IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 20 April 2024, accepted 10 May 2024, date of publication 15 May 2024, date of current version 18 June 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3401254

RESEARCH ARTICLE

Research on Early Fault Identification of Cables Based on the Fusion of MTF-GAF and Multi-Head Attention Mechanism Features

HAO WU, (Member, IEEE), DAN TANG, YUAN CAI, AND CHAOWEN ZHENG
School of Automation and Information Engineering, Sichuan University of Science and Engineering, Zigong, Sichuan 643000, China
Artificial Intelligence Key Laboratory of Sichuan Province, Zigong, Sichuan 643000, China

Corresponding author: Hao Wu (11305076@qq.com)

This work was supported in part by the Project of Sichuan Provincial Science and Technology Department under Grant 2022YFS0518 and Grant 2022ZHCG0035, in part by the Artificial Intelligence Key Laboratory of Sichuan Province Foundation under Grant 2023RYY06, in part by the Enterprise Informatization and Internet of Things Measurement and Control Technology Key Laboratory Project of Sichuan Provincial University under Grant 2022WYY04, in part by the Talent Introduction Project of Sichuan University of Science and Engineering under Grant 2021RC12, and in part by the Postgraduate Innovation Fund Project of Sichuan University of Science and Engineering under Grant Y2023278.

ABSTRACT The frequent occurrence of cable early faults can lead to permanent failure of cables, making the power grid damaged and unable to work normally. To avoid the cable early faults causing great damage to the power grid operation, in this paper, we propose a research method for cable early fault identification based on the fusion of Markov Transition Field (MTF)-Gramian Angular Field (GAF) and multi-head attention mechanism features to accurately identify the cable early faults. Firstly, the fault data are preprocessed by the least mean square algorithm optimized by the adaptive gradient method; then the preprocessed one-dimensional data are converted into two-dimensional (2D) images by using MTF and GAF, respectively, and then the two types of images are fused to serve as the input of the classification network; finally, a hybrid neural network for cable early fault identification composed of a deep convolutional neural network and dense convolutional network is established, the hybrid neural network is improved by using group convolution and Ghost convolution, and the output features of the hybrid neural network are fused and classified through the mechanism of multi-head attention, and the output results of the cable early fault identification are output. At the same time, the classification results of cable early faults are visualized using the t-distributed Stochastic Neighbor Embedding (t-SNE) method to visually observe the classification effect of the hybrid neural network. The experimental results show that the algorithm has a high recognition rate for cable early fault classification, and the least mean square algorithm optimized by the adaptive gradient method is more noise-resistant compared with other optimization methods.

INDEX TERMS Cable early faults, fault identification, MTF, GAF, AlexNet, DenseNet.

I. INTRODUCTION

Power cables are widely used in urban underground power grids, power plants, and other fields due to their small footprint and low failure rate during power transmission [1]. However, the harsh underground environment can lead to phenomena such as mechanical damage to cables, resulting in a reduction of cable insulation strength, which further leads to the occurrence of early cable faults [2], [3]. Early cable faults are intermittent arcing faults that can be categorized according to the duration of the fault into 1/4 cycle early faults and multi-cycle early faults [4], with the former lasting about

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

© 2024 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

1/4 cycle and the latter lasting about 1 to 4 cycles. Early faults in cables are also known as self-clearing faults because a short circuit or fault within the cable causes a change in the current and produces an electro-thermal effect, which causes the fault to repair itself. Early cable failures usually occur at the instant of voltage peak, accompanied by arcing, when cable insulation breakdown is more likely. Early faults in cables occur causing an instantaneous increase in the phase current of the fault, a decrease in the phase voltage, and a return to normal phase currents after the fault is over. However, since early cable faults are characterized by periodicity and repetition, and the frequency of early faults will increase over time, it will eventually lead to the occurrence of permanent cable faults [5]. Therefore, fast and accurate identification of cable early faults is crucial to ensure the safe and stable operation of the power grid [1], [2], [3], [4], [5], [6].

With the development of artificial intelligence technology, a variety of intelligent algorithms are gradually applied to the study of cable early fault identification. Ghanbari [2] used a Kalman filtering-based method to detect early faults in power cables that cannot be detected in time by conventional protective relays. The results show that the method can achieve reliable identification of early faults in cables and distinguish them from other similar faults on the basis of simulation and field measurement data. Shao et al. [7] proposed a deep learning (DL) network approach based on non-negative constrained autoencoder stacking. Primary features are extracted by wavelet transform, and then multilayer NCAE is used to construct a DL network, which finally realizes early fault identification by the Softmax classifier. The results show that the method is able to recognize early faults in cables more accurately than traditional pattern recognition methods when dealing with current waveform incompleteness and random uncertainty. Liu et al. [8] proposed a method to construct a deep neural network based on a sparse self-encoder and deep confidence network, which is able to effectively recognize early faults in cables without preprocessing fault signals. However, the method has high requirements for the accuracy and completeness of fault samples and requires a large number of fault samples for training. Wang and Deng [9] proposed a cable early fault identification method that combines a noise-reducing self-encoder, an improved particle swarm algorithm, and a support vector machine. The method can effectively extract current signal features and identify faults, and the recognition accuracy is improved compared with traditional methods. Wang et al. [10] proposed a recognition method based on S-transform feature extraction and Max-Relevance and Min-Redundancy (mRMR) feature selection. The initial feature set is extracted by S-transform and the best feature subset is selected by mRMR, and the SVM classifier with kernel function is utilized to realize cable fault recognition. Experimental results show that the method has high recognition accuracy and robustness under different noise environments. Lu et al. [11] proposed a fault identification method based on Stationary Wavelet Transform (SWT) and Dropout Deep Belief Network (DDBN). The method extracts shallow features by smooth wavelet transform and then uses the constructed DDBN model to identify early. Although the cable early faults in the above literature have a high recognition accuracy, when the cable early faults occur, the fault information is hidden in the phase currents and phase voltages, and the fault characteristics are relatively difficult to detect and capture. Therefore, this paper proposes the method of converting one-dimensional fault signals into 2D images to capture early fault information and realize the study of cable early fault recognition.

First, the cable early fault data are collected by the Power Systems Computer Aided Design/ Electromagnetic Transients including DC (PSCAD/EMTDC) simulation model, and the fault signals are preprocessed by the optimized Least Mean Square (LMS). Then, two image conversion methods are used to convert the one-dimensional fault data into two types of 2D images, and the images containing two types of fault information are fused into one feature image, and this feature image is used as the input of the classification network. Finally, the constructed hybrid neural network is improved using combined group convolution and Ghost convolution to realize the recognition study of early cable faults through multi-head attention mechanism feature fusion. The t-SNE visualization method is also utilized to demonstrate the classification results of the hybrid neural network used in this paper.

II. DATA PREPROCESSING AND FEATURE EXTRACTION

A. DATA PREPROCESSING METHOD BASED ON OPTIMIZED LMS

The core of the Least Mean Square (LMS) algorithm lies in adjusting the filter coefficients by constant iteration to minimize the mean square value of the output error signal [12], which is widely used in fields such as noise cancellation. Define the input signal as $X = \{x_1, x_2, \cdots, x_n\}$, the filter coefficients as $W(n)$, and the output of the filter as $Y(n)$. Then we have:

$$Y(n) = \sum_{i=0}^{M} W_i(n) X(n-i) \tag{1}$$

where M is the order of the filter.

The error signal $e(n)$ of an LMS filter is usually defined as the difference between the desired signal d(n) and the output signal Y(n) of the filter, i.e:

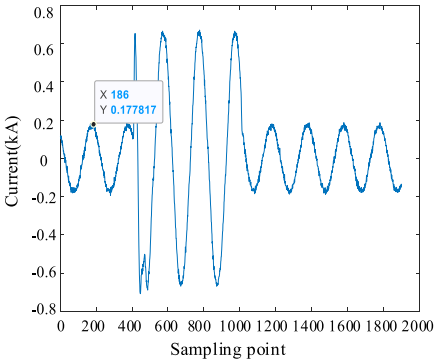
$$e(n) = d(n) - Y(n) \tag{2}$$

The formula for updating the coefficients of the filter by the LMS algorithm is as follows:

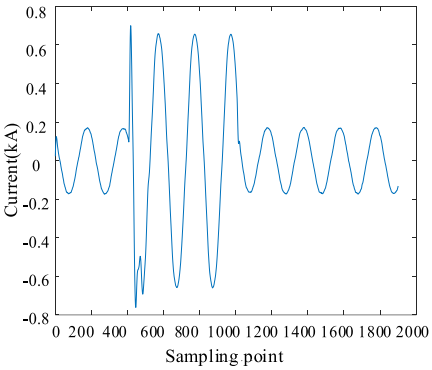
$$W(n+1) = W(n) + 2\mu e(n) X(n) \tag{3}$$

where the learning rate $0 < \mu < \frac{1}{\lambda_{\max}}$, $\lambda_{\max}$ is the maximum eigenvalue of the autocorrelation matrix of the input signal.

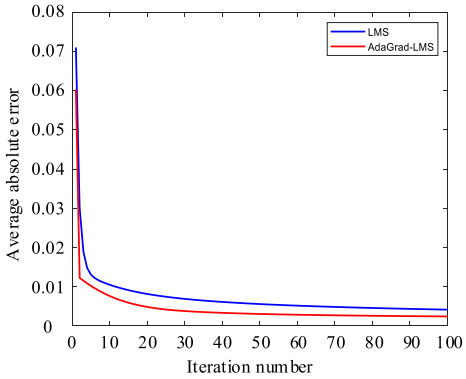
Since in the LMS algorithm, all parameters share a global learning rate, which can lead to slow learning or failure to converge to the optimal solution [13]. Therefore, in this paper,



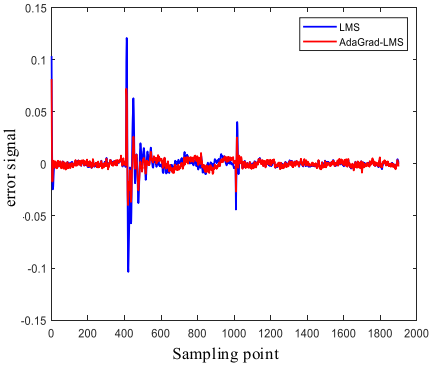
(a) Multi-cycle fault signal waveform of a cable with 30db of noise added



(b) AdaGrad-LMS pre-processed cable multi-cycle fault signal waveforms



(c) Average absolute error



(d) error signal *e(n)*

FIGURE 1. Comparison of error between LMS and improved LMS.

we use a modified LMS algorithm to filter the fault signal and introduce a personalized adjustment to the update of each parameter by introducing the adaptive gradient (AdaGrad) algorithm. The core idea of the AdaGrad algorithm is to maintain a cumulative gradient sum of squares for each parameter, and then based on this cumulative value, adjust each parameter's learning rate. This approach results in a faster decrease in the learning rate for frequently updated parameters and a slower decrease in the learning rate for infrequently updated parameters. Therefore, the weights of the improved LMS algorithm are updated as follows:

$$\begin{aligned} grad &= e(n) \cdot X(n) \\ r &= \alpha \cdot r + grad^2 \\ W(n+1) &= W(n) + \frac{\eta}{\sqrt{\varepsilon + r}} \cdot grad \end{aligned} \tag{4}$$

where grad is the gradient, r is the cumulative sum of all current gradient squares, η is the initial learning rate, α is the forgetting factor, which controls the effect of the historical gradient on the current learning rate, and ε is a small constant to prevent division by zero.

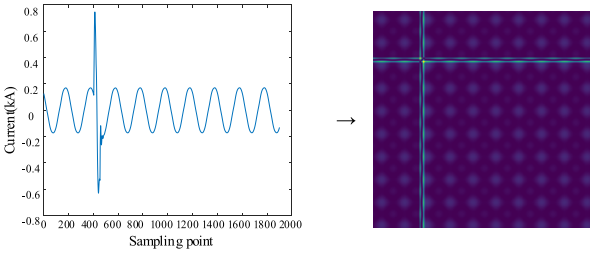
Early fault signals in cables may contain various types of noise, including interference from various sources such as sensors, electromagnetic interference, and the environment. To verify the denoising effect of the improved LMS algorithm, Fig. 1 below takes the cable multi-cycle fault with a signal length of 1900 sampling points as an example, preprocesses the fault signal with the addition of 30db of noise, and compares the denoising effect of the LMS algorithm and the improved LMS algorithm using the error signals and the average absolute error. As shown in Fig. (c), the optimized LMS algorithm has a lower average absolute error than the original LMS algorithm. As shown in Fig. (d), the optimized LMS algorithm performs better in terms of error effect during the fault signal occurrence. Therefore, AdaGrad's improved LMS algorithm improves the overall performance of the algorithm by introducing an independent adaptive learning rate for each weight.

B. FEATURE EXTRACTION

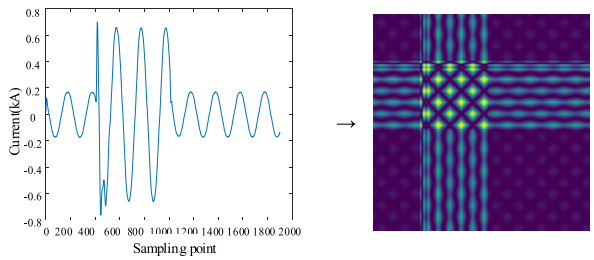
1) FEATURE EXTRACTION BASED ON GAF METHOD

Gramian Angular Field (GAF) is a representation that converts time series data into a 2D image, making time series data easier to understand and analyze [14], [15]. Given a set of time series data $X = \{x_1, x_2, \cdots, x_n\}$, the data is first normalized so that all its values are scaled in the interval $[-1,1]$, this is to eliminate the scale differences between the data and ensure the accuracy of the subsequent calculations.

$$\tilde{x}_i = \frac{(x_i - \max(X)) + (x_i - \min(X))}{\max(X) - \min(X)} \tag{5}$$

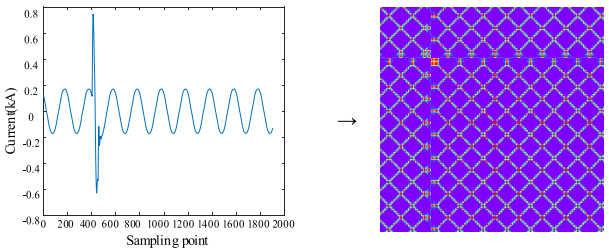


（a）2D image of GAF for early faults in 1/4 cycle of cable

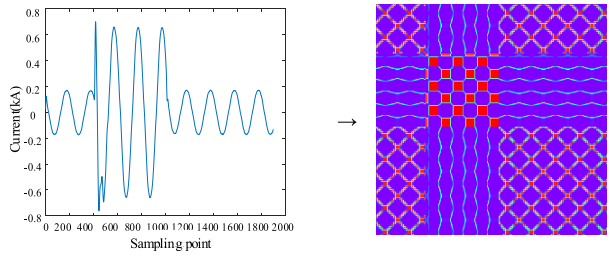


（b）2D image of GAF for early faults in multi cycle of cable

FIGURE 2. Feature extraction of fault data based on GAF.



（a）2D image of MTF for early faults in 1/4 cycle of cable



（b）2D image of MTF for early faults in multi cycle of cable

FIGURE 3. Feature extraction of fault data based on MTF.

The one-dimensional time series data is then rescaled in polar coordinates by encoding the values as angular cosines and setting the timestamp t_i to radius r. Then:

$$\begin{aligned} \phi &= \arccos \tilde{x}_i, \quad -1 \le \tilde{x}_i \le 1, \ \tilde{x}_i \in \tilde{X} \\ r &= \frac{t_i}{N}, \quad t_i \in \mathbb{N} \end{aligned} \tag{6}$$

where N is a constant factor of the regularized polar coordinate system generating space.

After transforming the one-dimensional time series, the Gram matrix is generated by considering the triangular sum between each point to realize the temporal correlation in different time intervals.

$$G = \begin{bmatrix} \cos(\phi_1 + \phi_1) & \cos(\phi_1 + \phi_2) & \cdots & \cos(\phi_1 + \phi_n) \\ \cos(\phi_2 + \phi_1) & \cos(\phi_2 + \phi_2) & \cdots & \cos(\phi_2 + \phi_n) \\ \vdots & \vdots & \ddots & \vdots \\ \cos(\phi_n + \phi_1) & \cos(\phi_n + \phi_2) & \cdots & \cos(\phi_n + \phi_n) \end{bmatrix} \tag{7}$$

In this paper, Gramain Angular Summation Fields (GASF) with angle summation inner product is selected for data feature extraction and 2D image acquisition. Fig. 2 below shows an example of a cable 1/4 cycle early fault and a multi-cycle early fault, where one-dimensional time series data with a fault signal length of 1×1900 is converted into a two-dimensional image of size 256×256.

2) FEATURE EXTRACTION BASED ON MTF METHOD

Markov Transition Field (MTF) is a method of converting time series data into a 2D image, aiming to make the features of time series data more intuitive and analyzable. The basic idea of MTF is to use the Markovian properties in time series data to construct an image [16]. Given a set of time series data $X = \{x_1, x_2, \ldots, x_n\}$, the given time series data is first divided into Q parts according to the range of values, and each part represents a different subspace. Then each data point in the time series is mapped into the defined subspace according to the numerical characteristics, and in this process, the calculation of the transfer probability w_{ij} is very important, which represents the probability that a data point in the time series is transferred from state i to state j. It can be estimated by observing the state change of consecutive data points in the time series. The specific calculation is as follows:

$$w_{ij} = p\left\{x_{t+1} \in q_j \,|x_t \in q_i\right\} \tag{8}$$

where x_t is the signal amplitude at moment t, P is the probability, qj is one state in the Markov chain, and qi is another state in the Markov chain.

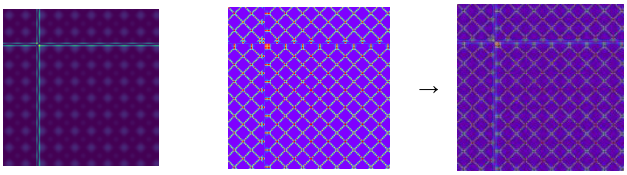
The Markov variation field M matrix is constructed for each probability by arranging them in time order:

$$M = \begin{bmatrix} w_{ij} \left| x_1 \in q_i, x_1 \in q_j \right. & \cdots & w_{ij} \left| x_1 \in q_i, x_n \in q_j \right. \\ w_{ij} \left| x_2 \in q_i, x_1 \in q_j \right. & \cdots & w_{ij} \left| x_2 \in q_i, x_n \in q_j \right. \\ \vdots & \ddots & \vdots \\ w_{ij} \left| x_n \in q_i, x_1 \in q_j \right. & \cdots & w_{ij} \left| x_n \in q_i, x_n \in q_j \right. \end{bmatrix} \tag{9}$$

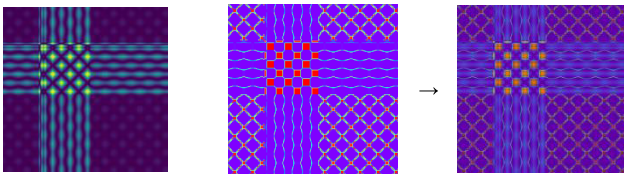
The one-dimensional time series data is converted to a 2D image by the above MTF calculation method, and Fig. 3 below shows the converted 256×256.2D image as an example of a cable 1/4 cycle early fault and a multi-cycle early fault.

3) COMBINATION FEATURE DATASETS

The cable early fault time series data are transformed into 2D image data by the GAF method and MTF method, and in order to preserve more features, this paper fuses the GAF image and MTF image. Since MTF captures the features by calculating the moments of the signal, focusing on the overall



(a) Early fault fusion image of cable half cycle wave



(b) Multi cycle early fault fusion image of cables

FIGURE 4. Construct feature image dataset.

shape and structure of the signal; while the features extracted by GAF pay more attention to the temporal evolution characteristics of the signal, capturing the angular information between the time series. Therefore, constructing the feature dataset through the fusion method can capture different aspects of the data more comprehensively, thus obtaining more comprehensive information. Meanwhile, different feature extraction methods have different sensitivities to noise and changes in the data, which can enhance the robustness of the model to noise. Each pixel of the fused image is obtained by taking the arithmetic average of the corresponding pixels of the GAF image and the MTF image. Let I_1 and I_2 be the pixel matrices of the GAF image and the MTF image, respectively, and for each pixel position (x, y) its fused pixel value F(x, y) is calculated by the following equation:

$$F(x, y) = \frac{I_1(x, y) + I_2(x, y)}{2} \tag{10}$$

Through the above formula, a new image of size 256 × 256 containing the intrinsic features of the two types of fault data is generated, constituting the feature image dataset, and Fig. 4 below shows a cable 1/4 cycle early fault and a multi-cycle early fault as examples.

III. A METHOD BASED ON THE IDENTIFICATION OF EARLY FAULTS IN CABLES

In order to better identify early cable faults and other cable overcurrent disturbances, and to guarantee the safe and stable operation of the power grid, this paper constructs a hybrid neural network based on AlexNet and DenseNet and fuses the output features of the hybrid neural network through the mechanism of multi-head attention to realize the identification study of early cable faults.

A. GROUP CONVOLUTION AND GHOST CONVOLUTION

1) GROUP CONVOLUTION PRINCIPLE

Group convolution is a technique that groups input channels and performs convolution operations independently; this method was first proposed as a solution to the problem of

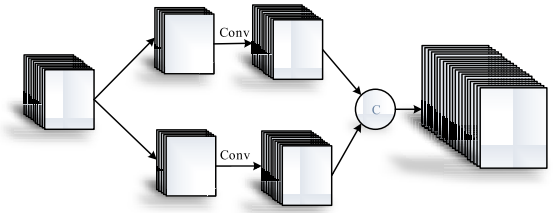


FIGURE 5. Group convolutional structure.

GPU memory constraints [17] and was later found to improve the computational efficiency of networks.

The structure of group convolution is shown in Fig. 5 above. In group convolution, the channels of the input feature map are first divided into several groups. The convolution operation is carried out independently for each group, and each group of convolution kernel acts only on the corresponding group of input channels. Finally, the output feature maps of the groups are combined to form the final output feature map. Define the number of channels of input data feature map and the number of channels of output data feature map as C_{input} and C_{output} respectively, and the size of convolution kernel is K × K, the group convolution divides the input feature map into N groups, then the number of channels of each group is $\frac{C_{input}}{N}$, and the number of channels of the output feature map is $\frac{C_{output}}{N}$, then the total number of convolution kernels is:

$$N \times \frac{C_{input}}{N} \times \frac{C_{output}}{N} \times K \times K \tag{11}$$

And the total number of convolution kernels should be. For normal convolution, the total number of convolution kernels should be:

$$C_{input} \times C_{output} \times K \times K \tag{12}$$

Comparing the two kinds of convolution kernel totals, it can be seen that the computation amount of group convolution is $\frac{1}{N}$ of the computation amount of normal convolution. Group convolution significantly reduces the number of parameters by reducing the number of connections between each convolution kernel and the input channel, and fewer parameters mean lower computational costs. Therefore, utilizing group convolution instead of normal convolution can reduce network computation, decrease model complexity, and reduce the occurrence of overfitting phenomenon.

2) GHOST CONVOLUTION PRINCIPLE

The lightweight Ghost convolution module is an efficient deep-learning architecture component that was proposed in 2019 [18]. The core idea of the Ghost convolution module is to generate more feature maps while reducing the computational cost. The module achieves this by generating a portion of the feature maps using fewer convolutional operations and then generating the remaining feature maps through linear operations.

The schematic structure of ordinary convolution and Ghost convolution is shown in Fig. 6 below. Ghost convolution first

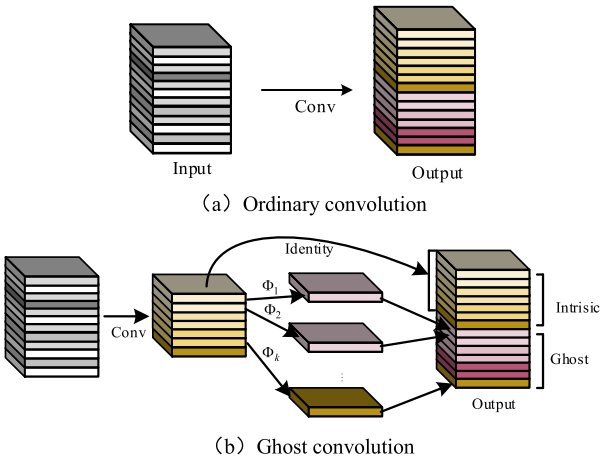


FIGURE 6. Schematic diagram of ordinary convolution and Ghost convolution.

uses ordinary convolution to generate a set of feature maps, which are called "native feature maps"; then it uses linear transformation operation to generate more feature maps from native feature maps, which are called These additional feature maps are called "Ghost feature maps"; finally, the native feature maps and Ghost feature maps are merged to form the final output feature maps.

Define the size of the input Input to be H × W × C, the number and size of the convolution kernels to be n and K × K, respectively, and the size of the output Output to be H'× W'× n. Then the computational and parametric quantities of the Ghost convolution are as follows:

$$\frac{n}{s}\times H'\times W'\times C\times k\times k+(s-1)\times\frac{n}{s}\times H'\times W'\times d\times d$$
$$\frac{n}{s}\times C\times k\times k+(s-1)\times\frac{n}{s}\times d\times d \tag{13}$$

where s (s > 1) is the scale factor for generating the Ghost feature map and d × d is the size of the linearly transformed convolutional kernel used to generate the Ghost feature map.

And the computational and parametric quantities of ordinary convolution are:

$$n\times H'\times W'\times C\times k\times k$$
$$n\times C\times k\times k \tag{14}$$

The computational and parametric quantities of the ordinary convolution and the Ghost convolution are compared, and the ratios are r_s and r_e. This is shown in the following equation, where the size of s ≪ c, d × d is similar to the size of k × k.

$$r_s=\frac{n\times H'\times W'\times C\times k\times k}{\frac{n}{s}\times H'\times W'\times C\times k\times k+(s-1)\times\frac{n}{s}\times H'\times W'\times d\times d}$$
$$=\frac{C\times k\times k}{\frac{1}{s}\times C\times k\times k+(s-1)\times\frac{1}{s}\times d\times d}\approx\frac{s\times C}{s+C-1}\approx s$$
$$r_c=\frac{n\times C\times k\times k}{\frac{n}{s}\times C\times k\times k+(s-1)\times\frac{n}{s}\times d\times d}\approx\frac{s\times C}{s+C-1}\approx s \tag{15}$$

From the above equation, it can be found that Ghost convolution is more efficient than ordinary convolution because it

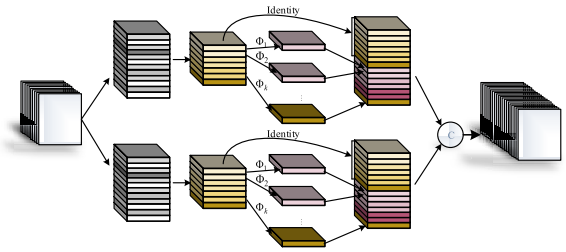


FIGURE 7. Combination structure diagram of group convolution and Ghost convolution.

reduces the number of convolution operations, which reduces the computational cost and ensures that the model is able to learn more feature information to a certain extent at the same time, realizes the model's lightweight and reduces the model's complexity.

3) COMBINING GROUP CONVOLUTION AND GHOST CONVOLUTION

In order to improve the computational efficiency of the model and reduce the number of parameters, this paper proposes a method to replace the ordinary convolution in the network model by combining the group convolution and Ghost convolution. This combination aims to fully utilize the advantages of the two convolution methods to achieve a more efficient network design and solve the problems of a large number of traditional convolutional parameters and high computational costs.

The structure of the combination method is shown in Fig. 7 below. In this combination method, the input feature maps are first divided into groups, and each group is subsequently processed by an independent Ghost convolution module. The Ghost convolution module effectively reduces the amount of computation required for each convolution by generating Ghost feature maps. This approach not only reduces the computational cost of a single convolution but also significantly reduces the number of parameters in the overall model as each group processes fewer input channels.

B. AlexNet MODEL PRINCIPLES AND IMPROVEMENTS

The AlexNet model is a key milestone in the history of deep learning, proposed by Alex Krizhevsky et al. in 2012. Its performance and effectiveness have been widely verified and recognized. The original AlexNet consisted of five convolutional layers followed by three fully connected layers [19], and the key to this architecture is the ability to automatically extract features from an image, which are then used for classification tasks.

Since the traditional AlexNet network model is relatively complex, but the cable early faults are relatively simple and easy to extract feature information, this paper chooses to use the simplified and improved AlexNet network model for the cable early fault recognition research. First, the five convolutional layers of the original network are simplified into three convolutional layers, and each convolutional layer is connected to a maximum pooling layer; then the three fully

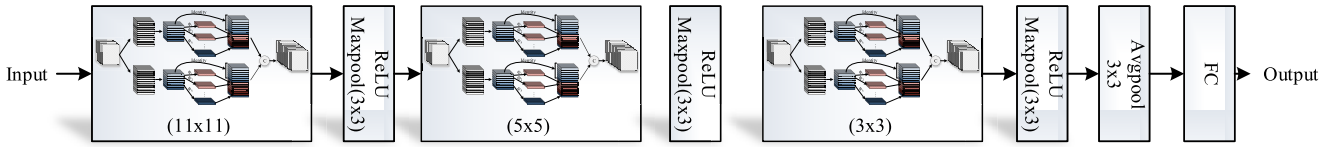


FIGURE 8. Improved AlexNet model structure diagram.

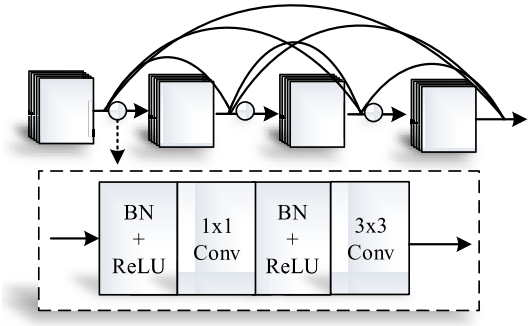


FIGURE 9. Dense block structure.

connected layers of the original network are simplified into two, and the first fully connected layer parameter is modified to 1024, and the second fully connected layer parameter is set to 7, i.e., the cable early faults are divided into 7 categories. Then, the ordinary convolutional layers in the AlexNet model are replaced by the combined group convolution and Ghost convolution. The improved AlexNet model reduces the computational requirements of the model while reducing the parameters, realizes lightweight, and also maintains the expressiveness of the model. The structure of the improved AlexNet model is shown in Fig. 8 below.

C. DenseNet MODEL PRINCIPLE AND IMPROVEMENT

DenseNet is an effective deep-learning architecture proposed by Huang et al. in 2016 [20]. Its core innovation lies in the reuse and transfer of features through dense layer-to-layer connections within Dense Blocks.

The DenseNet network mainly consists of several Dense Blocks and Transition Layers. In the Dense Block, each layer is connected to all previous layers in a cascade fashion and maintains the same size as the feature map. The Transition Layer contains the Batch Normalization Layer, Activation Layer, Convolution Layer, and Pooling Layer, whose main role is to connect different Dense Blocks and compress the feature map while connecting to reduce the dimensionality. The dense block structure is shown in Fig. 9 below.

In this paper, we utilize a combination of group convolution and Ghost convolution to replace the normal convolution in the densely connected layer of DenseNet, with the aim of generating ''Ghost'' feature maps using fewer convolution operations, thus reducing the computational complexity and model parameters. The improved DenseNet model structure is shown in Fig. 10 below.

D. INTEGRATION OF CHARACTERISTICS OF MULTI-HEAD ATTENTION MECHANISMS AND T-SNE VISUALIZATION

1) INTEGRATION OF CHARACTERISTICS OF MULTI-HEAD ATTENTION MECHANISMS

The multi-head attention mechanism is derived from the Transformer architecture [21] and has been shown to be effective in sequential tasks. The multi-head attention mechanism is the introduction of multiple attention heads in the attention mechanism, each of which independently learns different attention weights to capture relevant information at different locations in the input sequence. By introducing the multi-head attention mechanism, the model can parallelize the attention computation on the input from different perspectives, which improves the model's representational and generalization capabilities. Therefore, in this paper, we propose a feature fusion model based on the multi-head attention mechanism, which aims to integrate feature representations from two parallel neural networks.

The process of convergence of the characteristics of the multi-head attention mechanisms is as follows: first, two independent network models process the input images in parallel and output the processed feature vectors individually; then, the feature vectors from these two models are spliced together in terms of feature dimensions; then, a multi-head attention mechanism is introduced to process the spliced feature vectors, and the weights of the features are automatically adjusted according to the interrelationships among the input feature vectors, so as to highlight the features that are favorable for classification and suppress irrelevant noise; finally, a linear classification layer is used to map the attention-enhanced features to the final category labels to complete the fault classification task.

2) T-SNE VISUALIZATION

The core principle of the t-distributed Stochastic Neighbor Embedding (t-SNE) algorithm [22] is to compute the similarities between data points in a high-dimensional space, and then reproduce these similarities as probability distributions in a low-dimensional space. This visualization method enables us to observe the distribution and clustering of data points of different categories in the space, to verify whether the model can effectively distinguish different categories. Therefore, in this paper, we will visualize the feature representation of the hybrid network model by the t-SNE algorithm to show the network classification effect.

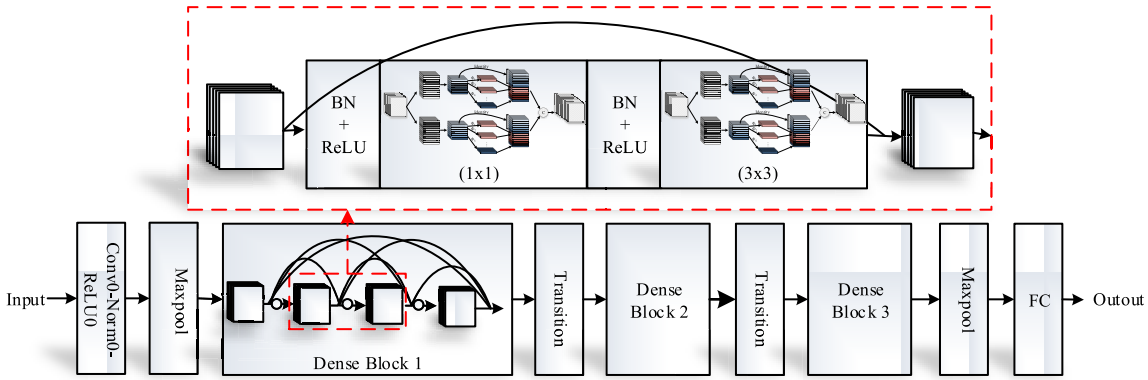


FIGURE 10. Improved DenseNet model structure diagram.

E. RECOGNITION NETWORK BASED ON EARLY FAULTS IN CABLES

Since frequent early cable faults can lead to grid damage problems, in order to avoid this problem this paper proposes a research method for early cable fault identification based on the fusion of features of multi-head attention mechanism. Firstly, the hybrid neural network of AlexNet and DenseNet is constructed, and the combined group convolution and Ghost convolution are used to improve AlexNet and DenseNet; then the output features of the hybrid model are fused by using the multi-head attention mechanism to realize the recognition study of early cable faults; finally, the t-SNE classification results are visualized to observe the fault classification results are visualized to more intuitively observe the effect of fault classification. The structure of the constructed fault identification network model is shown in Fig. 11 below.

IV. RESEARCH PROCESS BASED ON THE IDENTIFICATION OF EARLY FAULTS IN CABLES

The research process of cable early fault identification based on feature fusion of multi-head attention mechanisms is shown in Fig. 12 below. The specific research steps are as follows:

(1) Firstly, build a cable simulation model through PSCAD, set up cable early faults and cable overcurrent disturbance faults, and collect raw fault information.

(2) Data preprocessing of one-dimensional raw fault signals through AdaGrad optimized LMS filter.

(3) Convert the preprocessed 1D fault signal into two types of 2D image signals by GAF method and MTF method respectively, and then fuse the two types of 2D image signals into a new 2D image to construct the 2D image dataset and divide it.

(4) Input the divided training set into the feature fusion network based on the multi-head attention mechanism for training, adjust the network parameters, and train the classification network. Then the test set is input into the trained classification network to output the fault recognition results.

V. ALGORITHM ANALYSIS

A. DATA ACQUISITION

In order to validate the cable early fault identification method proposed in this paper, a 10kV cable early fault simulation model is constructed in PSCAD/EMTDC, and the structure is shown in Fig. 13 below. Since the cable early fault resistance is similar to the arc fault resistance characteristics and both have nonlinear time-varying characteristics, this allows the arc model to better simulate the waveforms of early cable faults. Therefore, most of the existing studies have used arc modeling for the identification and study of early faults in cables. Because the improved "cybernetic" arc model can more realistically reflect the characteristics of actual arcs, the improved "cybernetic" arc model is used in this paper to simulate early cable faults [23]. Early cable fault signals are usually characterized by weak fluctuations or changes in current. Therefore, cable early fault signals can be detected by monitoring and analyzing the current belief signals simulated in the simulation model. The simulation parameters of the arc are referred to in [24], and the voltage drop per centimeter of the arc gap is set to 75V; the amplitude of the steady-state current of metallic grounding fault in the small-current grounding system is 7.818A; the constant coefficient is 7.56×10^{-6}; and the length of the arc is 5.

In the PSCAD simulation model, the cable adopts XLPE tube type three-core cable, and the radius of the cable core, the radius of the main insulation of the cable, the radius of the metal sheath, and the radius of the outer insulation of the cable are set to 0.01175 m, 0.01275 m, 0.02325 m, and 0.02445 m. In building the cable early fault simulation model, not only the cable 1/4 cycle early fault and the cable multi-cycle early fault are built, but also five kinds of overcurrent disturbance faults, constant impedance grounding, transformer excitation inrush, motor switching, load switching, and capacitor switching, are built. The waveforms of the seven faults are shown in Fig. 14 below. During power system operation, various transient disturbances occur frequently, such as transformer excitation inrush, motor switching, load switching, and capacitor switching. These perturbations lead to changes in the current with characteristics similar to the

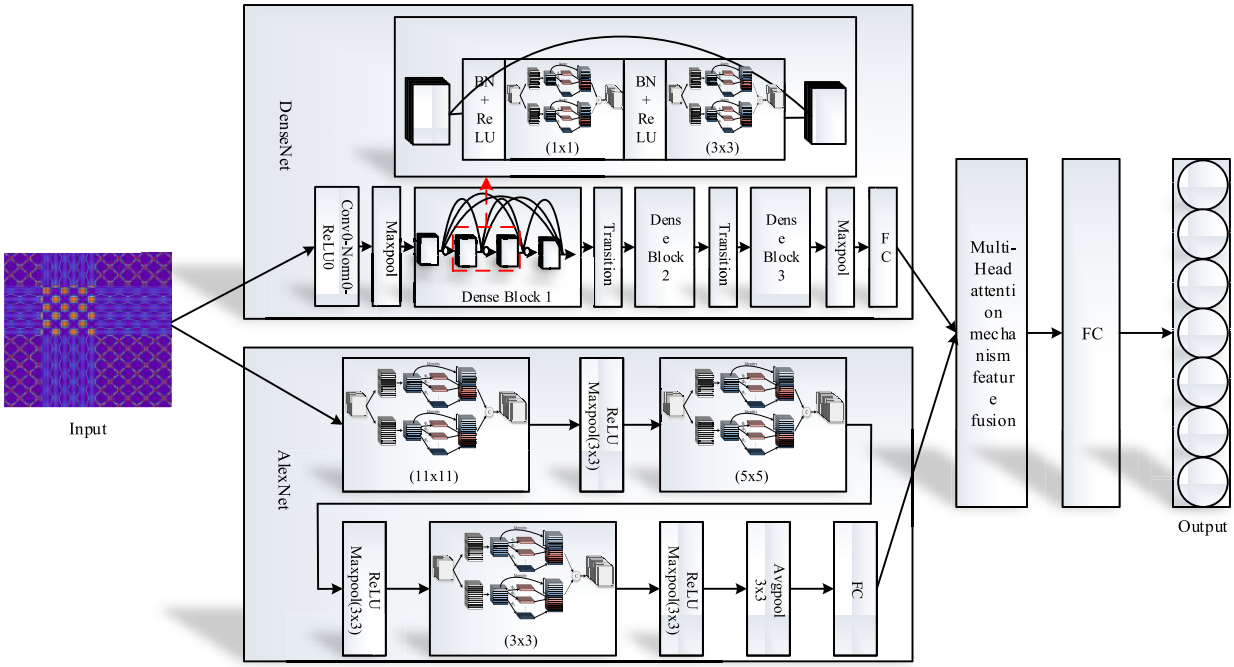


FIGURE 11. Recognition network based on early faults in cables.

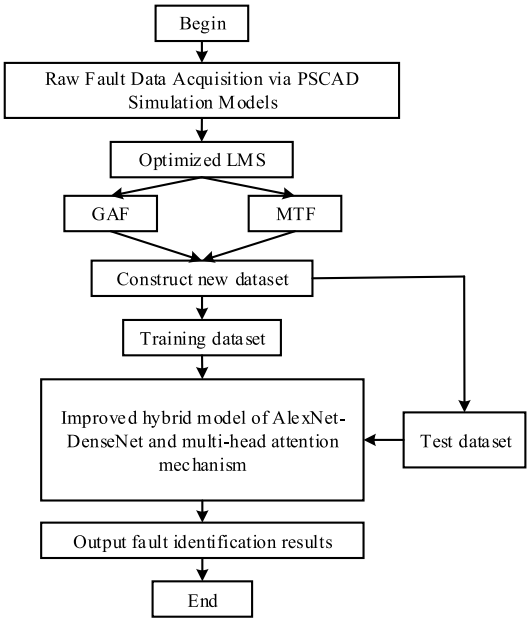


FIGURE 12. Cable early fault identification process.

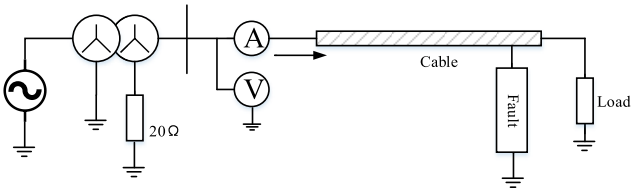


FIGURE 13. Structural diagram of early cable fault simulation model.

transient characteristics of the system when a short-circuit fault occurs. In addition, early faults are usually single-phase faults whose current transient process is similar to that of single-phase constant impedance grounding. Therefore, five cable disturbance overcurrent phenomena are considered in this paper for cable early fault classification study.

The cable early fault simulation is set up with two cases of grounding and ungrounding via an arcing coil, and the load is set up with three cases of light load, heavy load, and full load. For cable early fault types, such as cable 1/4-cycle early faults and multi-cycle early faults, the moment of occurrence is set as a random value within 0.001 seconds before and after the voltage peak. The grounding resistances for constant impedance faults are considered for 0.5 Ω, 50 Ω and 500 Ω cases and the angle of the fault is set in a random range of [0°, 360°]. The motor switching is set for 3 different power levels, while the load switching is set for 5 different cases. In addition, the capacitor switching also considered 3 different cases and set the fault angle in a random range of [0°, 360°]. Finally, for the transformer excitation inrush fault angle was also set to a random range of values within [0°, 360°]. The length of the cable line is set to 20km, and the fault simulation is done every 2km from the first 2km of the line. The sampling object of the data set is the cable fault current signal, and the sampling frequency is set to 10 kHz. the obtained cable early fault number set is divided into training set, validation set, and test set according to the ratio of 6:2:2, as shown in Table 1 below.

B. MODEL EVALUATION INDICATORS

In order to comprehensively evaluate the performance of the model, four key metrics, namely Accuracy, Precision, Recall, and F1 value, are selected as evaluation criteria in this paper.

TABLE 1. Dataset partitioning for early cable faults.

Fault Type	Number of samples	Training Samples	Validation Sample	Test Sample
1/4 cycle early faults	2700	1620	540	540
Multi-cycle early faults	2700	1620	540	540
Constant impedance grounding	2592	1556	518	518
Motor switching	2592	1556	518	518
Capacitor switching	2592	1556	518	518
Load switching	2700	1620	540	540
Transformer excitation inrush	2700	1620	540	540

The evaluation metrics are calculated as follows:

$$\begin{aligned} \text{Accuracy} &= \frac{TP + TN}{TP + FN + FP + TN} \\ \text{Precision} &= \frac{TP}{TP + FP} \\ \text{Recall} &= \frac{TP}{TP + FN} \\ \text{F}_1 &= \frac{2TP}{2TP + FP + FN} \end{aligned} \tag{16}$$

where TP is the true positive case, TN is the true negative case, FP is the false positive case, and FN is the false negative case.

Accuracy measures the proportion of samples correctly predicted by the classifier, which is the proportion of true and true-negative cases among all samples. Precision is the proportion of all samples classified as early cable faults that are actually early cable faults. Recall is the proportion of correctly classified cable early faults among all samples of actual cable early faults. The F1 value is the reconciled average of Precision and Recall, which combines the accuracy and coverage of the classifier.

C. NETWORK TRAINING AND FAULT IDENTIFICATION RESULTS

In this paper, we propose a cable early fault identification algorithm based on the fusion of features from multi-head attention mechanisms to classify fault feature maps. The cable early fault identification experiments used a computer equipped with an Intel Core i5-12600Kf 3.7GHz CPU, GeForce RTX3060Ti GPU, and 32 GB RAM running a 64-bit Windows operating system. The software environment was Python 3.10 using the PyTorch 2.0.0 library and CUDA version 11.7.

In the network model, the size of the feature picture is adjusted from 256×256 to 224×224 to ensure that it fits the input of the network model. This not only effectively reduces the computational complexity and storage requirements and speeds up the training of the model, but also the adjusted feature picture still retains the main fault characteristics and

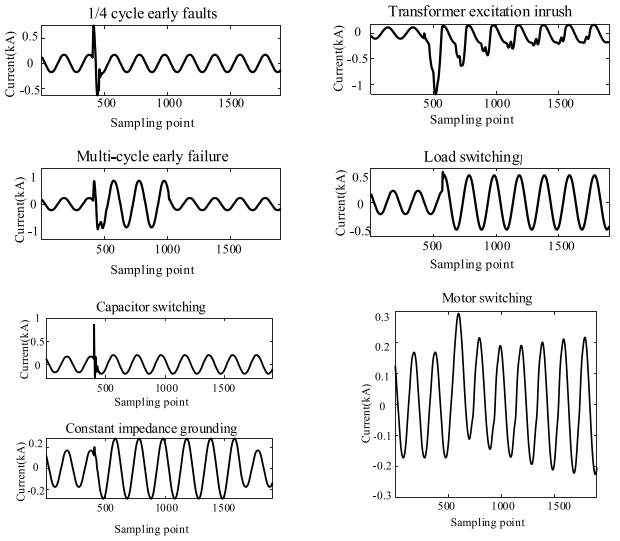


FIGURE 14. Fault line current waveform.

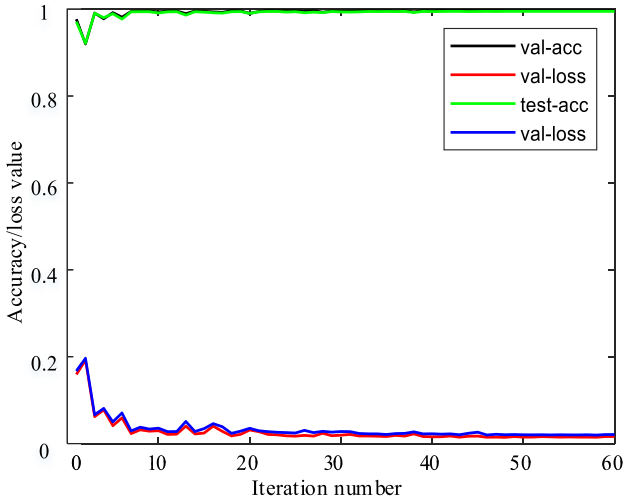


FIGURE 15. Fault classification curves for validation and testing sets.

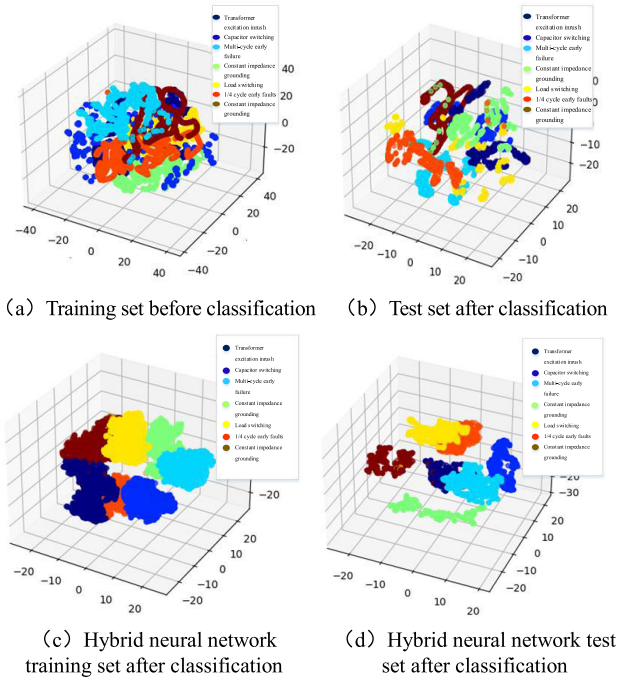
information. The training process of the whole network was set to 60 epochs with 32 samples per batch with an initial learning rate of 0.0001 and the learning rate was adjusted by cosine annealing strategy. AlexNet and DenseNet were trained separately and further training was done on the basis of fusing these two models. The Adam optimizer was used to tune the network parameters during training, and the cross-entropy loss function was employed to calculate the loss values of the models. At the end of each training cycle, we evaluated the performance of each model on the validation set and performed the final evaluation on the test set, and the validation and test set fault classification accuracy curves are shown in Fig. 15 below.

The cable early fault assessment parameters include accuracy, precision, recall and F1 score as shown in Table 2 below.

From the above change curves of fault classification accuracy and change curves of loss value for the validation and test sets, it can be found that the network has stabilized when the test network is iterated up to 30 times. From the Fig. and the

TABLE 2. Model classification indicators.

Accuracy	Precision	Recall	F1
99.49%	99.50%	99.49%	99.49%



(a) Training set before classification (b) Test set after classification

(c) Hybrid neural network training set after classification (d) Hybrid neural network test set after classification

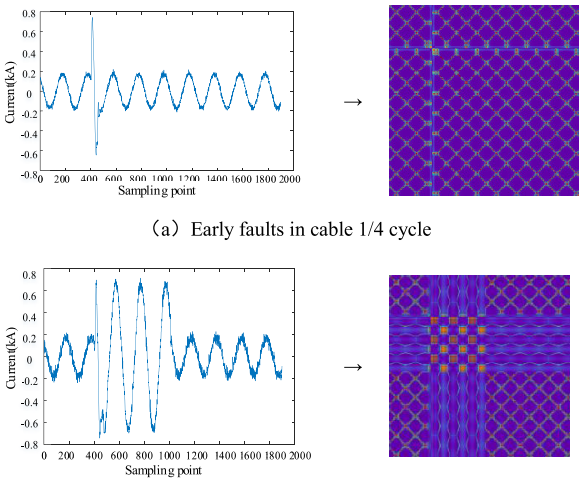
FIGURE 16. T-SNE visualization classification result.

table, it can be found that the accuracy of the test set for cable early fault identification reaches 99.49% and the loss value stays at 0.0202, which indicates that the proposed method in this paper can accurately categorize cable early faults. In addition, the t-SNE technique is also used to visualize the fault classification, and the fault classification on the training set and the test set is demonstrated by 3D graphs, as shown in Fig. 16 below. From the t-SNE visualization of classification results graphs on the test set, it can be found that the method used in this paper can accurately identify cable early faults.

D. COMPARISON EXPERIMENT

1) EFFECT OF DIFFERENT LEVELS OF NOISE ON FAULT IDENTIFICATION

In the field of early fault detection in cables, accurate fault identification is crucial, especially in the complex environment of power systems. The actual grid environment is full of unforeseen noise disturbances, which may come from the grid itself or from surrounding electronic devices, greatly affecting the accuracy and reliability of fault signals. In order to ensure the effectiveness of the cable early fault identification method, it is very necessary to examine its ability to resist noise interference. To this end, in this paper, common types of noise in the grid environment are simulated, such as Gaussian white noise, which is statistically normally distributed and has similar characteristics to the common noise in the grid. By introducing Gaussian white noise with different signal-to-noise ratios into the early fault signal of the cable, the noise



(a) Early faults in cable 1/4 cycle

(b) Early faults in cable multi cycle

FIGURE 17. Waveform and characteristic images of early cable faults after adding 20dB noise.

situation that the cable may encounter under actual grid operating conditions is simulated. Fig. 17 below shows a cable 1/4 cycle fault and a cable multi-cycle fault as an example, where 20dB noise is introduced into the original fault signal, and the fault waveforms and characteristic images are shown below.

In order to comprehensively evaluate and improve the performance of the algorithms in this paper in the face of grid noise, different data preprocessing methods are used. Specifically, the least mean square (LMS) algorithm optimized by AdaGrad was mainly used to preprocess early cable fault signals, and its performance is compared with several other optimization methods, including the LMS method using Adam optimization, RMSprop optimization, and the original LMS method. By testing these methods in Gaussian white noise with different signal-to-noise ratios, the effectiveness of various data preprocessing strategies in reducing noise and improving fault detection accuracy is explored in depth. The noise immunity of different data preprocessing methods is shown in Table 3 below.

Based on the results of the noise comparison experiments above, it can be seen that the noise level has a significant effect on fault classification. In low noise environments (40dB and 30dB), the AdaGrad optimized LMS method performs well with 99.43% and 99.41% accuracy, precision, and recall. In comparison, other methods such as Adam-LMS, RMSprop-LMS, and LMS performed slightly worse under these conditions. However, at higher noise levels (20 dB), the performance of all the methods suffers, but it can still be seen that the AdaGrad-LMS method is relatively stronger with an accuracy of 92.78%. Thus, the AdaGrad-optimized LMS method used in this paper performs well at all noise levels with better noise immunity.

2) COMPARATIVE ANALYSIS OF ABLATION EXPERIMENT RESULTS

In order to verify the effectiveness of the method proposed in this paper, different combinations of methods are set up

TABLE 3. Comparison results of noise resistance using different preprocessing methods.

Noise	Preprocessing method	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)
40dB	AdaGrad-LMS	99.43	99.43	99.43	99.44
	Adam-LMS	98.71	98.71	98.71	98.70
	RMSprop-LMS	99.38	99.40	99.37	99.38
	LMS	99.25	99.26	99.24	99.25
30dB	AdaGrad-LMS	99.41	99.42	99.40	99.41
	Adam-LMS	96.26	96.69	96.18	96.14
	RMSprop-LMS	98.87	98.89	98.86	98.86
	LMS	98.47	98.50	98.46	98.46
20dB	AdaGrad-LMS	92.78	92.49	92.76	92.77
	Adam-LMS	91.10	91.64	91.87	92.01
	RMSprop-LMS	88.99	91.13	88.87	98.10
	LMS	86.13	89.10	86.05	86.22

for ablation experiments to compare the cable early fault identification accuracy of different combinations of methods. The combinations are shown in Table 4 below, combination 1 is a densely connected network, combination 2 is a deep convolutional neural network, combination 3 is a model that introduces a multi-head attention mechanism to fuse a deep convolutional neural network and a densely connected network, combination 4 is a densely connected network that introduces a combination of group convolutions and Ghost convolutions, combination 5 is a deep convolutional neural network that introduces a combination of group convolutions and Ghost convolutions to replace the ordinary convolutional neural network, and combination 6 is the recognition method used in this paper.

The computation time in Table 4 above refers to the network running time for recognizing a set of fault data. It is obvious from the experimental results that the cable early fault diagnosis method proposed in this paper performs the best among all model combinations, with an identification accuracy of 99.49%, which is much higher than other combinations of methods. In addition, the improvement of DenseNet and AlexNet using combined group convolution and Ghost convolution, although slightly increasing the computation time for recognizing faulty data, the improved networks achieve 99.38% and 99.24% accuracy, respectively, which is a significant improvement over the original network recognition accuracy. Meanwhile, comparing the number of parameters and model size before and after the network improvement, it can be found that the combined group convolution and Ghost convolution instead of ordinary convolution can effectively reduce the number of parameters and model size of the network operation. The larger the number of parameters, the more fault features can be extracted from the network. In addition, the hybrid network identification method incorporating multi-head attention mechanism feature fusion has higher accuracy than the improved two networks, indicating that the multi-head attention mechanism feature fusion method is beneficial for categorizing early cable faults and identifying early cable faults and other cable overcurrent perturbation phenomena.

3) COMPARATIVE ANALYSIS OF FAULT IDENTIFICATION RESULTS OF DIFFERENT DIAGNOSTIC MODELS

In order to comprehensively evaluate the performance of the network model proposed in this study for early cable fault diagnosis, it is crucial to conduct a comparative analysis between different diagnostic models. This comparison not only highlights the advantages of the models in this paper but also provides insights for selecting the most suitable fault identification method for this application scenario. Specifically, we compare the network model in this paper with other algorithms. Five network models, ShufflenetV2, MobileNetV3, GoogleNet, Vggnet16, and ResNet18, are selected for comparison with this paper's method. This is because these five networks cover a wide range of network structures from lightweight to deep, while these network models have been extensively validated and applied in classification experiments. When the classification of different classification networks tends to a stable state, the classification accuracy results are compared as shown in Table 5 below.

According to the data in the table, it can be seen that the accuracy of this paper's method is much higher than that of the other five neural network models, and the loss value of the network is much lower than the loss value of the other methods, which indicates its excellent performance in the network classification process.

4) COMPARATIVE ANALYSIS OF DIFFERENT CLASSIFICATION METHODS

In order to further prove the superiority of this paper's method, this paper's method is compared and analyzed with other methods in the existing literature, as shown in Table 6 below.

The cable early fault classification includes seven cases of the cable 1/4 cycle early fault, the cable multi-cycle early fault, constant impedance grounding, transformer excitation inrush, motor switching, load switching, and capacitor switching. Literature [3] utilizes the Restricted Boltzmann machine (RBM) feature extraction method and Stacked Auto-Encoder (SAE) network to classify six fault cases. Literature [25] extracts energy, mean, and other features after wavelet transform and uses Convolutional Neural Networks (CNN) to classify five types of faults. In contrast, literature [26] utilizes DAE for feature extraction and combines it with CNN to classify six fault classifications. A comparative

TABLE 4. Classification performance of different combination models.

Combination	Dense Net	AlexNet	Group convolution and ghost convolution	multi-head Attention Mechanisms	Accuracy（%）	Calculation time (ms)	Total parameter（M）	Trainable parameter(M)	Model size（MB）
1	√				98.14	9.98	0.56	0.56	2.15
2		√			98.76	9.91	4.86	4.86	18.55
3	√	√		√	99.32	9.92	5.43	5.43	20.70
4	√		√		99.38	10.87	0.17	0.17	0.64
5		√	√		99.24	10.94	4.76	4.76	18.16
6	√	√	√	√	99.49	11.05	4.93	4.93	18.80

TABLE 5. Different neural network classification effects.

Different neural networks	Accuracy（%）	Loos
ShufflenetV2	89.45	0.2954
MobileNetV3	89.28	0.2770
GoogleNet	91.92	0.2336
Vggnet16	93.83	0.2289
ResNet18	98.38	0.1020
Methodology of this paper	99.49	0.0202

TABLE 6. Comparative analysis of different classification methods.

Identification methods	Feature extraction	Number of classifications	Accuracy（%）
SAE[3]	RBM	6	98.33
NCAE-DL[7]	SWT、Energy entropy, Information entropy	5	96.8
IPSO-SVM[9]	DAE	6	98.68
CNN[25]	Wavelet transform, Energy, Mean, RMS, Variance, etc.	5	97.9
IPSO-CNN[26]	DAE	6	99.05
Methodology of this paper	GAF、MTF	7	99.49

analysis of the different classification methods in Table 6 reveals that this paper's method has a higher classification accuracy for classifying seven cable early fault conditions.

VI. CONCLUSION

The study of cable early fault identification based on MTF-GAF and multi-head attention mechanism feature fusion proposed in this paper is based on preprocessing the raw fault data by AdaGrad-optimized LMS, then generating 2D feature images by using MTF and GAF methods, and finally diagnosing and classifying cable early faults by using improved hybrid neural network, which leads to the following conclusions:

(1) AdaGrad optimized adaptive filter is adopted to process the cable fault signal, this method can effectively reduce and eliminate extraneous noise in the signal, and compared with other optimization methods, AdaGrad optimized LMS demonstrates stronger noise resistance.

(2) In order to better preserve and analyze the information of early cable faults, MTF and GAF methods are used to transform one-dimensional fault data into 2D images. These two methods can reveal the intrinsic characteristics of fault data from different angles. By fusing the images generated by these two methods, a feature image containing rich fault information is obtained to realize cable early fault classification.

(3) A new convolutional structure is formed by combining the group convolution and Ghost convolution to replace the ordinary convolution in AlexNet and DenseNet, which can effectively reduce the computation and complexity of the model. The method of combining the constructed hybrid neural network with the multi-head attention mechanism effectively improves the recognition accuracy of early cable faults. At the same time, the hybrid neural network has a strong anti-interference ability and also has a better fault recognition effect compared with different network models.

The method proposed in this paper has a high recognition accuracy in processing cable early fault signals, but the adaptability of the method in the face of strong noise or other complex environments still needs to be improved. Also further validation of the methodology of this paper using actual field data is required later.

REFERENCES

[1] B. X. Du, Z. L. Li, and Z. R. Yang, "Application and research progress of high-voltage DC cross-linked polyethylene cables," *High Voltage Eng.*, vol. 43, no. 2, pp. 344–354, Feb. 2017.

[2] T. Ghanbari, "Kalman filter based incipient fault detection method for underground cables," *IET Gener., Transmiss. Distrib.*, vol. 9, no. 14, pp. 1988–1997, Nov. 2015.

[3] Y. Wang, H. Lu, X. Xiao, X. Yang, and W. Zhang, "Cable incipient fault identification using restricted Boltzmann machine and stacked autoencoder," *IET Gener., Transmiss. Distribution*, vol. 14, no. 7, pp. 1242–1250, Feb. 2020.

[4] Z. X. Zhou, X. Y. Xia, and P. Zhu, "Identification method for incipient intermittent arc ground fault of high-voltage cables," *Electr. Power*, vol. 53, no. 12, pp. 167–176, Aug. 2020.

[5] H. Samet, M. Tajdinian, S. Khaleghian, and T. Ghanbari, "A statistical-based criterion for incipient fault detection in underground power cables established on voltage waveform characteristics," *Electric Power Syst. Res.*, vol. 197, Aug. 2021, Art. no. 107303.
[6] H. Samet, S. Khaleghian, M. Tajdinian, T. Ghanbari, and V. Terzija, "A similarity-based framework for incipient fault detection in underground power cables," *Int. J. Electr. Power Energy Syst.*, vol. 133, Dec. 2021, Art. no. 107309.
[7] B. Z. Shao, S. H. Li, and X. Bai, "Application of nonnegative constraint autoencoder in cable incipient fault identification," *Power Syst. Protection Control*, vol. 47, no. 2, pp. 16–23, Jan. 2019.
[8] N. Liu, B. Fan, X. Xiao, and X. Yang, "Cable incipient fault identification with a sparse autoencoder and a deep belief network," *Energies*, vol. 12, no. 18, p. 3424, Sep. 2019.
[9] W. K. Wang and B. Deng, "Cable incipient fault identification method based on DAE-IPSO-SVM," *Foreign Electron. Meas. Technol.*, vol. 40, no. 8, pp. 29–35, Aug. 2021.
[10] S. Wang, J. Gong, and X. M. Yang, "An incipient cable failures identification method based on S-transform combined with mRMR feature selection," *Comput. Appl. Softw.*, vol. 39, no. 1, pp. 206–211&265, Jan. 2022.
[11] H. Lu, W.-H. Zhang, Y. Wang, and X.-Y. Xiao, "Cable incipient fault identification method using power disturbance waveform feature learning," *IEEE Access*, vol. 10, pp. 86078–86091, 2022.
[12] P. S. R. Diniz, "The least-mean-square (LMS) algorithm," in *Adaptive Filtering: Algorithms and Practical Implementation*. Germany, Nov. 2019, pp. 61–102.
[13] W. Zhang, X. Gu, and F. E. Liang, "ECG signal denoising using improved variable step size least mean square algorithm," *Chin. J. Med. Phys.*, vol. 40, no. 9, pp. 1135–1142, Sep. 2023.
[14] Z. Wang and T. Oates, "Encoding time series as images for visual inspection and classification using tiled convolutional neural networks," in *Proc. Workshops 29th AAAI Conf. Artif. Intell.* Menlo Park, CA, USA: AAAI, 2015, pp. 40–46.
[15] S. Qi, H. O. Shan, and L. Luo, "Application of deep feature learning with Gram's angle field for trace gas concentration identification," *Power Syst. Protection Control*, vol. 51, no. 15, pp. 55–65, Aug. 2023.
[16] L. Lu and Z. G. Wang, "Encoding temporal Markov dynamics in graph for time series visualization," Oct. 2016, *arXiv:1610.07273*.
[17] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 2, 2012, pp. 1097–1105.
[18] K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu, and C. Xu, "GhostNet: More features from cheap operations," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 1577–1586.
[19] J. D. Xue, H. Z. Ma, and H. S. Yang, "A fault diagnosis method for transformer winding looseness based on Gramian angular field and transfer learning-AlexNet," *Power Syst. Protection Control*, vol. 51, no. 24, pp. 154–163, Dec. 2023.
[20] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269.
[21] Y. Wang, G. Yang, S. Li, Y. Li, L. He, and D. Liu, "Arrhythmia classification algorithm based on multi-head self-attention mechanism," *Biomed. Signal Process. Control*, vol. 79, Jan. 2023, Art. no. 104206.
[22] W. Zhu, Z. T. Webb, K. Mao, and J. Romagnoli, "A deep learning approach for process data visualization using t-distributed stochastic neighbor embedding," *Ind. Eng. Chem. Res.*, vol. 58, no. 22, pp. 9564–9575, May 2019.
[23] Y. Xu, M. F. Guo, and B. Chen, "Research on modeling and simulation analysis of single-phase grounding arc in distribution networks," *Power Syst. Protection Control*, vol. 43, no. 7, pp. 57–64, Mar. 2015.
[24] S. R. Li, Y. D. Xue, and B. Y. Xu, "Modeling and overvoltage analysis of small current grounding fault arcs," *J. Electric Power Sci. Technol.*, vol. 34, no. 1, pp. 47–53, Mar. 2019.
[25] Y. Wang, J. F. Sun, and X. Y. Xiao, "Cable incipient fault classification and identification based on optimized convolution neural network," *Power Syst. Protection Control*, vol. 48, no. 7, pp. 10–18, Apr. 2020.
[26] Z. Xu, Z. Tang, T. Ji, and M. Li, "Detection and identification of underground cable incipient faults based on denoising autoencoder and optimized convolutional neural network," in *Proc. Int. Conf. Power Syst. Technol. (PowerCon)*, Sep. 2023, pp. 1–9.

HAO WU (Member, IEEE) received the B.S. degree in electrical engineering and automation from Southwest Jiaotong University, Chengdu, China, in 2003, the M.S. degree in pattern recognition and intelligent systems from Sichuan University of Science and Engineering, and the Ph.D. degree in power systems and automation from Southwest Jiaotong University. His research interests include big data, artificial intelligence, deep learning, image processing, modern signal processing and artificial intelligence technology in power systems, transmission (distribution) grid fault diagnosis and fault positioning technology, intelligent distribution grid protection, control and microgrid technology, and conditional intelligent monitoring technology for electrical equipment.

DAN TANG received the B.S. degree in automation and information engineering from Sichuan University of Science and Engineering, Zigong, China, in 2022. She is currently pursuing the master's degree with Sichuan Institute of Technology. Her research interests include signal processing and cable early fault diagnosis research.

YUAN CAI received the B.S. degree in electrical engineering and automation from Sichuan University of Science and Engineering, Yibin, China, in 2022, where he is currently pursuing the M.S. degree. His research interests include signal processing and photovoltaic power prediction studies.

CHAOWEN ZHENG received the B.S. degree in electrical engineering and automation from Sichuan University of Science and Engineering, Yibin, China, in 2022, where she is currently pursuing the M.S. degree. Her research interests include signal processing and fault diagnosis for flexible DC distribution grids.