Research on Fault Feature Extraction and Early Warning Method Based on MLP and Attention Mechanism CNN Fusion

Yanhua Shi
School of Information Engineering, Zhengzhou University of Technology, Zhengzhou 450064, China
E-mail: Yanhua_Shi@yeah.net

Keywords: fault feature extraction, attention mechanism, fusion technology, fault early warning method

Received: November 26, 2024

In modern industrial automation systems, fault feature extraction and early warning are the key technologies to ensure the stable operation of equipment. Traditional machine learning methods such as multi-layer perceptron often face the limitations of feature representation ability when dealing with such problems. In recent years, the attention mechanism combined with convolutional neural networks has become an effective way to improve the effect of feature extraction. CNN can effectively capture the spatial correlation in the image or signal by its local connection and weight sharing characteristics, and the attention mechanism can automatically focus on the most discriminative part among many features. The MLP is fused with CNN of attention matrix. Firstly, the original fault data is extracted by using CNN, and then the extracted features are weighted by attention module, emphasizing the most critical information for fault diagnosis. This fusion model not only inherits the nonlinear mapping ability of MLP, but also enhances the feature selection and representation ability of CNN in complex signal processing. Experiments show that the method can significantly improve the accuracy and robustness of fault feature extraction. Among 300 fault sample data, the S-network can correctly distinguish 295 fault types, and the early warning accuracy is more than 98%, which proves the effectiveness of the method. This study can achieve more effective early warning, reduce the cost of equipment maintenance, and improve the reliability of the system.

Povzetek: Raziskava združuje MLP in CNN z mehanizmom pozornosti za zgodnje opozarjanje na okvare v industriji. CNN izlušči prostorske značilnosti, pozornost poudari ključne signale, MLP izboljša nelinearno preslikavo.

1 Introduction

In the intricate tapestry of modern industrial systems, fault detection and early warning mechanisms loom large as a pivotal technology, ensuring the unwavering stability of equipment and augmenting production efficiency [1, 2]. Amidst the swift advancements in artificial intelligence, machine learning methodologies have come to occupy a paramount position in the realm of fault feature extraction and early warning systems. Notably, two deep learning paradigms, the Multilayer Perceptron (MLP) and the Convolutional Neural Network (CNN), have exhibited remarkable prowess across diverse domains, leveraging their distinctive advantages. Nonetheless, the inherent constraints of a solitary model have begun to surface, highlighting the need for a more comprehensive approach. Therefore, researchers began to explore the method of fusion of different models in order to obtain more powerful fault feature extraction capabilities.

As a feedforward neural network, MLP's basic structure is composed of input layer, multiple hidden layers and output layer, and it learns the internal laws of data by adjusting weights and biases [3]. It can deal with nonlinear problems, but it may face the problem of dimensional disaster and local optimal solution when dealing with high-dimensional complex data. In contrast, CNN performs well in the field of image processing, and it can effectively capture the characteristics of spatial levels through the combination of convolution layer, pooling layer and fully connected layer [4, 5]. The local perception and weight sharing mechanism of CNN make it have natural advantages in extracting local features, but its grasp of global information may not be as good as MLP.

In order to combine the advantages of the two, researchers propose a CNN model that combines MLP and attention mechanism [6]. This fusion model can not only take advantage of CNN's advantages in local feature extraction, but also enhance the understanding of global information through MLP. The introduction of attention mechanism enables the model to pay more attention to key feature areas when processing data, thereby improving the focusing ability and generalization performance of the model.

In the aspect of fault feature extraction, the fusion model can identify key fault-related signals from a large number of sensor data, even if these signals may be very weak or mixed with other interference signals. By learning and analyzing these key signals, the model can predict potential failure modes and issue early warning before failure occurs, thereby avoiding unexpected downtime

and production loss of equipment. In addition, the fusion model can adapt to changing industrial environments because it can automatically adjust its internal parameters to adapt to new data distributions [7, 8]. This means that the model can maintain high accuracy and reliability even when the equipment is aging or the working conditions change.

The CNN model combining MLP and attention mechanism provides a new and effective way for fault feature extraction and early warning. Through this integration strategy, we are expected to further improve the automation level of industrial systems, reduce human intervention, reduce operation and maintenance costs, and ultimately achieve the goal of intelligent manufacturing. Future research will continue to explore more efficient model fusion technologies and how to better apply these technologies to actual industrial scenarios.

2 Theoretical framework for the Fusion of MLP and attention mechanism CNN

2.1 Study on MLP model and CNN model

2.1.1 MLP model

MLP is a feedforward artificial neural network, which consists of at least three layers: input layer, one or more hidden layers and output layer. If there is more than one hidden layer, it is also called a deep artificial neural network [9, 10]. MLP models are trained using a back-propagation algorithm that minimizes prediction errors by adjusting weights and biases in the network. The MLP model's input layer processes incoming signals, while the output layer handles prediction and classification tasks. The hidden layer, situated between the two, constitutes the core computational unit. Data propagates from input to output, and neurons within the MLP learn through backpropagation. MLP neurons can employ various activation functions, incorporating weighted inputs and initial weights. Forward and backward propagation are outlined in Equation (1) and Equation (2).

$$a^{(l+1)} = \sigma(W^{(l)}a^{(l)} + b^{(l)}) \quad (1)$$

$$\Delta W^{(l)} = -\eta \frac{\partial J}{\partial W^{(l)}} \quad (2)$$

$a^{(l+1)}$ represents the output of the next layer, σ is the activation function, $W^{(l)}$ is the weight, $a^{(l)}$ is the output of the current layer, and $b^{(l)}$ is the deviation. W is the weight, η is the learning rate, and J is the loss function. In fault diagnosis, MLP model can be used to extract fault features from preprocessed data. For example, the characteristics of vibration amplitude, frequency and phase can be extracted, and then these characteristics can be used as inputs to establish MLP model for fault diagnosis. This method can help engineers quickly identify abnormal behavior of equipment, so as to take timely maintenance measures. In terms of fault early warning, the MLP model can be used to establish the operating state model of the equipment, and predict the future state of the equipment through real-time monitoring of the operating data of the equipment. For example, Shao et al. proposed a MLP-based residual life prediction method, which models the bearing operation state through a multi-layer perceptron, which can not only solve the residual life boundary problem, but also automatically adapt to changes in environmental factors [11]. In addition, Pan Lingyong proposed a method based on the combination of multi-feature fusion and MLP, which used the advantages of MLP in processing nonlinear data to classify and identify one-way valve faults, and realized the diagnosis of one-way valve faults of fracturing pumps.

2.1.2 CNN model

Convolutional neural network (CNN) is a deep learning model, which plays an important role in fault feature extraction and early warning. CNN can effectively extract fault-related features from time-series data, such as shock components in vibration signals, harmonic components in current signals, etc [12, 13]. These features are essential for identifying and predicting equipment failures.

$$S(i,j) = (I * K)(i,j) = \sum_m \sum_n I(i-m, j-n)K(m,n) \quad (3)$$

$$P(i,j) = max_{m,n \in R_{ij}} I(m,n) \quad (4)$$

Where $S(i,j)$ represents the result of convolution, I is the input data, K is the convolution kernel, and m and n are the indexes of the convolution kernel. The formula represents the calculation of the convolution result by sliding the convolution kernel over the input data and performing a weighted sum, which is the basic operation for extracting features in CNNs. $P(i,j)$ here denotes the result of pooling, and R_{ij} is the pooling area. In fault diagnosis, CNN can extract features with hierarchical structure from input data through convolution and pooling operations, which are shown in Equation (3) and (4). These features help to reveal the subtle changes in the operating state of the equipment, thereby providing strong support for fault early warning [14]. For example, in the analysis of vibration signals of mechanical equipment, CNN can capture characteristics such as vibration frequency and amplitude, which may indicate some failure modes of the equipment. In the quest for enhancing fault detection and early warning systems, the CNN stands out not only for its prowess in extracting fault features, but also for its capacity to collaborate with other state-of-the-art deep learning models [15, 16]. When paired with long-term and short-term memory networks (LSTM), the CNN-LSTM model emerges as a formidable contender. This hybrid approach harnesses the feature extraction process

of CNNs and the temporal modeling capabilities of LSTMs, adeptly navigating the intricate challenges posed by long-term dependencies in time series data. The resultant model is able to discern and learn the intricate evolution patterns of fault modes, thereby enabling precise and timely fault early warning.

2.2 Attention mechanism

The attention mechanism, a paradigm that emulates the intricate workings of human visual and cognitive systems, grants neural networks the ability to prioritize and focus on pertinent aspects within vast input data. By integrating this mechanism, neural networks are empowered to autonomously learn and selectively attend to salient information within the input, thereby elevating the model's performance and generalization capabilities [17]. In the realm of deep learning, attention mechanisms have found widespread application in the processing of sequential data, encompassing text, speech, and image sequences. Among the most notable varieties are the self-attention mechanism, which captures intrinsic dependencies, the spatial attention mechanism, which attends to spatial features, and the temporal attention mechanism, which focuses on temporal dynamics [18, 19].

$$e_t = v_a \tanh(W_s s_t + W_h h_t + b_a) \quad (5)$$

$$\alpha_t = \frac{exp(e_t)}{\sum_k exp(e_k)} \quad (6)$$

v_a, W_s, and W_h are the learnable parameters, s_t and h_t are information from different sources, and b_a is bias. The attention score is calculated using nonlinear activation (tanh). a_t is the normalized attention weight, e_t is the attention score calculated earlier, and K is the element. Attention weight calculation and attention weight normalization are shown in Equation (5) and Equation (6). A prevalent application in this realm involves the seamless integration of the attention mechanism within a bidirectional long-short-term memory neural network (BiLSTM), resulting in the BiLSTM-Attention model. This intricate architecture comprises an input layer, a BiLSTM layer, an attention layer, and an output layer. Specifically, the input layer is tasked with receiving preprocessed fault data, while the BiLSTM layer orchestrates the bidirectional processing of sequence information [20]. The attention layer, in turn, calculates the weight of each temporal step, performing a weighted summation of the BiLSTM layer's output to extract pivotal features from the fault data. Ultimately, the output layer yields the classification probability of the fault data.

Another noteworthy approach combines the prowess of CNN and long-short-term memory networks (LSTM) to forge the CNN-LSTM model. Here, the CNN component expertly extracts fault features from temporal data, while the LSTM segment models and classifies these time series. Leveraging CNN's feature extraction capabilities and LSTM's temporal modeling strengths, this hybrid model adeptly extracts fault features from time series data and categorizes faults accordingly.

The self-attention mechanism offers a direct mapping of the source text's word vector sequence X, generating the essential Q, K, and V components required by the attention mechanism [21, 22]. This approach calculates the attention weights among words within the source text, capturing word dependencies while simultaneously accomplishing encoding. The formulaic definition of the process is shown in Equation (7)-(9):

$$(Q,K,V) = Linear(X) = \begin{cases} Q = W^Q X \\ K = W^K X \\ V = W^V X \end{cases} \quad (7)$$

$$SelfAttention(Q,K,V) = softmax(\frac{Q^T K}{\sqrt{d_k}})V \quad (8)$$

$$e_t = v_a \tanh(W_s s_t + W_h h_t + b_a) \quad (9)$$

T stands for transpose operation. The sparsemax function is used instead of SoftMax, the formula definition of which is shown in formula (10), and the formal definition of the family of. α-entmax functions is shown in formula (11) and formula (12):

$$sparsemax(e) = \arg\min_{p \in \Delta^d} p - e^2 \quad (10)$$

$$\alpha - \text{entmax}(e) = \arg\max_{p \in \Delta^d} p^T e + H_\alpha^T(p) \quad (11)$$

$$H_\alpha^T(p) = \frac{1}{\alpha(\alpha - 1)} \sum_{i=1}^{d} (p_i - p_i^\alpha) \quad (12)$$

d denotes the dimension, and pi is the *i-th* element in the probability p distribution. For a given attention score matrix A, its orthogonal regularization formula is shown in Equation (13):

$$R_O = AA^T - I_F \quad (13)$$

2.3 Research method of CNN fusion of MLP and attention mechanism

In the burgeoning domain of deep learning, the MLP and CNN occupy pivotal positions as common network architectures. Of late, researchers have embarked on an intriguing exploration, delving into the integration of attention mechanisms within these networks to elevate their performance and expressive prowess [23]. Here are

some research methods on the fusion of MLP and attention mechanism CNN:

External fusion methods. The external fusion method refers to first applying the convolution and attention mechanisms separately, and then combining their outputs in different ways. For example, Squeeze-and-Excitation Networks (SE-Net) is an external fusion method that computes channel attention weights through global average pooling and multilayer perceptrons, and then applies these weights to convolution outputs to dynamically adjust the degree of activation of individual channels.

Intrinsic fusion approach. The intrinsic fusion method refers to the fusion of convolution and attention mechanisms into a single operation. For example, the Convolutional Block Attention Module (CBAM) is an intrinsic fusion method that sequentially deduces attention graphs along two independent dimensions of channel and space, and then multiplies the attention graphs to the input feature graph for adaptive feature refinement [24, 25]. CBAM can be seamlessly integrated into any CNN architecture and can be trained end-to-end with the underlying CNN. The feature fusion process is shown in Equation (14).

$$F = \alpha F_{MLP} + (1-\alpha) F_{CNN} \quad (14)$$

Table 1 shows the comparison between MLP and CNN fusion methods. The external-internal fusion method refers to the fusion of convolution and attention mechanisms into a single operator (internal), and then applies conventional convolution or attention operations (external) on this basis. This approach attempts to combine the advantages of both in order to achieve better performance. Other fusion methods, such as Non-Local Neural Networks, model the global morning and afternoon through the Self-Attention mechanism, effectively capturing long-distance feature dependencies. In addition, some studies have proposed multi-spectral channel attention Fca-Net from the perspective of frequency domain, which makes full use of information by introducing more frequency components.

Table 1: Comparison of MLP and CNN fusion methods

Contrast dimension	MLP	CNN	MLP and CNN Fusion
Feature extraction ability	Suitable for linear and nonlinear feature extraction	Especially good at extracting local features, such as edges and textures in images and time series data	Combining the characteristics of the two, more abundant features can be extracted
Model complexity	Relatively simple and easy to train	The model structure is complex and requires more computing resources	Through fusion, model complexity and performance can be balanced
Training speed	Generally faster	Training may be slow due to complex structure	Fusion may achieve faster training speed
Generalization ability	Depends on data distribution and may be at risk of overfitting	Through local receptive field and weight sharing mechanism, it has good generalization ability	After fusion, the generalization ability of the model may be improved
Application Scenario	Classification and Regression Problems for Static Data	Suitable for image recognition, video analysis and other spatio-temporal data processing	Applicable to a wider range of data types and application scenarios
Noise resistance	Generally, not as good as CNN	Due to local receptive field and weight sharing, it has strong anti-noise ability	The anti-noise ability may be improved after fusion
Pros and cons:	training speed fast, the ability to extract local features is weak	The bureau is particularly strong and generalized well. The speed of training is slow, and the demand for resources is large	It is widely applicable, difficult to set, requires resources, and has poor interpretation

In the field of fault feature extraction and early warning, it is important to continuously explore better models to improve the accuracy and efficiency of fault diagnosis. Table 2 provides a detailed multi-dimensional comparison of the MLP-attention-based CNN fusion model and other existing state-of-the-art technologies, such as traditional MLP, traditional CNN, and ordinary CNN-MLP fusion models. The feature extraction ability determines whether the model can accurately identify fault-related features, and the fusion model can comprehensively and deeply mine key information with its unique structural design, while the traditional model has different limitations in this regard. Generalization ability and noise immunity are the key indicators to measure the performance of the model in complex and changeable real-world scenarios, and the fusion model shows strong adaptability and anti-interference ability, compared with the traditional model in the face of new data and noise environment. In terms of fault warning accuracy and training efficiency, the fusion model also has significant advantages, which can accurately warn faults in advance and complete training in a short time, providing a reliable and efficient solution for practical applications. Through the comparison of these dimensions, it can be clearly seen that the model based on the fusion of MLP and attention mechanism CNN has better performance in fault feature extraction and early warning tasks.

Table 2: Performance comparison table of CNN fusion models based on MLP and attention mechanism with other advanced technologies

Compare dimensions	A model based on the fusion of MLP and attention mechanism CNNs	Traditional MLP, CNN, etc
Feature extraction	Combining the advantages of MLP and CNN, the attention mechanism is used to accurately extract key features, which is comprehensive and in-depth	Traditional MLP local feature extraction is weak, CNN key information is not focused enough, and ordinary fusion model has limited extraction ability
Generalization and noise cancellation	Strong generalization ability, stable performance under different working conditions and data; Outstanding anti-noise capability to identify real signals in noise	Traditional MLP is easy to overfit, has poor generalization, and has average noise resistance; CNNs have limitations in complex scenes and strong noise, and the improvement of ordinary fusion models is not obvious
Early warning and efficiency	The accuracy of fault warning is high, which greatly reduces false positives and false negatives; High training efficiency, combined with MLP speed advantages and optimized calculations	The accuracy of traditional MLP early warning is limited, and the training speed is fast but the function is weak. CNN early warning is not comprehensive enough, training is slow, and the early warning and efficiency improvement of ordinary fusion models are limited

The fault feature extraction and early warning method based on MLP and CNN fusion combines the advantages of both, and can improve the generalization ability and anti-noise ability of the model while maintaining high accuracy. This fusion method is suitable for a variety of data types and application scenarios, especially when dealing with complex time series data and image data, it can effectively extract features and perform fault warning. Therefore, in practical applications, according to the specific fault diagnosis requirements and data characteristics, choosing a suitable fusion strategy and model structure can significantly improve the effect of fault diagnosis.

This paper introduces the relevant theories and algorithms of fault feature extraction and early warning methods, and presents a development trend from single model to model fusion. Firstly, the application of MLP and CNN models in fault diagnosis and early warning is described, MLP can extract fault features and establish a running state model, and CNN can extract fault-related features from time series data, both of which have their own advantages in fault diagnosis and early warning. Then, the attention mechanism is introduced, which is widely used in sequence data processing and can improve the performance of neural networks. Finally, the fusion methods of MLP, CNN and attention mechanism are emphatically discussed, including external fusion and internal fusion, and the fusion method combines a variety of advantages, which is suitable for more data types and application scenarios, and can improve the fault diagnosis effect.

3 Construction of early warning system based on MLP and attention mechanism CNN

MLP relies on its nonlinear mapping capabilities to carry out preliminary processing of data, realize feature extraction and pattern recognition, and provide basic features for subsequent processes. In this process, as

mentioned in the article, the weights and biases will be carefully adjusted during the training period, such as dynamically changing the weight values according to the actual data features and training objectives to optimize the effect of feature extraction, and at the same time, the bias will be accurately adjusted to ensure the stability of the model when processing different data. Subsequently, attention mechanisms were introduced into CNNs. CNN extracts feature by sliding on the input data and weighting summing through the convolution kernel, while the attention mechanism processes the channel and spatial dimensions of the feature map, highlighting key information and suppressing secondary information. When calculating attention weights, a series of complex calculations are involved, and the relevant parameters are constantly adjusted according to the feedback during the training process, so that the attention mechanism can focus on important information more accurately. Finally, the data that has been preliminarily processed by MLP is fused with CNNs that are integrated into the attention mechanism. There are many fusion methods, such as external fusion methods, which first perform convolution and attention mechanism operations separately, and then combine their outputs in a specific way; The intrinsic fusion approach combines convolution and attention mechanisms into a single operation. In the whole integration process, the data processing procedures of each link have strict specifications, and the parameter settings will be continuously optimized according to the actual situation, which not only ensures the efficient operation of the model, but also enhances the replicability of the whole model construction process, so that other researchers can reproduce the whole model construction process more accurately.

Building an early warning system based on multi-layer perceptron and attention mechanism convolutional neural network is an innovative project combining advanced technology of deep learning. Firstly, the system uses MLP to process the original data, and utilizes its nonlinear mapping ability to perform preliminary feature extraction and pattern recognition on the input data. Subsequently, the attention mechanism is introduced into CNN, which enables the network to focus on the most critical information when processing image or sequence data, similar to the process that human eyes automatically pay attention to important details when observing complex scenes. In this way, the early warning system can more accurately capture potential risk signals, so as to respond quickly before problems occur. The construction of the whole system not only depends on efficient algorithm design, but also requires a large amount of data support and fine parameter adjustment to ensure its accuracy and reliability in practical applications.

3.1 MLP-Mixer

MLP-Mixer is a novel network architecture that uses MLP to replace the convolution operation in traditional CNN and the self-attention mechanism in Transformer [26]. MLP-Mixer enables fusion between features through two fusion structures: spatial fusion and channel fusion. Spatial fusion allows features at different spatial locations to communicate, while channel fusion allows features between different channels to communicate. The core of MLP-Mixer is Mixer Layer, which maps columns and rows through MLP to realize the information fusion of spatial domain and channel domain.

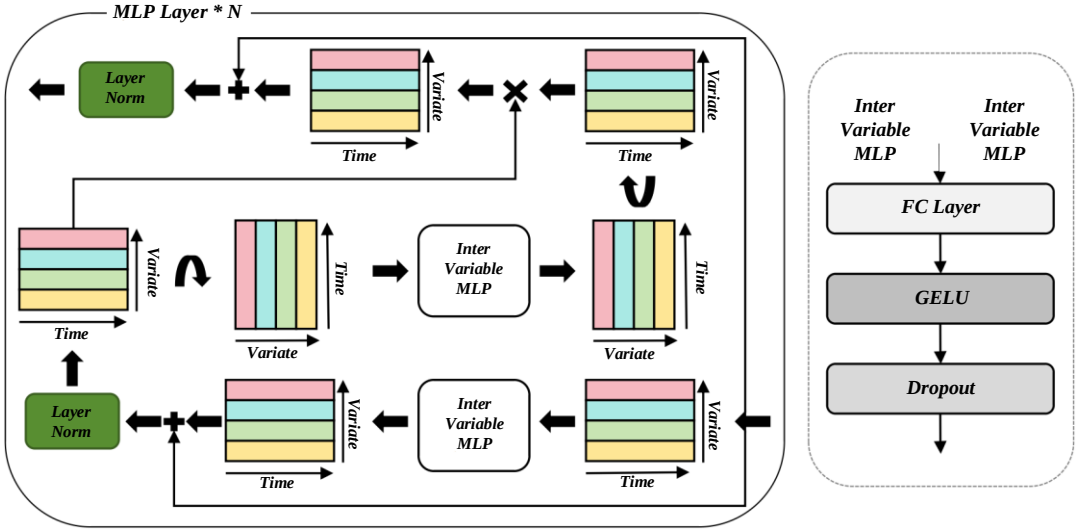

Figure 1: Converged MLP network architecture

Figure. 1 shows the fused MLP network architecture. When building an early warning system that fuses MLP and attention mechanism CNN, it is necessary to perform feature extraction, spatial fusion, channel fusion, processing of fused features, and predictive output. MLP-Mixer can be applied to image classification tasks, and its core idea is to use fully connected layers to replace the traditional CNN. Finally, the classification is carried out through the fully connected layer.

The new structure first uses MLP to preliminarily process the raw data to exert its nonlinear mapping capabilities, and then introduces the attention mechanism to the CNN, so that it can focus on key information and pay attention to important details like the human eye.

Compared with the previous one, the new structure data processing process is more refined, feature extraction and information attention are more accurate, the system performance and effect are better, the early warning can be more accurate and faster, and the data volume and parameter adjustment requirements are higher when building.

3.2 SE-Net and CBAM

SE-Net is an external fusion method that dynamically adjusts the importance of channels by obtaining feature maps through convolutional layers and then using global average pooling and multilayer perceptron (MLP) to capture the frequency and intensity distribution of each channel. SE-Net's Squeeze operation aggregates global information through global average pooling, while the Excitation operation dynamically adjusts the weight of channels according to the information obtained from the Squeeze operation, so as to strengthen the model's focus on key features [27, 28]. When building an early warning system that integrates SE-Net, MLP, and CNN, data preprocessing, CNN architecture design, SE-Net module integration, MLP integration, fusion strategy, training and verification, testing, and deployment are required.

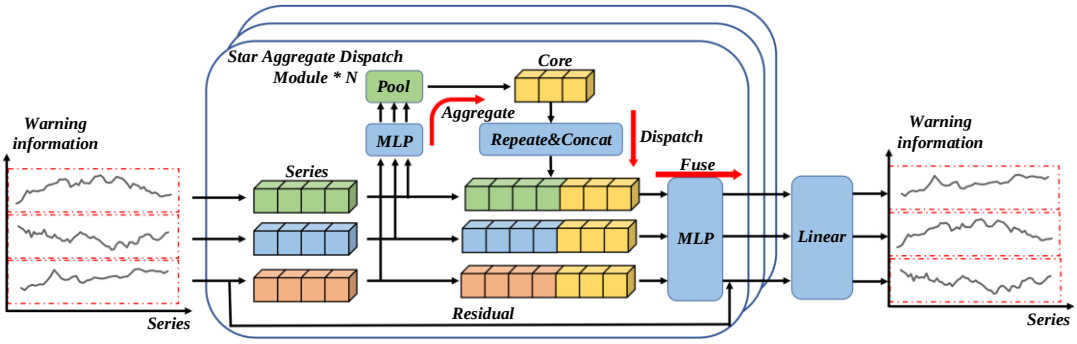

Figure 2: Early warning system fused by MLP and attention mechanism CNN

Figure. 2 shows the early warning system fused by MLP and attention mechanism CNN. By considering the channel and spatial dimensions of the feature graph, CBAM can highlight important features and suppress unnecessary features, thereby improving the performance of the model [29, 30]. In the early warning system where MLP and attention mechanism CNN are fused, CBAM can be used to strengthen the expression of features and help the system better understand and predict potential risks or anomalies. H stands for feature. The formal definition of the global attention layer is shown in Equation (15)-(19):

$$H' = mask_C(concat(H, P)) \quad (15)$$

$$\tilde{H} = \sigma(H'W_1 + b_1)W_2 + b_2 \quad (16)$$

$$att_g = softmax(\tilde{H}) \quad (17)$$

$$H_c = H'W_c + b_c \quad (18)$$

$$H_g = att_g^T H_c \quad (19)$$

3.3 System specifications for simulation

The construction of an early warning system based on multi-layer perceptron and attention mechanism convolutional neural network is an innovative project integrating advanced deep learning technology. The system first uses the nonlinear mapping ability of MLP to perform preliminary feature extraction and pattern recognition on the raw data, and then introduces the attention mechanism into the CNN, so that it can focus on key information when processing image or sequence data, so as to more accurately capture potential risk signals and respond quickly. The architecture of the system also adopts the novel architecture of MLP-Mixer, which replaces the convolution operation of traditional CNN and the self-attention mechanism of Transformer with MLP, realizes feature fusion through two structures: spatial fusion and channel fusion, and its core Mixer layer achieves information fusion between spatial domain and channel domain through MLP mapping columns and rows. In the data processing process, feature extraction is carried out first, and then the spatial and channel fusion is completed in the MLP-Mixer architecture, and then the fusion features are further processed. The system can be applied to image classification tasks, culminating in classification by fully connected layers. The construction of the whole system relies on efficient algorithm design, a large amount of data support and fine parameter adjustment to ensure the accuracy and reliability in practical applications.

4 Experimental design and analysis

4.1 Construction of program index system

In constructing the index system of fault feature extraction and early warning scheme fused by MLP and attention matrix CNN. Initially, a rigorous signal preprocessing procedure is enacted, encompassing denoising techniques to mitigate unwanted interference components, filtering

methods to refine signal clarity, and normalization processes to enhance overall signal quality. Following this, a feature extraction step is executed, leveraging the fast Fourier transform to transform the time-domain signal into its frequency-domain representation, thereby elucidating the spectral characteristics of the signal. At the same time, the time domain signal is extracted by convolutional neural network operation. In this way, the model can obtain the feature information in time domain and frequency domain at the same time. Secondly, on the basis of feature extraction, the time-domain and frequency-domain features are fused by using cross-attention mechanism. By calculating the attention weight, the model can pay attention to more important feature information and enhance the ability of fault feature recognition. The realization of cross-attention mechanism can be accomplished by multi-layer perceptron or self-attention mechanism.

The model uses cross-entropy loss to supervise the classification task, defined as Equation (20). The final training loss L is defined as in Equation (21):

$$L_{sc} = -\sum_{(s,a)\in D} y_{(s,a)} \log(\hat{y}_{(s,a)}) \quad (20)$$

$$L = L_{sc} + \alpha L_{con} + \varepsilon \square \theta \square^2 \quad (21)$$

In the experiment, a large number of parameters related to multilayer perceptron (MLP) and multilayer perceptron (CNN) and attention mechanism were defined. In MLP, $a^{(l+1)}$ represents the output of the next layer, σ is the activation function, $W^{(l)}$ is the weight of the l layer, $a^{(l)}$ is the output of the current layer, *b(l)* is the bias of the first layer, *W* is the weight, n is the learning rate, and J is the loss function. For CNN, *S(i,j)* is the convolution result, I is the input data, *K* is the convolution kernel *m*n* and *n* is the index of the convolution kernel, *P(i,j)* is the pooling result, and *Rij* is the pooling region. In the attention mechanism, v_a, W_s, and W_h are the learnable parameters, s_t and h_t are the information from different sources, b_a is the bias, e_t is the attention score, *t* is the normalized attention weight, *K* is the element, *T* is the transpose, *d* is the dimension, and *Pi* is the *i-th* element in the probability p distribution. In addition, there is the parameter *L*, which represents the loss function, and MSE, which is the mean square error. After that, the fused features are input into the classifier for fault classification. Classifiers can employ algorithms such as support vector machines, random forests, or deep learning models. By training and optimizing the classifier, the bearing fault can be identified accurately. Subsequently, the model's performance is rigorously assessed through the utilization of various evaluation metrics, including the confusion matrix, accuracy rate, precision rate, recall rate, and F1 score. These indices offer a comprehensive understanding of the model's efficacy across diverse categories, as well as its overall diagnostic accuracy. Figure. 3 depicts the outcomes of this performance evaluation. Guided by the results of this evaluation, the model undergoes adjustments and optimizations, such as the refinement of the network structure and the tuning of parameter settings. These modifications aim to enhance the diagnostic capabilities of the model. Ultimately, once the stability and reliability of the model's performance are assured, it is integrated into a practical fault prediction and early warning system, enabling real-time monitoring of equipment status and the timely detection and mitigation of potential faults.

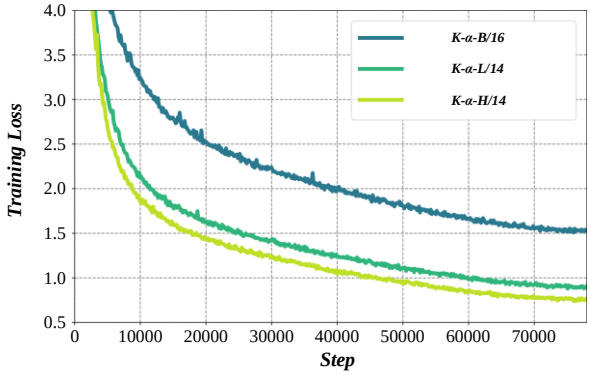

Figure. 3 Results of the performance evaluation

4.2 Scheme design and implementation

When building a machine learning model, the data is usually divided into a training set, a verification set, and a test set. These three data sets are used to train and optimize the model, and finally obtain the best effect and the best generalization ability we want. model. After the division of the training set, verification set and test set is completed, this paper first uses the training set to train the initially constructed model, and checks the prediction results of the fitted model, because the training set is used to fit the training set. At this time, the results obtained by the model should theoretically be high, that is, the effect of the model should be better; After the training of the training set is over, it is necessary to obtain different prediction accuracy rates by continuously adjusting the values of the parameters. The model with the highest accuracy rate obtained in the end is identified as the model with the best effect; After determining the optimal model through the verification set, it is necessary to use the test set to check whether the model is optimal or not. After determining the selected model, the training set is used to train the model.

5 Rationality test and implementation approach of cnn fusion scheme of MLP and attention mechanism

5.1 Model evaluation

When evaluating a fault feature extraction and early warning model of a convolutional neural network that combines a multi-layer perceptron and an attention mechanism, we mainly focus on its accuracy, robustness, real-time performance, and generalization capabilities.

This fusion model aims to improve the performance of fault detection through the nonlinear mapping ability learned by MLP and the spatial feature extraction ability of CNN, as well as the focus of attention mechanism on key information.

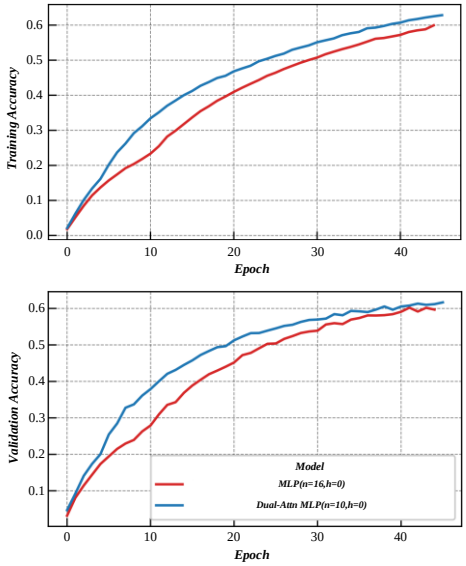

Figure 4: Accuracy results of the model

The accuracy of the model is measured by comparing the consistency of the predicted results with the actual fault state, as shown in Figure. 4. The high accuracy rate means that the model can effectively identify the real fault mode, while the low false positive rate and false negative rate indicate that the model has good discrimination in distinguishing normal and abnormal states. To verify this, cross-validation is often performed on a dataset containing multiple fault types to evaluate the model's performance in different scenarios. This figure intuitively compares the changes in the training and validation accuracy of different models, and evaluates the effectiveness of fault feature extraction and early warning methods based on the fusion of MLP and attention mechanism CNN. The comparison of the training and validation accuracy of the two models (MLP and Dual-Arm MLP) under multiple epochs shows that the training accuracy of the fusion model increases rapidly and is high, and the verification accuracy is also better, which proves that its performance and effect are better than those of the traditional MLP model, providing evidence for the advantages of the fusion method.

Figure. 5 is the robustness analysis of feature visualization. Robustness refers to the stability of the model in the face of noise interference or input data changes. A robust model can maintain high detection performance even in harsh environments, which can be tested by adding noise or data collected under different conditions. In addition, the real-time performance of the model is also an important indicator, especially in industrial applications, timely fault warning can avoid major losses. Therefore, evaluating the processing speed and response time of the model is very important to determine its practical application value.

The generalization ability reflects the adaptability of the model to unseen data or new types of failures. An excellent model must not only perform well on training data, but also be able to make accurate predictions on new and unknown fault samples. By evaluating the model on an independent test set, we can understand its generalization ability, and accordingly adjust the model parameters or improve the algorithm to improve its reliability in future applications.

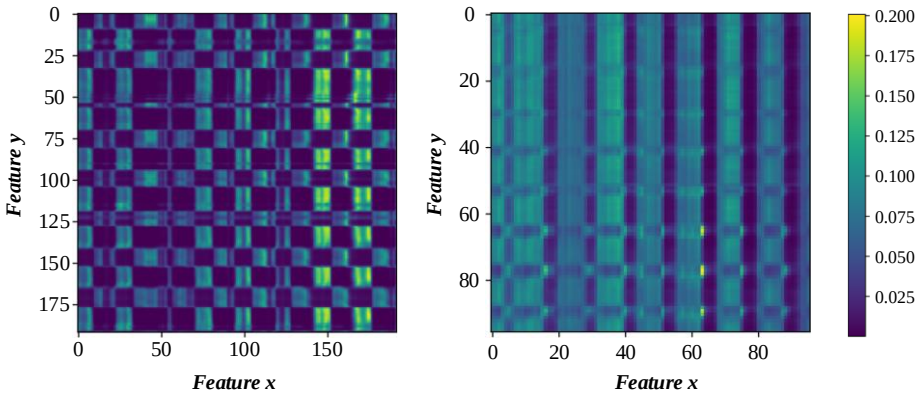

Figure 5: Robustness analysis of feature visualization

5.2 Comparison of effect between this scheme and other schemes

In modern industrial systems, fault feature extraction and early warning technology is the key to ensure the stable operation of equipment. Traditional feature extraction methods such as multilayer perceptron and convolutional neural network have their own advantages, but also have limitations. MLP is known for its powerful non-linear fitting capabilities, able to learn complex feature maps, but may encounter dimensional disasters when processing high-dimensional spatial data. In contrast, CNN can effectively capture local features of image data through local connection and weight sharing mechanisms, but it may not be sensitive enough to unstructured signal data.

A reasonable diagram can be made as follows: the purple curve is a model with only MLP (no attention mechanism); The red curve is the CNN fusion model of MLP and simple attention mechanism. The orange curve is a CNN fusion model of MLP and medium-complexity attention mechanism. The green curve is the CNN fusion model of MLP and complex attention mechanism. The blue curve is the optimized MLP and attention mechanism CNN fusion model to better understand the trend in the graph and the difference in performance between different models.

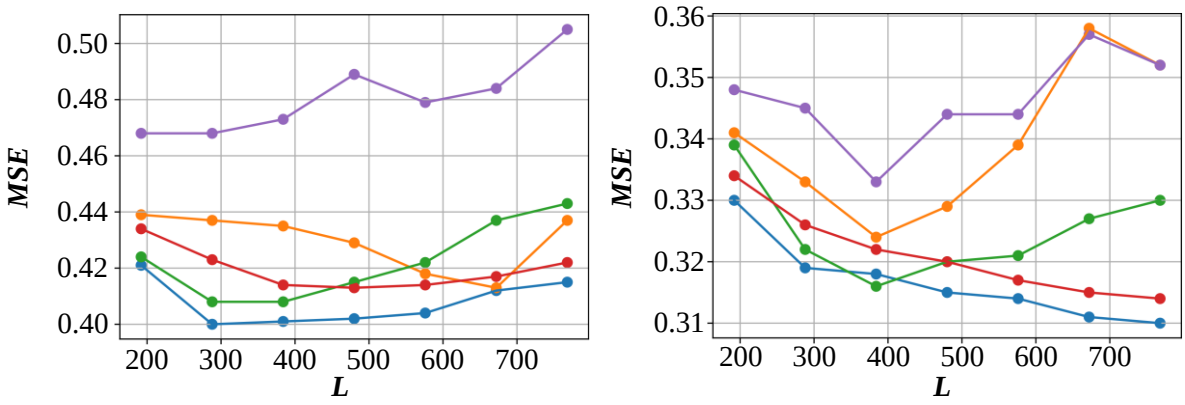

Figure: 6 MSE losses corresponding to different layers

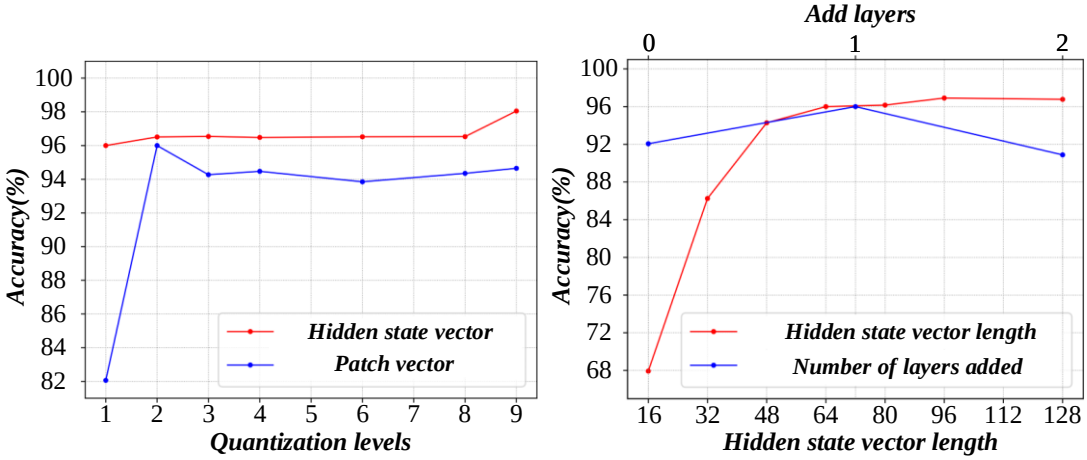

Figure 7: Corresponding accuracy rates under different patches

Figure. 6 and Figure. 7 show the MSE loss corresponding to different layers and the corresponding accuracy rate under different patches, respectively. In order to overcome the limitations of traditional methods, this paper proposes a CNN fusion scheme that combines MLP and attention mechanism. In this way, the fusion model can use the nonlinear mapping ability of MLP to identify and locate failure modes more accurately while retaining the spatial feature extraction ability of CNN. Among them, the early warning feature capture visualization is shown in Figure. 8.

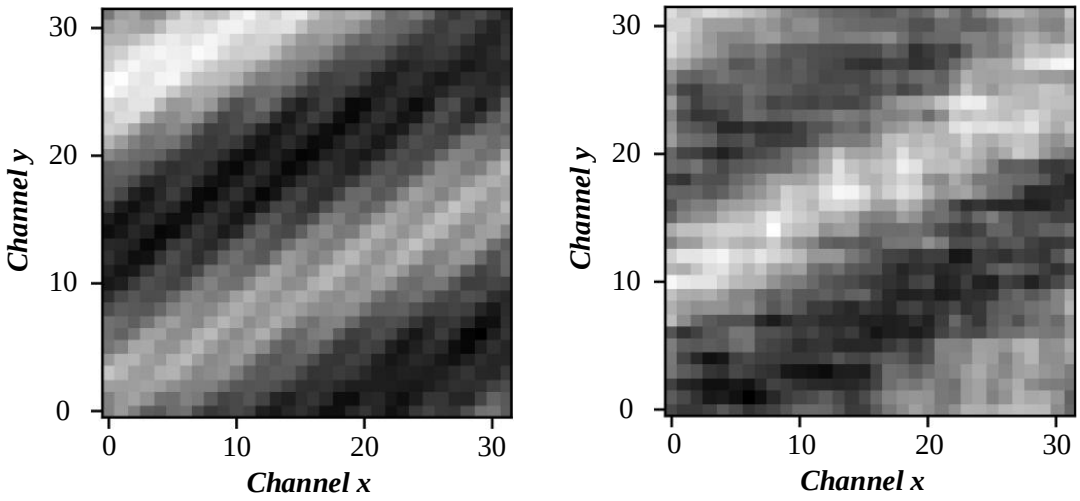

Figure 8: Visualization of early warning feature capture

Figure. 9 shows the corresponding accuracy and histogram analysis under different Table sizes. Compared with other schemes, the fault feature extraction and early warning method of MLP and attention matrix CNN fusion has significant advantages. Firstly, it can adaptively adjust the focus of feature extraction and improve the ability to identify complex fault modes. Secondly, by combining different types of neural networks, the scheme realizes complementary advantages and enhances the generalization ability and robustness of the model. Finally, due to the introduction of attention mechanism, the method is more efficient in processing large-scale data, which is helpful for real-time monitoring and rapid response to potential failure conditions, thereby improving the performance of the entire early warning system.

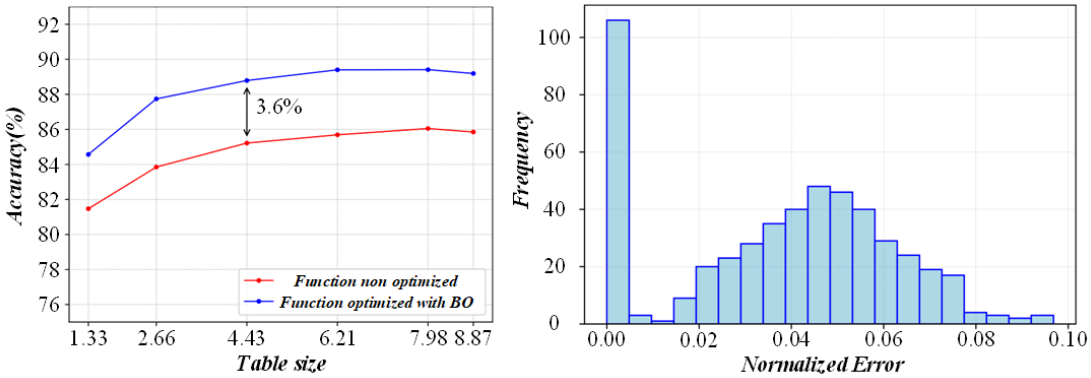

Figure 9: Corresponding accuracy rate and histogram analysis under different Table sizes

5.3 Programmed implementation approach

In the field of chemical process fault diagnosis, data-driven methods have been paid attention to because of their advantages of self-mining and building intrinsic relationships of data. Some studies have pointed out that the idea of combining multiple data-driven methods to solve chemical process problems has certain effectiveness. Combining LSTM and MLP to extract temporal features, and then classifying on SoftMax, this method can be applied to chemical processes with time-varying, non-linear, and high-dimensional characteristics.

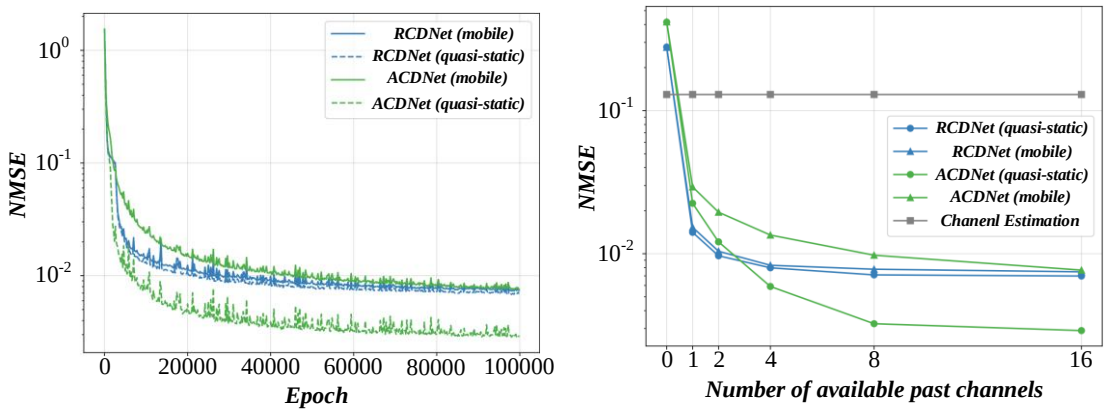

Figure 10: Comparative analysis of the method in this paper and other methods

In the airbag assembly process, this study combines BP neural network and SOM (self-organizing map) neural network to detect faults. The BP neural network is used to identify the state of each sensor, while the SOM neural network is used to determine the specific type of fault. In practical applications, the fault feature extraction and early warning method based on MLP and CNN fusion has shown good results. Figure. 10 shows the comparative analysis of the method in this paper and other methods. Among the 300 fault sample data, the SOM network can correctly distinguish the fault types of 295 groups, and the early warning correct rate exceeds 98%, which proves the effectiveness of the method.

6 Conclusion

In the field of modern industrial automation and intelligent manufacturing, the effectiveness of fault feature extraction and early warning systems is crucial to ensuring production efficiency and equipment safety. The fusion technology of multilayer perceptron's combined with attention mechanisms combined with convolutional neural networks provides an innovative solution to this challenge.

In order to ensure the reliability and validity of the research results, the researchers added statistical verification links on the basis of the original research. By carefully selecting the appropriate confidence level and conducting detailed statistical analysis of a large number of experimental data, the confidence interval of the model in key performance indicators such as fault warning accuracy and feature extraction accuracy is calculated. Then, according to the experimental data, the corresponding statistics are calculated and compared with the critical value, if the statistics exceed the critical value, the null hypothesis is rejected and the alternative hypothesis is accepted, so as to judge the significant advantages of the fusion model in performance from a statistical point of view, and fully confirm the importance of the research results.

As a feedforward neural network, MLP can learn complex nonlinear relations, while CNN is good at extracting spatial features from images or signals. The combination of these two networks can make full use of their respective advantages and realize the in-depth mining of fault features. Especially after adding attention mechanism, the model can focus more on key features, thereby improving the accuracy and robustness of fault detection. This fusion technology has shown its potential in many fields. In power systems, it can be used to identify early fault signs of transformers; In mechanical manufacturing, it helps to monitor the health of machine tools and predict potential failures. Through real-time monitoring and analysis, the system can issue an early warning before a failure occurs, thereby reducing downtime and maintenance costs. The method proposed in this paper can significantly improve the accuracy and robustness of fault feature extraction, and the correct rate of early warning exceeds 98%. This study can more effectively early warning and improve the reliability of the system.

In future research, we will attempt to extend the testing of the model to various industrial datasets to evaluate generalization ability and robustness. At the same time, exploring the impact of different attention mechanisms and other fusion strategies on model performance, integrating real-time data processing capabilities, evaluating the computational efficiency of different hardware settings, and continuously conducting in-depth research.

Funding

Project Type: Science and Technology Research Project of Henan Province： Research on the Intelligent Fruit Tree Spraying Robot System for Precision Agriculture (Project Number:242102110335).

References

[1] Yong Wang, Jianfei Pu, Duoqian Miao, L. Zhang, Lulu Zhang, and Xin Du, "SCGRFuse: An infrared and visible image fusion network based on spatial/channel attention mechanism and gradient aggregation residual dense blocks," Engineering Applications of Artificial Intelligence, vol. 132, pp. 107898, 2024. https://doi.org/10.1016/j.engappai.2024.107898

[2] Zhiyu Zhou, Yanjun Hu, Xingfan Yang, and Junyi Yang, "YOLO-based marine organism detection using two-terminal attention mechanism and

difficult-sample resampling," Applied Soft Computing, vol. 153, pp. 111291, 2024. https://doi.org/10.1016/j.asoc.2024.111291

[3] Li Jiang and Yifan Wang, "A wind power forecasting model based on data decomposition and cross-attention mechanism with cosine similarity," Electric Power Systems Research, vol. 229, pp. 110156, 2024. https://doi.org/10.1016/j.epsr.2024.110156

[4] Weirong Sun, Yujun Ma, and Ruili Wang, "k-NN attention-based video vision transformer for action recognition," Neurocomputing, vol. 574, pp. 127256, 2024. https://doi.org/10.1016/j.neucom.2024.127256

[5] Mingyang Ma, Lei Yang, Yanhong Liu, and Hongnian Yu, "An attention-based progressive fusion network for pixelwise pavement crack detection," Measurement, vol. 226, pp. 114159, 2024. https://doi.org/10.1016/j.measurement.2024.114159

[6] Kai Zhang, Dongxin Bai, Yong Li, Ke Song, Bailin Zheng, and Fuqian Yang, "Robust state-of-charge estimator for lithium-ion batteries enabled by a physics-driven dual-stage attention mechanism," Applied Energy, vol. 359, pp. 122666, 2024. https://doi.org/10.1016/j.apenergy.2024.122666

[7] Cong Hu et al., "A hybrid digital self-interference cancellation method with attention-based TCN-GRU for full-duplex systems," AEU-International Journal of Electronics and Communications, vol. 176, pp. 155144, 2024. https://doi.org/10.1016/j.aeue.2024.155144

[8] Borui Wu and Wenrui Zhao, "Fault prediction of electronic devices based on attention mechanism time-series point process," Measurement: Sensors, vol. 31, pp. 101023, 2024. https://doi.org/10.1016/j.measen.2023.101023

[9] Huali Yang et al., "MAHKT: Knowledge tracing with multi-association heterogeneous graph embedding based on knowledge transfer," Knowledge-Based Systems, vol. 310, pp. 112958, 2025. https://doi.org/10.1016/j.knosys.2025.112958

[10] Tao Ye, Haoran Chen, Hongbin Ren, Zhikang Zheng, and Zongyang Zhao, "LPT-Net: A Line-Pad Transformer Network for efficiency coal gangue segmentation with linear multi-head self-attention mechanism," Measurement, vol. 226, pp. 114043, 2024. https://doi.org/10.1016/j.measurement.2023.114043

[11] Lin Zhou et al., "multi-omics fusion based on attention mechanism for survival and drug response prediction in Digestive System Tumors," Neurocomputing, vol. 572, pp. 127168, 2024. https://doi.org/10.1016/j.neucom.2023.127168

[12] Juan Dong et al., "Estimating reference crop evapotranspiration using improved convolutional bidirectional long short-term memory network by multi-head attention mechanism in the four climatic zones of China," Agricultural Water Management, vol. 292, pp. 108665, 2024. https://doi.org/10.1016/j.agwat.2023.108665

[13] Zhiwu Shang and Zehua Feng, "Multiscale capsule networks with attention mechanisms based on domain-invariant properties for cross-domain lifetime prediction," Digital Signal Processing, vol. 146, pp. 104368, 2024. https://doi.org/10.1016/j.dsp.2023.104368

[14] Jing Li and XiaoMeng Wei, "Research on efficient detection network method for remote sensing images based on self attention mechanism," Image and Vision Computing, vol. 142, pp. 104884, 2024. https://doi.org/10.1016/j.imavis.2023.104884

[15] Sheng Shi, Dongsheng Du, Oya Mercan, Erol Kalkan, and Shuguang Wang, "A novel data-driven sensor placement optimization method for unsupervised damage detection using noise-assisted neural networks with attention mechanism," Mechanical Systems and Signal Processing, vol. 209, pp. 111075, 2024. https://doi.org/10.1016/j.ymssp.2023.111075

[16] Gang Liu, Aihua Ke, Xinyun Wu, and Haifeng Zhang, "GAN with opposition-based blocks and channel self-attention mechanism for image synthesis," Expert Systems with Applications, vol. 246, pp. 123242, 2024. https://doi.org/10.1016/j.eswa.2024.123242

[17] Mingdong Han and Lingyan Fan, "A short-term energy consumption forecasting method for attention mechanisms based on spatio-temporal deep learning," Computers and Electrical Engineering, vol. 114, pp. 109063, 2024. https://doi.org/10.1016/j.compeleceng.2023.109063

[18] Haiyang Jiang, Yuanyao Lu, Duona Zhang, Yuntao Shi, and Jingxuan Wang, "Deep learning-based fusion networks with high-order attention mechanism for 3D object detection in autonomous driving scenarios," Applied Soft Computing, vol. 152, pp. 111253, 2024. https://doi.org/10.1016/j.asoc.2024.111253

[19] Xinquan Liu et al., "An attention-based deep learning method for the detection of electrical status epilepticus during sleep from electroencephalogram waveform analysis in children," Biomedical Signal Processing and Control, vol. 91, pp. 105926, 2024. https://doi.org/10.1016/j.bspc.2023.105926

[20] Ling Chang, Kaijie Wu, Chaocheng Gu, and Cailian Chen, "A novel end-to-end chromosome classification approach using deep neural network with triple attention mechanism," Biomedical Signal Processing and Control, vol. 91, pp. 105930, 2024. https://doi.org/10.1016/j.bspc.2023.105930

[21] Zhe Yin et al., "Lightweight pig face feature learning evaluation and application based on attention mechanism and two-stage transfer learning," Agriculture, vol. 14, no. 1, pp. 156, 2024. https://doi.org/10.3390/agriculture14010156

[22] Hanwen Zhang, Hongyan Liu, and Chulsoo Kim, "Semantic and instance segmentation in coastal urban spatial perception: A multi-task learning framework with an attention mechanism," Sustainability, vol. 16, no. 2, pp. 833, 2024. https://doi.org/10.3390/su16020833

[23] Mohammad Irani Azad, Roozbeh Rajabi, and Abouzar Estebsari, "Nonintrusive Load Monitoring (NILM) Using a Deep Learning Model with a Transformer-Based Attention Mechanism and Temporal Pooling," Electronics, vol. 13, no. 2, pp.

407, 2024. https://doi.org/10.3390/electronics13020407
[24] Yi Deng, Lei Wang, Yitong Li, Hai Liu, and Yifei Wang, "EhdNet: Efficient Harmonic Detection Network for All-Phase Processing with Channel Attention Mechanism," Energies, vol. 17, no. 2, pp. 349, 2024. https://doi.org/10.3390/en17020349
[25] Chenhong Yan et al., "A Lightweight Network Based on Multi-Scale Asymmetric Convolutional Neural Networks with Attention Mechanism for Ship-Radiated Noise Classification," Journal of Marine Science and Engineering, vol. 12, no. 1, pp. 130, 2024. https://doi.org/10.3390/jmse12010130
[26] Dongjiang Niu, Lei Xu, Shourun Pan, Leiming Xia, and Zhen Li, "SRR-DDI: A drug–drug interaction prediction model with substructure refined representation learning based on self-attention mechanism," Knowledge-Based Systems, vol. 285, pp. 111337, 2024. https://doi.org/10.1016/j.knosys.2023.111337
[27] Dongdong Xu, Ning Zhang, Yuxi Zhang, Zheng Li, Zhikang Zhao, and Yongcheng Wang, "Multi-scale unsupervised network for infrared and visible image fusion based on joint attention mechanism," Infrared Physics & Technology, vol. 125, pp. 104242, 2022. https://doi.org/10.1016/j.infrared.2022.104242
[28] Sanghyuk Roy Choi and Minhyeok Lee, "Transformer architecture and attention mechanisms in genome data analysis: a comprehensive review," Biology, vol. 12, no. 7, pp. 1033, 2023. https://doi.org/10.3390/biology12071033
[29] Jiangxun Liu, Lei Zhang, Yanfei Li, and Hui Liu, "Deep residual convolutional neural network based on hybrid attention mechanism for ecological monitoring of marine fishery," Ecological Informatics, vol. 77, pp. 102204, 2023. https://doi.org/10.1016/j.ecoinf.2023.102204
[30] Honghui Wang et al., "A novel deep-learning model for detecting small-scale anomaly temperature zones in RDTS based on attention mechanism and K-Means clustering," Optical Fiber Technology, vol. 88, pp. 103969, 2024. https://doi.org/10.1016/j.yofte.2024.103969