IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 11 April 2023, accepted 4 May 2023, date of publication 9 May 2023, date of current version 17 May 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3274569

RESEARCH ARTICLE

Chemical Process Fault Diagnosis Based on Improved ResNet Fusing CBAM and SPP

XIAOCHEN YAN[1], YANG ZHANG[2], AND QIBING JIN[1]
[1]College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China
[2]College of Mechanical and Electrical Engineering, Beijing University of Chemical Technology, Beijing 100029, China
Corresponding author: Yang Zhang (2002500011@buct.edu.cn)

ABSTRACT This paper proposes a fault diagnosis method based on an improved residual network (ResNet) for complex chemical processes. The method can automatically and efficiently extract fault features from the extensive data generated by the chemical operation process. The improvement is carried out in three aspects. Firstly, 1D convolution is introduced in the construction of the model to reduce the number of parameters and training time, and shortcut connections are used to alleviate the network degradation problem of traditional deep neural networks. Second, a residual-CBAM module is proposed by combining residual networks with Convolutional Block Attention Module (CBAM). This module can effectively reduce the interference of invalid targets and improve the characterization ability of the model. Finally, based on the backbone path of the network, the branching path after spatial pyramid pooling (SPP) is introduced to enable the network to extract features from different angles of the feature map and further aggregation, which improves the robustness of the model. The Tennessee-Eastman (TE) process is used as the experimental object to compare the improved ResNet with several other deep learning models. The experimental results show that the improved ResNet model achieves the best fault diagnosis results. The t-SNE method was used to visualize the fault classification process by the improved ResNet model, and the effectiveness of the improved ResNet model was further analyzed and verified.

INDEX TERMS Fault diagnosis, chemical process, ResNet, CBAM, SPP, Tennessee Eastman process.

I. INTRODUCTION

Abnormal variations in chemical process variables that exceed normal thresholds are called faults. There are various methods for fault detection and diagnosis (FDD). However, fault diagnosis methods can be broadly classified into three categories qualitative knowledge-based methods [1], model-based methods [2], and data-driven methods [3], as summarized by Venkatasubramanian et al.

Qualitative knowledge-based and model-based methods can no longer meet the needs of modern complex industrial systems because of the extreme reliance on expert knowledge and experience and the difficulty of mathematical modeling [4], [5]. The data-driven methods only need to establish a fault diagnosis model based on the offline process data of the system and then substitute the online data into the model to achieve accurate fault detection and diagnosis through analysis and judgment, which has obvious advantages over the previous two methods and is gradually becoming a hot spot for research [6].

Data-driven fault detection and diagnosis methods can be further divided into statistical and machine-learning methods. The statistical methods extract features by analyzing the statistics of the system variables, which in turn enables fault diagnosis. Common statistical methods include principle component analysis (PCA) [7], [8], [9], partial least squares (PLS) [10], [11], independent component analysis (ICA) [12], [13], [14], kernel principle component analysis (KPCA) [15], linear discriminant analysis (LDA) [16], etc. Machine learning methods can be divided into shallow learning methods and deep learning methods. Among the shallow learning methods include support vector machines (SVM) [17], [18], K-nearest neighbors (KNN) [19], Gaussian mixture models (GMM) [20], artificial neural networks

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

ANN [21], [22], etc. However, the two methods are difficult to deal with data with strong nonlinear relationships. In order to further improve the accuracy of fault detection and diagnosis, another method of machine learning, deep learning, has achieved good results.

The essence of deep learning is the process of learning features through networks at multiple levels, which has a more robust feature extraction capability when dealing with nonlinear and non-normally distributed data compared to traditional machine learning methods, and the original data is transformed to obtain abstract features that will be learned autonomously [23]. Since Yann Le Cun proposed the Convolutional Neural Network (CNN) in 1998, it has been continuously optimized and applied to more practical models [24].

In 2013, Tamilselvan et al. were the first to propose applying deep learning methods to fault diagnosis [25]. Xie and Bai proposed a fault diagnosis method for chemical processes based on hierarchical DNN (HDNN), which can adaptively mine the available fault features from the measured signals [26]. Zhang and Zhao proposed a fault diagnosis model based on a scalable deep confidence network (DBN), which extracts each fault feature in the spatial and temporal domains through DBN sub-networks separately and performs fault classification after training with a full two-layer back propagation network [27]. For time-series fault data, Zhao et al. proposed a long short-term memory neural network (LSTM) method, which can adaptively learn the original data's dynamic temporal information. The results showed that the model has better fault diagnosis capability [28].

In this paper, firstly, to address the problem of a large number of model parameters and slow training time, the CNN and ResNet models are constructed by incorporating 1D convolution in the model instead of 2D convolution, which reduces the number of parameters of the model, thus improving the efficiency of modeling and reducing the time cost of the algorithm. Secondly, in order to reduce the interference of invalid targets and make the model more focused on several faults with low diagnostic accuracy, we continue to improve on the already constructed ResNet model by combining ResNet with the Convolutional Attention Module (CBAM) and propose the Residual-CBAM module, so that the characterization ability of the model can be improved. Finally, to enhance the generalization ability and robustness of the model, spatial pyramid pooling (SPP) is introduced to make the model extract features from multiple perspectives. Experiments are conducted for the TE process to evaluate the fault diagnosis performance of the proposed method.

The paper is organized as follows. Section II of this paper introduces the basic theory of the improved ResNet, including the basic structure of ResNet, the convolutional block attention module (CBAM), and the branch-and-stem path of spatial pyramidal pooling. Section III introduces the chemical process fault diagnosis method based on the improved ResNet. The comparative experimental results of different models for TE process fault diagnosis are given in Section IV. Finally, Section V concludes the paper.

II. RELEVANT THEORY

A. ResNet

With the emergence of convolutional networks, more and more scholars have devoted themselves to the study of convolutional networks. Therefore, several problems were soon exposed, one of which is that as the depth of the network increases, parameter optimization becomes more and more difficult, leading to more training and testing errors in multi-layer networks than in multi-layer networks, i.e., network degradation problems. To solve this problem, He et al. proposed the residual network (ResNet) in 2015 [29], where they introduced shortcut connections in the network structure, which can pass the input directly to the output across multiple layers of the network. The effect is that instead of fitting the underlying network, only the modified residuals based on the original input constant mapping can be fitted. If the input mapping is already optimal, then the residuals module must only fit a zero mapping. However, in practice, the constant mapping could be more optimal, and the later network only needs to fit the output of the previous network with the truly desired residuals. The results show that the problem of network degradation can be effectively mitigated. The structure of ResNet is similar to the results of the convolutional network and consists of the following components.

1) CONVOLUTIONAL LAYER

The convolutional layer is a core component of the CNN model and a fundamental component of the ResNet model. In general, the input to the convolution layer is three-dimensional data of HWC, where H denotes the height, W denotes the width, and C denotes the number of channels, which can be understood as the input consisting of C two-dimensional feature maps of size $H \times W$. The main purpose of the convolutional layer is to extract features from the input feature map by convolutional kernels, and one convolutional kernel extracts one kind of feature. The input data generally contains multiple features, so a complete convolutional layer usually contains multiple convolutional kernels and multiple output feature maps. The formula for the convolution layer is defined as (1).

$$Y_{i,j,0} = b_o + \sum_m \sum_n \sum_k X_{i-m,j-n,k} F_{m,n,k,o} \quad (1)$$

where X is the input feature map, F is the convolution kernel, b is the bias, Y is the output feature map, i represents the index of the height of the input feature map, j represents the index of the width of the input feature map, and k represents the index of the channel of the input feature map. m represents the index of the height of the convolution kernel, n represents the width of the convolution kernel, and o represents the index of the channel of the output feature map. This model has two 2D convolutional kernel sizes, 3×3 and 1×1, respectively. To process the timing data more conveniently, in addition

to 2D convolution, this model also uses 1D convolution, which differs from 2D convolution in that it only performs convolution operations in a one-dimensional direction with a convolution kernel size of 2×1. The expression for the 1D convolution operation is changed to (2).

$$Y_{i,0} = b_o + \sum_m \sum_k X_{i-m,k} F_{m,k,o} \tag{2}$$

At this point in the 1D convolution, i represents the index of the length of the input feature map, m represents the index of the length of the convolution kernel, k represents the index of the channel of the input feature map, and o represents the index of the channel of the output feature map.

2) CONVOLUTIONAL LAYER

This model uses the Rectified Linear Unit(ReLU) and Sigmoid activation functions.

ReLU is a segmentation function, which is formally defined as (3).

$$f(x) = \max\{0,x\} = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{3}$$

The Sigmoid function is a saturation activation function, which is formally defined as (4).

$$\sigma = \frac{1}{1 + e^{-x}} \tag{4}$$

3) SHORTCUT CONNECTIONS

Shortcut connections are the core part of the ResNet model and the key to the superior performance of the ResNet model over the ordinary CNN model. As the number of layers increases, the network degenerates. As the number of layers grows, the training set's loss rate decreases, then saturate, and increases when the depth of the network is increased. However, this is not overfitting because the training loss rate decreases all the time in overfitting. When the network degenerates, the shallow network can achieve better training results than the deep network, and if the features of the lower layer are passed to the higher layer, the results should be at least as good as the shallow network. In the forward transmission process, as the number of layers deepens, the feature information contained in the input feature map decreases layer by layer after several convolution operations, and the addition of constant mapping in the residual network ensures that the network at layer $l + 1$ must contain more feature information than that at layer l. The ResNet was born based on this idea of directly connecting different network layers using constant mappings. The ResNet model's main architecture is composed of several stacked residual blocks. The residual blocks can be expressed as (5).

$$y = F(x, \{W_i\}) + x \tag{5}$$

where x and y are the input and output of the residual block, $F(\cdot)$ is the residual mapping to be learned, and W_i is the parameter to be learned. $F(\cdot) + x$ is the summation of the elements in the corresponding positions. If the two dimensions are the same, they can be added directly, and if the dimensions are different, the method of averaging pooling and then padding "zero" is used to make the dimensions match.

4) POOLING LAYER

The pooling layer is used to downsample its previous layer. Its role is to reduce the size (length, width, number of channels) of the previous layer because the pooling layer does not contain parameters, so it can significantly reduce the amount of computation, memory usage, and number of parameters, and shorten the training time of the network, to achieve a certain scale, spatial invariance, and reduce overfitting. This model uses multiple pooling, max pooling, global max pooling, and global average pooling. Max pooling is partitioning the input image into multiple rectangular regions; for each subregion, a maximum value is an output. Global max pooling is a special kind of max pooling, except that instead of splitting out several rectangular regions, the maximum value is taken for all elements of the entire feature map and output to the next layer. Global average pooling is similar to global max pooling but differs from global max pooling in that it averages all elements of the entire feature graph before outputting them to the next layer..

5) FULLY CONNECTED LAYER

The final output layer of the improved ResNet model is a fully connected layer with a softmax classifier. The input of this layer is the output neurons of the previous maximum pooling layer, and the number of final output neurons is generally set to the number of work condition types. The softmax function is used as a classifier after the fully connected layer to evaluate the probability that the input data belongs to each type of working condition. This function is also a activation function, and its output can be considered a probability distribution, so it is generally used as a classifier only in the last layer. This function can be defined as (6).

$$y_j = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}}, \quad j = 1, \ldots, K \tag{6}$$

where z_j is the *jth* output feature of the fully connected layer, which is the *jth* input of the softmax function, y_j can be considered as the probability that the input data belongs to the *jth* category of working conditions, and K is the number of working conditions categories. Among these K probabilities, the model takes the working condition corresponding to the maximum value as the result of the diagnosis.

Meanwhile, the cross entropy loss function represents the error between the actual output and the desired output of the network output layer. The cross entropy loss function is defined as (7).

$$L(t_j, y_j) = -\sum_{j=1}^{k} t_j \ln(y_j) \tag{7}$$

where y_j is the probability that the input data belongs to the *jth* category of faults, t_j is the final output of the network, which

is the output of the softmax function, and K is the number of working condition types.

B. CONVOLUTIONAL BLOCK ATTENTION MODULE

The attention mechanism is essentially a resource allocation mechanism, which can change the resource allocation according to the importance of the target of attention so that the resources are tilted more toward the object that needs to be focused on. The Convolutional Block Attention Module (CBAM) incorporates both channel attention and spatial attention mechanisms to improve the overall detection accuracy of the model [30].

The channel attention operation is performed first, which can be calculated from (8). The input feature map $X(H \times W \times C)$ is subjected to global maximum pooling and global average pooling based on height and width to obtain two $1 \times 1 \times C$ feature maps, which are then sent to two fully connected layers respectively, with weights shared between the two fully connected layers. Then, the summation operation is performed on the features after the output of the two fully connected layers. The final channel attention feature map is generated by the sigmoid activation function so that the weights of each feature map are obtained.

The second step performs the spatial attention operation, which can be calculated from (9). The output feature map of the above step after the channel attention module is the input feature map of the spatial attention module. The input feature map is subjected to global max pooling and global average pooling based on channels, and 2 $H \times W \times 1$ feature maps are obtained. Then a channel splicing operation is performed on these 2 feature maps. At this point, the number of channels of the feature map is 2. After another convolution operation, the feature map is downscaled to 1 channel. Then, the sigmoid activation function is used to generate the output space attention mechanism feature map, so that the weights of each feature point on the feature map are obtained. Finally, the output feature map that has gone through the attention mechanism module is shortcut-connected with the output that initially went through two separate convolution operations to obtain the residual attention module.

$$Mc(F) = \sigma\left(MLP\left(AvgPool\left(F\right)\right) + MLP\left(MaxPool\left(F\right)\right)\right) \tag{8}$$

$$Ms(F) = \sigma\left(f\left(\left[AvgPool\left(F\right); MaxPool\left(F\right)\right]\right)\right) \tag{9}$$

C. THE SPATIAL PYRAMID POOLING LAYER

Spatial Pyramid Pooling (SPP) is to obtain multiple output feature maps of different sizes from an input feature map after the pooling operation, flatten these feature maps into 1D arrays, and then stitch these 1D arrays, and finally send the stitched feature maps to the fully connected layer for classification [31]. The spatial pyramid pooling layer has the following two advantages: first: it can solve the defects caused by the different sizes of input images. Second: it shows the robustness of the algorithm due to the feature extraction and aggregation of a feature map from different angles. The pooling kernel size and step size are calculated using the following formula: $windowsize = M/N(roundedup)$, $stride = M/N(roundeddown)$, where M is the input feature map size, and N is the output feature map size. The dimensions of the three feature maps of the spatial pyramid output used in this model are 4×4, 2×2, and 1×1, respectively. The feature map of this model is divided into two paths after three residual attention modules. One of the paths is to reconstruct the input feature map for feature map reconstruction and then perform a one-dimensional convolution operation, which is the backbone path. The other path is the spatial pyramid pooling branch, which turns the input feature map into 4×4, 2×2, and 1×1 by spatial pyramid pooling, then flattened and stitched into a one-dimensional array. This allows the features of the input feature maps to be extracted in two ways, after which the extracted feature maps are fused with the trunk road for feature fusion, which helps the subsequent model for fault diagnosis.

III. FAULT DIAGNOSIS METHOD BASED ON IMPROVED ResNet

The structure of the improved ResNet fault diagnosis model proposed in this paper consists of convolutional layers, activation function layers, residual-CBAM modules, pooling layers, a spatial pyramid pooling layer, and fully connected layers. The model can be divided into two main parts, as shown in Figure 1 and Figure 2.

The input data is originally a set of one-dimensional time-series data, so it needs to go through a fully-connected layer first to expand the dimensionality of the channels. Then the feature map is reconstructed into a two-dimensional feature map. Then, the structure of the proposed residual-CBAM module is shown in Figure 1. Each residual-CBAM module contains two convolutional layers as well as spatial attention and channel attention mechanisms. The input feature map is first passed through a convolutional layer with a convolutional kernel size of 3×3, followed by a convolutional layer with a convolutional kernel size of 1×1. The 1×1 convolutional kernel does not change the size of the current feature map. Its primary role is to follow the 3×3 large receptive field convolution kernel and further purify the features by a small receptive field convolution kernel to obtain a residual feature. After passing through the two convolutional layers, the output feature map passes through the spatial and channel attention mechanisms. Thus the output feature map of each layer is obtained. Then, each part of the output feature map is connected by shortcut connections to form the residual-CBAM module. The reconstructed feature maps go through 3 residual-CBAM modules.

After the three residual-CBAM modules, the model will be divided into two paths, as shown in Figure 2. One path reduces the obtained two-dimensional feature map into a one-dimensional array, and feature extraction is performed by

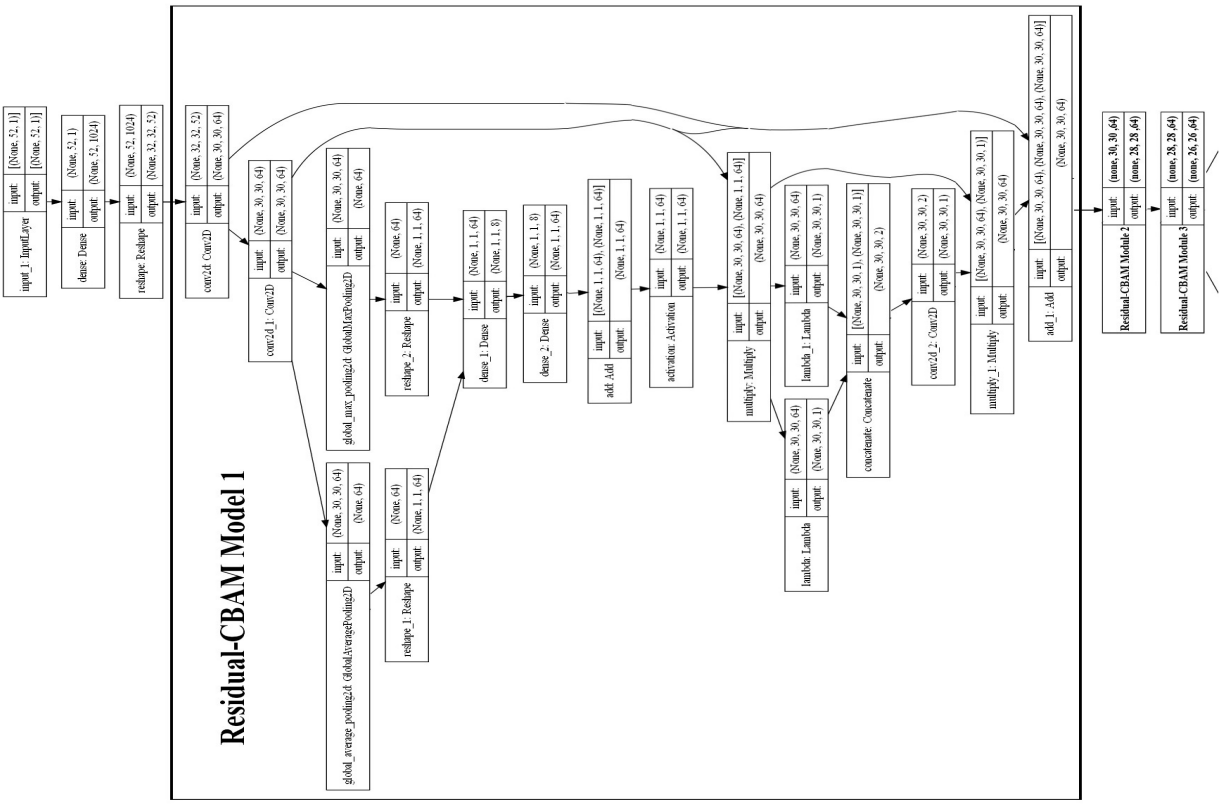


FIGURE 1. The first part of the improved ResNet model.

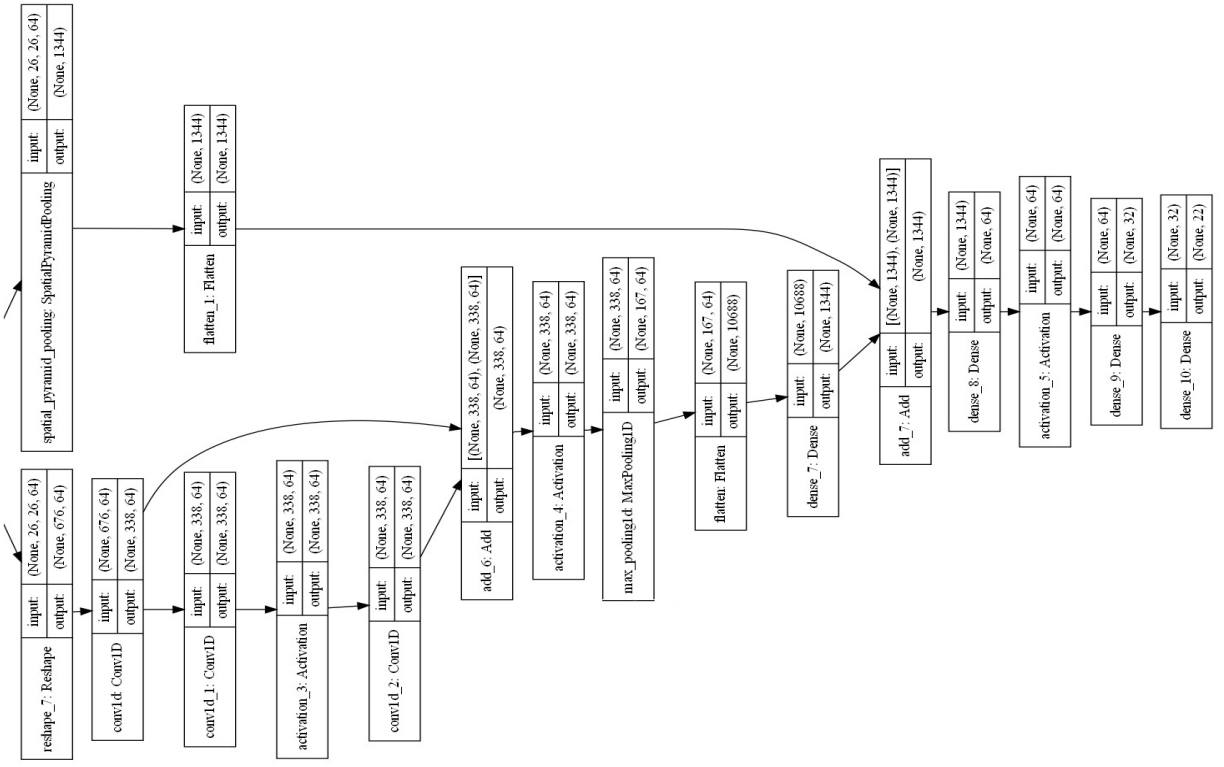


FIGURE 2. The second part of the improved ResNet model.

1D convolution in the network backbone. In contrast, the other path performs feature extraction by spatial pyramid pooling, and the output is also a one-dimensional array after spatial pyramid pooling. The features of the two paths are

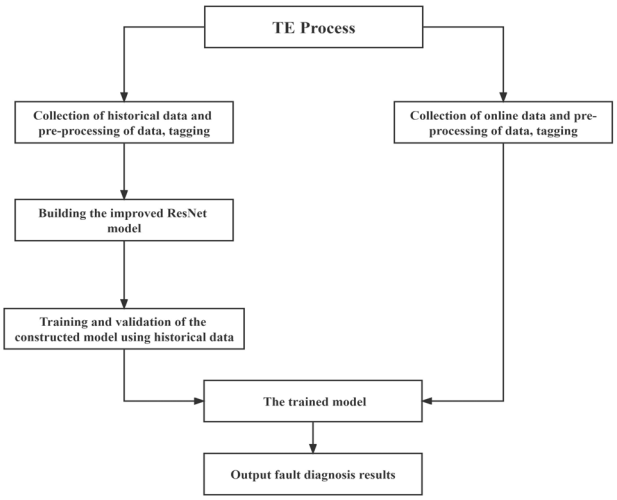


FIGURE 3. Fault diagnosis process based on improved ResNet model.

fused. The final step goes through a fully connected layer and a softmax classifier to output the fault classification results.

The fault diagnosis method based on improved ResNet can be divided into the following steps, and the specific process is shown in Figure 3.

1) Collection of historical data and pre-processing of data, tagging;
2) Building the improved ResNet model;
3) Training and validation of the constructed model using historical data;
4) Get the trained model;
5) Collection of online data and pre-processing of data, tagging;
6) Using the trained model for online real-time processing;
7) Output fault diagnosis results.

IV. SIMULATION EXPERIMENT AND ANALYSIS

A. DATASET INTRODUCTION AND PREPROCESSING

The TE data generated by the TE experimental platform is used as a data source to verify the performance of the proposed improved ResNet fault diagnosis method. Tennessee Eastman (TE) process is a model based on an actual chemical process and is widely used to evaluate the performance of process monitoring methods. A modified version of the TE process is used as the simulation platform, and the flow chart of the TE process is shown in Figure 4. Normal condition data and fault data can be collected from the Simulink simulation module of MATLAB. The collected TE data include normal data and 21 types of fault data for a total of 22 operating conditions. Each operating data set consists of 800 data sets, each consisting of 53 process variables, and of these 53 variables, there are 12 operational variables and 41 measured variables. However, the data set contains only 52 variables because one variable that was not manipulated, the stirring speed,

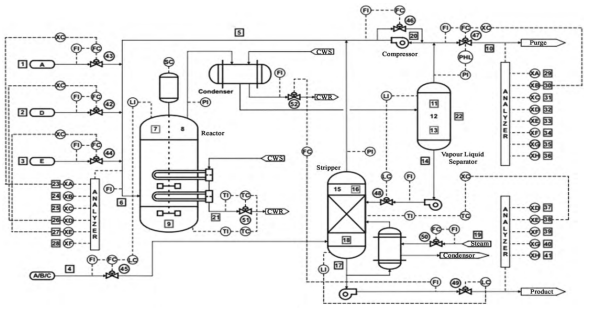


FIGURE 4. Flow diagram of TE process.

was removed. The details of all faults are shown in Table 1. Faults 1-7 are related to step changes in the process variables of the experimental process; faults 8-12 are related to random fluctuations in some of the process variables; fault 13 is related to changes in the kinetic characteristics of the reactor; faults 14 and 15 are related to valve sticking; and fault 21 is caused by the valve being in a constant position. This paper uses the z-score standardization method to pre-process the data. The formula is defined as (10).

$$Z(x) = \frac{x - \bar{x}}{\sqrt{\frac{\sum(x-\bar{x})}{n-1}}} \tag{10}$$

where x is the original data, $Z(x)$ is the standardized data, and $\bar{x}$ is the mean of the original data. To transform it into a data set with a mean zero variance of one. Finally, because the multi-classification problem is to be solved, the tags need to be processed for one-hot encoding, specifically by transforming these 22 working conditions into the corresponding 22 one-dimensional arrays for subsequent classification with the model.

B. MODEL PERFORMANCE COMPARISONS

This study implements this improved ResNet model on the open-source machine learning software library TensorFlow. For the hyperparameter settings, the batch size is 128, the number of iterations is 100, the learning rate is 0.0003, the loss function is cross entropy loss, and Adam is used as the optimizer.

To evaluate the model's performance, accuracy, precision, recall, and F1 score are used as evaluation metrics. Figure 5 shows the confusion matrix.

In the confusion matrix: True Positive(TP): an instance is positive and is predicted to be positive; False Negative(FN): an instance is positive but is predicted to be negative; False Positive(FP): an instance is negative but is predicted to be positive; True Negative(TN): an instance is a negative class, but is predicted to be a negative class. However, it is difficult to measure the performance of a model based on the number of TP and TN alone when faced with massive data. This led to the secondary evaluation indicators: accuracy, precision, and recall. Accuracy can be calculated from (11), which represents the proportion of all correct results in the classification

TABLE 1. TE process failure types.

Fault Number	Fault Description	Type
1	A/C feed flow ratio changes, component B content remains constant	Step
2	Component B changes, A/C feed flow ratio remains unchanged	Step
3	Temperature change of material D	Step
4	Temperature change of reactor cooling water inlet	Step
5	Temperature change of condenser cooling water inlet	Step
6	Material A loss	Step
7	Loss of pressure head of material C	Step
8	Change in the composition of materials A, B, C	Step
9	Temperature change of material D	Random fluctuation
10	Temperature change of material C	Random fluctuation
11	Temperature change of reactor cooling water inlet	Random fluctuation
12	Temperature change of condenser cooling water inlet	Random fluctuation
13	Change in reactor kinetic properties	Slow drift
14	Reactor cooling water valve sticking	Sticking
15	Condenser cooling water valve sticking	Sticking
16	Unknown	Unknown
17	Unknown	Unknown
18	Unknown	Unknown
19	Unknown	Unknown
20	Unknown	Unknown
21	Valve fixed in steady state position	Constant position

Confusion Matrix		Prediction	
		Positive	Negative
True	Positive	TP	FN
	Negative	FP	TN

FIGURE 5. Confusion matrix.

model to the total number of observations. Precision can be calculated from (12), which represents the proportion of correct predictions in the classification model among all results with Positive predictions. Recall can be calculated from (13), which represents the proportion of model prediction pairs among all the results for which the true value is Positive. By using the above secondary indicators, the results of the number of confusion matrices can be transformed into a ratio between 0 and 1, which can be easily measured in a standardized way. Then a tertiary index, F1 score, is derived from Equation (14), which combines precision and recall and

TABLE 2. Comparison of the results of different fault diagnosis models.

Model	Accuracy
Improved ResNet	96.69
ResNet	89.77
CNN	84.00
MLPC	79.91
Adaboost	34.13

is calculated by taking the harmonic mean of the two. The F1 score ranges from 0 to 1, with 1 representing the best model output and 0 representing the worst model output.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{11}$$

$$Precision = \frac{TP}{TP + FP} \tag{12}$$

$$Recall = \frac{TP}{TP + FN} \tag{13}$$

$$F1score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{14}$$

This paper selected two traditional machine learning models, MLPC, Adaboost, and modern deep learning models, CNN and ResNet, as fault diagnosis intelligence models and the improved ResNet model, for comparison experiments. The results obtained are shown in Table 2.

As can be seen from Table 2, the accuracy of the improved ResNet model proposed in this paper is 96.69% on the test set, which is significantly better than the accuracy of 89.77% of the original ResNet model and several other fault diagnosis models. The network structure of the CNN model is similar to that of the ResNet model. The size of the convolutional kernel and the number of convolutional layers remain the same, but the residual structure is missing. The accuracy on the test set is 84.00%. In contrast, the accuracy of models using traditional machine learning methods like MLPC is 79.91% and Adaboost 34.13%, both below 80%, indicating that these modern convolution-based deep learning models have higher fault diagnosis accuracy.

This paper verifies whether the introduction of 1D convolution can improve the model performance. We conducted experiments with the same dataset, using 1D dimensional convolution to construct CNN and ResNet models with the same number of network layers, and compared the experiments with the DCNN proposed in the paper [32]. The CNN and ResNet models used in the experiments have three convolutional layers, a max pooling layer, three fully connected layers, and a softmax classifier. They have the same number of layers, but ResNet has one more shortcut connection than CNN, and the network structure is shown in Figure 6 and Figure 7.

Since the paper [32] only experimented with the top twenty faults, we also selected the top twenty faults when comparing the three models. Also, as in the article [32], the fault diagnosis rate (FDR) is used, and the formula of FDR is shown

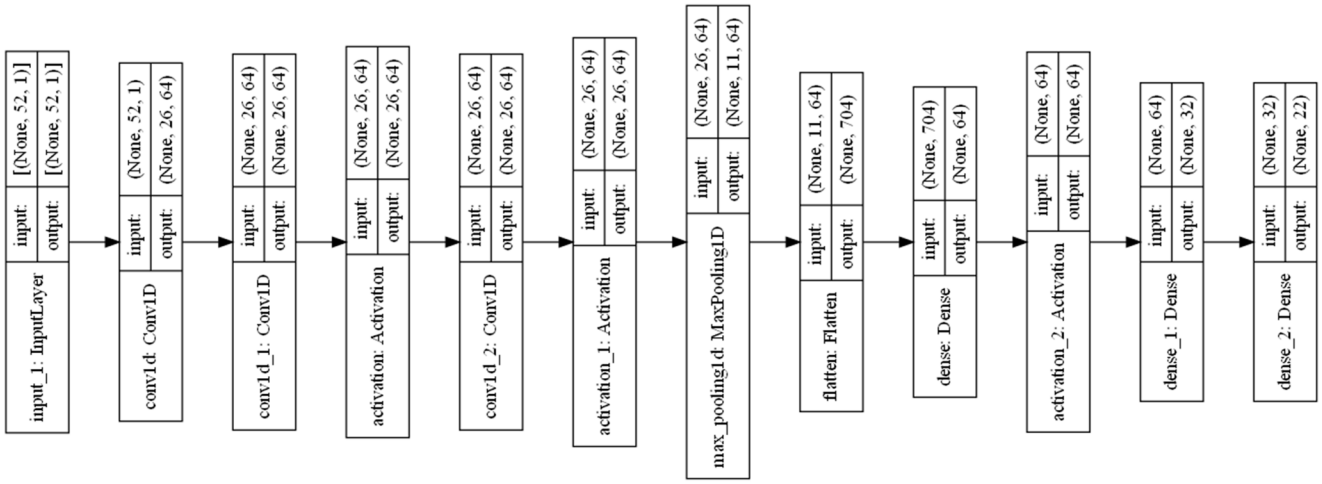


FIGURE 6. The structure of the CNN model.

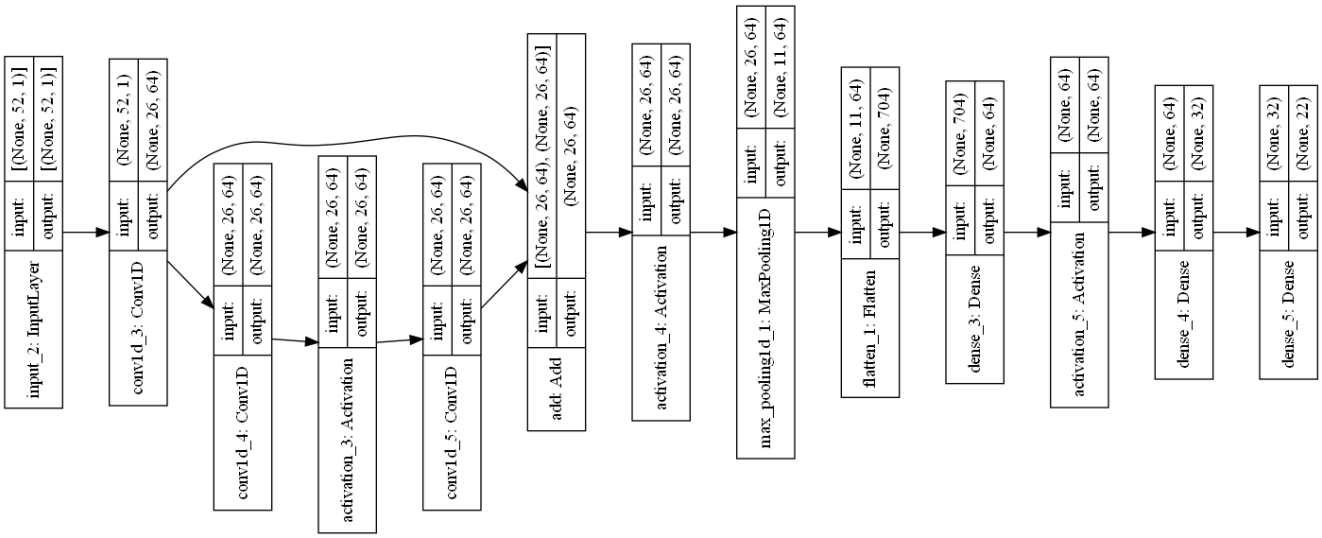


FIGURE 7. The structure of the ResNet model.

in (15), which is the same as the formula for the recall.

$$FDR = \frac{TP}{TP + FN} \tag{15}$$

The experimental results are shown in Table 3. According to the experimental results, the average FDR of the ResNet model constructed using 1D convolution in this paper is the highest among the three models, indicating that introducing shortcut connections is practical and can alleviate the network degradation problem. The average fault diagnosis rate of the CNN model is as high as 88.1%, which is almost the same compared to the 88.2% fault diagnosis rate of the DCNN model, with a difference of only 0.1%. However, Table 4 shows that the floating point operations of CNN and ResNet built with 1D convolution are much lower than those of DCNN, and the training time is also shorter, indicating that the introduction of 1D convolution can better handle the temporal data and effectively reduce the computation of the model. Nevertheless, the FDR for fault 3, fault 9, fault 15, and fault 19 need further improvement, so there is still room for improvement.

In this paper, to further verify whether the introduction of CBAM and SPP can improve the model performance, the Proposed ResNet in this paper is constructed based on ResNet constructed using one-dimensional convolution, and the CNN, ResNet, and Improved ResNet are used to conduct comparison experiments against the TE process. Precision and recall (also known as FDR) can evaluate the diagnostic performance of a model for a single fault, but they are usually contradictory. High precision leads to low recall, while high recall leads to low precision, which makes comparison difficult. Therefore, a three-level evaluation metric, the

TABLE 3. FDR of DCNN, CNN and ResNet.

Classes	DCNN	CNN	ResNet
Fault01	98.6	100	100
Fault02	98.5	100	100
Fault03	91.7	65.31	71.43
Fault04	97.6	97.87	97.87
Fault05	91.5	90	100
Fault06	97.5	97.87	100
Fault07	99.9	100	100
Fault08	92.2	100	100
Fault09	58.4	75	70.45
Fault10	96.4	90	92.5
Fault11	98.4	82.22	91.11
Fault12	95.6	97.44	100
Fault13	95.7	100	100
Fault14	98.7	100	100
Fault15	28	58.7	67.39
Fault16	44.2	75.68	78.38
Fault17	94.5	97.67	100
Fault18	93.9	97.5	97.5
Fault19	98.6	60	77.5
Fault20	93.3	76.67	90
Average	88.2	88.1	91.7

TABLE 4. FLOPs and training time of DCNN, CNN and ResNet.

Model	FLOPs	Training time /min
CNN	257452	1.23
ResNet	514904	1.33
DCNN	78000000	46.7

F1 score, is used as the judging criterion. F1 score combines precision and recall and can evaluate different models more effectively. Table 5 shows the F1 scores of the three models for one normal condition and 21 fault conditions. As we can see from the experimental results, the F1 scores of different faults also vary greatly, and the F1 scores of CNN are lower in the face of fault 3, fault 9, fault 15, and fault 19, and even less than 60% in faults 3 and 9, as shown in Figure 8. Nevertheless, it is still less than 80%. To solve this problem, the improved ResNet model introduced the attention mechanism and increased the depth of the network to make the model more focused on the diagnosis of these types of faults, and the results showed that the improved ResNet model with the residual-CBAM module achieved the highest F1 scores for faults 3, 9, 15, and 19, which are difficult to solve by CNN and ResNet. At the same time, in order to prevent the improved ResNet model from focusing too much on these local features, a branching path of spatial pyramidal pooling was introduced in addition to the main path of the network, enabling the network to extract richer features from multiple scales and focus on more comprehensive information, thus improving the generalization ability and robustness of the improved ResNet model. The results show that the F1 scores of the model are almost all above 90%, and although the F1 scores

TABLE 5. F1 score of CNN, ResNet and Improved ResNet.

Classes	CNN	ResNet	Improved ResNet
Normal	75.57	82.98	96.45
Fault01	100	100	100
Fault02	100	100	100
Fault03	56.64	68.63	95.05
Fault04	97.87	98.92	95.92
Fault05	93.51	100	98.73
Fault06	98.92	100	100
Fault07	100	100	100
Fault08	96.39	100	100
Fault09	54.55	72.09	94.62
Fault10	83.72	93.67	97.44
Fault11	89.16	93.18	94.38
Fault12	97.44	100	98.73
Fault13	97.22	100	98.55
Fault14	100	100	100
Fault15	63.53	72.94	88.37
Fault16	75.68	82.86	94.29
Fault17	98.82	100	100
Fault18	97.5	98.73	100
Fault19	67.61	72.94	88.61
Fault20	79.31	88.52	92.86
Fault21	78.57	85.25	93.94
Average	86.46	91.40	96.72

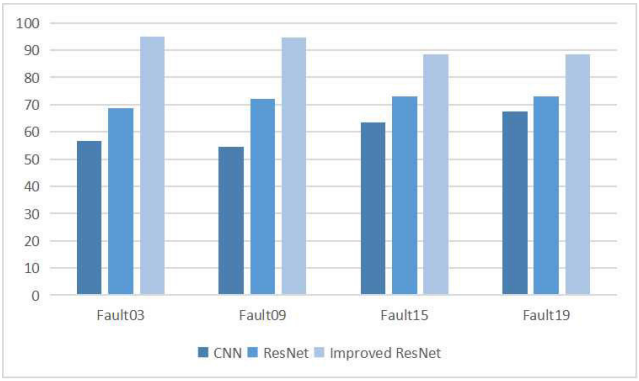


FIGURE 8. F1 scores of three models for faults 3, 9, 15, 19.

of several faults are not the highest, they are not very different from those of the other two methods. Figures 9 and 10 also show the three methods' accuracy and loss rate curves. The comparison can lead to the conclusion that the addition of the attention mechanism and spatial pyramid pooling in the model helps improve the model's classification performance, and the improved ResNet model is more suitable for fault diagnosis.

Finally, after studying the fault diagnosis effects of the three models for individual categories, the ROC curves of the CNN model, the ResNet model, and the improved ResNet model are plotted in this paper in order to show the overall classification results of the models more intuitively. The horizontal coordinate is the false positive

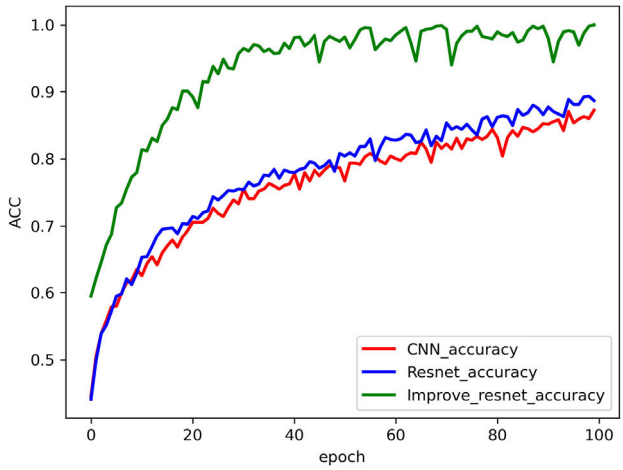


FIGURE 9. Accuracy curves of the three models.

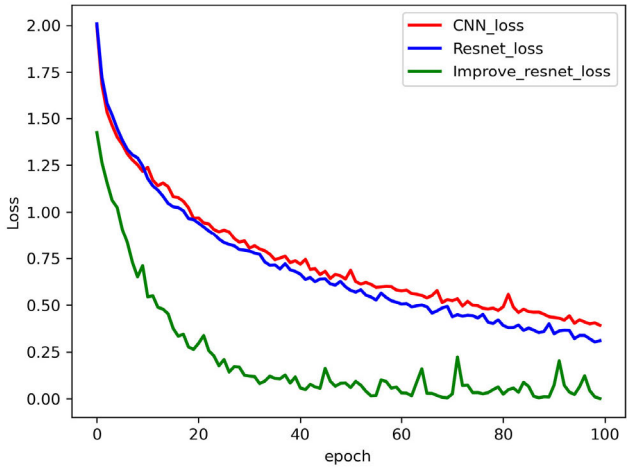


FIGURE 10. Loss curves for the three models.

rate (FPR), which represents the proportion of actual negative instances in the positive class predicted by the model to all negative instances. The formula for calculating FPR is as (16). The vertical coordinate is the true positive rate (TPR), which represents the proportion of the actual positive instances to all positive instances in the positive class predicted by the classifier. The formula for calculating TPR is as (17).

$$FPR = \frac{FP}{FP + TN} \tag{16}$$

$$TPR = \frac{TP}{TP + FN} \tag{17}$$

The closer the ROC curve is to the upper left corner, the better the diagnosis is. Figure 11 shows the average ROC curves for each category of the four models, where Figure 11(a)-(c) show the ROC curves of CNN, ResNet, and Improved ResNet. Another more straightforward way to judge the performance of a model is the area under the ROC curve (AUC), where the closer the value of AUC is to 1, the better the model performance. We can learn that the

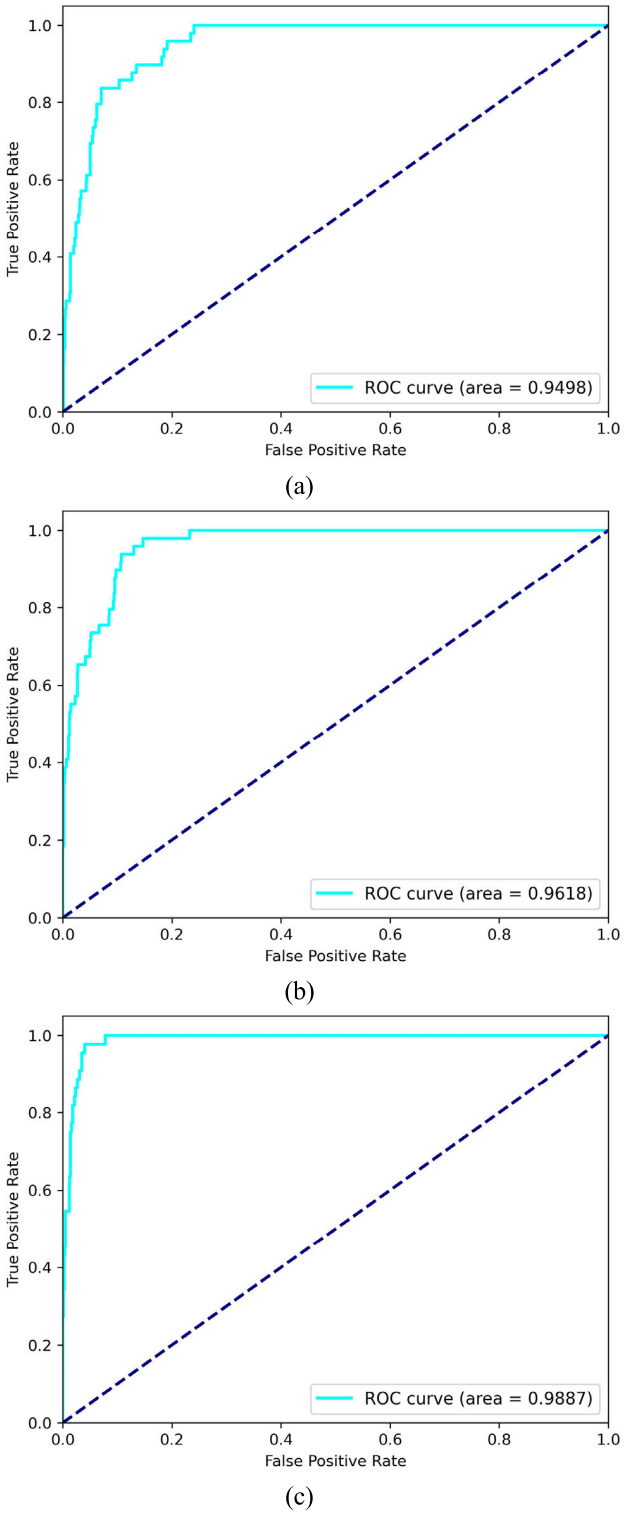


FIGURE 11. ROC curves for different models. (a) CNN. (b) ResNet. (c) Improved ResNet.

AUC value of the improved ResNet is 0.9887, which is closer to 1 than CNN with ResNet, so the overall classification performance of the ResNet model is also the best after the improvement.

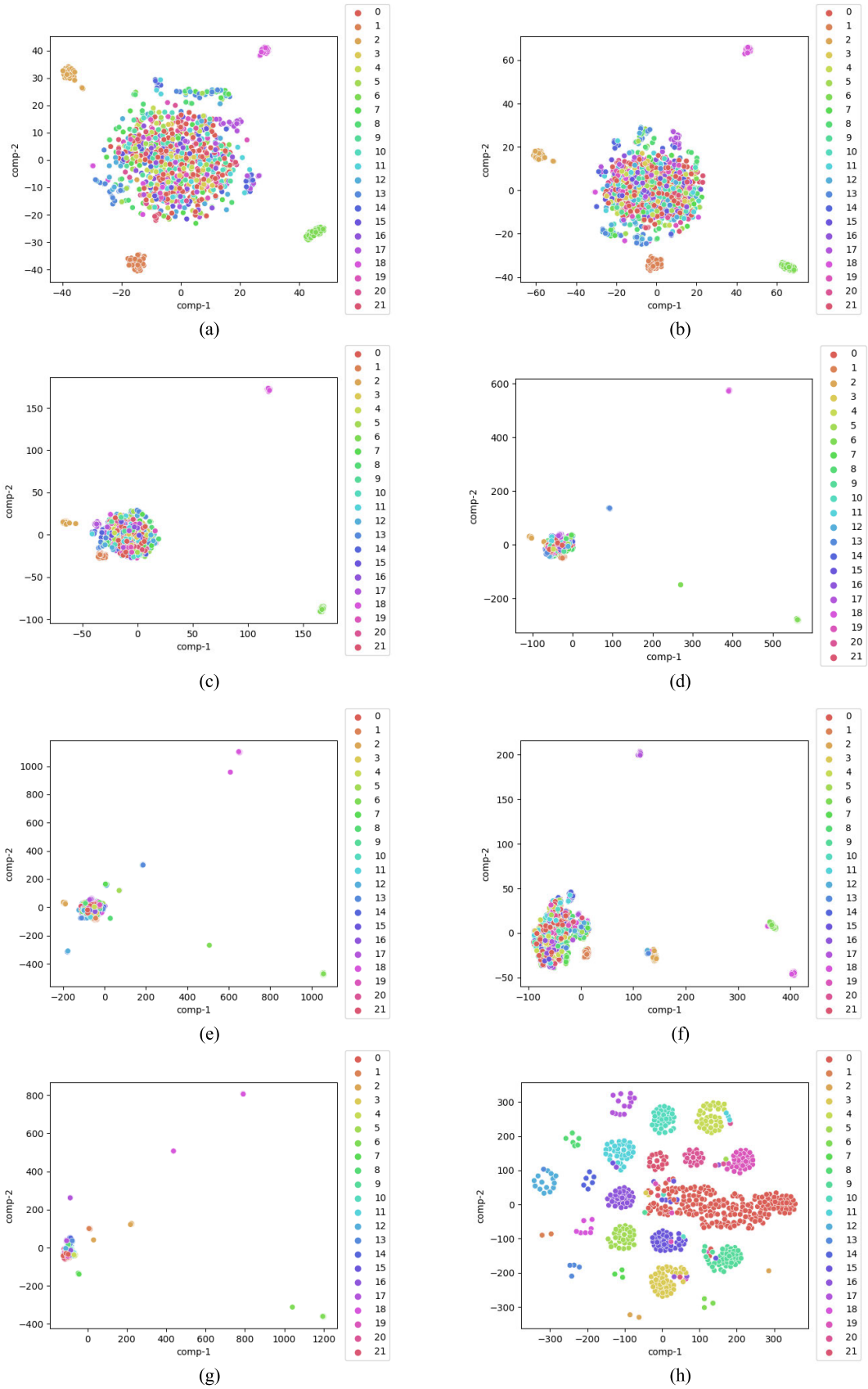


FIGURE 12. Visualizing the training process of the improved ResNet model using the t-SNE method. (a) input layer. (b)layer Conv2d_1. (c)layer Conv2d_3. (d)layer Conv2d_6. (e)layer Conv1d. (f)layer Dense_8. (g) layer Dense_9. (h)output layer.

C. HIERARCHICAL FEATURE LEARNING VISUALIZATION

In this paper, to facilitate the understanding of the classification process of the improved ResNet model, it is necessary to understand and recognize the feature learning process of the network. However, since the features learned by the model are high-dimensional, it is not easy to visualize the results obtained from the output of each layer. To address this issue, we use the t-distributed Stochastic Neighbor Embedding (t-SNE) method, which can be used to visualize the feature learning process of the improved ResNet model. The t-SNE algorithm serves to visualize the high-dimensional data in the form of two-dimensional coordinates. It is modeled as a Gaussian distribution in the original high-dimensional space, while it can be modeled as a t-distribution in the two-dimensional output space. The goal of the process is to find the transformation that maps high-dimensional space to two-dimensional space and to minimize the gap between the points in these two distributions. The input layer, layer Conv2d_1, layer Conv2d_3, layer Conv2d_6, layer Conv1d, layer Dense_8, layer Dense_9, and output layer are selected, and the high-dimensional output features of these network layers are embedded in a two-dimensional scatter plot using the t-SNE method and displayed as shown in Figure 12(a)-(h). Initially, the data samples of the input layer are displayed in a mixed state. After several convolutional layers with the attention mechanism module, the samples are gradually clustered into the corresponding classes. The two-dimensional scatter plot of the final output layer shows clear clustering of the samples, indicating the effectiveness of the improved ResNet model for fault diagnosis tasks.

V. CONCLUSION

In order to better diagnose the faults in chemical processes, we proposed a new ResNet model, and the improved ResNet model can accurately diagnose the types of chemical process faults. Performance evaluation experiments were conducted using the TE process. Based on the test data results of the TE process, the effects of 1D convolution, shortcut connection, residual- CBAM module, and spatial pyramid pooling on the model performance are discussed, and the following conclusions are obtained.

1. Aiming at the TE process data, which are time-series data, the CNN model is constructed by improving the convolutional kernel and using one-dimensional convolutional, which is more suitable for processing time-series data as the primary computational convolution. On top of it, we also constructed ResNet, which introduces shortcut connections, but the structure of each layer of the network is similar to prevent the network degradation problem. The constructed CNN and ResNet are used for TE process fault diagnosis and compared with DCNN for experiments. The results show that the ResNet fault diagnosis model achieves the best results. The model has excellent performance, few parameters, and a short training time.

2. Aiming at the poor fault diagnosis effect of a part of faults, an improved ResNet model is proposed based on ResNet. The improved ResNet model is compared with CNN and ResNet models for TE process fault diagnosis, and the experimental results show that the improved ResNet model not only improves the fault diagnosis effect of 3, 9, 15, and 19 using the attention mechanism but also uses spatial pyramid pooling to avoid the model from over-focusing on certain features so that the generalization ability and robustness of the whole model can be improved. The improved ResNet model achieves the best experimental results, indicating that introducing the attention module, and spatial pyramid pooling technique, can significantly improve the model's performance in fault diagnosis.

3. Using the t-distributed Stochastic Neighbor Embedding (t-SNE) method, we can reduce the high-dimensional features inside the improved ResNet model to two dimensions. The final two-dimensional scatterplot shows the improved model's effectiveness for fault diagnosis tasks.

REFERENCES

[1] V. Venkatasubramanian, R. Rengaswamy, and S. N. Kavuri, "A review of process fault detection and diagnosis: Part II: Qualitative models and search strategies," *Comput. Chem. Eng.*, vol. 27, no. 3, pp. 313–326, 2003.

[2] V. Venkatasubramanian, R. Rengaswamy, K. Yin, and S. N. Kavuri, "A review of process fault detection and diagnosis: Part I: Quantitative model-based methods," *Comput. Chem. Eng.*, vol. 27, no. 3, pp. 293–311, Mar. 2003.

[3] V. Venkatasubramanian, R. Rengaswamy, S. N. Kavuri, and K. Yin, "A review of process fault detection and diagnosis: Part III: Process history based methods," *Comput. Chem. Eng.*, vol. 27, no. 3, pp. 327–346, Mar. 2003.

[4] 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.*, vols. 72–73, pp. 303–315, May 2016.

[5] C. Zhang, L. Xu, X. Li, and H. Wang, "A method of fault diagnosis for rotary equipment based on deep learning," in *Proc. Prognostics Syst. Health Manage. Conf. (PHM-Chongqing)*, Oct. 2018, pp. 958–962.

[6] X. Dai and Z. Gao, "From model, signal to knowledge: A data-driven perspective of fault detection and diagnosis," *IEEE Trans. Ind. Informat.*, vol. 9, no. 4, pp. 2226–2238, Nov. 2013.

[7] B. M. Wise, N. Ricker, D. Veltkamp, and B. Kowalski, "A theoretical basis for the use of principal component models for monitoring multivariate processes," *Process Control Qual.*, vol. 1, no. 1, pp. 41–51, 1990.

[8] J. Chen and C.-M. Liao, "Dynamic process fault monitoring based on neural network and PCA," *J. Process Control*, vol. 12, no. 2, pp. 277–289, Feb. 2002.

[9] C. K. Lau, K. Ghosh, M. A. Hussain, and C. R. Che Hassan, "Fault diagnosis of Tennessee Eastman process with multi-scale PCA and ANFIS," *Chemometric Intell. Lab. Syst.*, vol. 120, pp. 1–14, Jan. 2013.

[10] J. F. MacGregor, C. Jaeckle, C. Kiparissides, and M. Koutoudi, "Process monitoring and diagnosis by multiblock PLS methods," *AIChE J.*, vol. 40, no. 5, pp. 826–838, May 1994.

[11] Y. Zhang, H. Zhou, S. J. Qin, and T. Chai, "Decentralized fault diagnosis of large-scale processes using multiblock kernel partial least squares," *IEEE Trans. Ind. Informat.*, vol. 6, no. 1, pp. 3–10, Feb. 2010.

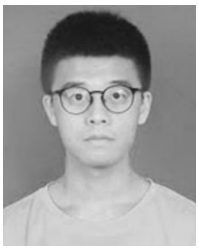
[12] M. Kano, S. Tanaka, S. Hasebe, I. Hashimoto, and H. Ohno, "Monitoring independent components for fault detection," *AIChE J.*, vol. 49, no. 4, pp. 969–976, Apr. 2003.

[13] H. Albazzaz and X. Z. Wang, "Statistical process control charts for batch operations based on independent component analysis," *Ind. Eng. Chem. Res.*, vol. 43, no. 21, pp. 6731–6741, Oct. 2004.

[14] J.-M. Lee, C. Yoo, and I.-B. Lee, "Statistical process monitoring with independent component analysis," *J. Process Control*, vol. 14, no. 5, pp. 467–485, Aug. 2004.

[15] Y. Zhang and C. Ma, "Fault diagnosis of nonlinear processes using multiscale KPCA and multiscale KPLS," *Chem. Eng. Sci.*, vol. 66, no. 1, pp. 64–72, Jan. 2011.

[16] B. Li, P.-L. Zhang, D.-S. Liu, S.-S. Mi, and P.-Y. Liu, "Classification of time–frequency representations based on two-direction 2DLDA for gear fault diagnosis," *Appl. Soft Comput.*, vol. 11, no. 8, pp. 5299–5305, Dec. 2011.
[17] A. Kulkarni, V. K. Jayaraman, and B. D. Kulkarni, "Knowledge incorporated support vector machines to detect faults in Tennessee eastman process," *Comput. Chem. Eng.*, vol. 29, no. 10, pp. 2128–2133, Sep. 2005.
[18] S. Mahadevan and S. L. Shah, "Fault detection and diagnosis in process data using one-class support vector machines," *J. Process Control*, vol. 19, no. 10, pp. 1627–1639, Dec. 2009.
[19] Z. Zhou, C. Wen, and C. Yang, "Fault detection using random projections and k-nearest neighbor rule for semiconductor manufacturing processes," *IEEE Trans. Semicond. Manuf.*, vol. 28, no. 1, pp. 70–79, Feb. 2015.
[20] J. Yu and S. J. Qin, "Multimode process monitoring with Bayesian inference-based finite Gaussian mixture models," *AIChE J.*, vol. 54, no. 7, pp. 1811–1829, 2008.
[21] J. C. Hoskins and D. M. Himmelblau, "Artificial neural network models of knowledge representation in chemical engineering," *Comput. Chem. Eng.*, vol. 12, nos. 9–10, pp. 881–890, Sep. 1988.
[22] M. A. A. Rad and M. J. Yazdanpanah, "Designing supervised local neural network classifiers based on EM clustering for fault diagnosis of Tennessee eastman process," *Chemometric Intell. Lab. Syst.*, vol. 146, pp. 149–157, Aug. 2015.
[23] Y. Bengio, "Learning deep architectures for AI," *Found. Trends Mach. Learn.*, vol. 2, no. 1, pp. 1–127, 2009.
[24] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
[25] P. Tamilselvan and P. Wang, "Failure diagnosis using deep belief learning based health state classification," *Rel. Eng. Syst. Saf.*, vol. 115, pp. 124–135, Jul. 2013.
[26] D. Xie and L. Bai, "A hierarchical deep neural network for fault diagnosis on Tennessee–Eastman process," in *Proc. IEEE 14th Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2015, pp. 745–748.
[27] Z. Zhang and J. Zhao, "A deep belief network based fault diagnosis model for complex chemical processes," *Comput. Chem. Eng.*, vol. 107, pp. 395–407, Dec. 2017.
[28] H. Zhao, S. Sun, and B. Jin, "Sequential fault diagnosis based on LSTM neural network," *IEEE Access*, vol. 6, pp. 12929–12939, 2018.
[29] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[30] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 3–19.
[31] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Sep. 2015.
[32] H. Wu and J. Zhao, "Deep convolutional neural network model based chemical process fault diagnosis," *Comput. Chem. Eng.*, vol. 115, pp. 185–197, Jul. 2018.



XIAOCHEN YAN received the B.E. degree in automation from the Beijing University of Chemical Technology, in 2020, where he is currently pursuing the master's degree in control science and engineering. His research interest includes fault diagnosis.

YANG ZHANG is currently an Associate Professor with the College of Mechanical and Electrical Engineering, Beijing University of Chemical Technology. Her main research interests include the development of fluid machinery, engineering mechanics, and the development of drilling and completion processes in the oil and gas extraction process.

QIBING JIN is currently a Professor and a Ph.D. Supervisor with the School of Information Science and Technology, Beijing University of Chemical Technology. He was selected as one of the "New Stars of Science and Technology in Beijing," in 2004, and one of the "New Century Talents" of the Ministry of Education, in 2008. In the past five years, he has presided more than 20 research projects, including 863 Program and National Natural Science Foundation of China as a Project Leader. He has published more than 150 scientific papers as the first author. The research results have been applied to more than ten large refining and chemical enterprises, such as Yanshan, Yangzi, and Daqing. His main research interest includes advanced control and its industrial application, especially based on the application of new control methods in production sites. He has been the first to receive two first-class and two third-class awards for scientific and technological progress at the provincial and ministerial level.