



Contents lists available at ScienceDirect

Chinese Journal of Chemical Engineering

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

Full Length Article

Adaptive multiscale convolutional neural network model for chemical process fault diagnosis

Ruoshi Qin¹, Jinsong Zhao^{1,2,*}¹ State Key Laboratory of Chemical Engineering, Department of Chemical Engineering, Tsinghua University, Beijing 100084, China² Beijing Key Laboratory of Industrial Big Data System and Application, Tsinghua University, Beijing 100084, China

ARTICLE INFO

Article history:

Received 1 July 2022

Received in revised form 14 September 2022

Accepted 10 October 2022

Available online 21 October 2022

Keywords:

Neural networks

Multiscale

Adaptive attention module

Triplet loss optimization

Fault diagnosis

Chemical processes

ABSTRACT

Intelligent fault recognition techniques are essential to ensure the long-term reliability of manufacturing. Due to the variations in material, equipment and environment, the process variables monitored by sensors contain diverse data characteristics at different time scales or in multiple operating modes. Despite much progress in statistical learning and deep learning for fault recognition, most models are constrained by abundant diagnostic expertise, inefficient multiscale feature extraction and unruly multimode condition. To overcome the above issues, a novel fault diagnosis model called adaptive multiscale convolutional neural network (AMCNN) is developed in this paper. A new multiscale convolutional learning structure is designed to automatically mine multiple-scale features from time-series data, embedding the adaptive attention module to adjust the selection of relevant fault pattern information. The triplet loss optimization is adopted to increase the discrimination capability of the model under the multimode condition. The benchmarks CSTR simulation and Tennessee Eastman process are utilized to verify and illustrate the feasibility and efficiency of the proposed method. Compared with other common models, AMCNN shows its outstanding fault diagnosis performance and great generalization ability.

© 2022 The Chemical Industry and Engineering Society of China, and Chemical Industry Press Co., Ltd. All rights reserved.

1. Introduction

The chemical industry is full of high risks with the complexity of manufacturing facilities and the interaction of reactive substances. A fault in the chemical production process which is incorrectly identified and not handled in time will evolve into a major safety accident [1]. Although a large number of measures and techniques have been applied in the actual industry to prevent system errors, such as distributed control system and safety instrumented system, abnormal situation management still relies on manual analysis and treatment by skilled operators [2]. Therefore, effective fault diagnosis of chemical process is crucial to guarantee the safety and reliability of a system. Thanks to the implementation of advanced information technologies, a large amount of measurement data is available to assist in fault recognition. Plentiful data-driven models are developed constantly in academic studies in the past decade.

Data-driven fault diagnosis models are generally divided into statistical models and deep learning models. Statistical models

evolve continually and succeed in processing the high-dimensional, noisy and highly correlated data from chemical processes [3]. Principal component analysis (PCA) [4] and its variants are one of the most popular dimensionality reduction techniques. Many researchers have improved this method and presented dynamic PCA [5], kernel PCA [6] and so on for application. Partial least squares (PLS) [7] is another powerful statistical tool which preserves the significant variability information of data. The updated versions including recursive PLS [8], and dynamic inner PLS [9] make progress in overcoming the nonlinear characteristics. Independent component analysis (ICA) [10], fisher discriminant analysis (FDA) [11], support vector machine (SVM) [12] and their developments are also widely used for monitoring. Apart from the above, some new statistical learning models are released lately. Zhang *et al.* [13] developed one kernel entropy component analysis classifier for each fault domain. Ziaei-Halimejani *et al.* [14] combined recurrence quantification analysis and unsupervised learning clustering methods such as k-means and density-based spatial clustering to analyze the features influenced by faults. However, the statistical models heavily depend on the manually selected features which depend on considerable expertise and extensive mathematical knowledge. The reliance on manual

* Corresponding author at: State Key Laboratory of Chemical Engineering, Department of Chemical Engineering, Tsinghua University, Beijing 100084, China.
E-mail address: jinsongzhao@tsinghua.edu.cn (J. Zhao).

experience imposes restrictions on generalizing these traditional models.

In recent years, a wide variety of deep learning models have been applied to diagnose chemical process faults. Alternate linear transformations and nonlinear activation functions enhance the nonlinear fitting performance of multi-layer artificial neural networks (ANN). Initially, some network structures derived from ANN were developed for fault recognition, such as hierarchical ANN [15] and ANN supplemented with expert knowledge [16]. Since the rapid development of deep learning mechanisms, deep belief network (DBN) [17], variational recurrent autoencoder (VRAE) [18] and convolutional stacked autoencoder (SAE) [19] have been published separately. Zhang *et al.* [20] employed bidirectional recurrent neural network (BiRNN) for complex chemical processes in both positive and negative directions. Wu *et al.* [21] innovatively proposed a process topology convolutional network (PTCN) that merged data information and process knowledge. The PTCN model integrated the process topology into a graph reflecting the relationships among different variables. Bi *et al.* [22] designed orthogonal self-attentive variational autoencoder (OSAVA) for process monitoring and performed fault detection and identification tasks simultaneously with interpretable results. Some novel generative adversarial network models were also applied to fault diagnosis tasks in current times [23].

Among those deep learning models, convolutional neural network (CNN) is a typical architecture based on feed-forward neural networks. The hierarchical units of CNN energize its abilities to perform translation-invariant classification of input so that CNN is demonstrated to be well adapted to fault diagnosis missions [24]. CNN was first proposed for two-dimensional (2D) image recognition. To meet the input requirements of the traditional CNN structure, the one-dimensional (1D) time-series data is usually generated into images. Wu transformed historical data into sample matrices with corresponding labels and constructed a deep convolutional neural networks (DCNN) diagnostic model which extracted features in spatial and temporal domains simultaneously [25]. The chemical process dataset contains multiple variables including temperature, pressure, liquid level and so on. Since different variables have different influenced factors, their fluctuation trends vary from each other under the regulation of the control system. Therefore, the features of a certain scale that reflect the variation of each variable cannot generalize the overall trend. More comprehensive features extracted at various scales can help fault diagnosis. However, the traditional 2DCNN is only capable of extracting features of a fixed scale, which prevents the multiscale feature learning ability and the adaptability of mode switching. Song created a multiscale CNN (MSCNN) model for state classification and converted the preprocessed data into matrix color diagrams as input [26]. But it is considered that the primarily measured signals contain the inherent fluctuation characteristics which reveal the periodic information of faults more accurately. Zhang *et al.* [27] stated that 1DCNN model worked better on raw rotating machinery signals of changing loads compared to 2DCNN without handcrafted data pre-processing. It's still rare to discover a practical chemical process fault diagnosis model so far that excellently captures the multiscale features from time-series data directly.

In addition to the problem of multiscale feature extraction, the variation of feed and environment brings about several operating modes in the chemical process. In order to capture more useful fault information from multiple data distributions, many creative multimode monitoring methods have been put forward. Wu *et al.* [28] delivered a self-adaptive deep learning method combining local adaptive standardization and variational auto-encoder bidirectional long short-term memory (LAS-VB) to achieve multimode feasibility. Zheng *et al.* [29] discovered a self-adaptive

temporal-spatial self-training algorithm (SATSST) to solve the challenge of switching working conditions and inadequate data labels. Transfer learning is also utilized to realize the domain adaptation from the source mode to the target mode [30]. Motivated by the successful application of attention mechanism in visual inspection, which is a recent topic of great interest to worldwide scholars, some improved approaches are adopted in the field of fault diagnosis. Miao presented adaptive densely connected convolutional auto-encoder (ADCAE) for gearbox signal feature learning [31]. Yao *et al.* [32] explored recursive attention mechanism (RAM) to strengthen the anti-noise acoustic-based classification performance.

Inspired by these observations, a fault diagnosis approach of chemical process based on Adaptive Multiscale Convolutional Neural Network (AMCNN) is proposed in this paper. The model designs a novel multiscale CNN to perform feature extraction and fault diagnosis on raw time-series data. A new adaptive attention module is embedded into the multiscale CNN structure to estimate the importance of extracted features so as to emphasize the useful ones. The entire deep network is optimized by a metric learning tool, the triplet loss optimization, which provides the model with strong power of discriminability and generalization in diverse operating conditions.

The remainder of this paper is organized as follows. In Section 2, the theoretical background and the overall fault diagnosis model are illustrated in detail. In Section 3, experimental studies are implemented to verify the effectiveness of AMCNN, and the presented model is compared with other published works in terms of fault classification performance. Conclusion and further discussion are indicated in Section 4.

2. Methodology

The proposed model includes three key parts: (1) multiscale convolutional neural networks for feature extraction from multivariate time-series data; (2) adaptive attention module for representative characteristics selection and (3) triplet loss mechanism for anomaly diagnosis optimization. The detailed principles of the above concepts are introduced in the following sections, followed by the summarization of the overall fault diagnosis model framework at the end of this section.

2.1. Multiscale convolution

As one of the most widely applied models in deep learning, CNN has been investigated in a variety of recognition tasks. Stacked convolutional layers and fully-connected layers endow the powerful model with automatic feature extraction and classification. CNN can deal with various types of signals and 1DCNN is able to address raw temporal data without any time-consuming preliminary conversion into matrix diagrams. The convolutional kernel is the key unit to extract data characteristics, and the size of the receptive field determines how much information can be obtained from the input signals. However, the general CNN models configured with the fixed kernel width are trapped in the same size of receptive, which means they can only take the single-scale features into account. As a result, it is undesirable to capture efficient features under variable operating conditions.

Motivated by the Inception modules of GoogLeNet network [33], multiple kernels of different sizes can be integrated to perform convolutional operations on one-dimensional signals, and multiscale features are extracted at the same time. Fig. 1 reveals the differences between multiscale convolution models and the traditional convolutional networks, where F stands for the feature information, W the kernel size and C the number of channels. In

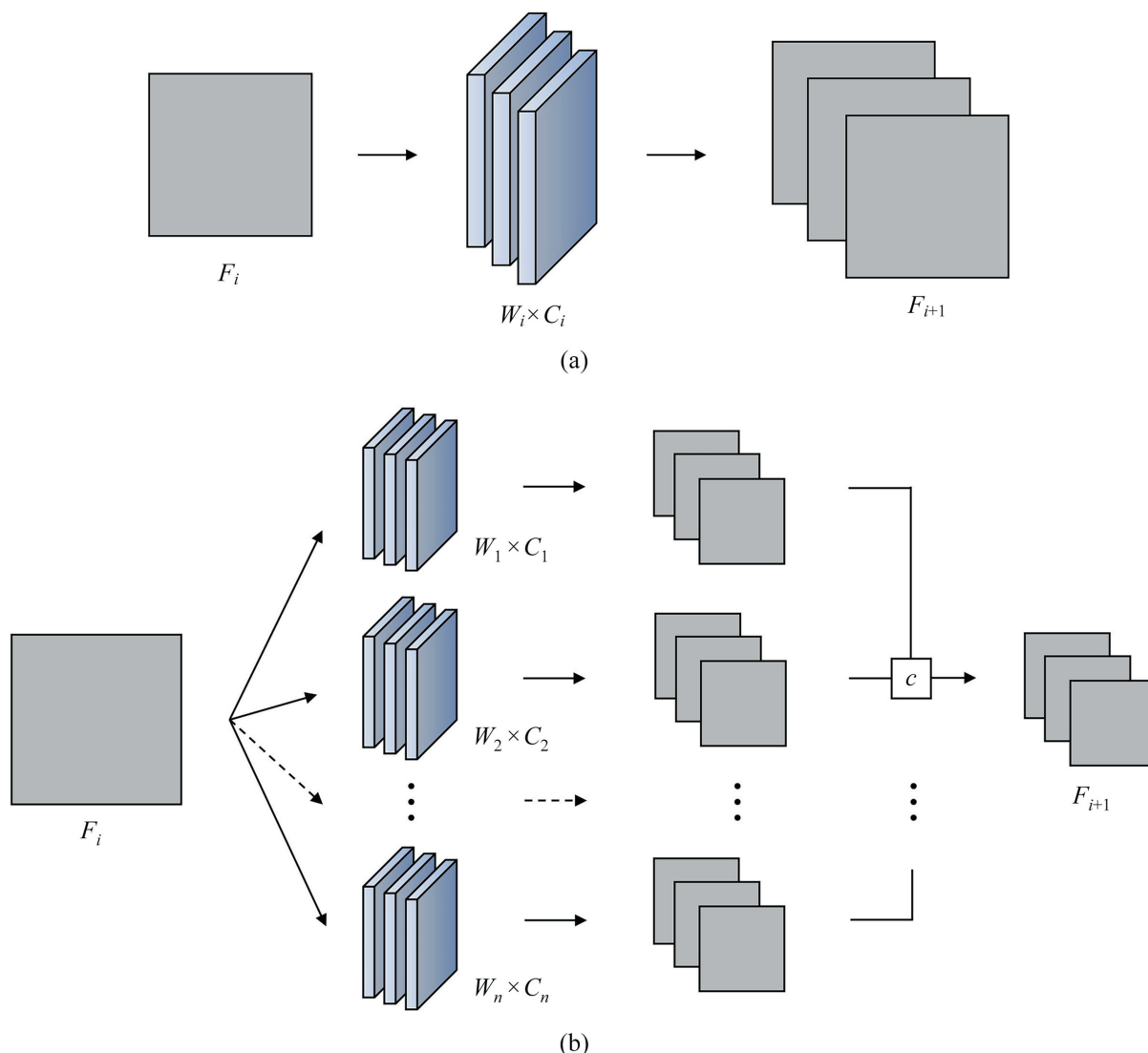


Fig. 1. Architectures of different convolutional layers in CNN. (a) One convolution. (b) Multiscale convolutions.

contrast with the basic convolutional layer, a multiscale convolutional block consists of multiscale convolutions and a flat convolutional layer. The filter concatenation achieves the fusion of multiscale features. Larger convolutional kernels allow more efficient extraction of slowly varying components at wider time scales, while the smaller ones can detect the transient behaviors of the data under fluctuation.

The designed multiscale convolution architecture example is illustrated in Fig. 2. A large-size convolutional kernel is introduced as the first layer of the network to seize the features of the low-frequency band. The number of channels is dynamically adjusted with the requirements of the task data. The model contains two multiscale convolution modules to obtain more information from the raw measurements. Three main kinds of filters with the sizes of 1×1 , 1×3 , and 1×5 are utilized to convolve with the characteristics accompanied by other large kernels as initial auxiliary. The multiple cascade channels facilitate the dimension reduction of the feature maps. The model robustness is further enhanced by embedding the maximum pooling operation with the subsampled vectors rescaled for further selection. In case of different classification tasks with multiple variables, the convolution kernel size and

the amount of input and output channels need to be enlarged along with the increasing dimension of input data.

2.2. Adaptive attention module

The multi-attention module aims to select useful features automatically and suppress useless information extracted from the deep networks. Two revised attention mechanisms, adaptive channel attention and adaptive spatial attention, are activated successively and combined as a whole to enhance feature learning.

2.2.1. Adaptive channel attention

The application of the attention mechanism in channel mode is regarded as a solution to select the most sensitive feature channel for fault diagnosis [34]. The adaptive channel attention algorithm shown in Fig. 3 receives the feature maps obtained from the multiscale convolutions, and the one generated through the i th kernel branch is defined as the feature map F_i . The global average pooling and max pooling are operated to acquire different channel descriptors F_{ave}^c and F_{max}^c , which make the element corresponding to the feature in the certain channel squeezed through a spatial

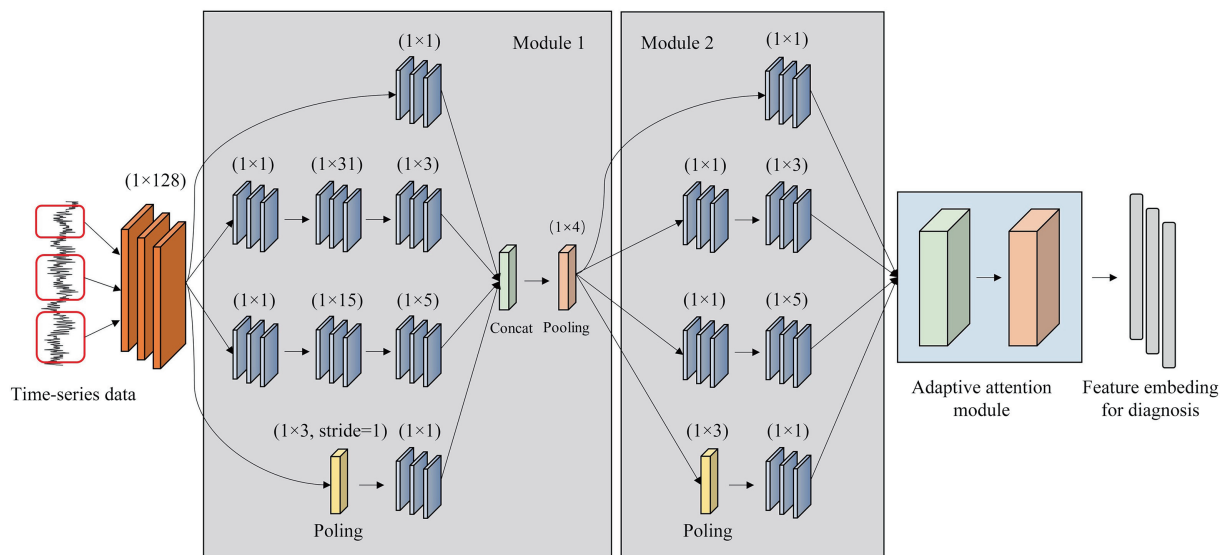


Fig. 2. Illustration of the multiscale convolution architecture.

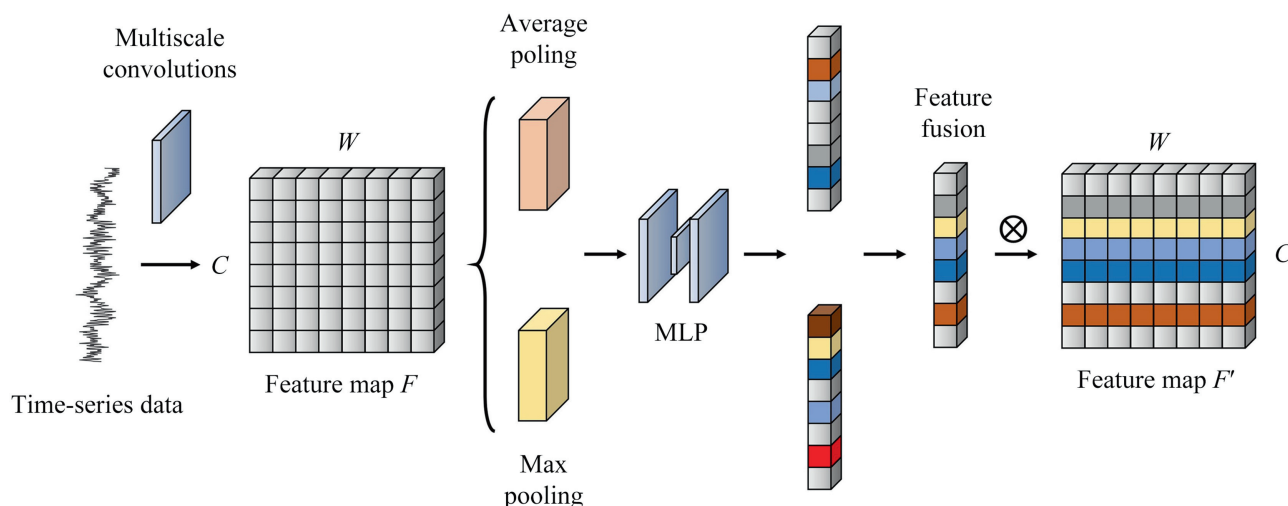


Fig. 3. Adaptive channel attention.

dimension $L \times 1$. These descriptors contain different ranges of information and denote diverse characteristics. Then, a shared network composed of the multilayer perceptron (MLP) with a hidden layer is embedded to generate the selective weights on channels and integrate them into a channel attention map $M_{\text{channel}} \in \mathbb{R}^{C \times 1 \times 1}$. The module computes the element-wise summation in Eq. (1):

$$M_{\text{channel}}(F) = \sigma(\text{MLP}(f_{\text{ave_pooling}}(F)) + \text{MLP}(f_{\text{max_pooling}}(F))) \\ = \sigma(W'(W(F_{\text{ave}}^c)) + W'(W(F_{\text{max}}^c))) \quad (1)$$

where σ performs the sigmoid function, $f_{\text{ave_pooling}}$ and $f_{\text{max_pooling}}$ launch the average-pooling and max-pooling operations, and W' and W are the weight matrixes of MLP.

The feature map F' output by the adaptive channel attention algorithm is presented in the following form:

$$F' = M_{\text{channel}}(F) \otimes F \quad (2)$$

where $\otimes$ represents the element-wise multiplication. Based on the excitation mechanism, the original feature maps extracted from different kernel widths are weighted-fused to get the novel feature maps with adaptive information scales.

2.2.2. Adaptive spatial attention

Different regions of the feature maps correspond to various variable relations and fault type characteristics. The adaptive spatial attention algorithm is employed to concentrate on the local information weighting to highlight significant features. The module combines the advantage of multiscale convolutions to handle the targeted features from all receptive fields. Fig. 4 displays the working procedure of adaptive spatial attention. Followed by the output of adaptive channel attention, the channel information is aggregated by two pooling operations to form the spatial context descriptors F_{ave}^s and F_{max}^s . The average-pooled and max-pooled features across the respective channels are forwarded to be concatenated and convolved by the convolutions at different scales. Next, the features are fused via an element-wise summation which is formulated in (3).

$$U = \sum_{i=1}^n u^i = \sum_{i=1}^n c^i(f_{\text{ave_pooling}}(F'), f_{\text{max_pooling}}(F')) \quad (3)$$

where c^i is the convolution operation of the i th kernel size and n is the number of convolution branches.

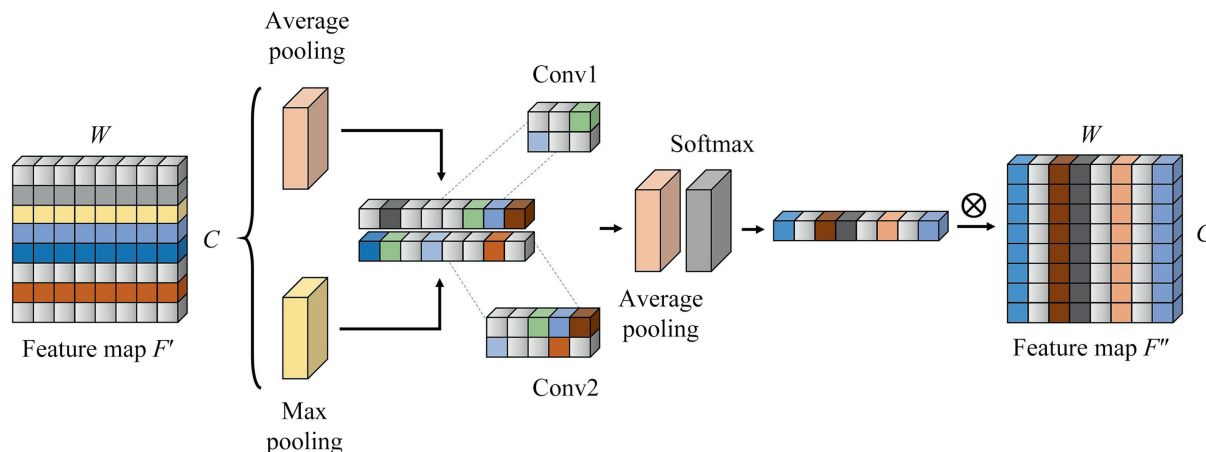


Fig. 4. Adaptive spatial attention.

Average pooling is executed along the branch axis to generate spatial statistics as $s \in \mathbb{R}^{1 \times L}$ and the softmax operator is followed close behind. The final spatial descriptor S is the compilation of each calculated element which is gained through the attention weights on various kernels.

$$S = \sum_{i=1}^n s_i \cdot u^i = \sum_{i=1}^n \varphi(f_{\text{ave_pooling}}(U)) \cdot u^i \quad (4)$$

Therefore, the final extracted feature map by adaptive spatial attention is computed as F'' .

$$F'' = S \otimes F' \quad (5)$$

To sum up, the adaptive attention module coalesced with the multiscale convolutions fulfills the adaptive feature fusion. The entire process of feature extraction and fusion is shown in Fig. 5. Compared to the traditional attention mechanism, adaptive channel attention and adaptive spatial attention are embedded for automatic weighted merge of multiscale features which boosts the feature learning of the AMCNN model.

2.3. Triplet loss optimization

The Triplet loss optimization is employed to achieve the generalization of fault features in different modes, which is first presented to learn better embeddings for face recognition in FaceNet model [35]. It allows the model to upgrade to a deep embedding metric network and get a better classification performance. The illustration of the triplet loss optimization is presented in Fig. 6.

The loss function defines a triplet of three samples, including the anchor x_i^a , the positive x_j^p and the negative x_k^n . The anchor x_i^a and the positive x_j^p belong to the same class, but the negative x_k^n has a different label from theirs. The optimization process encourages the distance between x_i^a and x_j^p with a margin to be closer than the one between x_i^a and x_k^n [36].

$$\|f(x_i^a) - f(x_j^p)\|_2^2 + \alpha < \|f(x_i^a) - f(x_k^n)\|_2^2, \forall (x_i^a, x_j^p, x_k^n) \in T \quad (6)$$

where α is a margin enforced between positive and negative pairs, T is the set of all possible triplets which has cardinality N , $f(\cdot)$

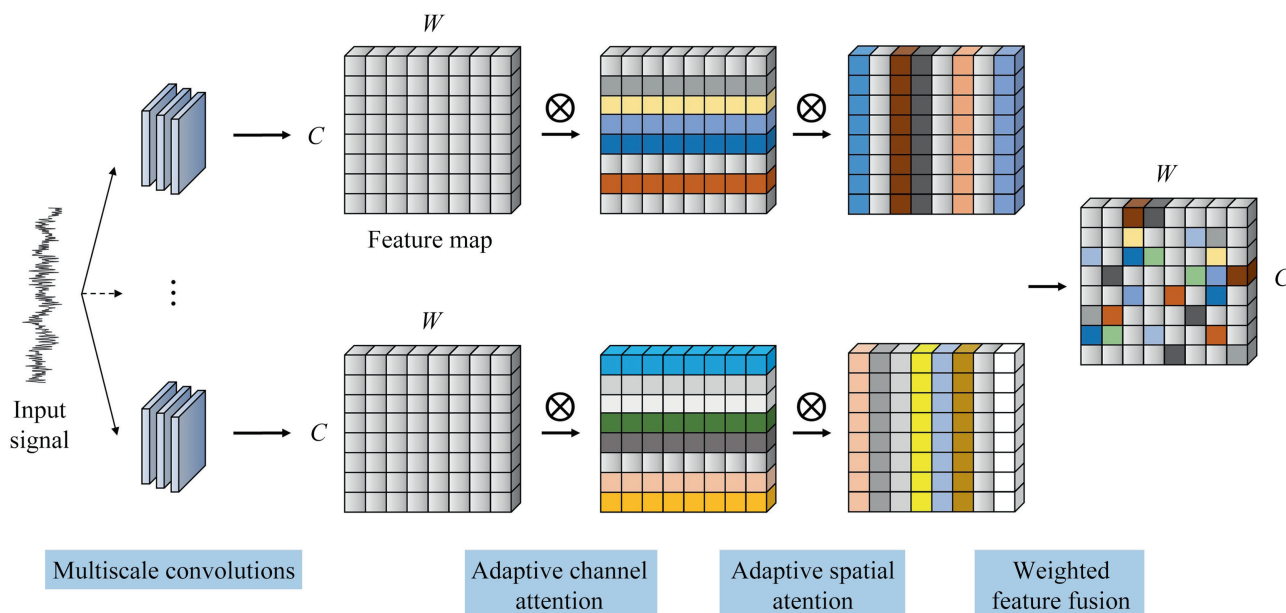


Fig. 5. Feature extraction and fusion with multiscale convolutions and the adaptive attention module.

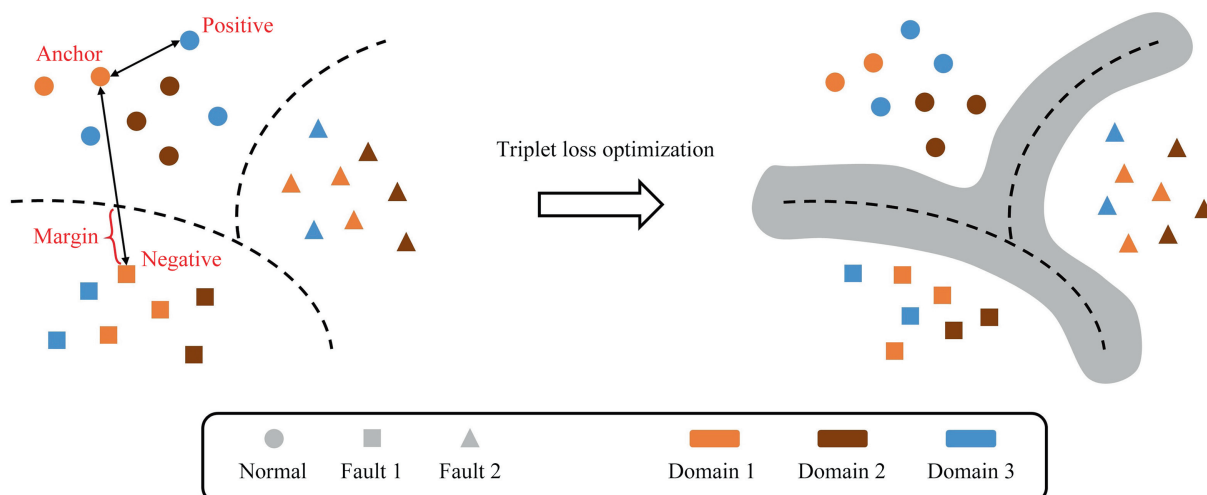


Fig. 6. Illustration of triplet loss optimization.

represents the feature extractor in a deep network to obtain the feature embedding.

It is crucial to select triplets that violate the triplet constraint in Eq.(7) to ensure fast convergence. Schroff *et al.* [35] define the triplets that fulfill Eq.(8) as easy triplets, and they do not need any optimization. Relatively speaking, hard triplets and semi-hard triplets are negative exemplars that are formulated as follows:

$$\|f(x_i^a) - f(x_j^p)\|_2^2 > \|f(x_i^a) - f(x_k^n)\|_2^2 \quad (7)$$

$$\|f(x_i^a) - f(x_j^p)\|_2^2 < \|f(x_i^a) - f(x_k^n)\|_2^2 < \|f(x_i^a) - f(x_j^p)\|_2^2 + \alpha \quad (8)$$

Semi-hard triplets are further away from the anchor than the positive exemplar, but the squared distance is still closer to the anchor-positive distance. These negatives lie inside the margin α .

As noted above, the triplet loss is designed to minimize the distance between x_i^a and x_j^p and maximize the distance between x_i^a and x_k^n . The formulation of the triplet loss is given as.

$$L_t = \frac{1}{N} \sum_i \left[\|f(x_i^a) - f(x_j^p)\|_2^2 - \|f(x_i^a) - f(x_k^n)\|_2^2 + \alpha \right]_+ \quad (9)$$

where $[\cdot]_+$ means $\max(\cdot, 0)$ [37]. Hard triplets and semi-hard triplets are selected to be optimized. With clustering the intrinsic distributions among multiple domains, the triplet loss optimization increases the discrimination of the diagnostic model and enhances its robustness under the multimode condition.

2.4. AMCNN-based fault diagnosis method

The detailed information of the proposed adaptive multiscale convolutional neural network model is unfolded in Fig. 7. The multiscale feature extraction module operates on the input one-dimensional data. The low-scale, mid-scale and high-scale kernels work in synergy. The adaptive attention module is embedded in the deep network which further improves the performance of feature selection. It adaptively pays more attention to the relevant information frames from the output of the convolutional layers. The selected features are input into two fully connected layers to recognize faulty conditions of the tasks. The classifier layer is constructed behind where softmax is utilized as activation function to output the probability of each class label. Triplet loss optimization is introduced to enhance the classification capability. The entire loss function of this network is defined in Eq.(10).

$$L = L_c + \lambda L_t + \mu L_{\text{weight}} \quad (10)$$

where L_c is the cross-entropy loss function of multi-classification, L_t is the optimized value of triplet loss, L_{weight} is the regularization of the model parameters to constrain the complexity of the network, λ and μ are the training coefficient set to 0.01 artificially.

The flowchart of the AMCNN-based fault diagnosis method is shown in Fig. 8. The procedures are described in detail below.

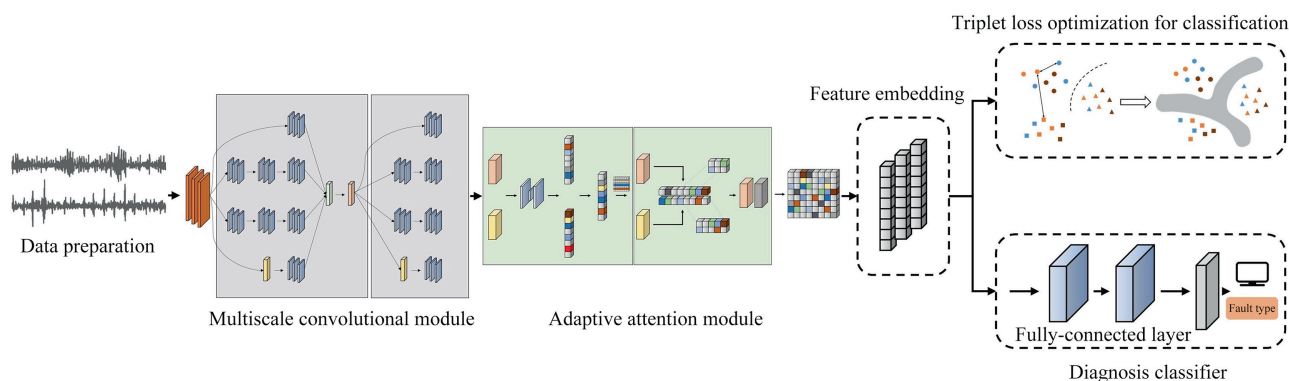


Fig. 7. Overview of the structure of the proposed AMCNN.

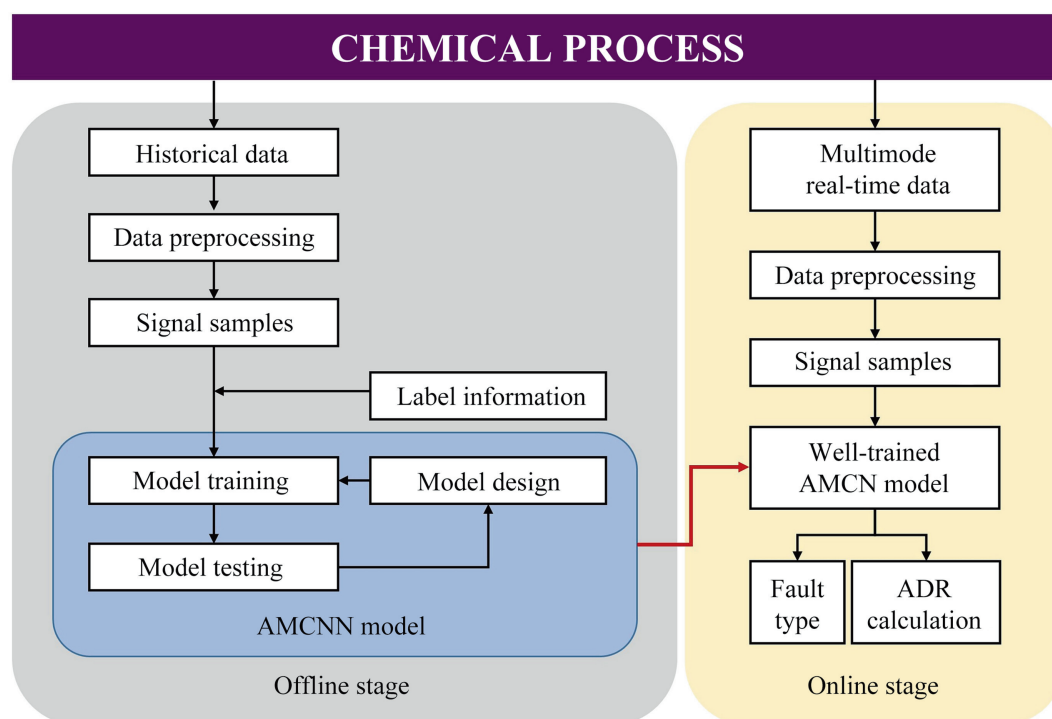


Fig. 8. Framework of AMCNN-based fault diagnosis method.

The offline stage can be divided into the following steps:

Step 1: A training dataset that contains normal and faulty data is generated for model building.

Step 2: Normalize the data of each variable with mean and standard deviation with label information.

Step 3: The AMCNN model is constructed based on the training dataset. The model hyperparameters are adjusted with the experimental requirements and continuously optimized according to the training results.

Step 4: The trained model is tested with the validation dataset, and the monitoring performance of the classifier is tested.

Step 5: If the testing results are satisfactory, the model is saved and prepared for further application. Otherwise, the feature extraction module and the classification module should be modified and retrained.

The online stage can be divided into the following steps:

Step 1: The AMCNN model saved at the off-line stage is loaded and prepared for real-time fault diagnosis.

Step 2: The real-time process signals under the multimode condition are collected from the task system.

Step 3: Data normalization with the same mean and standard deviation as training proceeding.

Step 4: The time-series data from the same working mode or other domains is put into the loaded diagnostic model for feature extraction in the new observation and testing.

Step 5: The diagnosis results are output and the accurately recognized rate is calculated.

3. Case Studies

In this section, the fault diagnosis performance of the proposed AMCNN model is evaluated by a non-isothermal continuous stirred tank reactor (CSTR) simulation and the Tennessee Eastman process (TEP), where AMCNN is compared with two classical statistical models, least-squares support vector machine (LSSVM) [38] and random forest (RF) [39], and other popular deep learning models,

Table 1
Confusion matrix for the *i*th class

	Amount of samples in the <i>i</i> th class (Predicted)	Amount of samples in the other classes (Predicted)
Amount of samples in the <i>i</i> th class (Actual)	<i>p</i> (true positive)	<i>b</i> (false negative)
Amount of samples in the other classes (Actual)	<i>q</i> (false positive)	<i>d</i> (true negative)

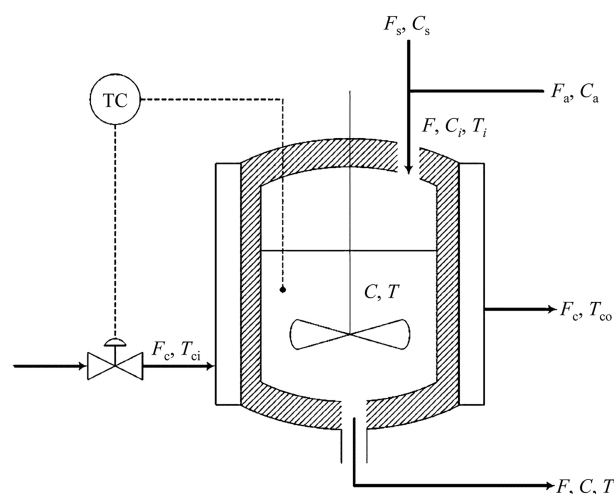


Fig. 9. Process flow diagram of the CSTR system.

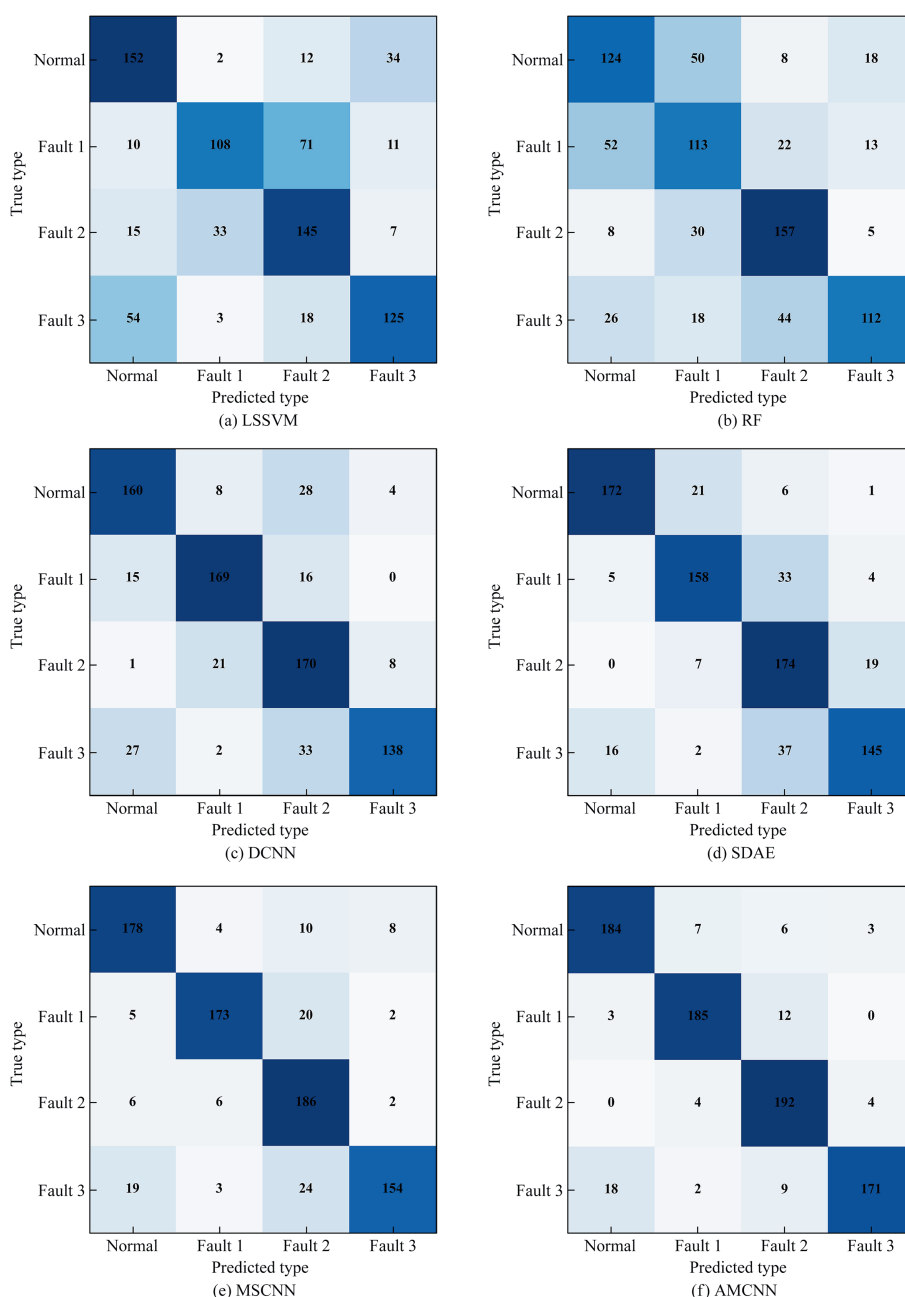


Fig. 10. Diagnosis results by different methods for CSTR.

including DCNN [25], stacked denoising autoencoders (SDAE) [19], and MSCNN [26] for ablation experiments.

In order to represent the experimental results directly, the confusion matrix for the i th class in multi-class classification tasks is defined as Table 1. The accurate diagnosis rate (ADR) is introduced in the following form:

$$ADR = \frac{p}{p + b + q + d} \quad (11)$$

ADR is a total classification accuracy of the fault diagnosis task which reflects the general performance of a model. The higher ADR is, the more accurately the model recognizes the fault type.

3.1. CSTR simulation

The AMCNN model is applied to the simulated non-isothermal CSTR system described by Yoon and MacGregor [40] and Cho

Table 2
Usage of different modules in the ablation experiment

Model	Multiscale convolution kernel	Adaptive attention module	Triplet loss optimization
1	×	×	×
2	✓	×	×
3	×	✓	×
4	×	×	✓
5	✓	✓	×
6	✓	×	✓
7	×	✓	✓
8	✓	✓	✓

et al. [6]. The reaction is the first order with the assumption that the reactor is well-mixed and the physical properties of the products are constant. The process flow diagram of the CSTR system is

Table 3

Comparison of the classification results of the ablation experiment

Model	1	2	3	4	5	6	7	8
ADR	62.25%	80.75%	70.25%	65.13%	89.50%	86.38%	76.13%	91.50%

Table 4

Setting information of four modes in the CSTR simulation

Variables	Reaction temperature T/K	Outlet concentration $C/\text{kmol}\cdot\text{m}^{-3}$	Cooling water flow $F_c/\text{m}^3\cdot\text{min}^{-1}$
Mode 1	368.25	0.8	15
Mode 2	370	0.75	6.8
Mode 3	368	0.9	9
Mode 4	371.65	0.5	12

shown in Fig. 9. The process has a feed stream which is the merged stream of the reactant A and the solvent S, a product stream, and a cooling water flow that controls the reaction temperature. The simulation system consists of the reaction component material balance and the energy balance, which are performed as follows:

$$\frac{dC_a}{dt} = \frac{F}{V} C_{a0} - \frac{F}{V} C_a - k_0 e^{-\frac{E}{RT}} C_a \quad (12)$$

$$V\rho c_p \frac{dT}{dt} = \rho c_p F(T_0 - T) - \frac{aF_c^{b+1}}{F_c + \frac{aF_c^b}{2\rho c_{pc}}} (T - T_{ci}) + (-\Delta H_{rxn}) V k_0 e^{-\frac{E}{RT}} C_a \quad (13)$$

where V is the volume of reacting mixture, k_0 is the first-order reaction rate constant, E is the activation energy, R is the gas constant, ρ and ρ_c are the density of reacting mixture in the feed stream and coolant, c_p and c_{pc} are the specific heat capacity of the reacting mixture and coolant, ΔH_{rxn} the reaction enthalpy.

There are nine process variables collected by the sensors, including the inlet temperature T_i , the cooling water temperature T_{ci} , the cooling water flow F_c , the reaction temperature T , the solvent flow F_s , the reactant flow F_a , the outlet concentration C , the inlet concentrations C_a and C_s . The simulation conditions are similar to Alcalá and Qin [41]. The setting information and the parameters are given in Appendix. Three scenarios of disturbances are introduced separately into the simulation system to describe the abnormal conditions:

- (1) a step bias of the reaction temperature T_i by adding 1.5 K.
- (2) a step bias of the inlet concentration C_a by adding 1.0 $\text{kmol}\cdot\text{m}^{-3}$.
- (3) a ramp change of the inlet concentration C_a by making $\frac{dC_a}{dt} = 0.1 \text{ kmol}\cdot\text{m}^{-3}\cdot\text{min}^{-1}$.

The simulations are performed under the normal and the three fault conditions in a certain mode called Mode 1. The variables are sampled every minute and the system is simulated for 400 min

with each disturbance made at $t = 151$ min. 200 samples containing 150 normal ones and 50 abnormal ones are used as the training dataset, while the last 200 samples are divided into the test dataset.

The confusion matrices of different fault diagnosis methods are shown in Fig. 10. The results depict that the AMCNN model achieves the best diagnosis performance. The two traditional statistical models cannot determine the categories of failures accurately and the diagnosis results of the other deep learning models are worse than the proposed model.

According to the point of view of Cho *et al.* [6], the first two faults are seen as simple faults which are caused by sensor biases and influence only one measurement heavily. Therefore, the above approaches diagnose the abnormality of CSTR simulation well in these scenarios. The third fault is a slow drift in the reaction kinetics. Catalyst poisoning leads to the exponential degradation of the reaction rate. It affects several other variables such as the output temperature T and the cooling water flow F_c so that it is regarded as a complex fault. On this occasion, AMCNN shows a pleasing result of fault identification while the other data-driven methods cannot recognize the fault type efficiently.

To further validate the optimal network architecture, the ablation experiment is applied based on the dataset mentioned above. The usage of different modules in the ablation experiment is displayed in Table 2. When the multiscale convolution kernels are not selected, the traditional convolution kernels will take place. The comparison of the classification results is shown in Table 3, which is the overall ADR of all categories. The three core modules of AMCNN have a cooperative relationship and work together to achieve the effectiveness of the entire framework.

In order to verify the robustness of the AMCNN model under the multimode condition, four modes are set in the CSTR simulation as Xiao *et al.* [42]. The mode setting information is presented in Table 4. The experiments are conducted across different operating situations with domain transfer between discrepant datasets. In each working mode, all monitored measurements are captured in the normal condition and three abnormal conditions in the same way as above. The diagnosis results are presented in Table 5. It is observed that AMCNN identifies the faults more precisely than the other models and MSCNN takes second place which implies that the adaptive attention module enhances the classification ability of the network.

The receiver operating characteristics (ROC) curve presents the model classification performance as a trade-off between selectivity and sensitivity. The false positive rate and the true positive rate are served as the axes of an ROC curve and the area under the curve is a quantitative way for classifier comparison. The model which gets the ROC curve closer to the upper left corner and the higher AUC

Table 5

ADR comparison of different methods under the multimode condition

Training dataset	Test dataset	LSSVM	RF	DCNN	SDAE	MSCNN	AMCNN
Mode 1,2	Mode 3,4	67.00%	68.38%	81.75%	86.13%	91.25%	94.50%
Mode 1,3	Mode 2,4	61.50%	65.00%	68.63%	75.38%	80.75%	84.00%
Mode 1,4	Mode 2,3	65.38%	69.88%	78.13%	77.25%	88.38%	91.63%
Mode 2,3	Mode 1,4	63.13%	65.75%	74.50%	76.63%	84.25%	89.38%
Mode 2,4	Mode 1,3	59.25%	63.88%	71.63%	74.25%	82.50%	86.75%
Mode 3,4	Mode 1,2	68.88%	72.75%	84.13%	82.50%	90.38%	92.88%
Average		64.19%	67.60%	76.46%	78.69%	86.25%	89.85%

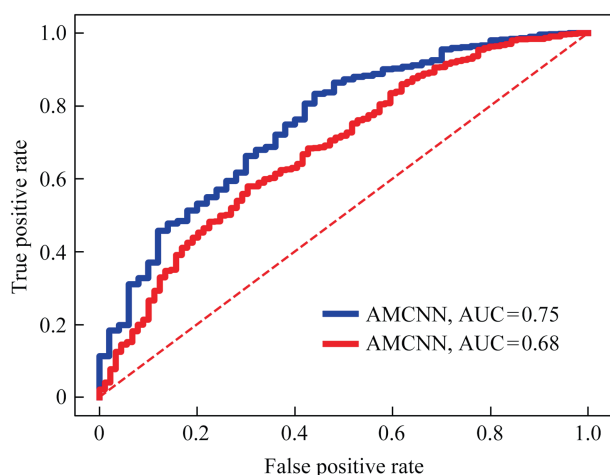


Fig. 11. ROC curves and AUC values compared by AMCNN and MSCNN.

value achieves the better diagnosis effect. Fig. 11 demonstrates the ROC curves and the AUC values of AMCNN and MSCNN under the multimode experimental condition. The proposed AMCNN model proves its versatility for different working circumstances with data domains migration from the source mode to the unknown target mode.

3.2. Tennessee Eastman process

The TEP which was first designed by Downs and Vogel [43] has been regarded as an industrial challenge for researchers in process monitoring. The revised version of the TEP is used in this experimental section which is available at <https://depts.washington.edu/control/LARRY/TE/download.html> [44]. It is presented as a chemical process simulator with 52 variables including 41 measured variables and 11 manipulated variables under observation. The redundant manipulated variable XMV(12) representing the agitator speed is excluded which keeps constant during the simulation. There are 28 process disturbances programmed in the revised simulator which make the process operate on abnormal conditions. The first 20 types defined in Table 6 are selected for comparison with other algorithms.

The process consists of five unit operations: a reactor, a condenser, a vapor–liquid separator, a recycle compressor and a stripper. The schematic of the TEP model is shown as Fig. 12. Two products are produced from four reactants with an inert and a byproduct involved. Except for the base case for initial setting, there are six modes of process operations at different product mass ratios or production rates. Mode 1 is utilized as a benchmark in the following experiments.

The data preparation and simulation process refers to [25]. The measurements are generated on MATLAB/Simulink shown in Table 7. The sampling interval is set to 3 min. The model firstly runs for 500 h in the normal state with 10,000 normal sample matrices collected. For 20 different faulty states, the process simulation repeats for 10 times. The model runs for 50 h in each simulation with the specific disturbance introduced after the initial 10-hour normal running. Therefore, there are 40 h of simulation data and 800 fault sample matrices collected for every run of a fault type. It should be noted that the simulation can only last for 7 h after the disturbance IDV (6) is inserted because some variables will exceed the allowable limits. For normal data, 80% of the sample matrices are randomly selected as the training set while the remaining as the testing set. For faulty data, eight runs of the simulation data are randomly chosen for training with the other two for testing.

The recognition rates of AMCNN are compared with the above mentioned diagnostic models in Table 8. For the overall performance of the models, AMCNN reaches the highest average accuracy which is 3.65% higher than MSCNN and 5.78% higher than DCNN. The corresponding recognition rates of 17 fault types are higher than 90%, and the proposed model displays the best diagnostic performance in 14 kinds of abnormal conditions. Considering the previous works, IDV(3) and IDV(16) make the diagnosis results unsatisfactory in some of the published models, but AMCNN overcomes these challenges. Fig. 13 illustrates the confusion matrix of the proposed model's diagnosis results, where the main diagonal numbers are the percentage of samples correctly identified and label 0 means the normal condition. As shown in the confusion matrix, IDV(9) and IDV(15) are still hard to recognize. The combined effect of the input disturbances and the internal control system makes the situation hard to be distinguished from the normal state and the other faulty states. The rest of the tested models perform better on some specific faults, such as RF on IDV(13), SDAE on IDV(17), and DCNN on IDV(10) and IDV(11).

Table 6
Process disturbances in the TEP

IDVs	Variable Description	Type
IDV(1)	A/C feed ratio, B composition constant (stream 4)	Step
IDV(2)	B composition, A/C ratio constant (stream 4)	Step
IDV(3)	D feed temperature (stream 2)	Step
IDV(4)	Reactor cooling water inlet temperature	Step
IDV(5)	Condenser cooling water inlet temperature	Step
IDV(6)	A feed sudden loss (stream 1)	Step
IDV(7)	C header pressure loss-reduced availability (stream 4)	Step
IDV(8)	A, B, C feed composition (stream 4)	Random variation
IDV(9)	D feed temperature (stream 2)	Random variation
IDV(10)	C feed temperature (stream 4)	Random variation
IDV(11)	Reactor cooling water inlet temperature	Random variation
IDV(12)	Condenser cooling water inlet temperature	Random variation
IDV(13)	Reaction kinetics	Slow drift
IDV(14)	Reactor cooling water valve	Sticking
IDV(15)	Condenser cooling water valve	Sticking
IDV(16)	Unknown	Unknown
IDV(17)	Unknown	Unknown
IDV(18)	Unknown	Unknown
IDV(19)	Unknown	Unknown
IDV(20)	Unknown	Unknown

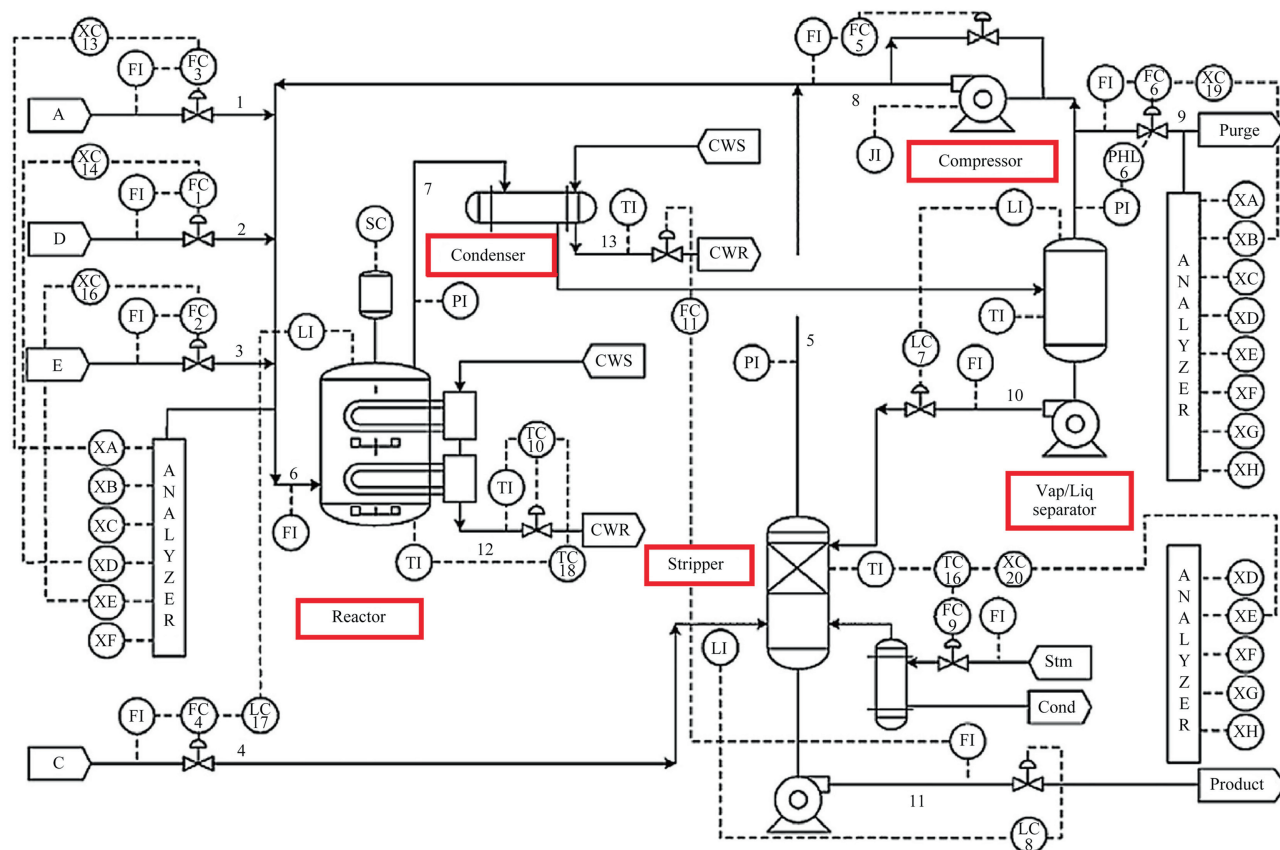


Fig. 12. Schematic of the TEP.

Table 7

The sample matrices collected from the TEP for training and testing

Status index	Simulation length (training)/h	Amount of data matrices (training)	Simulation length (testing)/h	Amount of data matrices (testing)
IDV (1)–(5) & (7)–(20)	40 × 19 × 8	800 × 19 × 8	40 × 19 × 2	800 × 19 × 2
IDV (6)	7 × 1 × 8	140 × 1 × 8	7 × 1 × 2	140 × 2
Fault	6136	122720	1534	30680
Normal	400	8000	100	2000

Table 8

ADR comparison on the TEP

IDVs	LSSVM	RF	DCNN	SDAE	MSCNN	AMCNN
Normal	86.95%	91.60%	97.10%	94.30%	95.25%	98.65%
IDV(1)	97.69%	98.81%	98.69%	99.00%	98.94%	99.69%
IDV(2)	97.88%	97.75%	98.25%	99.19%	99.38%	99.50%
IDV(3)	14.25%	24.06%	86.75%	55.81%	77.88%	89.56%
IDV(4)	84.00%	90.25%	96.44%	73.38%	95.69%	99.38%
IDV(5)	74.56%	68.63%	88.81%	58.88%	95.00%	96.69%
IDV(6)	96.43%	98.93%	97.50%	99.29%	100.00%	100.00%
IDV(7)	97.81%	98.81%	99.81%	99.81%	99.69%	99.75%
IDV(8)	96.38%	88.94%	90.69%	93.25%	96.63%	96.94%
IDV(9)	13.13%	26.63%	58.25%	42.75%	64.50%	71.44%
IDV(10)	60.56%	43.06%	94.94%	78.69%	93.06%	94.06%
IDV(11)	79.63%	63.81%	97.50%	57.69%	95.75%	97.44%
IDV(12)	81.69%	91.06%	94.38%	69.13%	95.56%	97.06%
IDV(13)	71.63%	95.56%	90.94%	87.63%	92.81%	94.44%
IDV(14)	85.31%	79.63%	98.75%	85.94%	99.25%	99.31%
IDV(15)	11.50%	17.06%	26.06%	54.31%	34.13%	49.25%
IDV(16)	52.63%	33.13%	47.00%	36.81%	81.06%	91.75%
IDV(17)	83.63%	94.25%	95.13%	97.75%	91.56%	94.94%
IDV(18)	78.06%	91.19%	92.25%	83.56%	95.38%	96.50%
IDV(19)	72.19%	80.25%	96.50%	85.88%	93.81%	97.88%
IDV(20)	82.69%	87.63%	92.31%	82.00%	88.75%	93.56%
Average	71.52%	73.55%	87.24%	77.19%	89.37%	93.02%

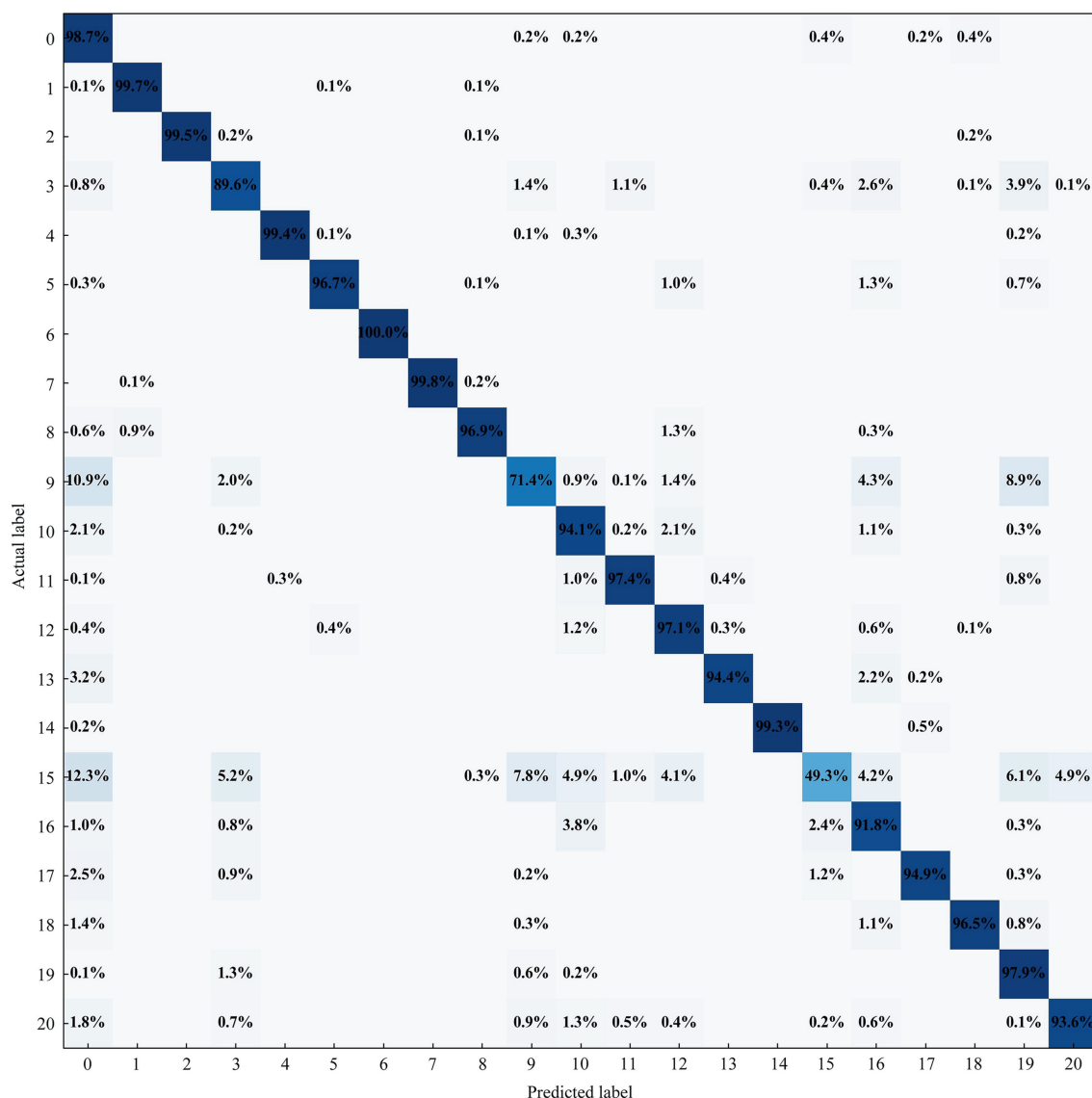


Fig. 13. Recognition rates of AMCNN in a confusion matrix.

In general, AMCNN reveals its powerful performance in classification and extends the processing capability of traditional convolutional networks for time-series data.

In order to demonstrate the generalizability of AMCNN in multimode fault diagnosis tasks, the other five different running modes are introduced as the experimental data. The descriptions of these working conditions are detailed in Table 9. The rules of the TEP simulation and data preparation are fulfilled in the same way. The training and testing sets are selected from mutually exclusive modes which indicates that the fault data of the source mode is applied to assist the diagnosis model training for the unseen target mode. When the simulation data of a certain mode is used for diag-

nosis, the data of the rest modes is used for training separately. Therefore, each of the methods to be tested undergoes a group of 30 pair experiments and then the comprehensive diagnosis rate is calculated to evaluate its performance.

Table 10 shows the overall ADR of the models involved in this paper. On account of the poor recognition precision of LSSVM, RF and SDAE, the detailed results of fault diagnosis for each mode pair are exhibited in Table 11 with only AMCNN, MSCNN and DCNN contrasted. It is observed that deep learning models are more robust than statistical models facing the industrial challenge of manufacturing mode switching. According to the compared outcomes, AMCNN still outperforms all the other models, certifying

Table 9
Descriptions of six running modes of the TEP

Mode	Product G/H mass ratio	G production rate/kg·h ⁻¹	H production rate/kg·h ⁻¹	Statement
1	50/50	7038	7038	Base mode
2	10/90	1408	12669	Optimized operation
3	90/10	10000	1111	Optimized operation
4	50/50	—	—	Maximum productivity
5	10/90	—	—	Maximum productivity
6	90/10	—	—	Maximum productivity

Table 10

Overall ADR comparison on the TEP under the multimode condition

	LSSVM	RF	DCNN	SDAE	MSCNN	AMCNN
Average	58.94%	63.15%	80.38%	67.52%	83.37%	87.29%

Table 11

Detailed results of fault diagnosis with AMCNN, MSCNN and DCNN

Training set		Testing set					
		Mode 1	Mode 2	Mode 3	Mode 4	Mode 5	Mode 6
(a) AMCNN	Mode 1	—	91.77%	89.78%	86.20%	85.44%	87.38%
	Mode 2	90.59%	—	90.04%	84.16%	90.45%	89.13%
	Mode 3	88.90%	90.81%	—	77.46%	85.73%	91.48%
	Mode 4	88.33%	82.05%	80.34%	—	85.30%	84.69%
	Mode 5	86.28%	89.55%	84.74%	86.93%	—	89.97%
	Mode 6	89.87%	86.92%	89.87%	86.32%	88.09%	—
	Mode 6	89.87%	86.92%	89.87%	86.32%	88.09%	—
Training set		Testing set					
		Mode 1	Mode 2	Mode 3	Mode 4	Mode 5	Mode 6
(b) MSCNN	Mode 1	—	88.93%	84.79%	83.30%	83.41%	82.21%
	Mode 2	88.76%	—	85.10%	79.35%	88.24%	86.12%
	Mode 3	87.46%	83.38%	—	70.80%	85.34%	88.57%
	Mode 4	83.82%	77.22%	73.12%	—	78.67%	84.75%
	Mode 5	79.96%	87.39%	86.11%	79.85%	—	84.98%
	Mode 6	79.17%	85.14%	86.65%	84.29%	84.32%	—
	Mode 6	79.17%	85.14%	86.65%	84.29%	84.32%	—
Training set		Testing set					
		Mode 1	Mode 2	Mode 3	Mode 4	Mode 5	Mode 6
(c) DCNN	Mode 1	—	86.36%	84.20%	72.47%	78.69%	80.41%
	Mode 2	85.83%	—	79.30%	78.13%	84.75%	82.36%
	Mode 3	82.52%	80.45%	—	66.03%	81.55%	84.70%
	Mode 4	75.06%	77.98%	69.88%	—	79.70%	78.32%
	Mode 5	81.39%	83.20%	80.62%	79.30%	—	84.91%
	Mode 6	82.11%	81.84%	85.05%	80.90%	83.29%	—
	Mode 6	82.11%	81.84%	85.05%	80.90%	83.29%	—

that the adaptive attention module and the triplet loss optimization achieve significant improvements on feature learning of the proposed model from process signals directly. The effectiveness of AMCNN under the multimode condition confirms that the features extracted at different scales can fully characterize the data distribution and the intrinsic interaction between multiple variables. The integrated model becomes more adaptive between the source mode and the target mode.

4. Conclusions

In this paper, an adaptive multiscale convolutional Neural Network (AMCNN) model is proposed for fault diagnosis of chemical processes. Different from the traditional deep CNN framework, the end-to-end learning model directly takes raw time-series data as input and captures multiscale features automatically. Adaptive attention mechanism constituted by channel-wise attention and spatial-wise attention stimulates key feature selection for classification across different working modes. Metric loss optimization through the triplet loss function facilitates the construction of generalized decision boundaries by narrowing the intra-class distance of features and expanding inter-class differences, which augments the discriminative capability of the network. The CSTR simulation and the TEP are utilized to highlight the applicability of the proposed model. AMCNN model creates higher-precision judgments on fault types compared with other commonly used diagnostic models no matter whether the process is under the multimode production situation. From the perspective of process safety, AMCNN can accurately recognize various types of faults in chemical processes and provide the basis of decision on the follow-up fault elimination.

Several issues also deserve further study in the future. First, AMCNN takes a little more time than the traditional CNN or statistical models for diagnosis due to the complexity of the model. It is a noteworthy task to make the model lightweight and accelerate the proceeding while maintaining its powerful identification accuracy. Subsequently, as a supervised diagnostic model, it is essential to further reduce the dependence on the labeled data. AMCNN is required to demonstrate its effectiveness with limited training data. Last but not least, the published fault diagnosis models including AMCNN still count on the knowledge of fault categories that have occurred before. These current models are likely to classify a brand new fault as an old type. A possible situation based on the application of AMCNN is to evaluate the contribution of each process variable or representative captured feature. The weight combination bestowed on the variables or features in the model is a symbol to define the attributes of the new fault.

Declaration of Competing Interest

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

Acknowledgements

The authors gratefully acknowledge support from the National Science and Technology Innovation 2030 Major Project of the Ministry of Science and Technology of China (2018AAA0101605) and the National Natural Science Foundation of China (21878171).

References

- [1] X.T. Bi, R.S. Qin, D.Y. Wu, S.D. Zheng, J.S. Zhao, One step forward for smart chemical process fault detection and diagnosis, *Comput. Chem. Eng.* 164 (2022) 107884.
- [2] M. Quiñones-Grueiro, A. Prieto-Moreno, C. Verde, O. Llanes-Santiago, Data-driven monitoring of multimode continuous processes: A review, *Chemom. Intell. Lab. Syst.* 189 (2019) 56–71.
- [3] Z.Q. Ge, Z.H. Song, F.R. Gao, Review of recent research on data-based process monitoring, *Ind. Eng. Chem. Res.* 52 (10) (2013) 3543–3562.
- [4] B.M. Wise, N.L. Ricker, D.F. Veltkamp, B.R. Kowalski, Theoretical basis for the use of principal component models for monitoring multivariate processes, *Process Contr. Qual.* 1 (1990) 41–51.
- [5] W.F. Ku, R.H. Storer, C. Georgakis, Disturbance detection and isolation by dynamic principal component analysis, *Chemom. Intell. Lab. Syst.* 30 (1) (1995) 179–196.
- [6] J.H. Cho, J.M. Lee, S. Wook Choi, D. Lee, I.B. Lee, Fault identification for process monitoring using kernel principal component analysis, *Chem. Eng. Sci.* 60 (1) (2005) 279–288.
- [7] J.F. MacGregor, C. Jaekle, C. Kiparissides, M. Koutoudi, Process monitoring and diagnosis by multiblock PLS methods, *AIChE J.* 40 (5) (1994) 826–838.
- [8] A.A. Khan, J.R. Moyne, D.M. Tilbury, Virtual metrology and feedback control for semiconductor manufacturing processes using recursive partial least squares, *J. Process. Control* 18 (10) (2008) 961–974.
- [9] Y.N. Dong, S.J. Qin, Dynamic-inner partial least squares for dynamic data modeling, *IFAC-PapersOnLine* 48 (8) (2015) 117–122.
- [10] M. Kano, S. Tanaka, S. Hasebe, I. Hashimoto, H. Ohno, Monitoring independent components for fault detection, *AIChE J.* 49 (4) (2003) 969–976.
- [11] L.H. Chiang, E.L. Russell, R.D. Braatz, Fault diagnosis in chemical processes using Fisher discriminant analysis, discriminant partial least squares, and principal component analysis, *Chemom. Intell. Lab. Syst.* 50 (2) (2000) 243–252.
- [12] L.H. Chiang, M.E. Kotanchek, A.K. Kordon, Fault diagnosis based on Fisher discriminant analysis and support vector machines, *Comput. Chem. Eng.* 28 (8) (2004) 1389–1401.
- [13] H.L. Zhang, Y.S. Qi, L. Wang, X.J. Gao, X.C. Wang, Fault detection and diagnosis of chemical process using enhanced KECA, *Chemom. Intell. Lab. Syst.* 161 (2017) 61–69.
- [14] H. Ziaei-Halimejani, R. Zarghami, S.S. Mansouri, N. Mostoufi, Data-driven fault diagnosis of chemical processes based on recurrence plots, *Ind. Eng. Chem. Res.* 60 (7) (2021) 3038–3055.
- [15] K. Watanabe, S. Hirota, L.Y. Hou, D.M. Himmelblau, Diagnosis of multiple simultaneous fault via hierarchical artificial neural networks, *AIChE J.* 40 (5) (1994) 839–848.
- [16] D. Ruiz, J. Nogués, Z. Calderón, A. Espuña, L. Puigjaner, Neural network based framework for fault diagnosis in batch chemical plants, *Comput. Chem. Eng.* 24 (2–7) (2000) 777–784.
- [17] Z.P. Zhang, J.S. Zhao, A deep belief network based fault diagnosis model for complex chemical processes, *Comput. Chem. Eng.* 107 (2017) 395–407.
- [18] F.F. Cheng, Q.P. He, J.S. Zhao, A novel process monitoring approach based on variational recurrent autoencoder, *Comput. Chem. Eng.* 129 (2019) 106515.
- [19] S.D. Zheng, J.S. Zhao, A new unsupervised data mining method based on the stacked autoencoder for chemical process fault diagnosis, *Comput. Chem. Eng.* 135 (2020) 106755.
- [20] S.Y. Zhang, K.X. Bi, T. Qiu, Bidirectional recurrent neural network-based chemical process fault diagnosis, *Ind. Eng. Chem. Res.* 59 (2) (2020) 824–834.
- [21] D.Y. Wu, J.S. Zhao, Process topology convolutional network model for chemical process fault diagnosis, *Process. Saf. Environ. Prot.* 150 (2021) 93–109.
- [22] X.T. Bi, J.S. Zhao, A novel orthogonal self-attentive variational autoencoder method for interpretable chemical process fault detection and identification, *Process. Saf. Environ. Prot.* 156 (2021) 581–597.
- [23] R.S. Qin, J.S. Zhao, High-efficiency generative adversarial network model for chemical process fault diagnosis, *IFAC-PapersOnLine* 55 (7) (2022) 732–737.
- [24] C.Y. Weng, B.C. Lu, Q. Gu, A multi-scale kernel-based network with improved attention mechanism for rotating machinery fault diagnosis under noisy environments, *Meas. Sci. Technol.* 33 (5) (2022) 055108.
- [25] H. Wu, J.S. Zhao, Deep convolutional neural network model based chemical process fault diagnosis, *Comput. Chem. Eng.* 115 (2018) 185–197.
- [26] Q.S. Song, P. Jiang, A multi-scale convolutional neural network based fault diagnosis model for complex chemical processes, *Process. Saf. Environ. Prot.* 159 (2022) 575–584.
- [27] W. Zhang, C.H. Li, G.L. Peng, Y.H. Chen, Z.J. 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.* 100 (2018) 439–453.
- [28] H. Wu, J.S. Zhao, Self-adaptive deep learning for multimode process monitoring, *Comput. Chem. Eng.* 141 (2020) 107024.
- [29] S.D. Zheng, J.S. Zhao, A self-adaptive temporal-spatial self-training algorithm for semisupervised fault diagnosis of industrial processes, *IEEE Trans. Ind. Inform.* 18 (10) (2022) 6700–6711.
- [30] H. Wu, J.S. Zhao, Fault detection and diagnosis based on transfer learning for multimode chemical processes, *Comput. Chem. Eng.* 135 (2020) 106731.
- [31] M.Q. Miao, C.H. Liu, J.B. Yu, Adaptive densely connected convolutional auto-encoder-based feature learning of gearbox vibration signals, *IEEE Trans. Instrum. Meas.* 70 (2021) 1–11.
- [32] Y. Yao, G. Gui, S.X. Yang, S. Zhang, An adaptive anti-noise network with recursive attention mechanism for gear fault diagnosis in real-industrial noise environment condition, *Measurement* 186 (2021) 110169.
- [33] C. Szegedy, W. Liu, Y.Q. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, Boston, MA, 2015, pp. 1–9.
- [34] T.D. Huang, S. Fu, H.N. Feng, J.F. Kuang, Bearing fault diagnosis based on shallow multi-scale convolutional neural network with attention, *Energies* 12 (20) (2019) 3937.
- [35] F. Schroff, D. Kalenichenko, J. Philbin, FaceNet: A unified embedding for face recognition and clustering, in: 2015 IEEE Conf. Comput. Vis. Pattern Recognit. CVPR, 2015, pp. 815–823.
- [36] T. Han, Y.F. Li, M. Qian, A hybrid generalization network for intelligent fault diagnosis of rotating machinery under unseen working conditions, *IEEE Trans. Instrum. Meas.* 70 (2021) 1–11.
- [37] Y. Zhuo, Z.Q. Ge, Auxiliary information-guided industrial data augmentation for any-shot fault learning and diagnosis, *IEEE Trans. Ind. Inform.* 17 (11) (2021) 7535–7545.
- [38] S.M. He, X.G. Liu, Y.L. Wang, S.H. Xu, J.G. Lu, C.H. Yang, S.W. Zhou, Y.X. Sun, W. H. Gui, W.Z. Qin, An effective fault diagnosis approach based on optimal weighted least squares support vector machine, *Can. J. Chem. Eng.* 95 (12) (2017) 2357–2366.
- [39] Y.Q. Zhang, L. Luo, X. Ji, Y.Y. Dai, Improved random forest algorithm based on decision paths for fault diagnosis of chemical process with incomplete data, *Sensors (Basel)* 21 (20) (2021) 6715.
- [40] S. Yoon, J.F. MacGregor, Fault diagnosis with multivariate statistical models part I: Using steady state fault signatures, *J. Process. Control* 11 (4) (2001) 387–400.
- [41] C.F. Alcalá, S.J. Qin, Reconstruction-based contribution for process monitoring, *Automatica* 45 (7) (2009) 1593–1600.
- [42] Y.T. Xiao, H.B. Shi, B.Y. Wang, Y. Tao, S. Tan, B. Song, Weighted conditional discriminant analysis for unseen operating modes fault diagnosis in chemical processes, *IEEE Trans. Instrum. Meas.* 71 (2022) 1–14.
- [43] J.J. Downs, E.F. Vogel, A plant-wide industrial process control problem, *Comput. Chem. Eng.* 17 (3) (1993) 245–255.
- [44] A. Bathelt, N.L. Ricker, M. Jelali, Revision of the Tennessee Eastman process model, *IFAC-PapersOnLine* 48 (8) (2015) 309–314.

Appendix. CSTR simulation parameters and initial conditions

The simulation parameters of the CSTR system are $V = 1 \text{ m}^3$, $\rho = 10^6 \text{ g} \cdot \text{m}^{-3}$, $\rho_c = 10^6 \text{ g} \cdot \text{m}^{-3}$, $E/R = 8330.1 \text{ K}$, $c_p = 1 \text{ cal} \cdot \text{g}^{-1} \cdot \text{K}^{-1}$, $c_{pc} = 1 \text{ cal} \cdot \text{g}^{-1} \cdot \text{K}^{-1}$, $b = 0.5$, $k_0 = 10^{10} \text{ m}^3 \cdot \text{kmol}^{-1} \cdot \text{min}^{-1}$, $a = 1.678 \times 10^6 \text{ cal} \cdot \text{min}^{-1} \cdot \text{K}^{-1}$, $\Delta H_{rxn} = -1.3 \times 10^7 \text{ cal} \cdot \text{kmol}^{-1}$.

The information of the temperature PI controller is $K_c = -1.5$, $T_i = 5.0$.

The initial conditions for the simulation are $T_i = 370.0 \text{ K}$, $T_c = 365.0 \text{ K}$, $F_c = 15 \text{ m}^3 \cdot \text{min}^{-1}$, $T = 368.25 \text{ K}$, $F_s = 0.9 \text{ m}^3 \cdot \text{min}^{-1}$, $F_a = 0.1 \text{ m}^3 \cdot \text{min}^{-1}$, $C = 0.8 \text{ kmol} \cdot \text{m}^{-3}$, $C_s = 0.1 \text{ kmol} \cdot \text{m}^{-3}$, $C_a = 19.1 \text{ kmol} \cdot \text{m}^{-3}$.

Measurement noises and disturbances are described as $x_t = \varphi x_{t-1} + \sigma_e e_t$, $x_{t,\text{meas}} = x_t + \sigma_m m_t$. Gaussian noise with small variance is added to the simulation, and the detailed setting information can be checked in [40].