IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 6, 2019, accepted October 7, 2019, date of publication October 21, 2019, date of current version October 31, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2948661

A New Deep Fusion Network for Automatic Mechanical Fault Feature Learning

YUMEI QI[1], CHANGQING SHEN[2], JUN ZHU[3], XINGXING JIANG[2], JUANJUAN SHI[2], AND ZHONGKUI ZHU[2]

[1]Wenzheng College, Soochow University, Suzhou 215104, China
[2]School of Urban Rail Transportation, Soochow University, Suzhou 215131, China
[3]Department of Industrial Systems Engineering and Management, National University of Singapore, Singapore 119077

Corresponding author: Changqing Shen (cqshen@suda.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 51875375 and Grant 51875376, in part by the China Postdoctoral Science Foundation funded project under Grant 2019T120456, and in part by the Suzhou Science Foundation under Grant SYG201802.

ABSTRACT Mechanical fault diagnosis is essential in ensuring the safety of production and economic development. In the field of fault diagnosis, deep learning has been extensively used due to its excellent feature learning ability. However, it still suffers from several issues; for example, 1) simultaneous requirements of features from multiple aspects, including sparsity and robustness, are hardly met due to the limited feature learning ability of a single model, and 2) most methods deal with preprocessed signals instead of original time domain signals because of the noise interference and deficiency of a single model. To solve these problems, this study proposes a new deep fusion network for fault feature learning, which combines two types of deep learning models, namely, sparse autoencoder and contractive autoencoder, which are respectively applied to enhance features' sparsity and robustness and thereby guarantee the representativeness of extracted features and gain strong anti-interference ability. Consequently, fault diagnosis with original time domain signals can be realized. Bearing and gearbox fault diagnosis experiments are conducted to verify the performance of the presented network. Results show that the diagnosis accuracies for two cases are higher than those of networks based on single contractive autoencoder and sparse autoencoder. These results demonstrate that the proposed fusion network has superior feature learning ability relative to single model networks and can deal with original time domain signals by simultaneously enhancing features' sparsity and robustness.

INDEX TERMS Deep fusion network, feature learning, fault diagnosis, robustness and sparsity enhancement.

I. INTRODUCTION

Health monitoring of rotating machines in engineering systems can guarantee sustainable economic development [1], [2]. Hen faults are not monitored, they may lead to economic losses or even casualties [3], [4]. Thus, machine fault diagnosis is crucial. The diagnostic process consists of two major steps: extracting features and recognizing faults. Methods based on signal analysis and processing, as well as artificial intelligence (AI), are commonly used fault diagnosis methods. There are various signal-analysis-and -processing-based fault diagnosis methods [5], [6]. Multivariate statistical process monitoring methods are effective in fault diagnosis. Wang *et al.* [7] gave a detailed discussion on these methods and pointed out several promising directions. In addition, to deal with the non-Gaussian features, Li *et al.* [8] proposed a weighted preliminary-summation-based principle component analysis method for fault detection, which improved the fault detection rate and shortened fault detection time. The latter, which is an intelligent fault recognition method, has recently attracted the attention of researchers. Support vector machine (SVM) [9] and artificial neural network (ANN) [10] are the commonly used traditional AI fault diagnosis methods. Soualhi *et al.* [11] preprocessed bearing signals with Hilbert-Huang transform and then used SVM to detect faults. Zheng *et al.* [12] applied multiscale fuzzy entropy and SVM for bearing fault detection. Li *et al.* [13] applied ANN for fault classification based on 12 artificially

The associate editor coordinating the review of this manuscript and approving it for publication was Youqing Wang.

extracted sensitive features. Moosavian *et al.* [14] detected four different bearing fault conditions by combining discrete wavelet transform and ANN. Despite the success of the applications of these AI fault diagnosis methods, they are still limited by their shallow architecture, which restricts them from learning distinctive features automatically. This limitation makes them rely heavily on preprocessing methods for feature extraction. However, selecting the most suitable features artificially is difficult and time-consuming [15].

As a solution, deep learning (DL) has been gradually developed, resulting in methods such as deep belief network (DBN), autoencoder (AE), and convolutional neural network (CNN), all of which can extract essential features automatically [16], [17] and are widely applied in fault diagnosis [18], [19]. Extensive research indicates that DL diagnosis methods have superior learning ability and better diagnosis performance than traditional AI methods. Zhang *et al.* [20] proposed a deep CNN for bearing fault detection under noisy environment and changing loads. Yuan *et al.* [21] successfully combined Hilbert-Huang transform for preprocessing and CNN for bearing fault recognition. Jia *et al.* [22] applied AE for mechanical fault diagnosis with frequency features obtained through Fourier transform. Guo *et al.* [23] and Tang *et al.* [24] proposed an adaptive learning rate algorithm for CNN and DBN to improve the convergence rates during training and achieved excellent fault diagnosis performance. However, most DL methods are still combined with preprocessing methods, that is, the inputs of the network are not original vibration signals [21]–[24]. Shao *et al.* [25] constructed a new DL network combining two types of DL models for fault diagnosis, and the results showed that the diagnosis accuracy of the proposed method is 8% and 5% higher than that of the network constructed with one of the types of DL models, respectively. It shows that a single DL model has a limited feature learning ability. Thus, preprocessing methods are usually combined with DL networks consisting of single type of DL model to help mine representative features. In addition, noise is a major threat for DL methods to extract discriminative features without preprocessing, especially when the DL networks are constructed with single type of DL model. Thus, the reasons for the above-mentioned DL methods' drawback can be summarized as follows: 1) a single DL model has limited feature learning ability and cannot extract features from different aspects, and 2) original vibration signals are overwhelmed by noise, and thus, applying a single DL model to extract discriminative features with original signals is difficult. Considering that the collected signals are original vibration signals in the time domain and command for timeliness of fault diagnosis, automatic fault diagnosis must be realized directly with original time domain signals. Research shows that among DL models, the contractive autoencoder (CAE) is superior in grasping internal factors that can directly extract robust features [26] and is thus suitable for dealing with noise-overwhelmed signals. Similarly, the sparse autoencoder (SAE) is claimed to be superior in extracting sparser features, highly discriminative, and useful for classification [27]. Thus, this study is based on AE and presents a deep fusion feature learning network that combines the CAE and SAE to extract robust and sparse features and then directly realize fault diagnosis with original collected signals overwhelmed with noise. First, multiple CAEs are stacked to deal with original signals for robust feature extraction. Second, multiple SAEs are stacked to deal with the extracted robust features for sparse and discriminative feature extraction. Finally, a softmax classifier is used for fault recognition on the basis of these learned features. Considering the important role of bearings and gearboxes in machinery, we investigate two fault diagnosis cases with bearing and gearbox signals to verify the validity and superiority of the proposed fusion fault diagnosis network relative to single models. Moreover, network architecture and parameter selection are fully discussed. The following are the summaries of the most contributions of this paper.

1) A new type of deep fusion fault feature learning and diagnosis network combining the CAE and SAE is proposed to automatically enhance the robustness and sparsity of the mined features.

2) The fusion fault diagnosis network can realize fault diagnosis with original time domain signals without preprocessing. Thus, it exhibits better diagnosis performance than networks based on traditional models.

The remainder of this paper is organized as follows. Section 2 briefly introduces the theoretical knowledge on SAE and CAE. Section 3 discusses the structure and diagnosis procedure of the proposed fusion network. Section 4 presents the cases of bearing and gearbox fault diagnosis to verify the effectiveness and preponderance of the proposed network. Section 5 comprises discussions about network architecture and parameter selection. Section 6 presents the conclusions.

II. BRIEF INTRODUCTION TO SPARSE AUTOENCODER AND CONTRACTIVE AUTOENCODER

A. BASIC AUTOENCODER

The AE is a typical DL model for unsupervised feature learning that contains three layers as shown in Fig. 1. The first two layers make up the encoder network, and the latter two layers comprise the decoder network. The second layer, which is also referred to as the hidden layer, aims to learn the compressed features of original inputs.

For datasets $\{\boldsymbol{x}^d\}_{d=1}^M$, if $\boldsymbol{x}^d$ denotes one of the input vectors, then the compressed vector $\boldsymbol{h}^d$ and reconstruction vector $\hat{\boldsymbol{x}}^d$ are calculated as in

$$\boldsymbol{h}^d = f(\boldsymbol{W}^{(1)}\boldsymbol{x}^d + \boldsymbol{b}^{(1)}) \tag{1}$$

$$\hat{\boldsymbol{x}}^d = f(\boldsymbol{W}^{(2)}\boldsymbol{h}^d + \boldsymbol{b}^{(2)}) \tag{2}$$

where $\boldsymbol{W}^{(1)}$ and $\boldsymbol{W}^{(2)}$ are weight matrixes, $\boldsymbol{b}^{(1)}$ and $\boldsymbol{b}^{(2)}$ are bias vectors, and f denotes the activation function.

$L(\boldsymbol{x}^d, \hat{\boldsymbol{x}}^d)$ represents the reconstruction error between $\hat{\boldsymbol{x}}^d$ and $\boldsymbol{x}^d$ and is expressed as in

$$L(\boldsymbol{x}^d, \hat{\boldsymbol{x}}^d) = \frac{1}{2}\left\|\boldsymbol{x}^d - \hat{\boldsymbol{x}}^d\right\|^2 \tag{3}$$

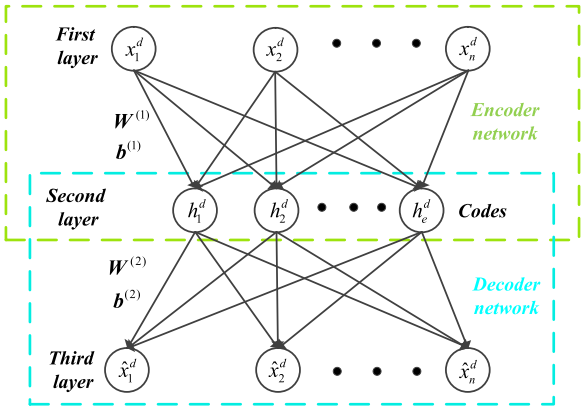


FIGURE 1. AE architecture.

Considering the over-fitting problem, weight decay penalty term is added to form the overall cost function, which can be obtained as in

$$J_{AE}(\boldsymbol{W},\boldsymbol{b}) = \left[\frac{1}{M}\sum_{d=1}^{M} L(\boldsymbol{x}^d, \hat{\boldsymbol{x}}^d)\right] + \frac{\lambda}{2}\sum_{l=1}^{n_l-1}\sum_{i=1}^{S_l}\sum_{j=1}^{S_{l+1}} (\boldsymbol{W}_{ji}^{(l)})^2 \tag{4}$$

where λ is the penalty coefficient; n_l and s_l are the numbers of network layer and neurons in layer l, respectively; and $\boldsymbol{W}_{ji}^{(l)}$ is the connecting weight between neuron i and j in layers $l+1$ and l.

Continuous encoding and decoding using the back propagation (BP) algorithm until the cost function is minimized to a certain acceptable extent results in the optimal parameters $\{\boldsymbol{W}^{(1)}, \boldsymbol{b}^{(1)}, \boldsymbol{W}^{(2)}, \boldsymbol{b}^{(2)}\}$. $\boldsymbol{h}^d$ calculated as in (1) is referred to as the feature of the original input extracted by the AE because it contains useful information about the original input vector $\boldsymbol{x}^d$. The minimization of the cost function is also called the training process of the AE.

B. SPARSE AUTOENCODER AND CONTRACTIVE AUTOENCODER

The AE simply copies the inputs. Although the features can perfectly reconstruct inputs without errors, they are not guaranteed as discriminative and representative enough for classification. Therefore, the SAE is proposed.

The SAE and AE share the same structure, but their cost functions differ. The cost function of the SAE adds a sparsity penalty term, which causes the inactivity of hidden neurons and defined as in

$$\sum_{g=1}^{S_2} KL(\rho||\hat{\rho}_g) = \sum_{g=1}^{S_2}\left(\rho\log\frac{\rho}{\hat{\rho}_g} + (1-\rho)\log\frac{1-\rho}{1-\hat{\rho}_g}\right) \tag{5}$$

where ρ is the sparsity parameter that denotes an extremely small value given artificially. $\hat{\rho}_g$ denotes the average activation value of hidden neuron g and is defined as in

$$\hat{\rho}_g = \frac{1}{M}\sum_{d=1}^{M} h_g^d \tag{6}$$

The cost function of the SAE can be calculated as in

$$J_{SAE}(\boldsymbol{W},\boldsymbol{b}) = J_{AE}(\boldsymbol{W},\boldsymbol{b}) + \beta\sum_{g=1}^{S_2} KL(\rho||\hat{\rho}_g) \tag{7}$$

where β is the penalty coefficient.

The training process of the SAE is the same as that of the AE. After the training process, the features learned by the SAE not only contain most of the valid information of the original inputs but also prevent redundancy to achieve better sparsity than those learned by the AE. Therefore, the SAE can extract more discriminative features by enhancing their sparsity.

The CAE is proposed to extract robust features. The cost function of the CAE is obtained as in

$$J_{CAE}(\boldsymbol{W},\boldsymbol{b}) = \frac{1}{M}\sum_{d=1}^{M}\left[L(\boldsymbol{x}^d,\hat{\boldsymbol{x}}^d) + \gamma\left\|J_f(\boldsymbol{x}^d)\right\|_F^2\right] \tag{8}$$

where $J_f(\boldsymbol{x}^d)$ denotes the Jacobian matrix defined as in

$$J_f(\boldsymbol{x}^d) = \begin{bmatrix} \frac{\partial h_1^d}{\partial x_1^d} & \cdots & \frac{\partial h_1^d}{\partial x_n^d} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_e^d}{\partial x_1^d} & \cdots & \frac{\partial h_e^d}{\partial x_n^d} \end{bmatrix} \tag{9}$$

where e denotes the number of hidden layer neurons as shown in Fig. 1. The second term in (8) denotes the contractive penalty term, which is equivalent to the Frobenius norm of $J_f(\boldsymbol{x}^d)$. γ is the penalty coefficient that controls the relative weight of the reconstruction error and contractive term.

During the training process, the CAE aims to grasp the internal factor to directly and automatically learn robust features [28], [29] by suppressing the impacts of inputs' changes on learned features.

In practical applications, multiple AEs, SAEs, or CAEs are stacked under deep architectures to transform inputs into abstract and essential features several times because a model only realizes feature transformation once. Fig. 2 shows the architecture of the stacked AE. The first two layers is the encoder network of AE1 for extracting feature1 at a low level. The second and third layer is the encoder network of AE2 for extracting feature2 and so on. Through multiple feature transformations, the most abstract and essential high-level features can be mined at the penultimate level and are then used for classification. Multiple SAEs or CAEs can also be stacked the same way to a stacked SAE or CAE network.

III. PROPOSED DEEP FUSION FEATURE LEARNING NETWORK

In this study, a new deep fusion fault feature learning and diagnosis network combining stacked SAE and CAE is proposed to enhance the sparsity and robustness of learned features for high-precision fault diagnosis with original time domain signals.

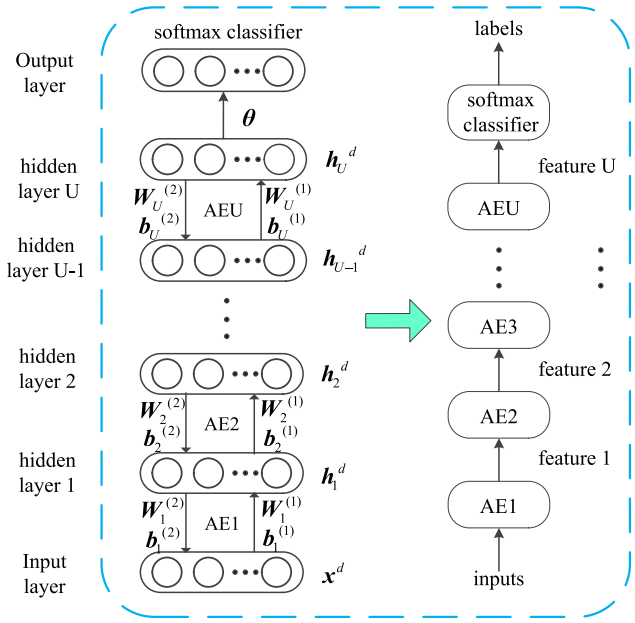


FIGURE 2. Architecture of stacked AE.

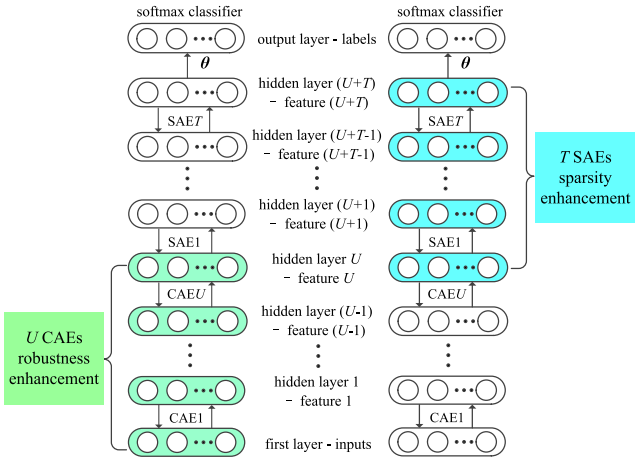


FIGURE 3. Architecture of feature-enhanced fusion diagnosis network.

The key of the network structure is to determine whether to enhance the sparsity or robustness of features first, that is, to place the stacked SAE or CAE in the front layers of the network. As inputs are original time domain signals containing a large amount of noise, CAEs are placed in the front layers of the network for robustness enhancement of features firstly, and then SAEs are placed behind for sparsity enhancement. Finally, a softmax classifier is stacked on the top layer for different health condition classification based on the mined features. Fig. 3 shows a fusion feature enhanced diagnosis network combining U CAEs and T SAEs.

Similar to other DL networks based on single type of model, the training process of the fusion network also involves pre-training and fine-tuning. Pre-training is applied for weight initialization, whereas fine-tuning is applied to learn the most suitable transformation relationship between extracted high-level features and ideal labels based on initialized weights, which is crucial for the precision of the classification. In detail, all CAEs and SAEs are individually pre-trained one by one. Similarly, the softmax classifier is pre-trained according to the error between the actual outputs and the ideal labels. Then, the weights are initialized, and the whole network is fine-tuned with BP algorithm to obtain optimal parameters, which can correctly reflect the transformation relationship between the inputs and the ideal output labels. So far, the proposed fusion network is now well-trained. After each iteration, the weight $W_{ij}^{(l)}$ and basis $b_i^{(l)}$ can be respectively updated as in

$$W_{ij}^{(l)} = W_{ij}^{(l)} - \alpha \frac{\partial J(W,b)}{\partial W_{ij}^{(l)}} \tag{10}$$

$$b_i^{(l)} = b_i^{(l)} - \alpha \frac{\partial J(W,b)}{\partial b_i^{(l)}} \tag{11}$$

where α denotes the learning rate.

In this study, the sigmoid activation function is used; thus, $\frac{\partial h_j^d}{\partial x_i^d}$ in (9) can be further expressed as in

$$\frac{\partial h_j^d}{\partial x_i^d} = h_j^d(1-h_j^d)W_{ji}^{(1)} \tag{12}$$

Thus, the Frobenius norm of $J_f(\boldsymbol{x}^d)$ is simplified as in

$$\left\|J_f(\boldsymbol{x}^d)\right\|_F^2 = \sum_{i=1}^{n}\sum_{j=1}^{e}\left[h_j^d(1-h_j^d)W_{ji}^{(1)}\right]^2 \tag{13}$$

For the hidden layer of the SAE, the inhibition or activation of the hidden neuron g can be determined by sparsity parameter ρ. If the activation value h_g is greater than ρ, then this neuron can be considered as active; otherwise, it is inactive. Thus, the sparsity of the features learned in the hidden layer can be defined as in

$$Sp = \frac{No_{inactive}}{No_{all}} \tag{14}$$

where $No_{inactive}$ and No_{all} respectively denote the number of inactive neurons and all neurons of the hidden layer.

Fig. 4 shows the diagnosis flow chart of the proposed fusion diagnosis network. The inputs of the network are original signals collected in the time domain without preprocessing. First, these inputs are randomly divided into two parts, one part for training and another for testing the constructed network. Second, the proposed fusion network is trained with training datasets through pre-training and fine-tuning. Finally, the diagnosis performance of the well-trained fusion network can be validated using testing datasets.

IV. EXPERIMENTAL VALIDATION

A. CASE 1: BEARING FAULT DIAGNOSIS

1) DATA INTRODUCTION

Datasets containing different health conditions collected from a bearing fault simulation test rig are used to verify the

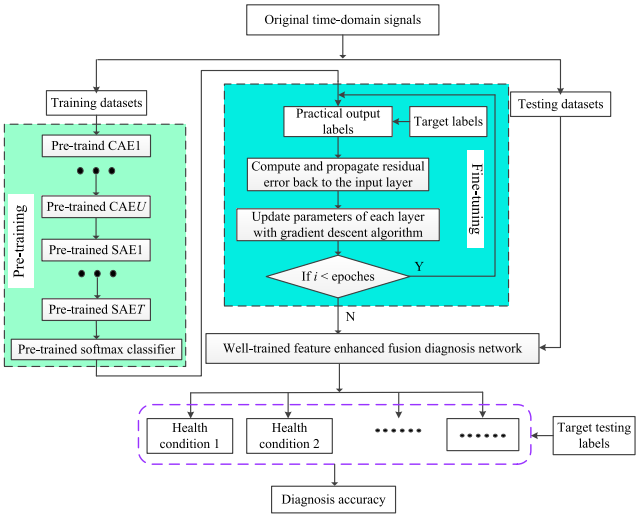


FIGURE 4. Diagnosis flow chart of the proposed fusion network.

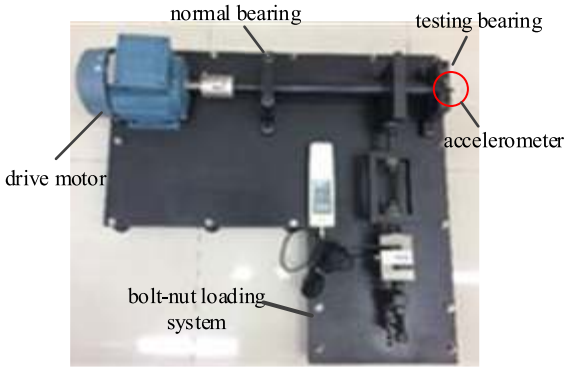


FIGURE 5. Bearing test rig.

TABLE 1. Description of bearing datasets.

Datasets	Training sample No.	Testing sample No.	Labels
IF	200	100	1
OF	200	100	2
BF	200	100	3
N	200	100	4

performance of the presented diagnosis network. Fig. 5 shows the test rig, which includes a drive motor, bolt-nut loading system, a normal bearing, a testing bearing, an accelerometer, and a data acquisition system. The defect with a width of 0.4 mm is artificially set at the inner race, outer race, and rolling element through wire cutting. During sampling, the accelerometer is placed on the test bearing pedestal and located in the direction of 12 points. The speed of the motor without load is 961 rpm, the sampling frequency is 10 kHz, and the sampling points are 1000. Four health condition signals are collected: inner race fault (IF), outer race fault (OF), ball fault (BF), and normal condition (N) as shown in Fig. 6. The detailed data information and label settings are shown in Table 1.

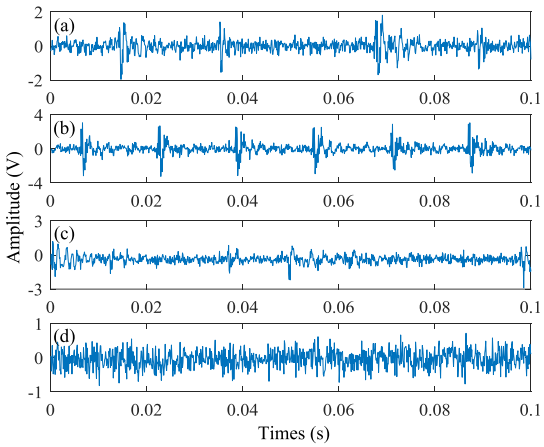


FIGURE 6. Bearing time domain signals: (a) IF, (b) OF, (c) BF, and (d) N.

2) VALIDATION RESULTS AND ANALYSIS

The inputs of the fusion network are original time domain signals, thus the input size is 1000. It is finally determined that the fusion network consists of three hidden layers, including two CAEs and one SAE. Contractive coefficient and sparsity parameter are respectively set to 10^{-4} and 0.1. Table 2 shows other specific network parameters, and Fig. 7 shows the diagnostic result after testing. By observation, only four IF samples are misdiagnosed as normal. The diagnosis accuracy for the IF testing samples is 96%, whereas the other three health condition testing samples are all classified correctly. Thus, the overall testing accuracy is 99%.

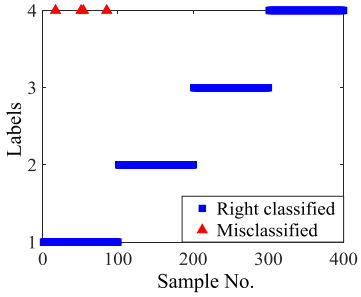


FIGURE 7. Bearing fault diagnosis result of fusion network.

TABLE 2. Parameter settings of the fusion network.

Input size	Output size	Hidden layer size
1000	4	600-200-100
Learning rate of hidden layer	Learning rate of classifier	Learning rate of fine-tuning
0.2	0.2	0.2
Iteration No. of hidden layer	Iteration No. of classifier	Iteration No. of fine-tuning
10000	30000	40000

Comparisons between our proposed fusion network and the standard deep SAE and CAE network, as well as other

commonly used diagnosis methods such as deep AE network and traditional AI methods with shallow architecture are conducted to demonstrate the advantage of our proposed fusion network in feature enhancement and fault diagnosis. The architecture and parameter settings of the deep SAE, CAE, and AE networks are the same as those of the proposed network. ANN is a typical traditional AI method with shallow architecture; thus, it is set to be a single hidden layer network with a hidden size of 600. In addition, ANN with 3 hidden layers is also used for comparison. Epsilon-SVR and radial basis function kernel are adopted in this study. In addition, Fast Fourier Transform (FFT) is used as the preprocessing method and combined with one-hidden-layer ANN and SVM for comparison. Table 3 presents the diagnosis accuracies of the abovementioned methods. In sum, the proposed fusion network achieves the highest diagnosis accuracy when dealing with original time-domain signals.

Making a concrete analysis, the first four methods are DL methods, whereas the remaining five belong to traditional AI methods with shallow architecture. When dealing with original collected signals, the diagnosis accuracies achieved by the DL methods are at least 20% higher than those achieved by SVM and ANN because deep architectures are conducive to the multi-transformation of features. In conclusion, DL methods possess a powerful feature learning ability, and the features automatically mined by these methods are essential and representative enough for classification.

In addition, deep CAE and SAE networks both perform better than the deep AE network, especially deep CAE. The diagnosis accuracies achieved by the deep CAE and SAE networks are 8.25% and 2.25% higher than that achieved by the deep AE network, respectively. This result indicates that adding specific robustness or sparsity constraints to the AE is beneficial to enhance the robustness or sparsity of features and extract more expressive features. Moreover, for time domain signals with much noise, the robustness enhancement of features is sometimes more important. Meanwhile, the diagnosis accuracy of the fusion network consisting of CAE and SAE is 6.5% and 12.5% higher than that of the networks based on single CAE and SAE model, respectively, which validates that networks based on a single type of DL model has the limited feature learning ability. Therefore, the proposed fusion diagnosis network has the stronger feature learning ability and is able to combine the advantages of both CAE and SAE so as to enhance the robustness and sparsity of mined features and achieve a higher diagnosis accuracy than deep CAE and SAE networks when dealing with original signals. When the original signals are preprocessed by FFT, the diagnosis accuracies of SVM and one-hidden-layer ANN increase about 40%, which fully validates that preprocessing methods are beneficial for discriminative feature extraction. This is the reason why most DL methods are combined with preprocessing methods. Besides, it also shows the difficulty of feature extraction without preprocessing. Although the one-hidden-layer ANN combined with FFT achieves a 1% higher diagnosis accuracy than our proposed method, it takes some time for preprocessing and cannot get rid of human interference. In general, the proposed fusion network achieves superior diagnosis performance when dealing with original collected time-domain signals.

To visualize the sparsity enhancement of mined features, Fig. 8 illustrates the sparsity of the features learned in the third hidden layer of the proposed network and that of deep AE and CAE networks, that is, features extracted by SAE, AE, and CAE. The sparsity of the features is calculated using (14). About 60% of the extracted features in our proposed network are inactive, and this percentage is higher than that of the other two networks. Therefore, the SAE is proven to extract sparser features than the AE and CAE. The proposed network containing SAE is able to carry this characteristic of enhancing the sparsity of features to make them increasingly representative.

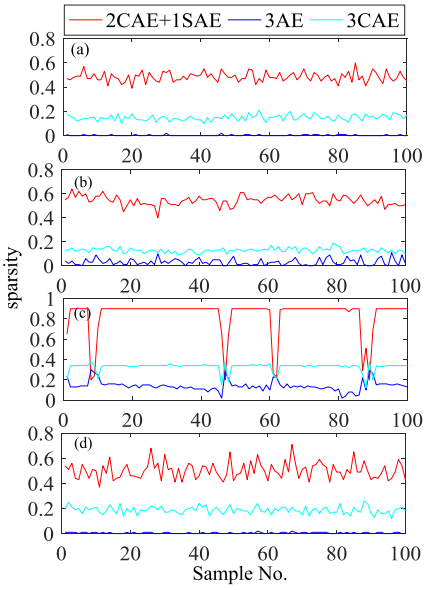


FIGURE 8. Sparsity of features extracted by different networks: (a) IF, (b) OF, (c) BF, and (d) N.

3) FEATURE VISUALIZATION

The features of original signals and those extracted in different hidden layers are visualized using principle component analysis (PCA) to validate the feature learning ability of the proposed fusion network. Fig. 9 shows the visualization result. By observation, the features of original signal cluster have the lowest quality among all features. However, by transforming the features thrice, the features learned in the third hidden layer cluster improve and become beneficial for classification. This phenomenon fully proves that the proposed network based on DL can automatically extract discriminative features. Moreover, this finding illustrates that deep architecture is beneficial in learning essential features with multiple feature transformations.

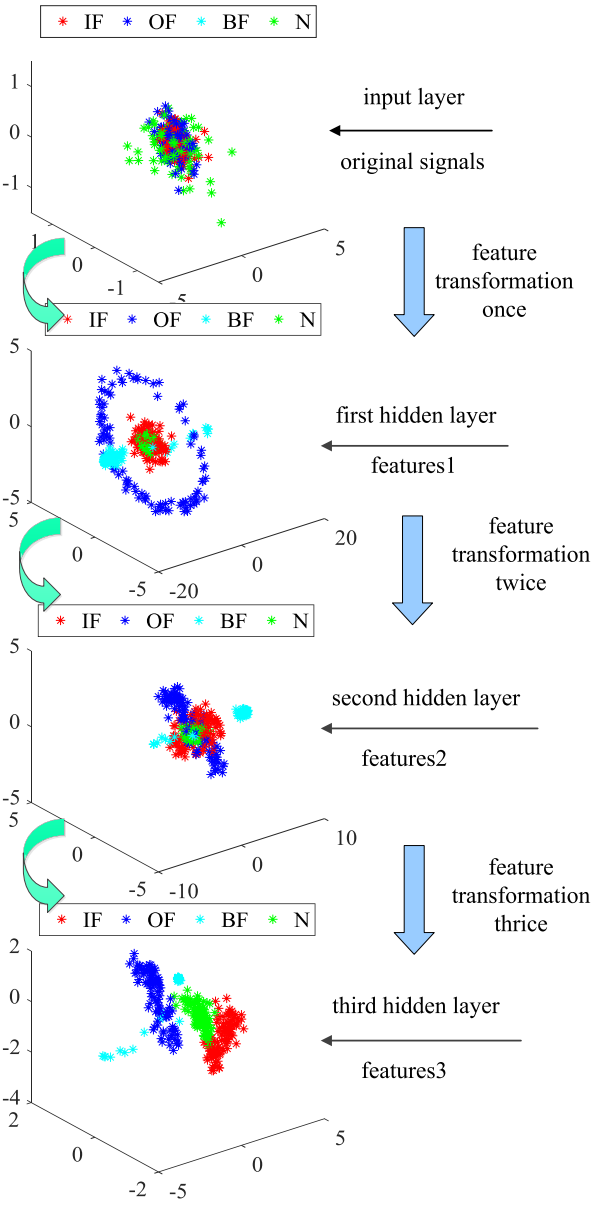


FIGURE 9. Feature visualization with bearing datasets.

B. CASE 2: GEARBOX FAULT DIAGNOSIS

1) DATA INTRODUCTION

Another scenario that can verify the performance of the proposed fusion network is the gearbox fault diagnosis case. Gearbox signals are collected from an experimental platform shown in Fig. 10. The platform comprises five forward gears and one reverse gear. The third gear is used as the testing gear. By installing an acceleration sensor on the shell and setting the speed to 1600 rpm, vibration signals, including normal (N), slight wear (SW), medium wear (MW), and broken tooth (BT), are collected during the whole running cycles as shown in Fig. 11. The sampling frequency is 3 kHz, and the sample size is 300.

FIGURE 10. Gearbox platform.

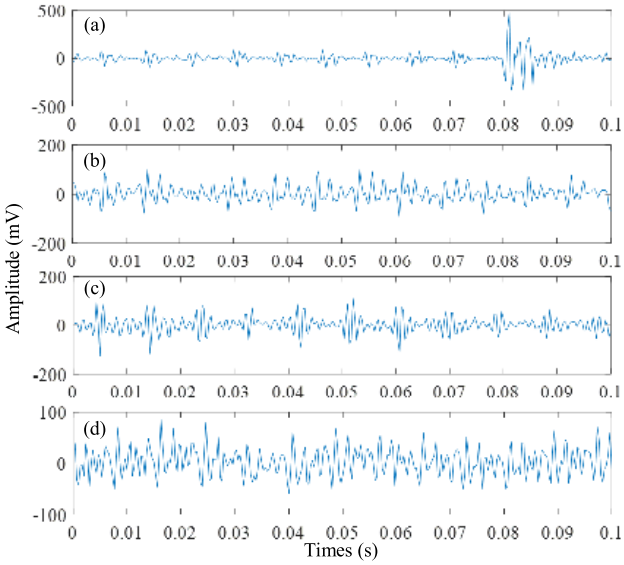


FIGURE 11. Gearbox time domain signals: (a) SW, (b) MW, (c) BT, and (d) N.

2) VALIDATION RESULTS AND ANALYSIS

The fusion network consists of one CAE and two SAEs, and the respective sizes of each layer from the input to the output layer are 300, 200, 100, 50, and 4. The contractive coefficient and sparsity parameter are set to 10^{-4} and 0.1, respectively. The learning rate and iteration number of each layer in Table 2 are applied in this case. Table 4 shows the detailed information of the gearbox datasets. After training and testing, the proposed fusion network misdiagnoses only four samples, and the testing accuracy is 98.75%. Fig. 12 shows the detailed diagnostic result.

Table 5 shows the testing results of the other comparative methods conducted in Case 1. The deep CAE, SAE, and AE have the same parameters and structure as the fusion network used in this case. The hidden size of the one-hidden-layer ANN is 150.

By observation, the proposed fusion network performs the best in gearbox fault diagnosis when dealing with original collected signals, achieving an accuracy that is about 8%, 5%, and 10% higher than that of the deep CAE, SAE, and AE network, respectively. It fully validates that the fusion network

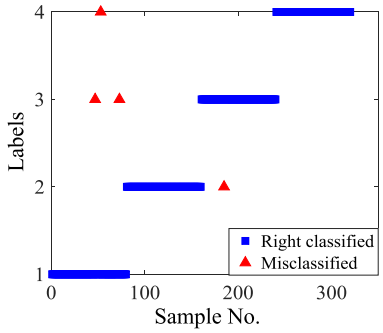


FIGURE 12. Gearbox fault diagnosis result of fusion network.

TABLE 3. Diagnosis accuracies of different methods.

Methods	IF	OF	BF	N	Overall
2CAE+1SAE	96.00%	100%	100%	100%	99.00%
3CAE	78.00%	95.00%	100%	97.00%	92.50%
3SAE	51.00%	100%	99.00%	96.00%	86.50%
3AE	59.00%	79.00%	99.00%	100%	84.25%
SVM	49.00%	51.00%	21.00%	98.00%	54.75%
ANN (one)	52.00%	35.00%	92.00%	49.00%	57.00%
ANN (three)	50.00%	35.00%	99.00%	63.00%	61.75%
FFT+SVM	82.00%	100%	96.00%	90.00%	92.00%
FFT+ANN	100%	100%	100%	100%	100%

TABLE 4. Description of Gearbox datasets.

Datasets	Training sample No.	Testing sample No.	Labels
SW	200	80	1
MW	200	80	2
BT	200	80	3
N	200	80	4

TABLE 5. Testing accuracies of different methods.

Methods	SW	MW	BT	N	Overall
1CAE+2SAE	96.25%	100%	98.75%	100%	98.75%
3CAE	78.75%	100%	86.25%	98.75%	90.94%
3SAE	86.25%	98.75%	90.00%	100%	93.75%
3AE	76.25%	100%	80.00%	95.00%	87.81%
SVM	53.75%	97.50%	32.50%	86.25%	67.50%
ANN (one)	37.50%	47.50%	43.80%	31.30%	40.00%
ANN (three)	42.50%	40.00%	43.75%	36.25%	40.63%
FFT+SVM	80.00%	100%	88.75%	100%	92.19%
FFT+ANN	96.25%	100%	98.75%	100%	98.75%

consisting of different DL models is beneficial to strengthen the feature learning ability as a single DL model has limited the feature learning ability, especially when dealing with original signals. Without preprocessing, the performances of SVM and ANN are poor because of their shallow architecture and limited feature learning ability. Generally, DL methods are superior to traditional AI methods in automatic feature learning with original signals. Moreover, adding robustness or sparsity constraints to traditional AE is beneficial for robust and discriminative feature extraction and improved diagnosis performance. In general, the proposed fusion network combining CAE and SAE has a stronger feature learning ability than other networks and is superior in extracting more robust and sparser features. The automatically extracted features are beneficial for dealing with original time domain signals overwhelmed by noise. Although the one-hidden-layer ANN combined with FFT achieves the same diagnosis accuracy as our proposed method, it cannot meet the real-time requirements. In sum, the proposed fusion network has practical application value.

In addition, signals under different signal-to-noise ratios (SNRs) are used for further validation. As Fig. 13 shows, deep CAE and SAE network both perform better than deep AE network under different SNRs, which illustrate that adding robustness or sparsity constraints is beneficial for feature extraction and fault diagnosis. Our proposed fusion network can combine the advantages of the CAE and SAE and achieves the highest diagnosis accuracy under different SNRs.

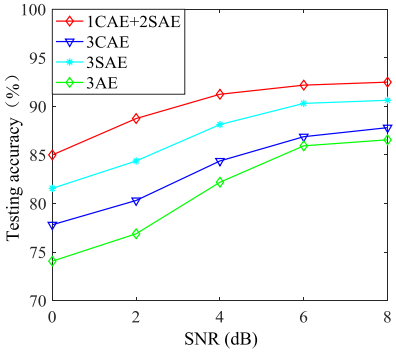


FIGURE 13. Diagnosis accuracies of different DL methods under different SNRs with gearbox datasets.

Fig. 14 shows the sparsity of the features extracted in the second hidden layer of the proposed network and the deep AE and CAE networks. About 60% of the features extracted with our proposed network are inactive, and this percentage is about 30% higher than the sparsity of the features extracted by the other two networks. This result fully demonstrates the ability of SAE to enhance the sparsity of learned features.

3) FEATURE VISUALIZATION

Fig. 15 shows the feature visualization results with gearbox datasets. Similar to that in Case 1, the clustering situation improves with the increase in the number of hidden layers. Moreover, the features mined from the top layer are clear enough to be classified to some extent. Therefore, the same conclusions can be drawn with regard to the capability of the proposed fusion network to automatically learn representative features through feature transformation. In conclusion, multiple feature transformations are highly beneficial for discriminative feature extraction.

V. DISCUSSION

A. NETWORK ARCHITECTURE DETERMINATION

The architecture of the fusion network is a key factor that affects its performance in feature learning and fault diagnosis,

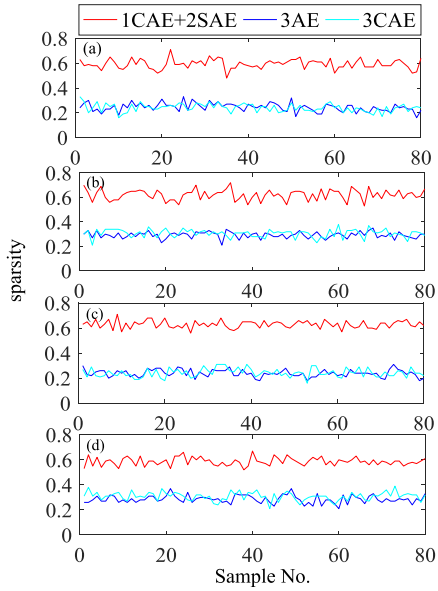


FIGURE 14. Sparsity of features extracted by different networks: (a) SW, (b) MW, (c) BT, and (d) N.

including the combination order of the CAE and SAE and their individual quantities.

Take bearing datasets as an example, different combination orders and quantities under different numbers of hidden layers are analyzed, and the testing accuracies are shown in Fig. 16. "1CAE + 1SAE" denotes that the fusion network consists of one CAE and one SAE, with the SAE stacked after the CAE. The inputs are initially transformed by the CAE to enhance the robustness of the features and then by the SAE to enhance the sparsity. "1SAE + 1CAE" denotes that the CAE is stacked after the SAE and that the inputs are first transformed by the SAE and then by the CAE to enhance the sparsity first and then robustness of the features. Other representations have similar indications.

By observation, with the increase in the number of hidden layers, the highest diagnosis accuracy over different combinations increases first and then decreases. Thus, it is not the more hidden layers, the higher diagnosis accuracy. In addition, it is observed that the highest diagnosis accuracy achieved by the combination order "CAE + SAE" is higher than that achieved by "SAE + CAE" under the same number of hidden layers. This result implies that using the CAE first to enhance robustness is more suitable when dealing with original time domain signals under immense noise interference. According to the highest diagnosis accuracy, the architecture of the fusion network for bearing fault diagnosis is determined to be "2CAE + 1SAE". Fig. 17 shows the testing accuracies under different combinations for gearbox datasets. When the hidden layer number is 2 or 3, the highest diagnosis accuracy achieved by the combination order "CAE + SAE" is higher than that achieved by "SAE + CAE". When the hidden layer number is increased to 4 or 5, the highest diagnosis

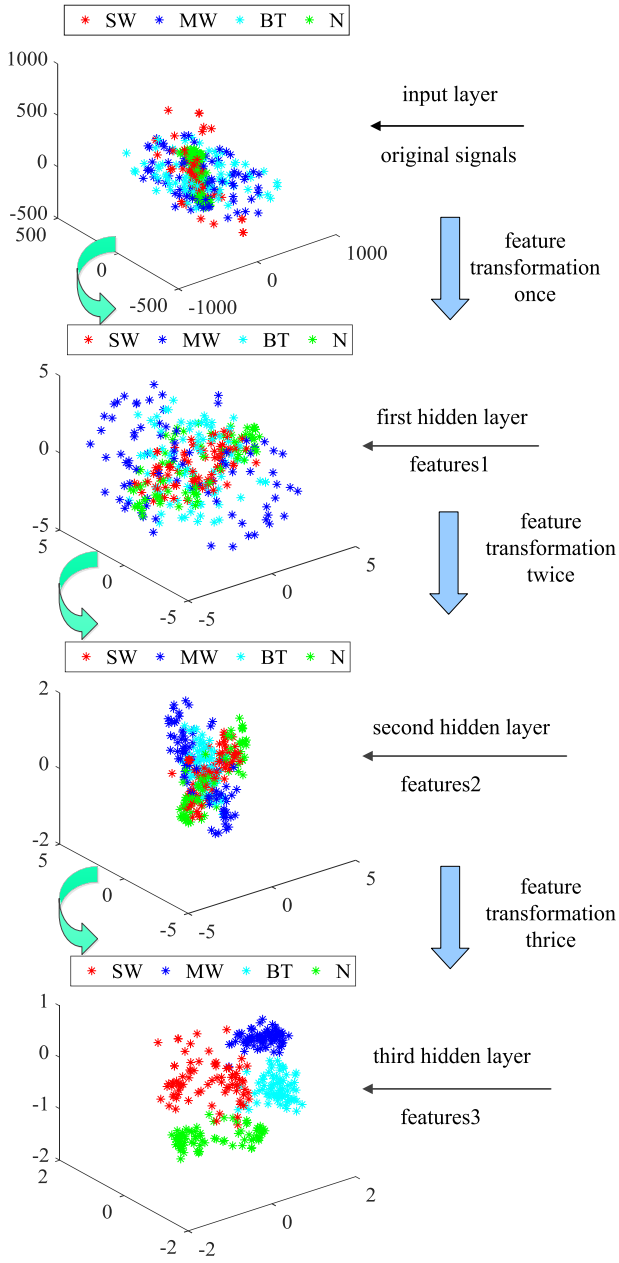


FIGURE 15. Feature visualization with gearbox datasets.

accuracy achieved by "CAE + SAE" is about equal to that achieved by "SAE + CAE"; however, the highest testing accuracy decreases. Thus, according to the highest diagnosis accuracy shown in Fig. 17, "1CAE + 2SAE" is the suitable combination order for gearbox fault diagnosis, which also illustrates that applying CAE first is more suitable when dealing with noise-overwhelmed signals.

B. PENALTY PARAMETER SELECTION

Penalty parameters are also important factors that affect the performance of feature learning and diagnosis. Table 6 and

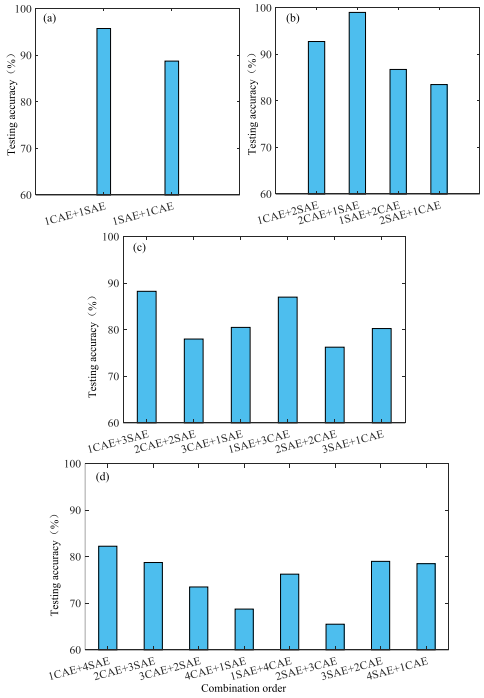


FIGURE 16. Diagnosis accuracies of different combinations with bearing datasets: (a) 2 hidden layers, (b) 3 hidden layers, (c) 4 hidden layers, and (d) 5 hidden layers.

TABLE 6. Diagnosis accuracies under different penalty parameters with bearing datasets.

ρ \ γ	10^{-6}	10^{-5}	10^{-4}	10^{-3}	10^{-2}
0	82%	82.5%	91.5%	74.25%	55.5%
0.05	98.5%	98.75%	95%	77.5%	58.75%
0.10	97.5%	98%	99%	77.25%	61%
0.15	96%	97%	97.5%	77.75%	55.5%
0.20	96%	97%	96%	77%	56.5%
0.25	90%	98%	95.25%	75.75%	56.75%
0.30	89.75%	93.75%	87.5%	78.75%	56.5%

Fig. 18 show the diagnosis accuracies for bearing datasets under different sparsity parameter ρ from 0 to 0.3 with a step size of 0.05 and contractive penalty coefficient γ from 10^{-6} to 10^{-2} with a step size of 10.

By observation, when γ is greater than 10^{-4}, the diagnosis accuracies are lower than 80% regardless of how the sparsity parameter is selected, that is, the contractive penalty coefficient dominates at this time. It is can be explained that a large γ will weaken the weight of the data fidelity term and cannot extract useful features. Therefore, γ should not be extremely large. When γ is between 10^{-6} and 10^{-4}, the diagnosis accuracies are around 85%; particularly in the large marked area, the diagnosis accuracies can exceed 95%, that is, the sparsity parameter dominates. When ρ is set to be large (0.3) or small (0), the diagnosis accuracies considerably decrease because extremely redundant or sparse features are not conducive for fault diagnosis. Finally, the model achieves the highest diagnosis accuracy when the sparsity parameter and contractive penalty coefficient are set to 0.1 and 10^{-4}, respectively. Similarly, different penalty parameters are discussed with gearbox datasets in Fig. 19 and the same conclusion can be drawn.

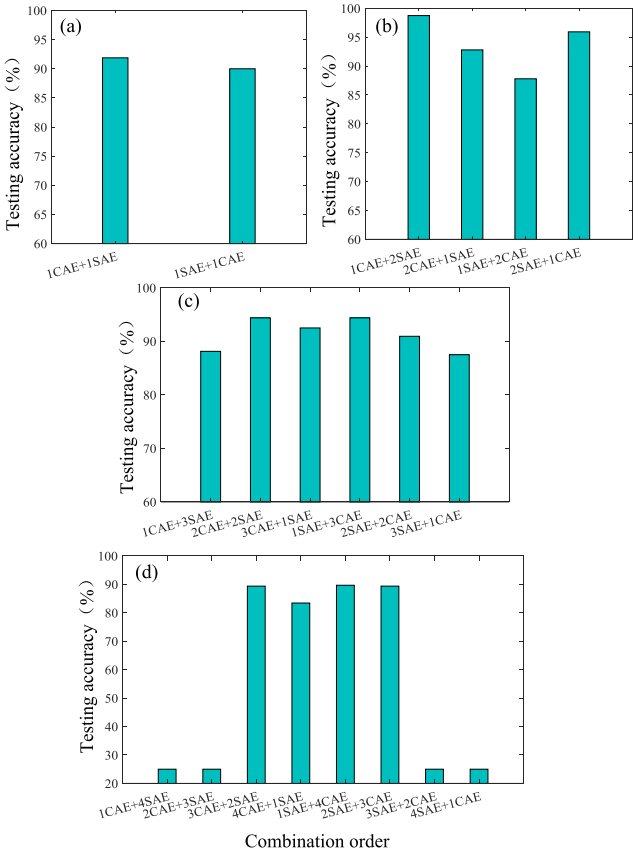


FIGURE 17. Diagnosis accuracies of different combinations with gearbox datasets: (a) 2 hidden layers, (b) 3 hidden layers, (c) 4 hidden layers, and (d) 5 hidden layers.

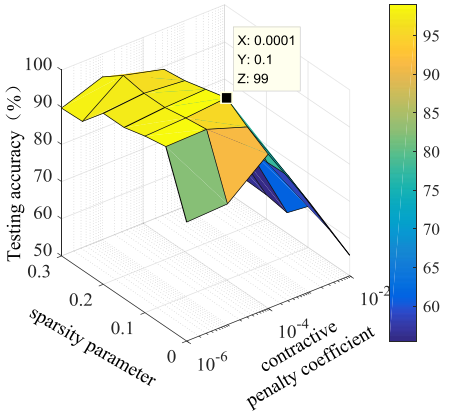


FIGURE 18. Diagnosis accuracies under different penalty parameters with bearing datasets.

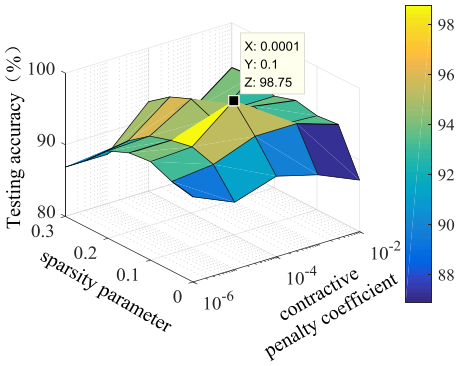


FIGURE 19. Diagnosis accuracies under different penalty parameters with gearbox datasets.

VI. CONCLUSION

A deep fusion network that combines the CAE and SAE is presented in this work for automatic feature learning and fault diagnosis with original time domain signals. The fusion network is capable of learning more robust and sparser features automatically. First, a stacked CAE is used to transform the original signals overwhelmed by noise to extract more robust features automatically. Second, the stacked SAE further transforms the extracted features into sparser and more discriminative ones. Finally, the extracted high-level features are classified by a classifier for fault recognition. The proposed fusion network is verified with bearing and gearbox fault diagnosis cases and achieves diagnosis accuracies of 99% and 98.75%, respectively, which are about 6.5% and 8% higher than those achieved by networks based on single CAE network and about 12.5% and 5% higher than those of networks based on single SAE network. The results fully demonstrate that compared with single CAE and SAE networks, the proposed fusion network has stronger feature learning ability and can effectively deal with time domain signals under noise interference because of its induced enhancement of features' robustness and sparsity. Moreover, the discussion on network architecture and parameter selection can provide further guidance for future research.

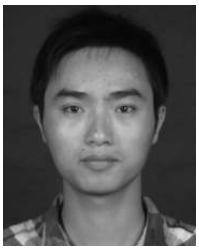
REFERENCES

[1] G. Qian, S. Lu, D. Pan, H. Tang, Y. Liu, and Q. Wang, "Edge computing: A promising framework for real-time fault diagnosis and dynamic control of rotating machines using multi-sensor data," *IEEE Sensors J.*, vol. 19, no. 11, pp. 4211–4220, Jun. 2019.
[2] D. Wang, X. Zhao, L.-L. Kou, Y. Qin, Y. Zhao, and K.-L. Tsui, "A simple and fast guideline for generating enhanced/squared envelope spectra from spectral coherence for bearing fault diagnosis," *Mech. Syst. Signal Process.*, vol. 122, pp. 754–768, May 2019.
[3] H. Wang, S. Li, L. Song, and L. Cui, "A novel convolutional neural network based fault recognition method via image fusion of multi-vibration-signals," *Comput. Ind.*, vol. 105, pp. 182–190, Feb. 2019.
[4] L. Cui, X. Wang, Y. Xu, H. Jiang, and J. Zhou, "A novel switching unscented Kalman filter method for remaining useful life prediction of rolling bearing," *Measurement*, vol. 135, pp. 678–684, Mar. 2019.
[5] Q. He, E. Wu, and Y. Pan, "Multi-scale stochastic resonance Spectrogram for fault diagnosis of rolling element bearings," *J. Sound Vib.*, vol. 420, pp. 174–184, Apr. 2018.
[6] Z. Mo, J. Wang, H. Zhang, and Q. Miao, "Weighted cyclic harmonic-to-noise ratio for rolling element bearing fault diagnosis," *IEEE Trans. Instrum. Meas.*, to be published.
[7] Y. Wang, Y. Si, B. Huang, and Z. Lou, "Survey on the theoretical research and engineering applications of multivariate statistics process monitoring algorithms: 2008–2017," *Can. J. Chem. Eng.*, vol. 96, no. 10, pp. 2073–2085, Oct. 2018.
[8] N. Li, S. Guo, and Y. Wang, "Weighted preliminary-summation-based principal component analysis for non-Gaussian processes," *Control Eng. Pract.*, vol. 87, pp. 122–132, Jun. 2019.
[9] Q. W. Gao, W. Y. Liu, B. P. Tang, and G. J. Li, "A novel wind turbine fault diagnosis method based on intergral extension load mean decomposition multiscale entropy and least squares support vector machine," *Renew. Energ.*, vol. 116, pp. 169–175, Feb. 2018.
[10] S. Haykin and N. Network, "A comprehensive foundation," *Neural Netw.*, vol. 2, no. 2004, p. 41, Feb. 2004.
[11] A. Soualhi, K. Medjaher, and N. Zerhouni, "Bearing health monitoring based on Hilbert-Huang transform, support vector machine, and regression," *IEEE Trans. Instrum. Meas.*, vol. 64, no. 1, pp. 52–62, Jan. 2015.
[12] J. Zheng, H. Pan, and J. Cheng, "Rolling bearing fault detection and diagnosis based on composite multiscale fuzzy entropy and ensemble support vector machines," *Mech. Syst. Signal Process.*, vol. 85, pp. 746–759, Feb. 2017.
[13] C. Li, J. V. de Oliveira, M. Cerrada, F. Pacheco, D. Cabrera, V. Sanchez, and G. Zurita, "Observer-biased bearing condition monitoring: From fault detection to multi-fault classification," *Eng. Appl. Artif. Intell.*, vol. 50, pp. 287–301, Apr. 2016.
[14] Moosavian, S. M. Jafari, M. Khazaee, and H. Ahmadi, "A comparison between ANN, SVM and least squares SVM: Application in multi-fault diagnosis of rolling element bearing," *Int. J. Acoust. Vib.*, vol. 23, no. 4, pp. 432–440, Dec. 2018.
[15] C. Lu, Z. Wang, and B. Zhou, "Intelligent fault diagnosis of rolling bearing using hierarchical convolutional network based health state classification," *Adv. Eng. Inform.*, vol. 32, pp. 139–151, Apr. 2017.
[16] F. Jia and Y. Lei, "A neural network constructed by deep learning technique and its application to intelligent fault diagnosis of machines," *Neurocomputing*, vol. 272, pp. 619–628, Jan. 2018.
[17] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, "Deep learning and its applications to machine health monitoring," *Mech. Syst. Signal Process.*, vol. 115, pp. 213–237, Jan. 2019.
[18] C. Sun, M. Ma, Z. B. Zhao, and X. Chen, "Sparse deep stacking network for fault diagnosis of motor," *IEEE Trans. Ind. Informat.*, vol. 14, no. 7, pp. 3261–3270, Jul. 2018.
[19] H. Shao, H. Jiang, X. Zhang, and M. Niu, "Rolling bearing fault diagnosis using an optimization deep belief network," *Meas. Sci. Technol.*, vol. 26, no. 11, Sep. 2015, Art. no. 115002.
[20] W. Zhang, C. Li, G. Peng, Y. Chen, and Z. Zhang, "A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load," *Mech. Syst. Signal Process.*, vol. 100, pp. 439–453, Feb. 2018.
[21] Z. Yuan, L. Zhang, L. Duan, and T. Li, "Intelligent Fault Diagnosis of Rolling Element Bearings Based on HHT and CNN," in *Proc. PHM*, Chongqing, China, Oct. 2018, pp. 292–296.
[22] F. Jia, Y. Lei, J. Lin, X. Zhou, and N. Lu, "Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data," *Mech. Syst. Signal Process.*, vol. 72, pp. 303–315, May 2016.
[23] X. Guo, L. Chen, and C. Shen, "Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis," *Measurement*, vol. 93, pp. 490–502, Nov. 2016.
[24] S. Tang, C. Shen, D. Wang, S. Li, W. Huang, and Z. Zhu, "Adaptive deep feature learning network with Nesterov momentum and its application to rotating machinery fault diagnosis," *Neurocomputing*, vol. 305, pp. 1–14, Aug. 2018.
[25] H. Shao, H. Jiang, F. Wang, and H. Zhao, "An enhancement deep feature fusion method for rotating machinery fault diagnosis," *Knowl.-Based Syst.*, vol. 119, pp. 200–220, Mar. 2017.
[26] C. Shen, Y. Qi, J. Wang, G. Cai, and Z. Zhu, "An automatic and robust features learning method for rotating machinery fault diagnosis based on contractive autoencoder," *Eng. Appl. Artif. Intell.*, vol. 76, pp. 170–184, Nov. 2018.
[27] Z. Chen and W. Li, "Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 7, pp. 1693–1702, Jul. 2017.
[28] S. Rifai, P. Vincent, X. Müller, X. Glorot, and Y. Bengio, "Contractive auto-encoders: Explicit invariance during feature extraction," in *Proc. ICML*, Jul. 2011, pp. 833–840.
[29] B. Poole, J. Sohl-Dickstein, and S. Ganguli, "Analyzing noise in autoencoders and deep networks," 2014, *arXiv:1406.1831*. [Online]. Available: https://arxiv.org/abs/1406.1831

YUMEI QI received the B.S. and M.S. degrees in measurement techniques and instruments from Soochow University, in 2015 and 2018, respectively. Her research interests include fault diagnosis and intelligent machine learning.

CHANGQING SHEN received the B.S. and Ph.D. degrees in instrument science and technology from the University of Science and Technology of China, in 2009 and 2014, respectively, and the Ph.D. degree in systems engineering and engineering management from the City University of Hong Kong, in 2014. He is currently an Associate Professor with the School of Rail Transportation, Soochow University, China. His research interests include machinery data mining and intelligent fault diagnosis.

JUN ZHU received the B.S. degree in mechanical engineering from Huazhong Agricultural University, Wuhan, China, in 2013, and the M.S. degree in mechatronic engineering from the University of Science and Technology of China, Hefei, China, in 2016. He is currently pursuing the Ph.D. degree in industrial engineering with the National University of Singapore, Singapore. His research interests include signal processing and data mining for machinery prognostics and diagnosis.



XINGXING JIANG received the B.S. and Ph.D. degrees in vehicle engineering from the Nanjing University of Aeronautics and Astronautics, China, in 2008 and 2016, respectively. He is currently an Associate Professor with the School of Rail Transportation, Soochow University, China. His current research interests include machinery condition monitoring and fault diagnosis, and time–frequency analysis.

JUANJUAN SHI received the B.S. and M.S. degrees in mechanical engineering from Northwest A&F University, in 2008 and 2011, respectively, and the Ph.D. degree in mechanical engineering from the University of Ottawa, in 2015. She is currently an Associate Professor with the School of Rail Transportation, Soochow University, China. Her research interests include rotating machinery signal processing and condition monitoring.

ZHONGKUI ZHU received the B.S. and M.S. degrees in vehicle engineering from Hefei Polytechnic University, in 1997 and 2002, respectively, and the Ph.D. degree in instrument science and technology from the University of Science and Technology of China, in 2005. He is currently a Professor with the School of Rail Transportation, Soochow University. His research interest includes fault diagnosis of mechanical equipment.