosis of rotating machinery

Yiyuan Gao [a,b], Mang Chen [a], Dejie Yu [a,*]

[a] State Key Laboratory of Advanced Design and Manufacturing for Vehicle Body, Hunan University, Changsha 410082, PR China
[b] Henan Key Laboratory of Intelligent Manufacturing of Mechanical Equipment, Zhengzhou University of Light Industry, Zhengzhou 450002, PR China

ARTICLE INFO

Keywords:
Graph convolutional network
Semi-supervised learning
Rotating machinery
Fault diagnosis

ABSTRACT

Aiming at the difficulty of mechanical fault diagnosis with small samples, an intelligent fault diagnosis method for rotating machinery is proposed based on semi-supervised graph convolutional network (SSGCN). SSGCN has a good application in analyzing the naturally formed graph data, but it has not been researched for the complex mechanical vibration data. Besides, SSGCN is only applicable to graph data, but the collected vibration signals are one-dimensional time series. To well reflect the local geometry property between all vibration samples, we construct all vibration samples into an undirected and weighted *k*-nearest neighbor graph. The detailed parameter analysis is also carried out for SSGCN. Experimental results indicate that our proposed method can adaptively extract the available fault features from the raw vibration signals. Even if the label ratio is only 0.05 for each gear or bearing condition, our proposed method can still obtain an average accuracy of more than 98%.

1. Introduction

Gears and bearings are key components of rotating machinery, and their operating conditions will directly affect the overall performance of the mechanical equipment [1–4]. Therefore, the fault diagnosis of gears and bearings has very important practical significance. The fault diagnosis methods based on vibration signals generally include three steps of signal acquisition, feature extraction and pattern recognition. At present, feature extraction is mainly achieved by designing some statistical indicators in the time domain, frequency domain and time–frequency domain, but these indicators often lack adaptability and are subject to human experience [5–7]. On the other hand, the high-dimensional feature set established by feature extraction from multiple perspectives often contains a large amount of redundant information, so it is necessary to perform dimensionality reduction, such as principle component analysis (PCA) [8] and locally linear embedding (LLE) [9]. All in all, the above-mentioned process of feature extraction is still limited by human experience, and the automatic extraction of effective features cannot be achieved.

In recent years, deep learning algorithms have been widely used in the fault diagnosis of rotating machinery. Their advantage lies in the end-to-end fault diagnosis process from original signals to feature extraction, dimensionality reduction and pattern recognition. Classical deep learning algorithms include convolutional neural network (CNN), deep belief network (DBN) and deep auto-encoder (DAE). Janssens et al. [10] first used CNN to identify the outer-raceway faults and lubrication degradation of bearings, and the diagnosis accuracy is significantly improved in comparison to the classical manually-designed features based approach. Zhang et al. [11] proposed an improved method named deep convolutional neural networks with wide first-layer kernels (WDCNN) for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals. Shao et al. [12,13] successfully applied the improved DBN and DAE to identify different bearing conditions without manual feature extraction. However, the above methods all have common application limitations, that is, the training process of the entire model is supervised, and usually the number of labeled samples must be sufficient, otherwise it is prone to overfitting. In real-world applications, once the mechanical equipment fails, it will be shut down immediately, and the samples under the faulty condition will be less. Therefore, it is particularly necessary to study the diagnosis of mechanical faults under limited labeled samples.

Graph convolutional network (GCN) [14–16] is an emerging and effective deep learning algorithm for the non-Euclidean data like graph structure data. In the graph domain, the geometry structure of the data can provide additional information, including not only the values of the nodes but also the relationships between them (edges) [17]. Hence, GCN

* Corresponding author.
E-mail address: djyu@hnu.edu.cn (D. Yu).

https://doi.org/10.1016/j.measurement.2021.110084
Received 27 April 2021; Received in revised form 17 August 2021; Accepted 24 August 2021
Available online 28 August 2021

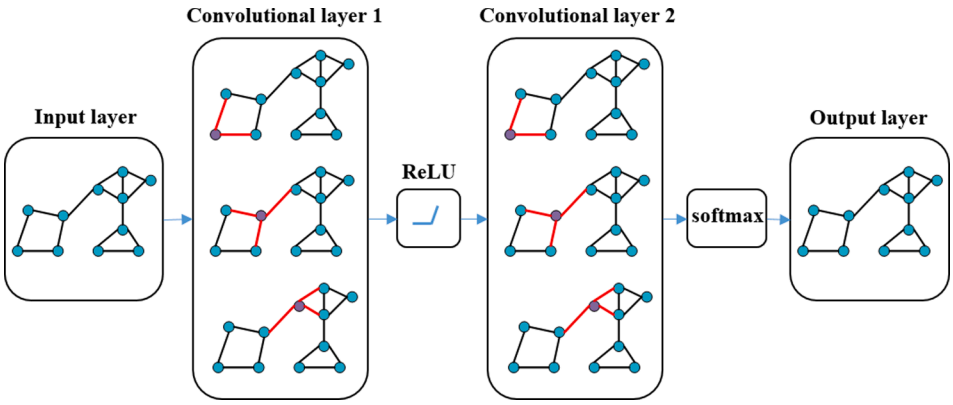

Fig. 1. Schematic diagram of the model.

can use the attributes of nodes and edges to improve the abilities of extracting features and resisting noise, and the fault diagnosis techniques based on GCN have gradually become popular. Zhang et al. [18] proposed a deep GCN to achieve the intelligent acoustic-based fault diagnosis of roller bearings. Yu et al. [19] proposed a novel fast deep GCN to diagnose faults in the gearbox of wind turbines, in which the wavelet coefficient matrices of the vibration signals are regarded as graphs. Li et al. [20] proposed a multi-receptive field GCN for mechanical fault diagnosis, in which the raw vibration signals are converted into affinity graphs directly. But the fly in the ointment is that these proposed methods are still supervised fault diagnosis methods and require a large number of labeled samples for training the models. Kipf and Welling [21] pointed out that the GCN can be used for semi-supervised classification, which is known as semi-supervised graph convolutional network (SSGCN) in this paper. The potential of SSGCN has been proven in a number of experiments, including semi-supervised document classification in citation networks and semi-supervised entity classification in a bipartite graph extracted from a knowledge graph [21,22]. Although SSGCN has a good application in analyzing the naturally formed graph data, it has not been researched for the complex vibration data of rotating machinery.

In this paper, an intelligent fault diagnosis method based on SSGCN is proposed for rotating machinery. SSGCN can naturally be used for semi-supervised classification of graph structure data, but the vibration signals of rotating machinery we collected are not a graph structure. Thus, the key to successfully introducing SSGCN is to construct all vibration samples into a suitable graph structure. The study found a k-nearest neighbor (KNN) graph can better reflect the local geometry property between all vibration samples [23]. In some graph-based algorithms, such as the Laplacian eigenmaps [24], Laplacian score [25] and spectral clustering [26], the heat kernel is usually used to define the edge weights to express the difference between nodes. So a KNN graph weighted by the heat kernel is constructed by treating each vibration sample in frequency domain as a node. The influence of different parameters of SSGCN on diagnosis accuracy is also analyzed in detail. More importantly, the two sets of experimental results indicate that our proposed method can adaptively extract the available fault features from the vibration signals in frequency domain. Even if the label ratio is only 0.05 for each gear or bearing condition, our proposed method can still obtain an average accuracy of more than 98%.

The main contributions of this paper are as follows.

1) Based on SSGCN, an intelligent fault diagnosis method is proposed for rotating machinery. Although SSGCN has a good application in analyzing the naturally formed graph data, it has not been researched for the complex vibration data of rotating machinery.
2) SSGCN is only applicable to graph data, but the collected mechanical vibration signals are time series. In order to introduce SSGCN, we use the weighted KNN to establish a vibration data graph, which can reflect the local geometry property between vibration samples well.
3) The detailed parameter analysis is carried out for SSGCN. Two experimental datasets of gears and bearings in different conditions verify the effectiveness and superiority of our proposed method.

The rest of this paper is arranged as follows. First of all, the related theory is introduced in Section 2. Then, our proposed method is described in Section 3. Next, the experimental results are analyzed in Section 4. Finally, the conclusions of this paper are drawn in Section 5.

2. Semi-supervised graph convolutional network (SSGCN)

SSGCN is a semi-supervised deep learning algorithm, which can utilize the category information of labeled samples, the graph connection relationship between all samples and the operation of spectral graph convolution to realize the automatic feature extraction and pattern recognition. In this paper, a two-layer SSGCN model is built, and its schematic diagram is shown in Fig. 1.

In Fig. 1, the core is the two convolutional layers with an activation function between them. In each convolutional layer, the operation of spectral graph convolution is used to adaptively extract the available features from the input high-dimensional feature data. In order to obtain better diagnosis effect, the KNN graph is constructed to fully reflect the local geometry property between all samples, and its weighted adjacency matrix are taken as the input of the two-layer SSGCN model. Therefore, in this section, the KNN graph and spectral graph convolution are first described, and then the two-layer SSGCN model is defined.

2.1. K-nearest Neighbor (KNN) graph

Suppose that the sample set containing some labeled and unlabeled samples is denoted as$X = \{x_i\}_{i=1}^{N}$, and its label ratio is $lr = N_1/N$, where x_i is the ith sample, N is the total number of samples, and N_1 is the number of labeled samples. The KNN graph G can be constructed by treating each sample as a node and using the Euclidean distance to judge the adjacency relationship between all nodes. If the node x_j is one of the k nearest distance nodes of the node x_i, then there is an undirected edge connection between x_i and x_j. However, there is no established rule for the selection of the neighbor parameter k, and we take k as$log_2(N)$according to the empirical formula in the literature [26]. After determining the adjacency relationship between all nodes in the graph G, the corresponding adjacency matrix $\mathbf{A}$ can be obtained as follows. The element a_{ij} or a_{ji} of the adjacency matrix$\mathbf{A} \in \mathbb{R}^{N\times N}$represents the adjacency and similarity relationships between the nodes x_i and x_j. If there is no edge connection between the nodes x_i and x_j, the values of a_{ij} and a_{ji} equal 0, otherwise, the values of a_{ij} and a_{ji} are defined by the heat kernel shown in Eq. (1).

$$a_{ij} = a_{ji} = e^{-\frac{\left\|x_i - x_j\right\|_2^2}{2t^2}}, \tag{1}$$

where t is the width of heat kernel, and is defined as the mean value of the Euclidean distances between all nodes, namely $t = \frac{2}{N(N-1)}\sum_{i=1}^{N}\sum_{j=i+1}^{N}\left\|x_i - x_j\right\|_2$. If x_i and x_j are connected by an undirected edge and their distance is closer, the values of a_{ij} and a_{ji} are closer to 1. Since the values of a_{ij} and a_{ji} are equal, the adjacency matrix $\mathbf{A}$ is a symmetric matrix in the KNN graph.

2.2. Spectral graph convolution

In CNN, the convolution operation is used to extract the spatial features, but it is not suitable for the non-Euclidean structure data like graph structures data. In order to realize the feature extraction of graph structure data, the spectral graph convolution can be defined by means of the graph Fourier transform (GFT) and inverse graph Fourier transform (IGFT) [15].

Assume that $f \in \mathbb{R}^N$(N is the number of nodes) is the signal defined on the graph G, $f(i)$ is the signal value of the ith node in the graph G. Similar to the definition of traditional Fourier transform, the GFT can be defined as the expansion of graph signals with respect to the eigenvectors of the graph Laplace matrix [27,28]. The symmetric normalized Laplacian matrix of the graph G is defined as Eq. (2).

$$\mathbf{L} = \mathbf{I}_N - \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}, \tag{2}$$

where $\mathbf{I}_N$ is the identity matrix, and $\mathbf{D}$ is the degree diagonal matrix with diagonal elements $d_{ii} = \sum_j a_{ij}$, indicating the number of nodes connected to each node. The spectral decomposition of $\mathbf{L}$ is shown in Eq. (3).

$$\mathbf{L}u_l = \lambda_l u_l, \cdots, N, \tag{3}$$

where l is the order, u_l is the eigenvector corresponding to the order l in the eigenvector matrix $\mathbf{U} = [u_1\, u_2 \cdots u_N]$, and λ_l is the eigenvalue corresponding to the order l in the eigenvalue matrix $\Lambda = diag([\lambda_1\, \lambda_2 \cdots \lambda_N])$. Therefore, the matrix form of the spectral decomposition is $\mathbf{L}\mathbf{U} = \Lambda\mathbf{U}$, and the matrix forms of GFT and IGFT can be expressed as Eq. (4).

$$\begin{aligned} F_G(f) &= \widehat{f} = \mathbf{U}^T f \\ F_G^{-1}\left(\widehat{f}\right) &= \mathbf{U}\widehat{f} \end{aligned} \tag{4}$$

In the vertex domain of a graph, the operation of graph convolution between the graph signal $f \in \mathbb{R}^N$(a scalar for each node) and the convolution kernel g is expressed as Eq. (5).

$$f * g = F_G^{-1}\{F_G(f)F_G(g)\} \tag{5}$$

By means of the GFT and IGFT in Eq. (4), the operation of graph convolution can be further expressed as Eq. (6).

$$f * g = \mathbf{U}\left(\left(\mathbf{U}^{\mathrm{T}}\mathrm{f}\right) \odot \left(\mathbf{U}^{\mathrm{T}}\mathrm{g}\right)\right) = \mathbf{U}\begin{bmatrix} \widehat{g}(\lambda_1) & & \\ & \ddots & \\ & & \widehat{g}(\lambda_N) \end{bmatrix}\mathbf{U}^{\mathrm{T}}\mathrm{f} = \mathbf{U}\widehat{\mathrm{g}}(\Lambda)\mathbf{U}^{\mathrm{T}}\mathrm{f}. \tag{6}$$

Eq. (6) is known as spectral graph convolution in [21]. It can be seen from Eq. (6) that the complexity of this convolution calculation is very high for large graphs, because it involves solving the eigenvectors and eigenvalues of graph Laplacian matrix. In order to reduce the computational complexity of spectral graph convolution, the finite K terms of Chebyshev polynomials $T_k(\cdot)$ can be used to approximate the convolution kernel $\widehat{g}(\Lambda)$ [29], as shown in Eq. (7).

$$\widehat{g}(\Lambda) \approx \sum_{k=0}^{K}\theta_k T_k\left(\widetilde{\Lambda}\right), \tag{7}$$

with a rescaled $\widetilde{\Lambda} = (2/\lambda_{max})\Lambda - \mathbf{I}_N$, where $\lambda_{\max}$ denotes the largest eigenvalue of $\mathbf{L}$ and θ_k is the Chebyshev coefficient. As can easily be verified by noticing that $\mathbf{U}\Lambda^k\mathbf{U}^{\mathrm{T}} = \left(\mathbf{U}\Lambda\mathbf{U}^{\mathrm{T}}\right)^{\mathrm{k}} = \mathbf{L}^{\mathrm{k}}$, the spectral graph convolution shown in Eq. (6) can be approximated as Eq. (8).

$$f * g \approx \sum_{k=0}^{K}\theta_k T_k\left(\widetilde{\mathbf{L}}\right) f, \tag{8}$$

with $\widetilde{\mathbf{L}} = (2/\lambda_{max})\mathbf{L} - \mathbf{I}_N$. When $K = 1$, the spectral graph convolution can be expanded into a linear formulation, at the same time, we further approximate $\lambda_{max} \approx 2$. In order to reduce the training parameters, with a single parameter $\theta = \theta_0 = -\theta_1$, the spectral graph convolution shown in Eq. (8) can be further expressed as Eq. (9).

$$f * g \approx \theta\left(\widetilde{\mathbf{D}}^{-1/2}\widetilde{\mathbf{A}}\widetilde{\mathbf{D}}^{-1/2}\right) f. \tag{9}$$

Here, $\widetilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}_N$ is the adjacency matrix of the graph G with added self-connections, and $\widetilde{\mathbf{D}}$ is the degree diagonal matrix corresponding to $\widetilde{\mathbf{A}}$.

Of course, the definition of spectral graph convolution can be generalized to the graph signal $\mathbf{F} \in \mathbb{R}^{N\times M}$ with M input channels (i.e. a $M-$dimensional feature vector for every node) as Eq. (10).

$$\mathbf{Z} = \widetilde{\mathbf{D}}^{-1/2}\widetilde{\mathbf{A}}\widetilde{\mathbf{D}}^{-1/2}\mathbf{F}\mathbf{W}, \tag{10}$$

where $\mathbf{W} \in \mathbb{R}^{M\times C}$ is the parameter matrix of convolution kernels and $\mathbf{Z} \in \mathbb{R}^{N\times C}$ is the convolved signal matrix. C is the total number of categories.

2.3. The two-layer SSGCN model

From Fig. 1, it can be seen that the activation function between the two convolutional layers is $\mathrm{ReLU}(\cdot) = max(0, \cdot)$, so based on Eq. (10), the layer-wise propagation rule can be constructed as Eq. (11).

$$\mathbf{H}^{l+1} = \mathrm{ReLU}\left(\widetilde{\mathbf{D}}^{-1/2}\widetilde{\mathbf{A}}\widetilde{\mathbf{D}}^{-1/2}\mathbf{H}^{l}\mathbf{W}^{l}\right), \tag{11}$$

where $\mathbf{H}^l$ is the matrix of activations in the lth layer with $\mathbf{H}^0 = \mathbf{F}$, and $\mathbf{W}^l$ is a layer-specific trainable weight matrix. The output layer uses the softmax classifier defined as $\mathrm{softmax}(\mathrm{f_i}) = exp(\mathrm{f_i})/\sum_i exp(\mathrm{f_i})$. Eventually, the two-layer SSGCN model is formulated as Eq. (12).

$$\mathbf{Z} = \mathrm{softmax}\left(\widehat{\mathbf{A}}\ \mathrm{ReLU}\left(\widehat{\mathbf{A}}\mathbf{F}\mathbf{W}^0\right)\mathbf{W}^1\right) \tag{12}$$

Here, $\widehat{\mathbf{A}} = \widehat{\mathbf{D}}^{-1/2}\widehat{\mathbf{A}}\widehat{\mathbf{D}}^{-1/2}$, $\mathbf{W}^0 \in \mathbb{R}^{M\times h}$ is the input-to-convolutional weight matrix for a convolutional layer with h feature maps, and $\mathbf{W}^1 \in \mathbb{R}^{h\times C}$ is the convolutional-to-output weight matrix, where M is the feature dimensions of the input signal and h is the number of neurons in the convolutional layers. There is a specific analysis in the experimental section about the setting of parameter h.

For the two-layer SSGCN model, the cross-entropy loss function of labeled nodes (samples) is used as the objective function to perform back propagation update on the weight parameters $\mathbf{W}^0$ and $\mathbf{W}^1$, which is specifically expressed as Eq. (13).

$$LF = -\sum_{i\in V_{\mathrm{label}}}\sum_{j=1}^{F} Y_{ij}^{\mathrm{r}} ln\left(Y_{ij}^{\mathrm{p}}\right), \tag{13}$$

where Y_{ij}^{r} is the real label value of the ith labeled node, Y_{ij}^{p} is the predicted label value of the ith labeled node, and V_{label} is the set of labeled nodes.

All in all, the automatic extraction of spatial features is realized through the operation of spectral graph convolution in SSGCN. The local geometry property of sample data is represented by the weighted adjacency matrix of a KNN graph, and the category information of labeled samples is used as the loss function to iteratively train the network weight parameters, and finally the semi-supervised classification is

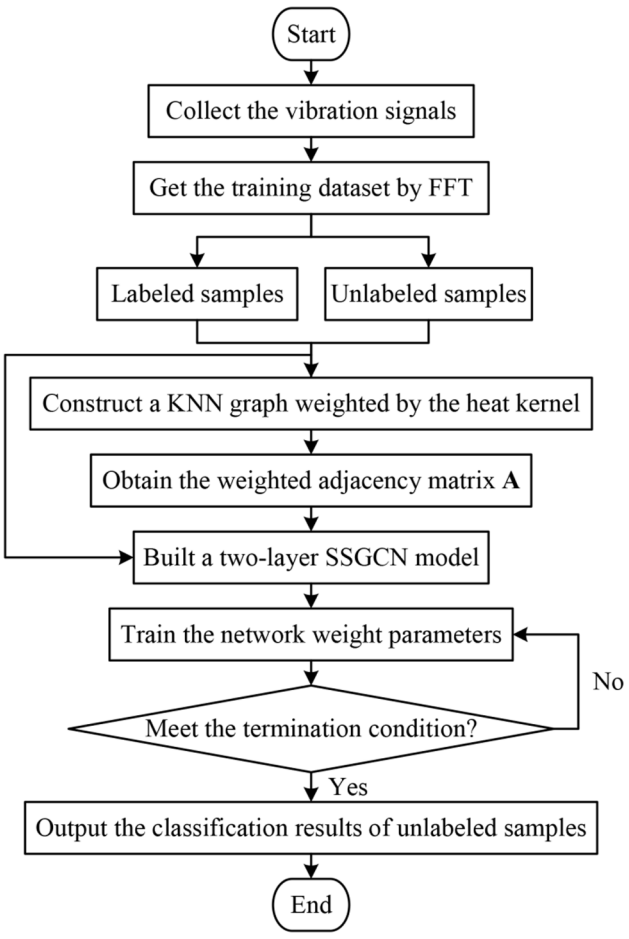

Fig. 2. Training process of the two-layer SSGCN model.

realized.

3. Rotating machinery fault diagnosis using SSGCN

The SSGCN can use the category information of labeled samples, the graph adjacency relationship between all samples and the graph convolution operation to realize the automatic feature extraction and pattern recognition. Based on SSGCN, an intelligent fault diagnosis method is proposed for rotating machinery. In our proposed method, the vibration signals in frequency domain are first used to form samples without manual feature extraction, and then a KNN graph weighted by the heat kernel is constructed by treating each sample as a node. Next, all samples and the graph adjacency matrix are taken as the input of the two-layer SSGCN model. Finally, the cross-entropy loss function of labeled samples is regarded as the objective function to update the network weight parameters, so as to our proposed method can effectively realize the automatic feature extraction and semi-supervised classification for rotating machinery. The training process of the two-layer SSGCN model is shown in Fig. 2 and the detailed implementation steps of our proposed method are described as follows:

1) Acquisition and pre-processing of the vibration data: The vibration data of rotating machinery under different conditions are acquired, and the fast Fourier transform (FFT) is performed on the vibration data to obtain their spectrum data.
2) Obtaining of the sample dataset: The spectrum data is directly intercepted into a certain number of training and testing samples with a certain number of sampling points. For each condition, some training samples are randomly selected as the labeled samples with a certain label ratio of lr, and the remaining training samples are regarded as the unlabeled samples. All the labeled and unlabeled samples form the training dataset $\mathbf{X}$ containing N samples.
3) Construction of the KNN graph: The KNN graph is constructed by treating each training sample as a node and using the Euclidean distance to judge the adjacency relationship between all the nodes. The nearest neighbor parameter k is determined by the empirical formula $log_2(N)$. According to Eq. (1), the weighted adjacency matrix $\mathbf{A}$ corresponding to the KNN graph is calculated.
4) Building of the SSGCN model: The two-layer SSGCN model is built according to Eq. (12), and the training dataset $\mathbf{X}$ and weighted adjacency matrix $\mathbf{A}$ are directly used as the input of the two-layer SSGCN model. The cross-entropy loss function shown in Eq. (13) is used as the objective function to train the weight parameters $\mathbf{W}^0$ and $\mathbf{W}^1$. When the termination condition is met, the training of weight parameters is stopped. The testing samples are input into the trained SSGCN model, and then the condition of all testing sample can be determined based on the output.

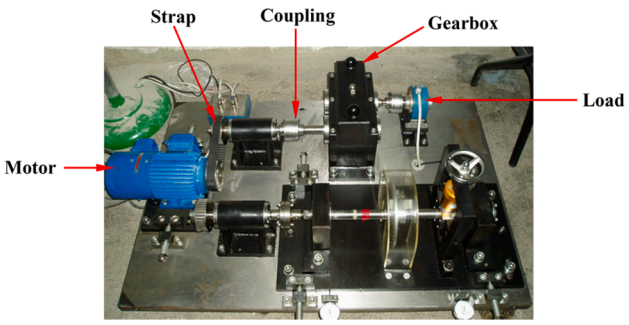

Fig. 3. Test bench.

4. Experimental validations

4.1. Case 1: Gear fault diagnosis

4.1.1. Data description

In order to verify the effectiveness of our proposed method, the QPZZ-II rotating machinery vibration and failure simulation test bench produced by Jiangsu Qianpeng Diagnostic Engineering Co., Ltd. is used for gear experiments. The test bench shown in Fig. 3 has a single-stage parallel shaft speed-increasing gearbox. The number of teeth of the driving gear and driven gear is 75 and 55 respectively, the modulus is both 2, and the tooth surface material is S45C. In order to simulate a variety of gear faults in the gearbox, the pitting fault or broken tooth fault is arranged on the driving gear, and the wear fault is arranged on the driven gear. Considering the health, single fault and compound faults of the driving gear and driven gear, a total of six gear conditions

Table 1
Description of the six gear conditions.

Gear conditions		Number of training samples	Number of testing samples	Label
Driving gear	Driven gear			
Broken tooth	Health	400	400	1
Broken tooth	Wear fault	400	400	2
Health	Health	400	400	3
Pitting fault	Health	400	400	4
Pitting fault	Wear fault	400	400	5
Health	Wear fault	400	400	6

Table 2
Description of the nine training datasets D1-D9.

Training datasets	Number of labeled samples	Number of unlabeled samples	Label ratios
D1	120 (20*6)	2280	0.05 (120/2400)
D2	360 (60*6)	2040	0.15 (360/2400)
D3	600 (100*6)	1800	0.25 (600/2400)
D4	840 (140*6)	1560	0.35 (840/2400)
D5	1080 (180*6)	1320	0.45 (1080/2400)
D6	1320 (220*6)	1080	0.55 (1320/2400)
D7	1560 (260*6)	840	0.65 (1560/2400)
D8	1800 (300*6)	600	0.75 (1800/2400)
D9	2040 (340*6)	360	0.85 (2040/2400)

are designed. The acceleration sensor is placed on the output shaft to collect the vibration signals with a sampling frequency of 5120 Hz, and the rotational speed of the input shaft is 880 rpm.

For each gear condition, four different loads (0, 0.05A, 0.1A and 0.2A) are loaded, and 800 samples with a length of 2048 sampling point are collected (200 samples per load), of which the first 400 are used as the training samples and the remaining 400 are regarded as the testing samples. For fully testing the diagnosis performance of our proposed method on the six gear conditions under different number of labeled samples, the label ratio *lr* changes from 0.05 to 0.85 and the step size is 0.1. That is to say, the *lr**400 samples are randomly selected from the 400 training samples of each gear condition as the labeled samples, and the remaining training samples as the unlabeled samples. Eventually, the nine training datasets D1-D9 with different label ratios can be obtained, and the detailed information of the six gear conditions and nine training datasets D1-D9 is described in Tables 1 and 2 respectively. Compared with time domain signals, frequency domain signals can not only reduce the data length by half, but also contain more fault information [30]. Therefore, in order to improve the diagnosis efficiency and accuracy, the raw amplitudes of the frequency spectrum of the vibration data are used as the input features of SSGCN.

4.1.2. Parameter optimization of SSGCN

Considering the construction of the two-layer SSGCN model, there are some critical hyper-parameters, including the learning rate L, weight decay w, dropout d and the number of neurons h in the convolutional layers. Among them, L is the most important and sensitive parameter. So the parameter optimization using the 5-fold cross validation (5-CV) method [31,32] is employed on the candidate sets: L is from {0.0005, 0.001, 0.005, 0.01, 0.05, 0.1} or {0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1}, w is from {0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05}, d is from {0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8} and h is from {96, 128, 160, 192, 224, 256, 288}. The 5-CV results of the parameters (L, w), (L, d) and (L, h) for the training dataset D1 are described as Figs. 4–6.

It can be seen from Figs. 4–6 that the SSGCN model is very sensitive to the selection of the learning rate L. If the learning rate L is selected between 0.01 and 0.1, the SSGCN model can achieve a lower recognition error rate. Otherwise, a too large value will lead to over fitting of the model, and a too small value will make the convergence speed of the model very slow. For the parameters w, d and h, when the learning rate L is low, they have a greater impact on the diagnosis performance of SSGCN, but when the learning rate L is large, they have a smaller impact on the diagnosis performance of SSGCN.

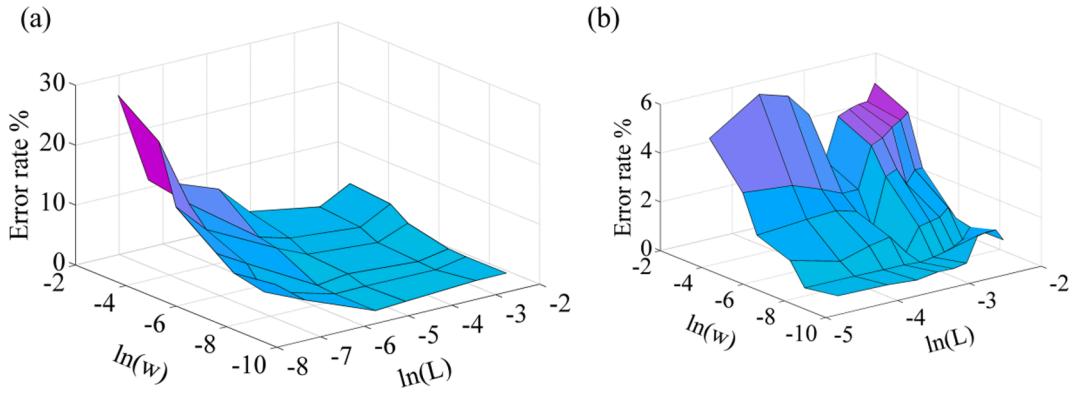

Fig. 4. 5-CV results of the parameters (L, w).

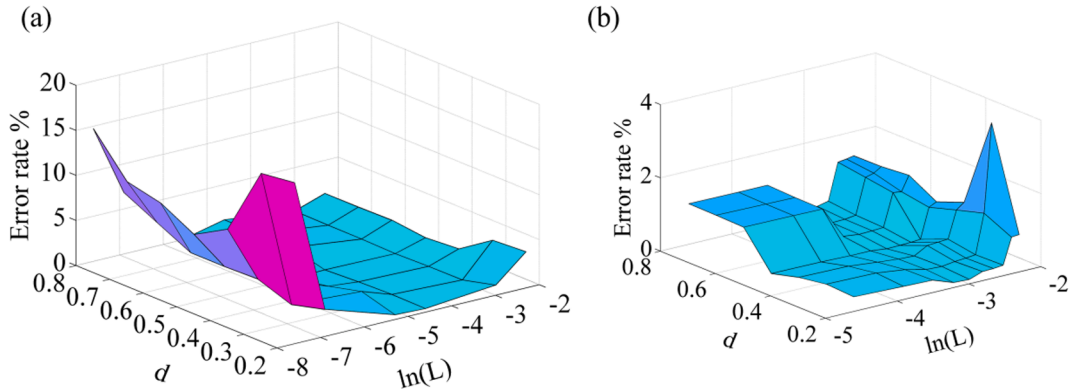

Fig. 5. 5-CV results of the parameters (L, d).

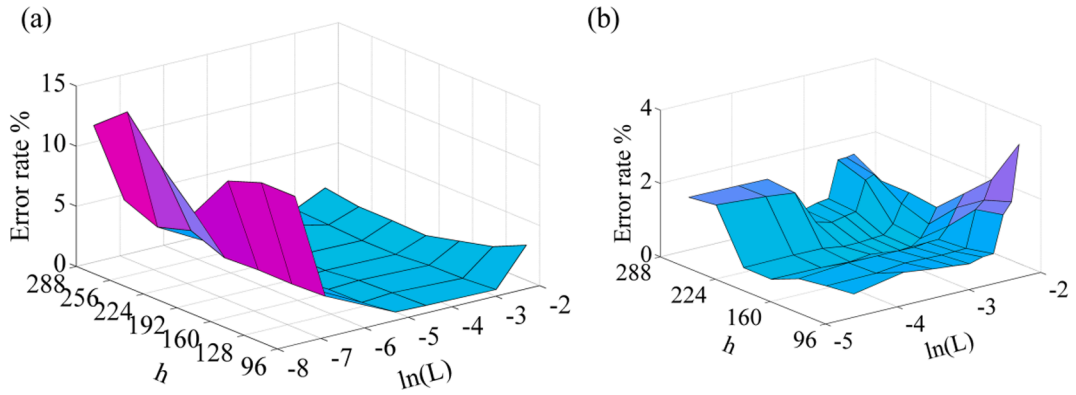

Fig. 6. 5-CV results of the parameters (L, h).

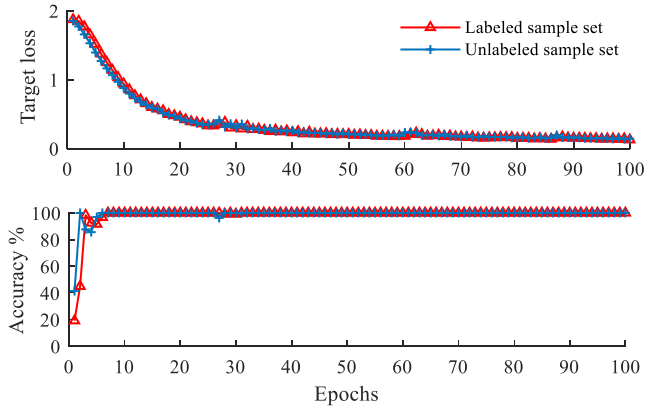

Fig. 7. Iteration curve of target loss and diagnosis accuracy.

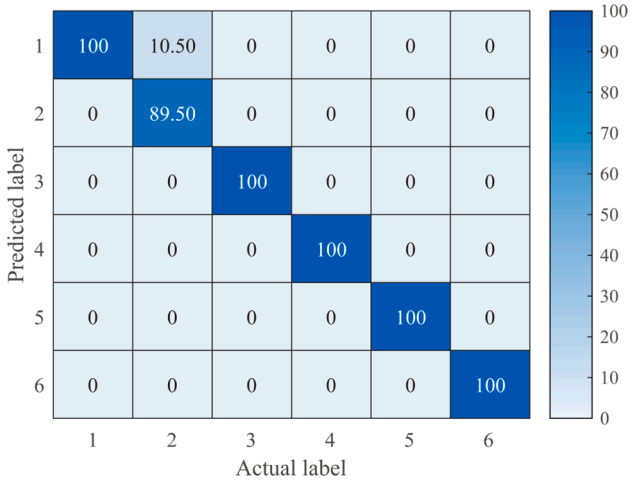

Fig. 8. Multi-class confusion matrix of SSGCN when label ratio $lr = 0.05$.

From Fig. 6(a), it appears that when the parameter h is selected between 96 and 192, the recognition error rate of SSGCN decreases slightly with the increase of the parameter selection, while when the parameter h is selected between 192 and 288, the recognition error rate of SSGCN increases slightly with the increase of the parameter selection. This is because too few neurons in the convolutional layers will lead to underfitting, and too many neurons (that is, with too much information processing capability) are prone to over-fitting, which both affect the diagnosis performance. Roughly the same changing rules also appear in other training datasets. Thus, the optimal parameter set$\{L, w, d, h\} = \{0.05, 0.0005, 0.5, 192\}$is determined in the two-layer SSGCN model.

4.1.3. Results and analysis

To test the diagnosis performance of our proposed method on different gear conditions under low label ratio and different working loads, the training dataset D1 containing 120 labeled samples and 2280 unlabeled samples is first used to train the two-layer SSGCN model. According to the above-mentioned network parameter settings, the two-layer SSGCN model is built by Eq. (12). The cross-entropy loss function shown in Eq. (13) is used as the objective function to train the weight parameters. The termination condition is set to stop training when the total number of iterations reaches 100. The target loss and diagnosis accuracy of the labeled sample set and unlabeled sample set during the training process are shown in Fig. 7.

As can be observed from Fig. 7, the target loss of the labeled sample set and unlabeled sample set basically remains stable after 40 rounds of epochs, and the diagnosis accuracies of the labeled sample set and unlabeled sample set basically remain stable after 7 rounds of epochs. These indicate that the two-layer SSGCN model not only has a fast convergence speed but also has good stability. In addition, for the testing samples, the multi-class confusion matrix of our proposed method is plotted in Fig. 8, and it manifests the diagnosis results of each gear condition in detail, including the diagnosis accuracy and error ratio. More concretely, the diagonal elements of the multi-class confusion matrix represent the accuracy of each gear condition being correctly diagnosed, and the other elements represent the ratio of each gear condition being incorrectly diagnosed as other gear conditions.

As can be seen from Fig. 8, there are a few testing samples that are misidentified between the conditions (1) and (2). Specifically, 10.50% of the testing samples in condition (2) are misidentified as condition (1). For other gear conditions, the diagnosis accuracies are both 100%. On

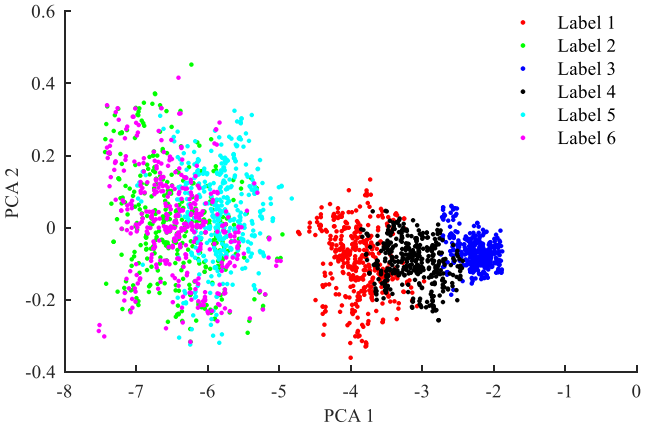

Fig. 9. Input dataset visualization.

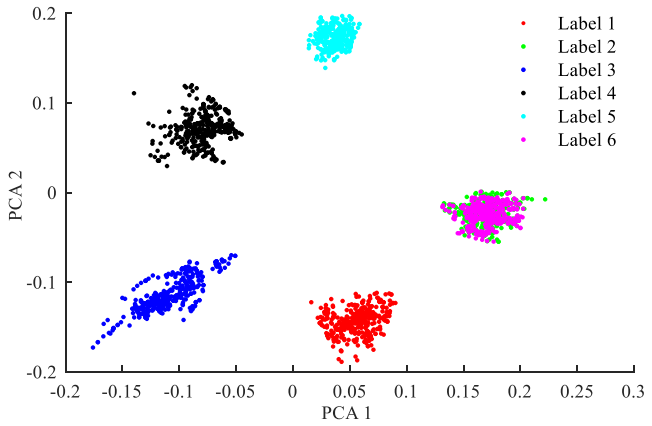

Fig. 10. First convolutional layer output result visualization in SSGCN.

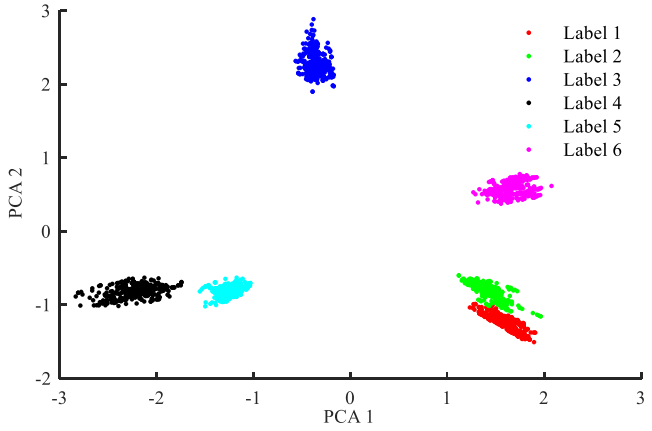

Fig. 11. Output layer output result visualization in SSGCN.

the whole, our proposed method still has a very high diagnosis accuracy of 98.25%. This is because the vibration signals of conditions (1) and (2) are extremely similar, so it is difficult to correctly diagnose under very few labeled samples. When the label ratio is increased from 0.05 to 0.35, all the testing samples in condition (2) can be always diagnosed correctly, and the overall diagnosis accuracy of our proposed method is stable at 100%. These indicate that our proposed method can effectively realize the semi-supervised classification of different gear conditions under a small number of labeled samples. In order to further verify the effectiveness of our proposed method in adaptively extracting the available fault features from the raw vibration signals in frequency domain, the initial features (1024 dimensions), the features in the first convolutional layer (192 dimensions) and the features in the output layer (6 dimensions) are visualized by PCA. In Figs. 9–11, PCA1 and PCA2 represent the first two principle components, which are usually

Table 3
5-CV results of SSGCN for the training dataset D1 under different sample length.

Sample length	5-CV results	
	Recognition error rate (%)	Running time (s)
256	16.67	20.32
512	3.95	39.28
1024	1.75	55.14
2048	1.54	153.59
4096	1.36	468.97

the most useful information hidden in the high-dimensional features.

As shown in Fig. 9, most of the testing samples overlap with each other, that is to say, there is not much discrimination between the testing samples in the initial features. It can be seen from Fig. 10 that the features in the first convolutional layer are more recognizable than the initial features. However, they are still not good enough to distinguish each gear condition well. From Fig. 11, it can be observed that the discrimination ability of all gear conditions is further improved after the feature extraction of the two-layer SSGCN model, only a few testing samples in conditions (1) and (2) overlap each other, which is consistent with the diagnosis result of the confusion matrix in Fig. 8. The above results show that even if there are a few labeled samples, the two-layer SSGCN model can still learn and extract the useful information hidden in the original high-dimensional features through the operation of spectral graph convolution for gears.

In our proposed method, the KNN graph is the premise. The influence of the quality of the KNN graph on the diagnosis results is first studied. The value of neighbor parameter k and the definition of edge weights are two key points in the KNN graph. In the graph-based algorithms, the edge weights of a graph are often taken as a constant value of 1 or defined by heat kernel. To successfully extract the fault impulse signal of rolling bearings by the graph Fourier transform, Ou et al. [33] also used Euclidean distance to define the edge weights of the path graph. When the KNN graph is weighted by 1, heat kernel and Euclidean distance, the 5-CV recognition error rate of SSGCN for the training dataset D1 is 12.28%, 1.75% and 7.81%, respectively. The deep reason why the KNN graph weighted by heat kernel obtains the lowest recognition error rate is that it can better reflect the local geometry property between all samples. When the KNN graph is weighted by heat kernel, Fig. 12 presents the 5-CV recognition error rate of SSGCN for the training dataset D1 under different k value.

It can be observed from Fig. 12 that the values around 11 are a better choice, which is consistent with the empirical formula$log_2(N)$, namely the base-2 logarithm of the total number of samples. Then, the influence of the length of all samples input into SSGCN on the diagnosis results is studied. The sample length is also set as 256, 512, 2048, 4096 in addition to 1024, and the 5-CV recognition error rate of SSGCN for the training dataset D1 are listed in Table 3. When all samples are extended from 256 to 1024 points, the recognition error rate is reduced significantly, while when all samples are extended from 1024 to 4096 points, the recognition error rate is reduced slightly. With the continuous

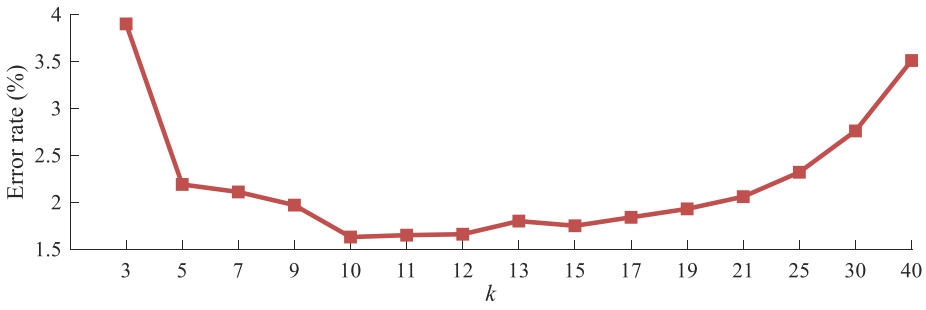

Fig. 12. 5-CV recognition error rate of SSGCN for the training dataset D1 under different k value.

Table 4
5-CV results of SSGCN for the training dataset D1 under different layers.

The number of layers	5-CV results	
	Recognition error rate (%)	Running time (s)
1	24.91	25.47
2	1.75	55.14
3	1.84	101.55
4	3.51	160.29
5	12.28	300.71
6	16.67	360.78

Table 5
Diagnosis results of the five methods under different label ratios.

Label ratios	Diagnosis methods				
	Method 1 (Our proposed method)	Method 2 (CNN: LeNet-5)	Method 3 (Proposed in [11])	Method 4 (Proposed in [12])	Method 5 (Proposed in [13])
0.05	98.25 ± 2.46	53.63 ± 8.86	56.13 ± 7.21	65.21 ± 5.73	62.92 ± 6.28
0.15	99.50 ± 2.18	78.54 ± 8.12	79.46 ± 6.48	86.50 ± 4.58	84.00 ± 5.02
0.25	99.88 ± 0.56	82.79 ± 7.36	84.54 ± 5.26	90.21 ± 4.09	86.33 ± 4.46
0.35	100 ± 0	88.00 ± 6.12	90.42 ± 3.12	93.21 ± 3.51	91.71 ± 2.98
0.45	100 ± 0	93.08 ± 3.04	93.29 ± 2.36	96.63 ± 2.16	94.54 ± 1.81
0.55	100 ± 0	93.67 ± 2.12	98.33 ± 1.78	99.17 ± 0.72	98.88 ± 0.95
0.65	100 ± 0	97.00 ± 1.32	99.96 ± 0.08	100 ± 0	100 ± 0
0.75	100 ± 0	98.08 ± 0.89	100 ± 0	100 ± 0	100 ± 0
0.85	100 ± 0	100 ± 0	100 ± 0	100 ± 0	100 ± 0

Note: The format of the diagnosis result is average diagnosis accuracy (%) ± standard deviation.

increase of sample length, the running time keeps growing rapidly. For the comprehensive consideration of recognition error rate and running time, the length of all samples input into SSGCN is taken as 1024 in this paper.

In addition, the influence of different layers of SSGCN model on the diagnosis results is also studied. The number of layers is also set as 1, 3, 4, 5, 6 in addition to 2, and the 5-CV recognition error rate of SSGCN for the training dataset D1 are listed in Table 4. It can be found from Table 4 that a 2-layer SSGCN model obtains the lowest recognition error rate, and the 3-layer SSGCN model also obtains a recognition error rate similar to that of the 2-layer SSGCN model. With the continuous increase of model depth, the running time keeps growing rapidly. Considering the recognition error rate and running time, a two-layer SSGCN model is built in this paper.

4.1.4. Comparison with deep learning methods

For further illustrating the superiority of our proposed method, a classical CNN model called LeNet-5 and the three state-of-the-art deep learning methods are also applied to process the same gear datasets, which are recorded as the method 1 (Our proposed method), method 2 (CNN: LeNet-5), method 3 (Proposed in [11]), method 4 (Proposed in [12]) and method 5 (Proposed in [13]). To make the comparison more fair and convincing, the same vibration signals in frequency domain are used as input in the four comparison methods and ten trials are carried out for each methods. The average diagnosis accuracy is obtained by calculating all testing samples and comparing their predicted labels with their real labels. In practice, the acquisition of labeled samples tend to be very time-consuming, expensive and difficult, so the number of labeled samples required to achieve satisfactory results is a very important point

Table 6
Description of the ten bearing conditions.

Bearing conditions	Fault diameter (inch)	Number of training/testing samples (0/1/2/3 HP)	Label
Normal	0	100/100 (25/25/25/25)	1
Rolling elements	0.007	100/100 (25/25/25/25)	2
Rolling elements	0.014	100/100 (25/25/25/25)	3
Rolling elements	0.028	100/100 (25/25/25/25)	4
Inner ring	0.007	100/100 (25/25/25/25)	5
Inner ring	0.014	100/100 (25/25/25/25)	6
Inner ring	0.028	100/100 (25/25/25/25)	7
Outer ring	0.007	100/100 (25/25/25/25)	8
Outer ring	0.014	100/100 (25/25/25/25)	9
Outer ring	0.028	100/100 (25/25/25/25)	10

to measure the quality of a classification model. For this reason, our proposed method and the four comparison methods are trained by the training datasets D1-D9, respectively. Their diagnosis results are presented in Table 5.

From Table 5, we can observe that with the increase of label ratio, the average diagnosis accuracies of the four comparison methods are significantly improved, and the average diagnosis accuracy of our proposed method is stable above 98%. This indicates that our proposed method requires only a small number of labeled samples, while the four comparison methods require a large number of labeled samples to obtain the same satisfactory results. In addition, the standard deviation of our proposed method is smaller than that of other methods in diagnosis accuracy, which means that the two-layer SSGCN model is more stable in gear fault diagnosis. The deeper reason for these advantages is that our proposed method, a semi-supervised fault diagnosis method, can make full use of the relationship between samples.

4.2. Case 2: Bearing fault diagnosis

4.2.1. Data description

In order to verify the effectiveness of our proposed method in bearing fault diagnosis, the bearing data collected from the Case Western Reserve University Bearing Data Center [34,35] is also used. The bearing experiments have been operated under the four different loads (0, 1, 2 and 3 HP), and the test bearing is 6205-2RS JEM SKF deep groove ball bearing which is placed on the drive end of the motor. The accelerometer is placed on the drive end to collect the vibration signals at a sampling frequency of 12 kHz. Single point fault with different diameters is arranged on the inner ring, outer ring and rolling elements of the test bearings by electric spark technology, and the fault diameters

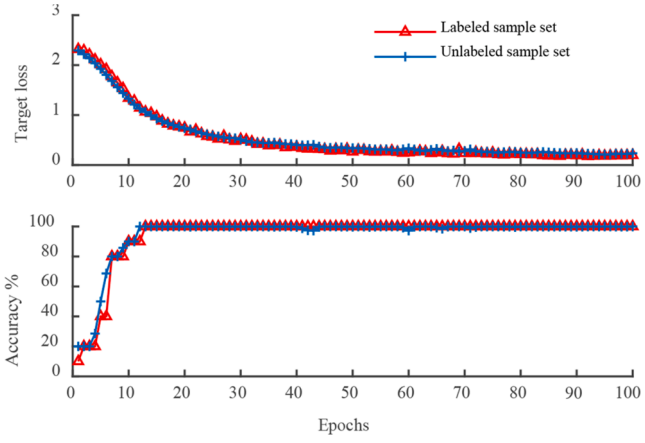

Fig. 13. Iteration curve of target loss and diagnosis accuracy.

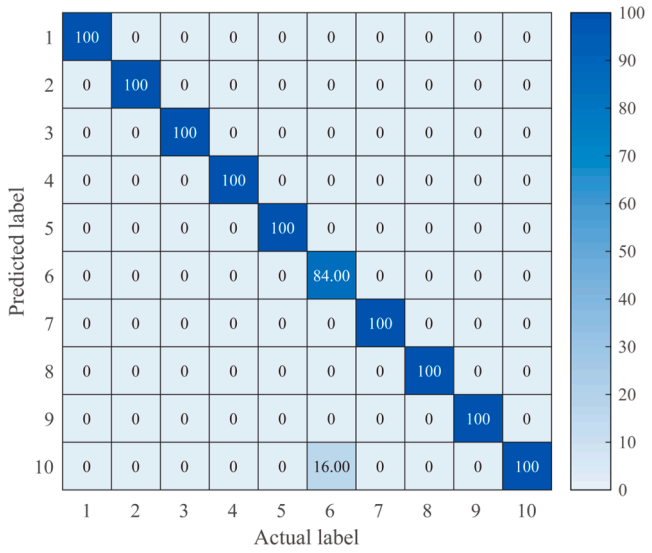

Fig. 14. Multi-class confusion matrix of SSGCN when label ratio $lr = 0.05$.

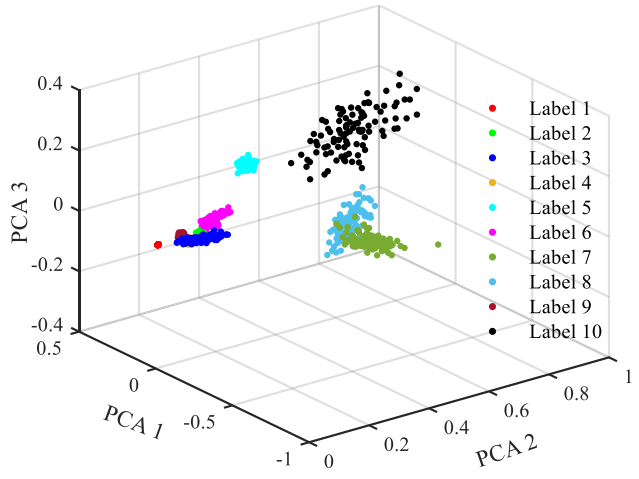

Fig. 15. Input dataset visualization.

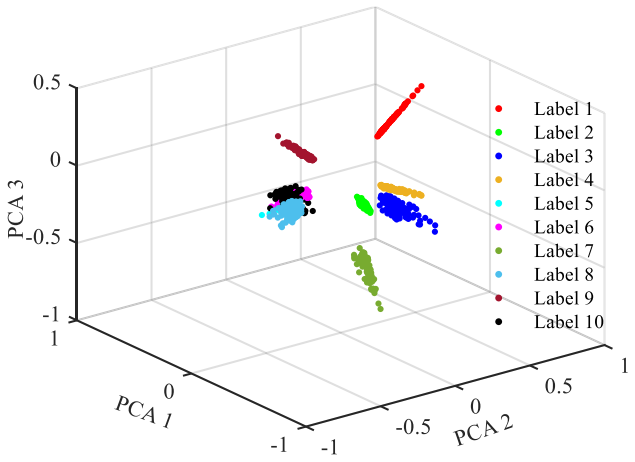

Fig. 16. First convolutional layer output result visualization in SSGCN.

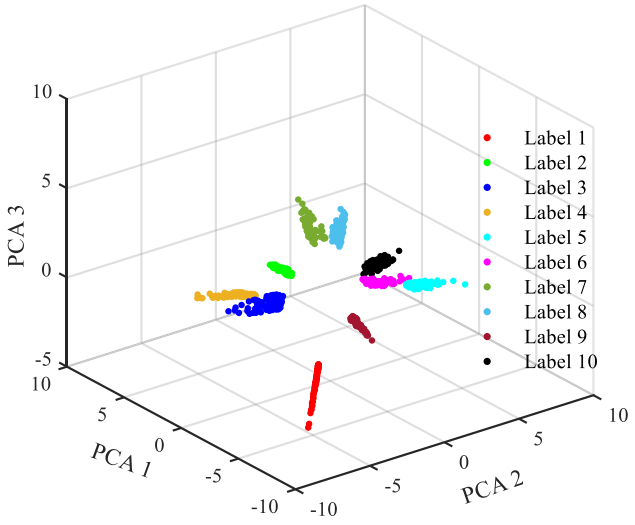

Fig. 17. Output layer output result visualization in SSGCN.

are 0.007, 0.014 and 0.021 in., respectively. Therefore, there are ten different bearing conditions including the normal condition. Each bearing condition contains 200 samples under the four different loads, of which the first 100 are used as the training samples and the remaining 100 are regarded as the testing samples. Each sample contains 2048 sampling points. The detailed description of the ten bearing conditions is shown in Table 6.

4.2.2. Results and analysis

For testing the diagnosis performance of our proposed method on different bearing conditions under a few labeled samples and different working loads, the 5 samples are randomly selected from the 100 training samples of each bearing condition as the labeled samples, and the remaining training samples as the unlabeled samples. Our proposed method is applied to process the training dataset containing 50 labeled samples and 950 unlabeled samples, and the target loss and diagnosis accuracy of the labeled sample set and unlabeled sample set during the training process are shown in Fig. 13, which again indicate that the SSGCN model has a fast convergence speed and good stability. Same as the gear experiments, the termination condition is set as the total number of iterations reaching 100 in the bearing experiments.

For the testing samples, the multi-class confusion matrix of our proposed method is plotted in Fig. 14. Only 16.00% of the testing samples in condition (6) are misidentified as condition (10), and the diagnosis accuracies are both 100% for other bearing conditions. On the whole, our proposed method still has a very high diagnosis accuracy of 98.40%. When the label ratio is increased from 0.05 to 0.45, all the testing samples in condition (6) can be always diagnosed correctly, and the overall diagnosis accuracy of our proposed method is stable at 100%. Thus seen, our proposed method can effectively realize the semi-supervised classification of different bearing conditions under a small number of labeled samples. As shown in Figs. 15–17, the initial features (1024 dimensions), the features in the first convolutional layer (192 dimensions) and the features in the output layer (10 dimensions) are also visualized by the PCA. Comparing Fig. 16 with Fig. 15, we can see that the features in the first convolutional layer are more recognizable than the initial features. Comparing Fig. 17 with Fig. 16, we can see that the discrimination ability of all bearing conditions is further improved after the feature extraction of the two-layer SSGCN model, and only a few testing samples in conditions (6) and (10) overlap each other, which is consistent with the diagnosis result of the confusion matrix in Fig. 14. These demonstrate that even if there are a few labeled samples, the two-layer SSGCN model can still learn and extract the useful information hidden in the original high-dimensional features through the operation of spectral graph convolution for bearings.

Table 7
Diagnosis results of the five methods under different label ratios.

Label ratios	Diagnosis methods				
	Method 1 (Our proposed method)	Method 2 (CNN: LeNet-5)	Method 3 (Proposed in [11])	Method 4 (Proposed in [12])	Method 5 (Proposed in [13])
0.05	98.20 ± 2.13	56.60 ± 8.13	58.10 ± 6.57	65.10 ± 5.66	64.40 ± 6.12
0.15	99.00 ± 1.35	79.50 ± 6.13	80.50 ± 4.18	88.80 ± 4.09	84.30 ± 4.79
0.25	99.80 ± 1.11	84.80 ± 5.13	87.60 ± 2.56	92.70 ± 3.45	88.30 ± 3.85
0.35	99.90 ± 0.27	90.00 ± 2.73	92.40 ± 2.29	95.20 ± 3.03	93.90 ± 3.24
0.45	100 ± 0	94.20 ± 2.16	95.30 ± 2.08	97.10 ± 2.77	96.60 ± 2.98
0.55	100 ± 0	94.70 ± 1.93	98.40 ± 1.33	99.20 ± 0.84	98.70 ± 1.01
0.65	100 ± 0	97.00 ± 1.14	100 ± 0	100 ± 0	100 ± 0
0.75	100 ± 0	98.10 ± 0.51	100 ± 0	100 ± 0	100 ± 0
0.85	100 ± 0	100 ± 0	100 ± 0	100 ± 0	100 ± 0

Note: The format of the diagnosis result is average diagnosis accuracy (%) ± standard deviation.

4.2.3. Comparison with deep learning methods

For fully testing the diagnosis performance of our proposed method on the ten bearing conditions under different number of labeled samples, the label ratio *lr* also changes from 0.05 to 0.85 and the step size is 0.1. That is to say, the *lr**100 samples are randomly selected from the 100 training samples of each bearing condition as the labeled samples, and the remaining training samples as the unlabeled samples. Same as the previous experiment, our proposed method and the four comparison methods are also used to process the nine bearing datasets with different label ratios. Their diagnosis results are displayed in Table 7.

From Table 7, we can also observe that with the increase of label ratio, the average diagnosis accuracies of the four comparison methods are significantly improved, and the average diagnosis accuracy of our proposed method is stable above 98%. More importantly, the lower the label ratio, the more obvious the advantage of our proposed method in diagnosis accuracy. In addition, the standard deviation of our proposed method is still smaller than that of other methods in diagnosis accuracy, which indicates that the two-layer SSGCN model is also more stable in bearing fault diagnosis. Likewise, the deeper reason for these advantages is that our proposed method, a semi-supervised fault diagnosis method, can make full use of the relationship between samples.

5. Conclusions

The SSGCN is a semi-supervised deep learning algorithm, which can use the category information of labeled samples, the graph adjacency relationship between all samples and the graph convolution operation to realize the automatic feature extraction and pattern recognition. In this paper, an intelligent fault diagnosis method based on SSGCN is proposed for rotating machinery. To ensure the intelligence and improve the diagnosis efficiency and accuracy, the vibration signals in frequency domain are used to form samples without manual feature extraction. To better reflect the local geometry property between all samples, a KNN graph weighted by the heat kernel is constructed by treating each sample as a node. Finally, all samples and the graph adjacency matrix are taken as the input of the two-layer SSGCN model. In addition, the cross-entropy loss function of labeled samples is used as the objective function to update the weight parameters, and other critical hyper-parameters are determined by the 5-CV. The performance of our proposed method in the fault diagnosis of rotating machinery is comprehensively evaluated through two experimental cases, and the following conclusions can be drawn.

1) Even if there are a small number of labeled samples, the two-layer SSGCN model can adaptively extract the available fault features from the raw vibration signals in frequency domain for gears and bearings, and has a fast convergence speed and good stability.
2) Without manual feature extraction, SSGCN has better application in gear and bearing fault diagnosis compared with the state-of-the-art deep learning methods.
3) There are some regular relations between some critical hyper-parameters and the diagnosis performance of our proposed method. For example, when the most sensitive parameter, namely learning rate *L*, is selected between 0.01 and 0.1, our proposed method will achieve a lower recognition error rate. In the future, the parameter settings of SSGCN will be further explored.

CRediT authorship contribution statement

Yiyuan Gao: Conceptualization, Methodology, Software, Writing – original draft. **Mang Chen:** Data curation. **Dejie Yu:** Supervision.

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 study was supported by the National Natural Science Foundation of China (No. 51875182) and Foundation for University Key Teacher of Henan Educational Department (No. 2018GGJS091). The valuable and constructive comments from the reviewers are sincerely appreciated for their help to further improve this paper.

References

[1] Z.Y. He, H.D. Shao, X. Zhong, et al., Ensemble transfer CNNs driven by multi-channel signals for fault diagnosis of rotating machinery cross working conditions, Knowl.-Based Syst. 207 (2020), 106396.
[2] H.Y. Pan, J.D. Zheng, Y. Yang, et al., Nonlinear sparse mode decomposition and its application in planetary gearbox fault diagnosis, Mech. Mach. Theory 155 (2021), 104082.
[3] M. Zeng, Z. Chen, SOSO boosting of the K-SVD denoising algorithm for enhancing fault-induced impulse responses of rolling element bearings, IEEE Trans. Ind. Electron. 67 (2020) 1282–1292.
[4] X. Li, Y. Yang, H.D. Shao, et al., Symplectic weighted sparse support matrix machine for gear fault diagnosis, Measurement 168 (2021), 108392.
[5] D.Y. Dou, S.S. Zhou, Comparison of four direct classification methods for intelligent fault diagnosis of rotating machinery, Appl. Soft. Comput. 46 (2016) 459–468.
[6] J.D. Zheng, H.Y. Pan, Use of generalized refined composite multiscale fractional dispersion entropy to diagnose the faults of rolling bearing, Nonlinear Dyn. 101 (2020) 1417–1440.
[7] G.F. Bin, J.J. Gao, X.J. Li, et al., Early fault diagnosis of rotating machinery based on wavelet packets-Empirical mode decomposition feature extraction and neural network, Mech. Syst. Sig. Process. 27 (2012) 696–711.
[8] D. Wang, K-nearest neighbors based methods for identification of different gear crack levels under different motor speeds and loads: Revisited, Mech. Syst. Sig. Process. 70–71 (2016) 201–208.
[9] Z.Q. Su, B.P. Tang, J.H. Ma, et al., Fault diagnosis method based on incremental enhanced supervised locally linear embedding and adaptive nearest neighbor classifier, Measurement 48 (2014) 136–148.
[10] O. Janssens, V. Slavkovikj, B. Vervisch, et al., Convolutional neural network based fault detection for rotating machinery, J. Sound Vib. 377 (2016) 331–345.
[11] W. Zhang, G.L. Peng, C.H. Li, et al., A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals, Sensors 17 (2) (2017) 425–445.
[12] H.D. Shao, H.K. Jiang, H.Z. Zhang, et al., Electric locomotive bearing fault diagnosis using a novel convolutional deep belief network, IEEE Trans. Ind. Electron. 65 (2018) 2727–2736.
[13] H.D. Shao, M. Xia, J.F. Wan, et al., Modified stacked auto-encoder using adaptive Morlet wavelet for intelligent fault diagnosis of rotating machinery, IEEE-ASME Trans. Mech. doi: 10.1109/TMECH.2021.3058061.

[14] M. Niepert, M. Ahmed, K. Kutzkov, Learning convolutional neural networks for graphs, in: Proceedings of the 33rd annual international conference on machine learning. ACM, 2016.
[15] J. Bruna, W. Zaremba, A. Szlam, et al., Spectral networks and locally connected networks on graphs, 2013. [Online.] Available: https://arXiv.1312.6203.
[16] M. Defferrard, X. Bresson, P. Vandergheynst, Convolutional neural networks on graphs with fast localized spectral filtering, in: Advances in neural information processing systems, 2016, pp. 3844–3852.
[17] A. Ortega, P. Frossard, J. Kovacevic, et al., Graph signal processing: overview, challenges, and applications, Proc. IEEE 106 (5) (2018) 808–828.
[18] D.C. Zhang, E. Stewart, M. Entezami, et al., Intelligent acoustic-based fault diagnosis of roller bearings using a deep graph convolutional network, Measurement 156 (2020), 107585.
[19] X.X. Yu, B.P. Tang, K. Zhang, Fault diagnosis of wind turbine gearbox using a novel method of fast deep graph convolutional networks, IEEE Trans. Instrum. Meas. 70 (2021) 1–14.
[20] T.F. Li, Z.B. Zhao, C. Sun, et al., Multi-receptive field graph convolutional networks for machine fault diagnosis, IEEE Trans. Ind. Electron. doi: 10.1109/TIE.2020.3040669.
[21] T.N. Kipf, M. Welling, Semi-supervised classification with graph convolutional networks, 2016. [Online.] Available: https://arXiv.1609.02907.
[22] Q. Li, Z. Han, X.-M. Wu, Deeper insights into graph convolutional networks for semi-supervised learning, in: Proceedings of the AAAI Conference on Artificial Intelligence, 2018, pp. 1929–1958.
[23] Y.Y. Gao, D.J. Yu, Intelligent fault diagnosis for rolling bearings based on graph shift regularization with directed graphs, Adv. Eng. Inform. 47 (2021), 101253.
[24] Q.S. Jiang, M.P. Jia, J.Z. Hu, et al., Machinery fault diagnosis using supervised manifold learning, Mech. Syst. Signal Process. 23 (2009) 2301–2311.
[25] J.D. Zheng, H.Y. Pan, S.B. Yang, et al., Generalized composite multiscale permutation entropy and Laplacian score based rolling bearing fault diagnosis, Mech. Syst. Signal Process. 99 (2018) 229–243.
[26] U.V. Luxburg, A tutorial on spectral clustering, Stat. Comput. 17 (2007) 395–416.
[27] D.I. Shuman, S.K. Narang, P. Frossard, et al., The emerging field of signal processing on graphs: extending high-dimensional data analysis to networks and other irregular domains, IEEE Signal Proc. Mag. 30 (2013) 83–98.
[28] Y.Y. Gao, D.J. Yu, H.J. Wang, Fault diagnosis of rolling bearings using weighted horizontal visibility graph and graph Fourier transform, Measurement 149 (2020) 12.
[29] D.K. Hammond, P. Vandergheynst, R. Gribonval, Wavelets on graphs via spectral graph theory, Appl. Comput. Harmon. A. 30 (2009) 129–150.
[30] X. Zhao, M. Jia, Z. Liu, Semi-supervised deep sparse auto-encoder with local and non-local information for intelligent fault diagnosis of rotating machinery, IEEE Trans. Instrum. Meas. 70 (2021) 1–13.
[31] X. Li, Y. Yang, N.Q. Hu, et al., Discriminative manifold random vector functional link neural network for rolling bearing fault diagnosis, Knowl.-Based Syst. 211 (2021), 106507.
[32] H.Y. Pan, J.D. Zheng, An intelligent fault diagnosis method for roller bearing using symplectic hyperdisk matrix machine, Appl. Soft. Comput. 105 (2021), 107284.
[33] L. Ou, D.J. Yu, H.J. Yang, A new rolling bearing fault diagnosis method based on GFT impulse component extraction, Mech. Syst. Sig. Process. 81 (2016) 162–182.
[34] http://csegroups.case.edu/bearingdatacenter/pages/welcome-case-western-reserve-university-bearing-data-center-website.
[35] W.A. Smith, R.B. Randall, Rolling element bearing diagnostics using the Case Western Reserve University data: a benchmark study, Mech. Syst. Sig. Process. 64–65 (2015) 100–131.