J. COMBIN. MATH. COMBIN. COMPUT. 127b (2025) 1363--1381

Journal of Combinatorial Mathematics and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Combinatorial Press

Identification of Key Quality Defects and Intelligent Operation and Maintenance of Distribution Networks Based on Transformer Framework

Lijuan Yan[1], Ming Wang[2], Yan Zeng[2], Wensen Li[2], Yu Zou[2,✉]

[1] *Guangxi Power Grid Co., Ltd, Nanning, Guangxi, 530022, China*

[2] *Qinzhou Power Supply Bureau of Guangxi Power Grid Co., Ltd, Qinzhou, Guangxi, 535000, China*

ABSTRACT

In this paper, OpenCV technology is used to produce the distribution network defects dataset, which can be used as a training set, validation set, and test set in the ratio of 6:2:2. Combining the dataset and the Transformer framework, the S-Transformer based distribution network key quality defect identification model is constructed together. At this level, the degree of equipment deterioration is fitted, the distribution network intelligent operation and maintenance optimization strategy is formulated, and the experimental method is applied to evaluate the distribution defect identification and intelligent operation and maintenance. The identification rate of S-Transformer network for the six collected distribution network equipment defects is 0.9~0.95, which accurately controls the potential dangers, and is conducive to the subsequent intelligent equipment operation and maintenance of the distribution grid and its management and control, compared to the Compared with the traditional operation and maintenance program, the operation and maintenance program in this paper can reduce the operation and maintenance time by 52 hours per month, which greatly provides the efficiency of operation and maintenance labor.

Keywords: Transformer, distribution network, defect identification, intelligent operation and maintenance

1. Introduction

With the rapid development of economy and science and technology and the increasing needs of people's life, the requirements of users for power supply reliability are also increasing. As the last link of smart grid for users, distribution network is directly related to the reliability of smart grid power supply to end users [1]. The State Grid Corporation of China (SGC) has started to promote the further enhancement of the smart grid, and has applied the distribution automation system in a large number

✉ Corresponding author.
E-mail address: ZY_1106@outlook.com (Y. Zou).
Received 20 December 2024; Revised 05 February 2025; Accepted 25 March 2025; Published Online 16 April 2025.
DOI: 10.61091/jcmcc127b-075

of distribution grids. Therefore, it is the primary task to improve the power supply reliability of distribution grids, and to improve the ability of self-discovery and self-processing of faults in order to build smart distribution grids [2-3].

Smart distribution grid power equipment is characterized by a wide variety, large number, and difficult operation and maintenance [4]. With the continuous development of the power system and the deepening of the smart grid construction, the power data stored in the database of power grid enterprises shows explosive growth with the operation and maintenance of the smart grid. These data are usually stored in the form of non-institutionalized data such as images and text, and contain a large amount of information about the operating status of power grid equipment [5-7]. However, due to the large number of power equipment and the complexity of defect types, existing specifications cannot achieve a comprehensive summary [8]. In addition, due to the human-recorded characteristics of defective texts, phenomena such as text mutilation or even errors due to the inexperience of recorders occur from time to time, resulting in the complexity of defective text mining [9-11]. Therefore, error identification and quality improvement is one of the key technologies of power equipment defective text mining, and through the deep mining and identification of power equipment defective text, it can realize real-time monitoring of grid operation status, fault location and equipment maintenance, and provide guidance for the reliable operation of the power grid [12-15].

In order to deeply analyze large-scale textual information, most scholars use knowledge graph (KG) to realize the functions of retrieval and error checking in the process of distribution network operation and maintenance. Gao, Y. et al. proposed a deep learning-based entity recognition method for distribution network, which uses textual data of smart distribution network operation and maintenance scheduling to train the model, and it can effectively solve the recognition work of proper nouns in large-scale distribution network operation and maintenance data , which is of guiding significance for the construction of knowledge graph of smart distribution grid [16]. Zhang, L. et al. design a named entity algorithm that combines the attention mechanism and word-word joint embedding vectors for the large-scale operation and maintenance text data generated by the smart distribution grid, and experiments show that the algorithm can significantly improve the recognition accuracy of fault location and fault type of the distribution grid, and provide support for the construction of the knowledge graph in the field of operation and maintenance of the smart grid [17]. Liu, W. et al. constructed an automated defect knowledge graph for distribution networks and used named entity recognition model and relationship extraction model for distribution network diagnostic data extraction and analysis. The automated defect knowledge graph effectively solves the various problems existing in manual recording of defects and promotes the development of intelligent operation and maintenance of distribution networks [18]. Li, K. et al. designed a knowledge graph-based multimodal neural network (KGMNN) approach, which can effectively improve the defect detection performance of distribution networks by integrating multi-source data in distribution network operation and maintenance into a knowledge graph and interpreting it using KGMNN, which is of practical value for improving the reliability and operational efficiency of power grids [19]. Zhu, L. et al. examined the application of short text mining techniques in power defect text recognition and power equipment management and the designed short text mining architecture effectively enhances the operation of smart distribution grids for power text and management activities of power equipment [20]. Zhao, Z. et al. combined a bidirectional long and short-term memory neural network and a conditional random field to construct an automatic feature recognition model for electric power entities, and input additional features from distribution network operation and maintenance data into the neural network for training, which significantly improved the efficiency of named entity recognition in the electric power domain [21].

Knowledge graph can effectively show the relevant connections between knowledge, and constructing the knowledge graph of power equipment defects based on distribution network operation and maintenance data can realize the data retrieval and visualization functions required by the auxiliary entry system, but the knowledge graph has very high requirements for the accuracy of the relationship extraction, which may result in the misidentification in the face of diversified as well as complex operation scenarios. If deep learning frameworks such as Transformer are constructed, the error recognition effect of defective text can be further improved, which is also a possible direction for defect recognition in distribution network operation and maintenance in the future.

The article produces the dataset through three steps of data labeling, data augmentation, and division of the dataset, extracts local features from the dataset by using the residual U-Net, and combines local features with multiscale depth features through the activation function ReLU to achieve better preservation of structural information. After extracting the multimodal features using the U-Net network model, the S-Transformer is constructed using the intra-domain self-attention unit and inter-domain cross-attention mechanism, and these features are fused to finalize the modeling of the key quality defect identification model of the distribution network. The model parameters are determined and the model is applied to identify and analyze the dataset. Taking the results of equipment defect identification as an entry point, the relative deterioration degree values of the equipment are fitted, and more intelligent operation and maintenance strategies are formulated in order to promote the intelligent development of high quality and high level of distribution networks.

2. Transformer-based Distribution Network Defect Recognition

2.1. Transformer Theory

2.1.1. Overview of the Transformer network. Transformer is mainly composed of an encoder and a decoder, which is specifically composed of a self-attentive layer and a fully-connected layer, and its network structure is shown in Fig. 1, with the encoding part on the left and the decoding part on the right. After inputting the data into the network, the first step is to carry out the use of input embedding to adjust the data information, and then add it with the position encoding information, so that the sequence contains the relative position information between the data [22-23]. Then the data is fed into the encoding part on the left side, the encoder mainly consists of a multi-head self-attention layer, a normalization layer, a feed-forward layer, and a residual connection, and multiple encoders are stacked for encoding; on the right side is the decoding part, and the internal structure is similar to that of the encoding part. Finally, a linear layer with Softmax classification layer is used for output.

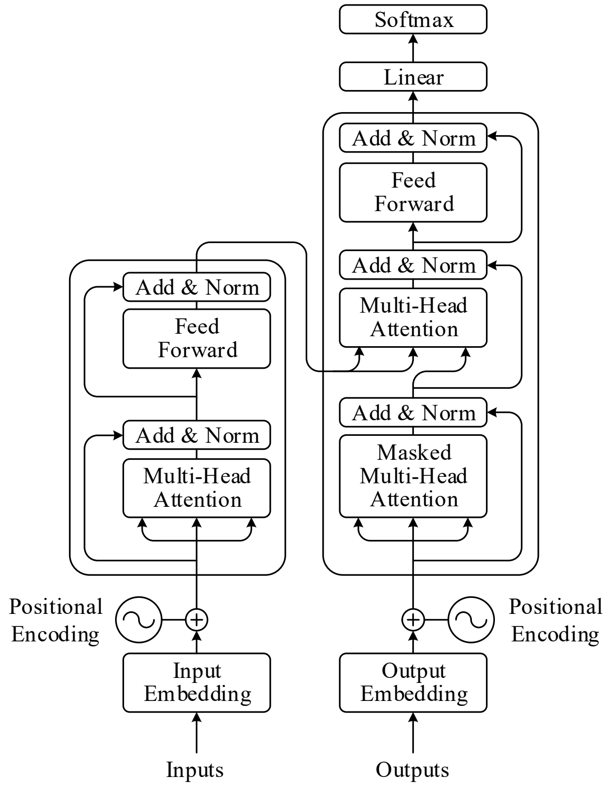


Fig. 1. Transformer network structure diagram

ViT (Vision Transformer), which uses Transformer for image recognition, works basically similarly, and the network structure of ViT is shown in Fig. 2. Firstly, the two-dimensional image is first divided into multiple subgraphs, then the two-dimensional subgraphs are linearly mapped into one-dimensional vectors, and then the trainable position embedding and category information are added; then the processed one-dimensional vectors are input into the Transformer to extract features [24-26]. Finally, after encoding by the encoder, one-dimensional feature vectors are output, and the category vectors are obtained after processing using MLP to predict the image categories. Since the image is encoded by the encoder, the output information can be used to predict the image information by using for Patchs, the encoder is mainly used for regression tasks, and the use of the decoder module will lead to excessive computational cost, therefore, in ViT and the subsequent improvement of Transformer used for visual recognition tasks, only the encoding part is used, and in this paper, we follow the same practice and use only the encoder part structure.

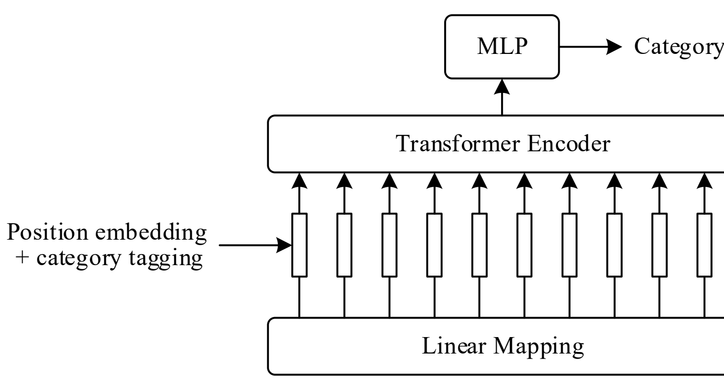


Fig. 2. Vision Transformer network structure diagram

2.1.2. Transformer Network Architecture:

1) Embedding layer. Since the Transformer is used for one-dimensional sequence processing and the input picture is two-dimensional, the picture first needs to be converted into one-dimensional data. Assuming that the size of the image is $H \times W$ and the size of the subgraph after division is $p \times p$, the image can be divided into $(H/p) \times (W/p)$ subgraphs, and then the subgraphs are linearized to obtain one-dimensional vectors. It is also necessary to embed the category information and location information of the image, both of which are parameters that can be trained and randomly initialized when embedding. The final processed feature vector is input to the encoder.

2) Encoder. The embedded subgraph is fed into the encoder, the encoder module mainly consists of normalization layer, multi-attention layer, MLP layer, and it contains two residual connections to avoid the gradient vanishing problem. And multiple encoder modules are stacked to form the encoder:

(1) Layer normalization layer. Layer Normalization (LN) is used in Transformer, the LN layer is mainly used in Natural Language Processing, and its formula is similar to the BN layer, but they act in different dimensions. The BN layer is normalized for each channel of each batch of data. The LN layer has nothing to do with the batch of data being processed, but normalizes the specified dimensions of the data, which solves the problem of the BN layer's poor results when the batch is small.

(2) Multiple head self-attention layers. The head self-attention layer is stacked by multiple self-attention layers. The picture becomes N+1 one-dimensional feature vectors after passing through the Embedding layer, and the self-attention mechanism needs to be made to establish the connection between different parts of the picture to capture the internal information of the picture. For example, for the input N one-dimensional vectors represent N sub-pictures on the picture, then the role of the self-attention mechanism is to establish the connection between each sub-picture and other N-1 sub-pictures, and the formula of the self-attention mechanism is:

$$Attention(Q,K,V) = SoftMax\left(\frac{QK^T}{\sqrt{d}}\right)V \tag{1}$$

Where: Q represents the query $(Query)$ matrix, K represents the key (Key) matrix, and V represents the value $(Value)$ matrix. Multiplying Q and K matrices can calculate the correlation between two subgraphs, V represents the information extracted from the feature vector, and the larger value obtained by multiplying Q and K matrices means that the two subgraphs are more correlated, and the more weight is given to V. However, the value obtained when calculating the correlation may be very large, resulting in a small change in the gradient of SoftMax, which is not conducive to the training of the network, so after calculating the correlation, it is also necessary to divide by $\sqrt{d}$, where d represents the length of the vector in K. Q, K, and V are generated after three transformation matrices, i.e., W_q, W_k, and W_v, and the values of these three transformation matrices can be learned during network training. The formulae for Q, K, V are:

$$Q = XW_q, K = XW_k, V = XW_v \tag{2}$$

Where: x is the feature matrix after Embedding.

The multi-head attention mechanism is to divide Q, K, V three matrices into h attention heads on average, each attention head is independent of each other, respectively, to carry out self-attention computation, and finally the results obtained from each head computation are stacked, and the output of the stacked feature layer. The computation process of the multi-head attention mechanism is shown in Fig. 3. Relative to the ordinary self-attention mechanism, the multi-head attention mechanism divides the input feature information into multiple subspaces and performs linear transformation within each subspace, which enhances the attention information within the subspace, and facilitates the network to extract different information and improve the model fitting ability.

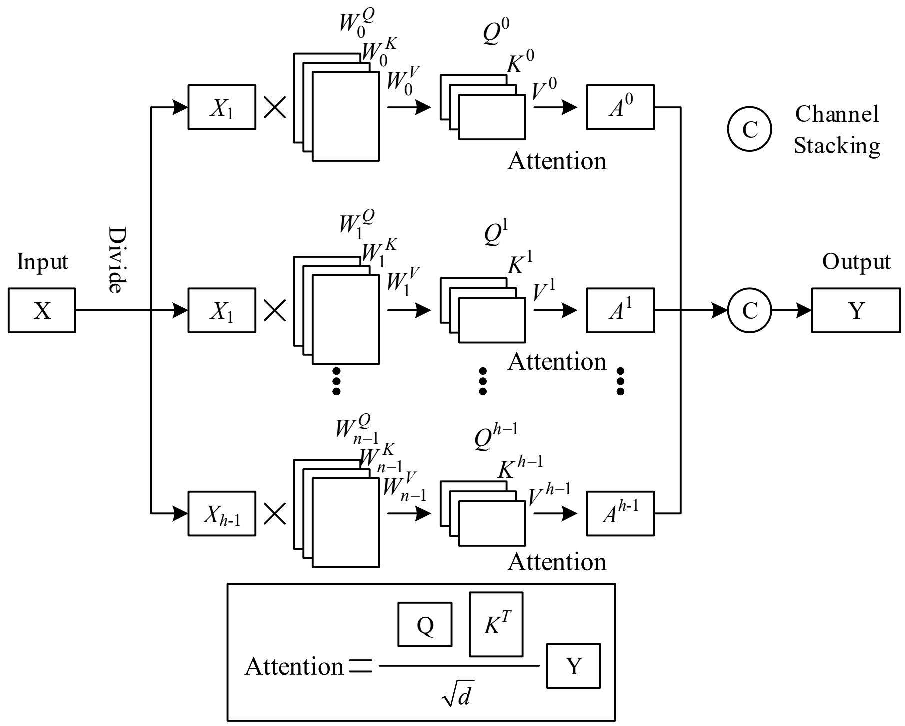


Fig. 3. Schematic diagram of the multi-head attention mechanism

3) MLP (Multilayer Perceptron) Layer. The structure of the MLP layer is relatively simple, and it mainly consists of a Linear transformation layer (Linear), a Dropout layer (Dropout) and a GELU activation function layer. The main role of the MLP layer is to further improve the feature extraction ability of the network, extract data features and adjust the data dimensions through the linear transformation layer; and improve the generalization ability and fitting ability of the network through the regularization operation of the network by the abstention layer and the GELU activation function layer to improve the performance of the network. It is mainly used in the field of NLP, which can improve the nonlinear fitting ability and generalization ability of the network, and improve the model training robustness.
4) Linear transformation layer with Softmax function. After the data is encoded by the encoder, it is output to the linear transformation layer, which is the fully connected neural network layer, to perform dimensionality reduction on the data. In the image classification problem, assuming that the image category is C, the dimensionality reduced data is turned into a C-dimensional vector, and then the probability that a value in the vector corresponds to a label is calculated using the Softmax classifier.The Softmax function is given by:

$$S(x) = \frac{1}{1+e^{-x}} \tag{3}$$

The above equation shows that the Softmax function can restrict the values in the vector to between $(0,1)$. By comparing the category labels, the category of the image containing the object can be predicted.

2.2. *Data set production*

The dataset must fulfill the following two requirements. When there is insufficient data, the model is prone to overfitting problems. A large amount of data training can make the model generalization better, especially the deep network model, which needs a large number of samples for training in order to exert the expression ability of the model. Second, the samples should be balanced. Large differences in the sample size of various types of features can cause serious bias in the model, and the classification results vary greatly in accuracy, which seriously affects the model effect. In the study, data sets are created through three steps: data labeling, data augmentation, and data set division.

2.2.1. Data tagging. Distribution network defects are categorized into two main groups: functional and structural defects. Functional defects are defects that affect the distribution network by changing the overwater section, such as faulty short circuits X1, line breaks X2, phase-to-phase short circuits, etc.X3 Structural defects are caused by damage to the structure, such as weak architectures X4, low level of automation X5, and poor basic data X6, etc., and the long-term existence of these defects can cause extremely serious damage to the pipeline.

In the study, OpenCV was used for category labeling and 130 image sequences were obtained. Each sequence consists of 8 consecutive frames, totaling 1040 images. Combined with the distribution network inspection realities and data characteristics, 936 defect-containing sequences (156 each of faulty short-circuit X1, line break X2, phase-to-phase short-circuit X3, weak architecture X4, low automation level X5, and poor basic data X6) and 464 defect-free sequences were selected.

2.2.2. Data enhancement. For the Transformer model, a sufficient amount of data can both avoid overfitting during model training and enrich the network feature learning to move the model towards the global optimum. Insufficient data will lead to overfitting problems in the model network, thus making the model recognition ability poor and accuracy low. However, data is not easy to obtain and the collection cost is high, which makes it difficult to directly collect a large amount of data for training in practical research. Data augmentation is a method that changes the geometric features of the existing data to augment the data, and is widely used in the production of various types of data sets. For example, in the well-known MINIST handwritten digit picture dataset, the researcher augmented it into a dataset including 80,000 training pictures and 16,000 test pictures for training, and the recognition error rate was reduced from 0.9623% to 0.9213%. In order to improve the generalization ability of the model, this paper uses data augmentation to obtain more data, and after rotating the original data by 90°, 180°, and 270° through OpenCV, the images are mirror transformed, scaled, and finally 1040 224×224 images are obtained.

2.2.3. Data set partitioning. Dividing the data according to the appropriate ratio is the key to guarantee the sample balance. In this study, the distribution network dataset is named D, which contains three sets: training set, validation set and test set. The training set accomplishes model training. The validation set is used for model parameter correction and monitoring whether overfitting occurs in the model. The test set is used for model generalization ability evaluation, which can check the model learning effect. After experimental verification, in the distribution network pipeline dataset, the ratio of the three is maintained at about 6:2:2 is most appropriate, so the settings of the training set, validation set and test set are shown in Table 1.

Table 1. Data set partitioning

Data set	Defective	No defect
Training set	600	280
Validation set	187	93
Test set	186	94

2.3. Defect Recognition Model Based on Transformer Framework

2.3.1. General framework. Defect identification of grid devices based on a multimodal Transformer model the model will be constructed separately as a multimodal feature extractor to extract features from data of different modalities. The Transformer-based multimodal model fusion method proposed in this paper consists of three main parts: global feature extraction, feature fusion and feature reconstruction. The model interactively fuses the information by encoding data from different modalities and uses Transformer to process the sequence data for defect recognition.

2.3.2. Feature extraction. RUB can extract more global information from shallow, high-resolution feature maps. The residual U-Net block can be defined as $RUB-L(C_{in},M,C_{out})$, where L is the number of U-Net layers, C_{in} and C_{out} denote the input and output channels, respectively, and M denotes the number of internal channels. The structure of the U-Net network is shown in Fig. 4, and the composition of the RUB can be divided into 3 parts:

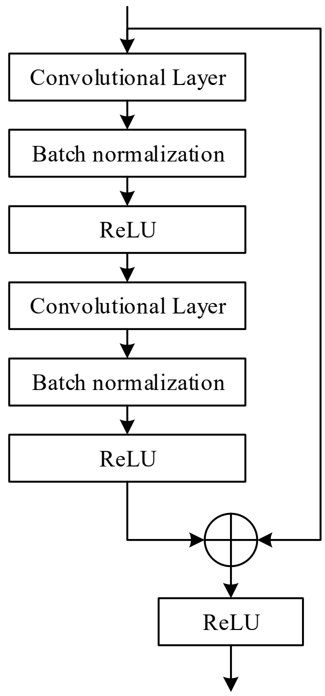


Fig. 4. U-Net Network Model Structure

Input layer: performs local feature extraction and transforms the input feature map $x(H\times W\times C_{in})$ into intermediate features $F_1(x)$ with C_{out} channels. Convolution kernel size is 3×3 and the activation function used is ReLU. U-Net encoder and decoder: the structure is similar to U-Net, encoding multiscale contextual information from $F_1(x)$ and acquiring decoder features $\mu(F_1(x))$.

Residual Connection: combines local features with multiscale depth features for better preservation of structural information. It is specifically represented as:

$$G_1(x)=F_1(x)+\mu(F_1(x)) \tag{4}$$

where G_1 is the output of RUB.

The comparison between the normal residual block and RUB is shown in Fig. 5. The extraction operation in the normal residual block usually consists of one or more convolutional layers and can be represented as:

$$G_2(x)=F_2(F_1(x))+F_1(x) \tag{5}$$

where G_2 is the expected output.The main difference between RUB and the normal residual block is that RUB uses a U-Net structure instead of single-stream convolution. This allows the network to directly extract global information with very low computational overhead. In view of this, in this paper, RUB is used as an encoder for the proposed method to thoroughly extract the essential features at each scale.

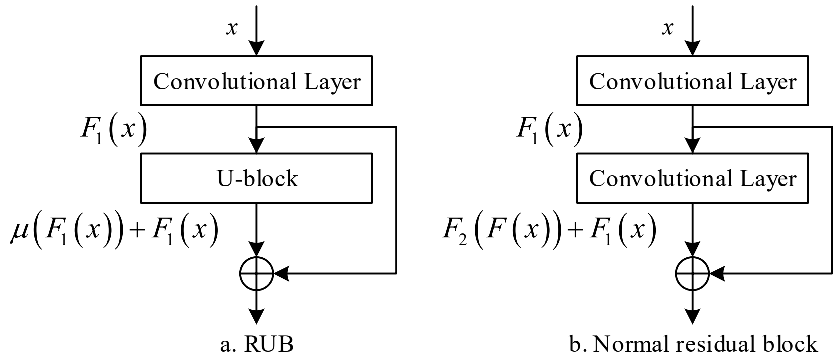


Fig. 5. Com parison of RUB and Ordinary Residual Blocks

2.3.3. Feature Fusion. After extracting multimodal features using the U-Net network model, the S-Transformer was constructed using intra-domain self-attention units and inter-domain cross-attention mechanisms to fuse these features. First, the multi-head self-attention (MSA) mechanism considers the global feature distribution, which helps the model to capture information from multiple coding subspaces. Then, in order to optimize the MSA-generated feature markers, a feed-forward network (FFN) consisting of two multilayer perceptron (MLP) layers and one GELU activation function layer is applied. Subsequently, layer normalization (LN) is executed after the MSA and FFN runs. Finally, the residuals are deployed after processing in both modules. The process of in-domain self-attention can be expressed as follows:

$$\{Q,K,V\}=\left\{XW^{Q},XW^{\kappa},XW^{V}\right\} \tag{6}$$

$$O = Attention\left(Q,K,V\right) \tag{7}$$

After intra-domain perception, inter-domain cross-attention interaction is realized to further explore the common information between different domains The basic modules of inter-domain perception and intra-domain perception are similar, the main difference is that inter-domain perception employs the mechanism of Multi-head Cross-attention (MCA) to realize the interaction of global content. The whole process of cross-domain cross-attention interaction can be defined as follows:

$$\left\{Q_1,K_1,V_1\right\}=\left\{X_1W_1^{Q},X_1W_1^{\kappa},X_1W_1^{v}\right\} \tag{8}$$

$$\left\{Q_2,K_2,V_2\right\}=\left\{X_2W_2^{Q},X_2W_2^{K},X_2W_2^{V}\right\} \tag{9}$$

$$O_1 = Attention\left(Q_1,K_2,V_2\right) \tag{10}$$

$$O_2 = Attention\left(Q_2,K_1,V_1\right) \tag{11}$$

where: $\left\{Q_1,K_1,V_1\right\}$ queries features in domain 2 that are similar to Q_1 in domain 1; similarly, $\left\{Q_2,K_1,V_1\right\}$ queries features in domain 1 that are similar to Q_2 in domain 2. After the S-Transformer processing, a convolutional layer is employed to integrate local information from different domains. Ultimately, intra- and inter-domain hybrid attention can be cascaded for effective fusion of complementary multimodal features.

2.3.4. Feature reconstruction. For global features $\Phi_{GF}^{ir}(i,j)$ and $\Phi_{GF}^{vis}(i,j)$ (where the infrared image is denoted as ir and the visible image is denoted as vis), first, the weights of the row vectors are computed by the L1 paradigm, and then the Softmax function is used to obtain the activity levels $\varphi_{row}^{ir}(i)$ and $\varphi row_{row}^{vis}(i)$. This process can be represented by the following equation:

$$\varphi_{row}^{ir}(i)=\frac{\exp\left(\left\|\Phi_{GF}^{ir}(i)\right\|\right)}{\exp\left(\left\|\Phi_{GF}^{ir}(i)\right\|_1\right)+\exp\left(\left\|\Phi_{GF}^{vis}(i)\right\|_1\right)} \tag{12}$$

$$\varphi_{row}^{vis}(i)=\frac{\exp\left(\left\|\Phi_{GF}^{vis}(i)\right\|\right)}{\exp\left(\left\|\Phi_{GF}^{ir}(i)\right\|_1\right)+\exp\left(\left\|\Phi_{GF}^{vis}(i)\right\|_1\right)} \tag{13}$$

where: $\|\cdot\|_1$ denotes the L1 paradigm computation.

Then, the fused global feature is obtained from the row vector dimension by directly multiplying its activity level with the corresponding global feature, called $\Phi_F^{Fow}(i,j)$. This process can be represented by the following equation:

$$\Phi_F^{row}(i,j)=\Phi_{GF}^{ir}(i,j)\times\varphi_{row}^{ir}(i)+\Phi_{GF}^{vis}(i,j)\times\varphi_{row}^{vis}(i) \tag{14}$$

Subsequently, the activity levels of the column vectors were measured to obtain $\varphi_{cd}^{ir}(i)$ and $\varphi_{col}^{vis}(i)$. Next, the fused global features of the column vectors were extracted $\Phi_F^{col}(i,j)$.

Finally, the features of the row and column vectors were fused by an element-by-element addition operation to obtain the final fused global feature, which was used to reconstruct the fused image through the convolutional layer.

$$\Phi_F(i,j)=\Phi_F^{row}(i,j)+\Phi_F^{col}(i,j) \tag{15}$$

3. Distribution Network Defect Identification and Intelligent Operation and Maintenance Strategy

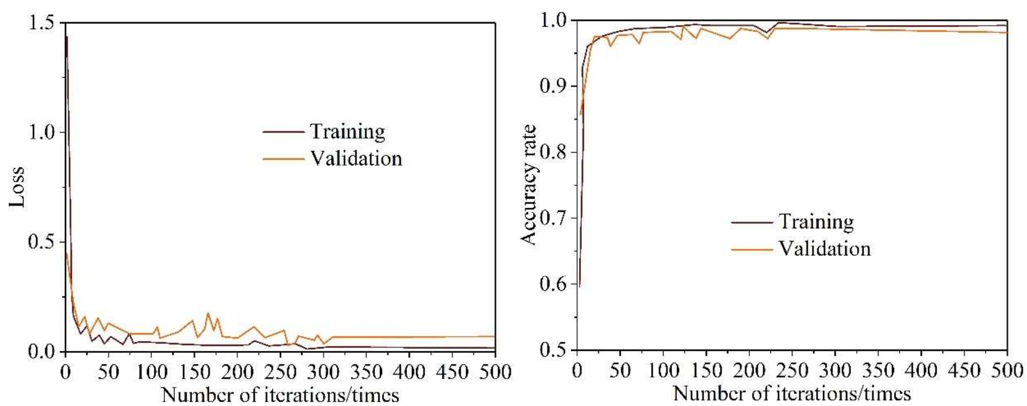
3.1. Distribution network defect identification analysis

3.1.1. Model parameterization. The structure and parameters of the S-Transformer neural network model are shown in Table 2. The S-Transformer neural network is used as the main parameter of the feature extraction network, and the number of iterations is set to be 500, the batch size to be 64, the learning rate to be 0.001, the input size to be 128*128, the number of encoder layers to be 6, the hidden layer to be 512 dimensions, the polytope to be 8, and the dimension of the matrix to be 64 (512/8).

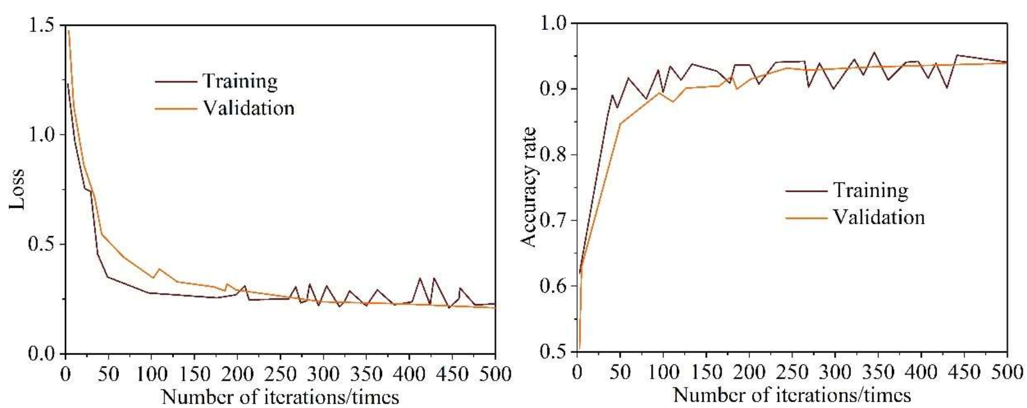
Table 2. S-Transformer neural network model structure and parameters

Sequence	Network layer	Output size	Standardization
1	Input layer	128×128×1	
2	Embedding layer	128×128×32	
3	Position coding layer	128×128×32	
4	Encoder layer	128×128×1	Yes
5	Encoder layer	128×128×1	Yes
6	Encoder layer	128×128×1	Yes
7	Encoder layer	128×128×1	Yes

Feature extraction is the second step of the attribute description-based defect recognition method for distribution networks. The complete S-Transformer model consists of an encoder and a decoder, and considering the characteristics of the pipeline defect recognition task, only a continuous representation of the one-dimensional time-series signal needs to be computed to obtain the features of different defect types. So only the encoder of the S-Transformer model is used to extract the features. The model consists of an input signal matrix, a position encoding, a Transformer feature extraction layer, and a fully connected layer. In order to verify the ability of S-Transformer neural network to extract features, Transformer and S-Transformer models are used to set up comparative experiments respectively, and to ensure the credibility of the experiments, all algorithms use the same input data. When setting up the comparison experiments, the relevant parameters of the selected comparison algorithms need to be set. Set up three pairs of convolutional and pooling layers, 3 layers in parallel, the size of the convolutional kernel is 32-64-128, and the size is 3-57, respectively, and set up the Transformer model as a comparative experiment of the S-Transformer model, and the classification accuracy plots and the loss function of the training set and the test set are shown in Fig. 6, in which (a) to (d) are respectively the classification accuracy plots and the loss function of the S Transformer and Transformer's loss function, accuracy. For S-Transformer, in the training stage, when the iteration reaches 30 times, the classification recognition accuracy of the model on the training set has reached 98.72%, and the Loss value has dropped to 0.119, after which the classification accuracy and Loss value of the model training reaches 121 iterations, and the change trend is gradually flattened. However, as shown in Fig. 6(c) and Fig. 6(d) for the Transformer model, in the testing phase, when the 30th iteration is carried out, the classification accuracy of the model can only reach 89.9%, and the Loss value has dropped to 0.327, after which the accuracy and the Loss value fluctuate substantially. It indicates that the improved Transformer model establishes the connection between defective features of pipes of different shapes in a global view parallel manner, has a better ability to capture features at a distance from the input time series, and uses a self-attention mechanism to distinguish the degree of contribution of different samples to the results.



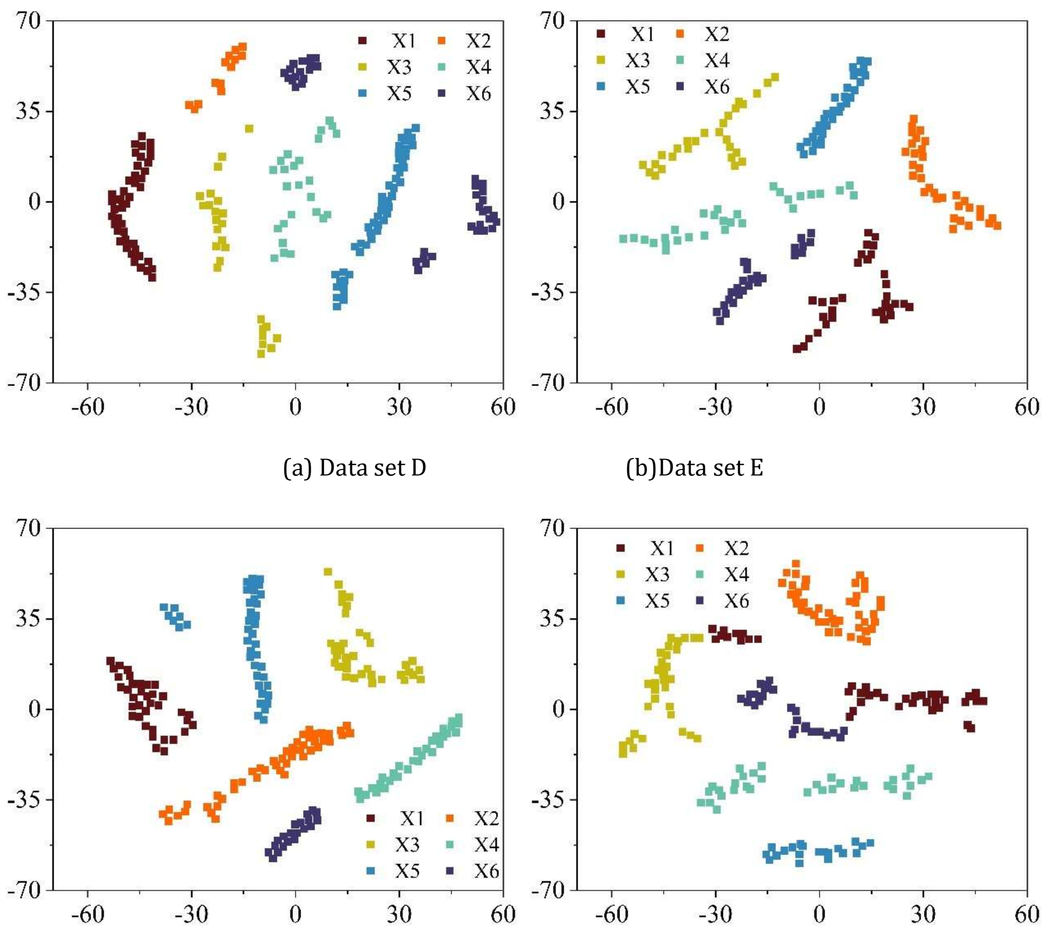
(a)Loss function of the S-Transformer (b)Accuracy of S-Transformer



(c)Loss function of the Transformer (d)Accuracy of Transformer

Fig. 6. Identify accuracy and loss function graphs

3.1.2. Data analysis. In order to avoid too few datasets resulting in less credibility of the results, this paper adds three datasets E, F and G on the basis of the above dataset D. Figure 7 shows the visualization results of the output features of the defects of the distribution network, in which (a)~(d) are the datasets D~G, respectively. The Transformer model of the six types of defects in the dataset D, dataset E, dataset F and dataset G (X1, X2, X3, X4, X5, X6) output features of the visualization results, the figure between the various types of samples can be divided into very good, the features are not confusing, the classification boundary is clear, with a good aggregation, the features are more centralized and can be divided into good. It shows that the multi-head self-attention mechanism of the S-Transformer neural network model encoder adopts a parallel approach to defect feature extraction, which improves the defect recognition accuracy and solves the problem that the single-layer convolution can not capture the long-range features.

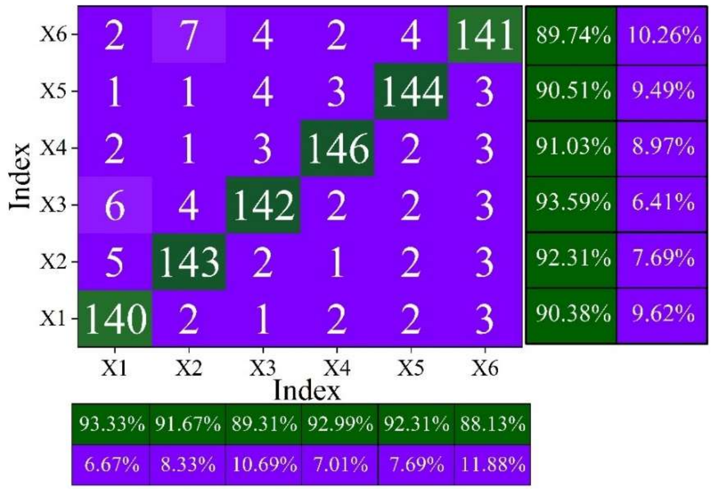


(a) Data set D (b)Data set E

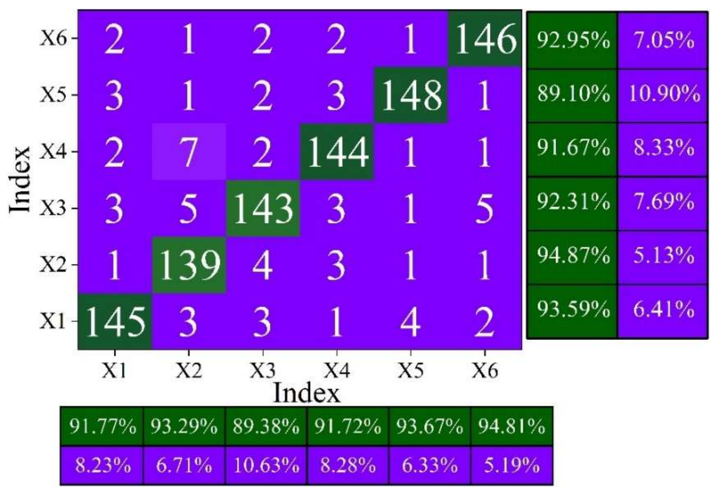
(c)Data set F (d)Data set G

Fig. 7. Visualization results of distribution network defect output characteristics

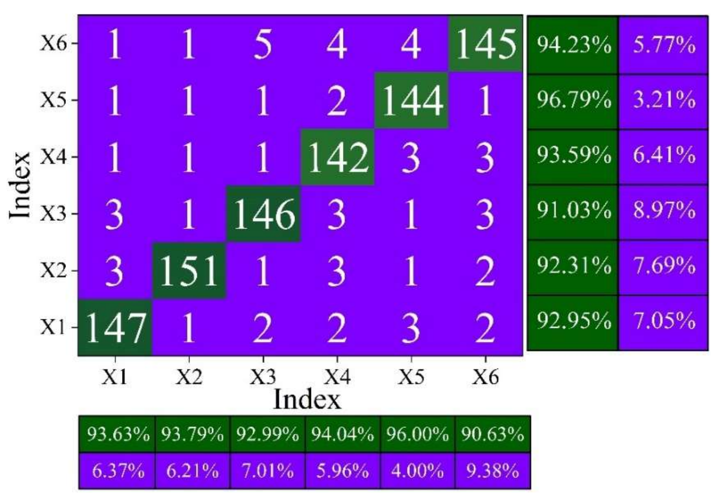
Fig. 8 demonstrates the confusion matrix for distribution network defect identification using U-Net network as an attribute learner after feature extraction using S-Transformer for datasets D, E, F, and G. The confusion matrix is shown in Fig. 8. The vertical coordinates represent the actual labels of the test samples, the horizontal coordinates represent the predicted labels predicted by the model, and the intersection of the horizontal and vertical coordinates represent the predicted accuracy, which is verified by setting up different datasets to verify the accuracy of the distribution network defect recognition based on the S-Transformer network. It can be seen that the average value of S-Transformer network's recognition rate for all six defects in distribution networks is greater than 0.9, which accurately detects the potential dangers in distribution networks and facilitates the intelligent remote maintenance and control of distribution networks.



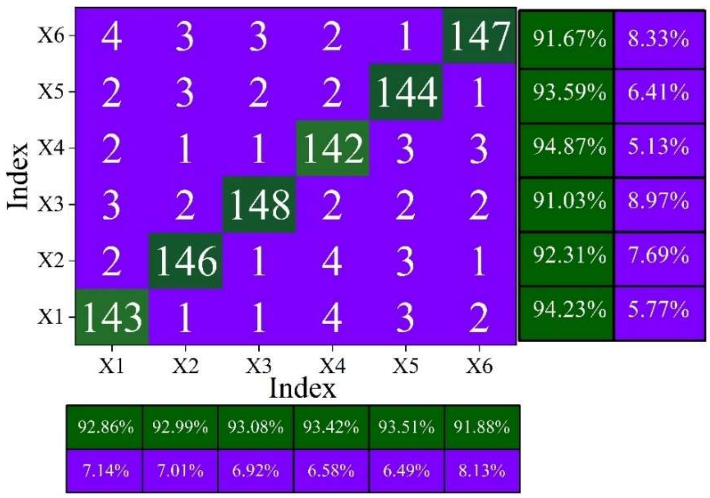
(a)Data set D



(b)Data set E



(c)Data set F



(d)Data set G

Fig. 8. Confusion matrix of distribution network defect identification

3.2. Intelligent O&M Optimization Analysis

3.2.1. Basic idea. By detecting the problem of equipment operation status through the defect identification mentioned above, compared with the fixed time period for the patrol of distribution network equipment, the operation and maintenance personnel are more willing to grasp the real-time operation information of the substation operation equipment, and then formulate a reasonable patrol cycle and patrol plan. When the equipment is in operation, the degree of deterioration of the equipment can be described by the linear fitting method, and then issue early warning information. As a key link in the optimization process of equipment operation and maintenance strategy, it is to issue early warning information based on equipment information. Firstly, the relative deterioration degree of the equipment is calculated based on the historical data of the equipment, and secondly, the deterioration degree is fitted by matlab, and the residual life of the equipment as well as the patrol cycle are calculated, and the predicted value of the equipment elaboration is used to judge whether to issue the warning information. In order to further improve the efficiency and the accuracy of obtaining the shortest path for substation inspection, O&M personnel usually utilize linear fitting to realize the inspection scheme. First create an electronic map of the substation, obtain the distribution map of primary equipment and label the distance between primary equipment. Input the input points, path variables and other parameters into the matlab software, and output the optimal inspection route through the shortest path method.

3.2.2. Optimization program evaluation. The relative deterioration of equipment parameters is calculated as shown in equation (16):

$$x_i = \frac{X_i - X_0}{X_c - X0} \tag{16}$$

Where X_i is the value of the characteristic parameter of the equipment for the i nd time, X_0 is the initial value of the characteristic parameter, and X_c is the threshold value of the characteristic parameter.

When the relative deterioration of the equipment reaches 1, it indicates that the equipment is in urgent need of inspection. Therefore, when the equation of relative deterioration degree and relative time is obtained, the relative life time of the equipment can be found:

$$t_r = \frac{-b+\sqrt{b^2-4a(c-1)}}{2a} \tag{17}$$

where t_r is the relative lifetime of the equipment. Therefore, the time from the current time to the next tour of the equipment is:

$$T_r = \frac{-b+\sqrt{b^2-4a(c-1)}}{2a} - t_m \tag{18}$$

where t_m is the current operating time of the equipment.

The capacitor relative deterioration fitting curve is shown in Fig. 9, and the distribution network equipment status is shown in Table 3. From the main transformer relative degradation fitting curve and capacitor relative degradation fitting curve, it can be known that the main transformer relative degradation value is 0.87, and the predicted time for the next inspection is 0.6 days later. The capacitor relative deterioration value is 0.33 and the predicted next inspection time is 6 days later. In addition, the larger the relative deterioration value, the more urgent the equipment needs to be overhauled.

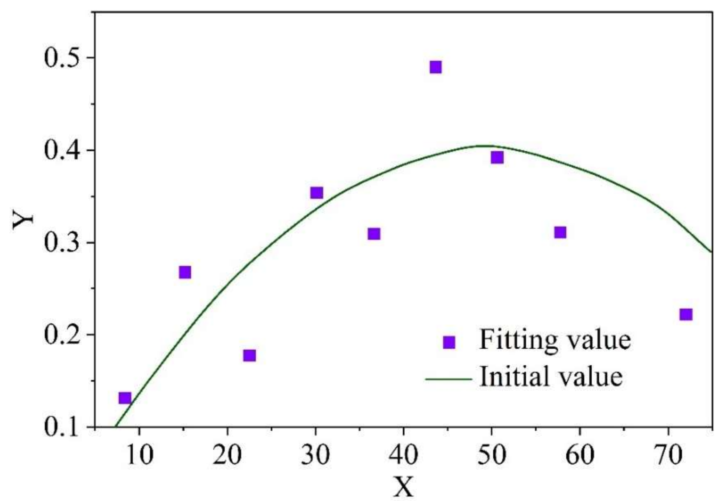


Fig. 9. Capacitor relative deterioration fitting curve

Table 3. Distribution network device status

Relative deterioration value	Device status	Inspection cycle
0.8~1	Urgent inspection	Reference forecast cycle
0.6~0.8	Reliability degradation	Reference forecast cycle
0.4~0.6	Need to step up inspections	Reference forecast cycle
0.2~0.4	General	Make regular rounds
0~0.2	Good	Normal inspection, can be appropriately increased inspection cycle

The traditional substation O&M program is that at least 8 inspections need to be scheduled per month, and the monthly inspection time is about 250 hours. The intelligent O&M roadmap for distribution networks is shown in Figure 10, with V1-V11 being switches and circuit breakers, transformers, capacitors, electronic transformers, integrated power supplies, protection devices, consolidation units, intelligent terminals, switches and other equipment.

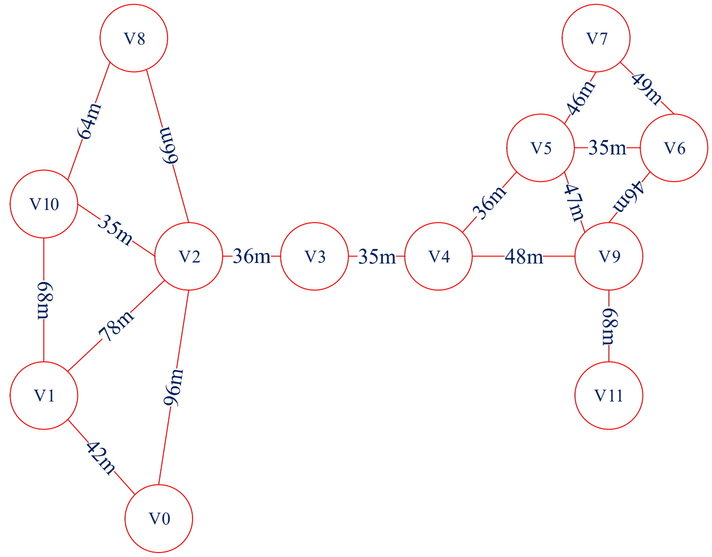


Fig. 10. Intelligent operation and maintenance roadmap of distribution network

According to the intelligent operation and maintenance route of the distribution network in Fig. 10. Matrix Q can be constructed. As shown below:

$$Q=\begin{bmatrix}
0 & 42 & 96 & \infty & \infty & \infty & \infty & \infty & \infty & \infty & \infty & \infty \\
42 & 0 & 78 & \infty & \infty & \infty & \infty & \infty & \infty & \infty & 68 & \infty \\
96 & 78 & 0 & 36 & \infty & \infty & \infty & \infty & 66 & \infty & 35 & \infty \\
\infty & \infty & 36 & 0 & 35 & \infty & \infty & \infty & \infty & \infty & \infty & \infty \\
\infty & \infty & \infty & 35 & 0 & 36 & \infty & \infty & \infty & 48 & \infty & \infty \\
\infty & \infty & \infty & \infty & 36 & 0 & 35 & 46 & \infty & 47 & \infty & \infty \\
\infty & \infty & \infty & \infty & \infty & 35 & 0 & 49 & \infty & 46 & \infty & \infty \\
\infty & \infty & \infty & \infty & \infty & 46 & 49 & 0 & \infty & \infty & \infty & \infty \\
\infty & \infty & 66 & \infty & \infty & \infty & \infty & \infty & 0 & \infty & 49 & \infty \\
\infty & \infty & \infty & \infty & 48 & 47 & 46 & \infty & \infty & 0 & \infty & 68 \\
\infty & 68 & 35 & \infty & \infty & \infty & \infty & \infty & 49 & \infty & 0 & \infty \\
\infty & \infty & \infty & \infty & \infty & \infty & \infty & \infty & \infty & 68 & \infty & 0
\end{bmatrix}$$

After 12 iterations, the shortest path is output:

$$R(12)=\begin{bmatrix}
0 & 2 & 2 & 6 & 8 & 10 & 6 & 6 & 16 & 10 & 4 & 5 \\
4 & 0 & 4 & 6 & 8 & 10 & 6 & 6 & 16 & 10 & 4 & 5 \\
6 & 6 & 0 & 6 & 8 & 10 & 6 & 6 & 6 & 10 & 6 & 5 \\
6 & 6 & 8 & 0 & 8 & 10 & 6 & 6 & 6 & 10 & 6 & 5 \\
6 & 6 & 8 & 10 & 0 & 10 & 6 & 6 & 6 & 10 & 6 & 5 \\
6 & 6 & 8 & 10 & 6 & 0 & 6 & 6 & 6 & 10 & 6 & 5 \\
6 & 6 & 8 & 10 & 6 & 14 & 0 & 14 & 6 & 14 & 6 & 5 \\
6 & 6 & 8 & 10 & 6 & 4 & 4 & 0 & 6 & 6 & 6 & 5 \\
4 & 16 & 5 & 3 & 8 & 10 & 6 & 6 & 0 & 10 & 4 & 5 \\
6 & 6 & 8 & 10 & 5 & 5 & 5 & 6 & 6 & 0 & 6 & 5 \\
4 & 6 & 6 & 6 & 8 & 10 & 6 & 6 & 6 & 10 & 0 & 5 \\
6 & 6 & 8 & 10 & 5 & 5 & 5 & 3 & 6 & 3 & \infty & 0
\end{bmatrix}$$

Based on the linear fitting method to optimize the distribution network equipment, the statistics of inspection time within 30 days are shown in Table 4. Compared with the traditional scheme, based on the defect identification results of Transformer, the use of linear fitting for O&M inspection of equipment can save about 52 hours per month, which reduces the workload of O&M personnel to a certain extent. This helps to improve the rationality of the inspection program, reduce the workload, and improve the efficiency of distribution network O&M personnel.

Table 4. Effort saving statistics

Statistical time/day	Routine inspection/h	post-optimization/h	Save time/h
1	8	5	3
2	7	5	2
3	7	5	2
4	8	6	2
5	9	9	0
6	9	6	3
7	8	6	2
8	10	7	3
9	8	6	2
10	7	6	1
11	8	6	2
12	10	8	2
13	8	5	3
14	10	8	2
15	9	9	0
16	8	6	2
17	8	7	1
18	9	8	1
19	10	9	1
20	8	6	2
21	7	6	1
22	7	7	0
23	7	6	1
24	7	6	1
25	9	8	1
26	9	6	3
27	7	5	2
28	9	7	2
29	10	8	2
30	9	6	3
Total	250	198	52

4. Conclusion

In this paper, the data set is first produced, and the S-Transformer model is used to construct the distribution network defect recognition model, and after passing the recognition test, the relative deterioration degree of its equipment is fitted, and then the intelligent operation and maintenance optimization strategy is formulated. Based on the S-Transformer framework, the defect recognition rate of the distribution network is greater than 0.9, which can effectively meet the user's demand for intelligent detection of defects in the distribution network, and also proves the feasibility of the S-Transformer framework in the distribution network defects. Compared with the traditional solution, the O&M maintenance of the equipment can save 52 minutes, which minimizes the efficiency and time of the O&M staff.

Funding

Supported by Science and Technology Project of Guangxi Power Grid Co., Ltd.: Research on Automatic Acceptance Technology of Distribution Network UAV Based on Visual Fusion Tracking and Image Recognition (040600KC24010001).

References

[1] Zhaoyun, Z., & Linjun, L. (2022). Application status and prospects of digital twin technology in distribution grid. Energy Reports, 8, 14170-14182.

[2] Yan, Y., Liu, Y., Fang, J., Lu, Y., & Jiang, X. (2021). Application status and development trends for intelligent perception of distribution network. High Voltage, 6(6), 938-954.

[3] Mahmoud, M. A., Md Nasir, N. R., Gurunathan, M., Raj, P., & Mostafa, S. A. (2021). The current state of the art in research on predictive maintenance in smart grid distribution network: Fault's types, causes, and prediction methods—A systematic review. Energies, 14(16), 5078.

[4] Ge, L., Li, Y., Li, Y., Yan, J., & Sun, Y. (2022). Smart Distribution Network Situation Awareness for High-Quality Operation and Maintenance: A Brief Review. Energies 2022, 15, 828. Situation Awareness for Smart Distribution Systems, 41.

[5] Wang, P., Poovendran, P., & Manokaran, K. B. (2021). Fault detection and control in integrated energy system using machine learning. Sustainable Energy Technologies and Assessments, 47, 101366.

[6] Li, S., Cao, B., Li, J., Cui, Y., Kang, Y., & Wu, G. (2023). Review of condition monitoring and defect inspection methods for composited cable terminals. High Voltage, 8(3), 431-444.

[7] Dashti, R., Daisy, M., Mirshekali, H., Shaker, H. R., & Aliabadi, M. H. (2021). A survey of fault prediction and location methods in electrical energy distribution networks. Measurement, 184, 109947.

[8] Zhao, S., & Zhao, H. (2024). Intelligent Power Equipment for Autonomous Situational Awareness and Active Operation and Maintenance. Journal of Modern Power Systems and Clean Energy.

[9] Fan, P., Shen, H. M., Zhao, C., Wei, Z., Yao, J. G., Zhou, Z. Q., ... & Hu, Q. (2021, February). Defect identification detection research for insulator of transmission lines based on deep learning. In Journal of Physics: Conference Series (Vol. 1828, No. 1, p. 012019). IOP Publishing.

[10] Sarathkumar, D., Srinivasan, M., Stonier, A. A., Samikannu, R., Dasari, N. R., & Raj, R. A. (2021, February). A technical review on classification of various faults in smart grid systems. In IOP conference series: Materials science and engineering (Vol. 1055, No. 1, p. 012152). IOP Publishing.

[11] Ghasemi, H., Farahani, E. S., Fotuhi-Firuzabad, M., Dehghanian, P., Ghasemi, A., & Wang, F. (2023). Equipment failure rate in electric power distribution networks: An overview of concepts, estimation, and modeling methods. Engineering Failure Analysis, 145, 107034.

[12] Tiwari, G., & Saini, S. (2024). Optimizing Fault Identification in Power Distribution Systems by the Combination of SVM and Deep Learning Models. Journal of Operation and Automation in Power Engineering.

[13] Sudhakar, A. V. V., Goswami, C., Neeraja, B., Jain, A. K., Gupta, S., & Gowri, G. (2023, December). Wireless sensor network for fault detection using block chain technology based smart grid security. In International Conference on Data Science, Machine Learning and Applications (pp. 528-537). Singapore: Springer Nature Singapore.

[14] Stefenon, S. F., Singh, G., Yow, K. C., & Cimatti, A. (2022). Semi-ProtoPNet deep neural network for the classification of defective power grid distribution structures. Sensors, 22(13), 4859.

[15] Yang, Y., Wang, S., Ren, J., & Zhang, Y. (2023, July). Abnormal Scene Image Recognition Method for Intelligent Operation and Maintenance of Electrical Equipment in Substations. In 2023 3rd International Conference on Energy Engineering and Power Systems (EEPS) (pp. 298-305). IEEE.

[16] Gao, Y., Kang, B., Zhao, T., Xiao, H., Li, J., Xu, Z., ... & Wang, Z. (2022, November). Deep Learning-Based Text Entity Recognition Method for Distribution Network Operation and Maintenance. In 2022 9th International Forum on Electrical Engineering and Automation (IFEEA) (pp. 1096-1100). IEEE.

[17] Zhang, L., Hu, E., Zhou, X., Chen, T., Yang, S., & Zhang, Y. (2022, October). Named Entity Recognition for Smart Grid Operation and Inspection Domain using Attention Mechanism. In 2022 International Conference on Machine Learning, Control, and Robotics (MLCR) (pp. 134-140). IEEE.

[18] Liu, W., Gu, Y., Zeng, Z., Qi, D., Li, D., Luo, Y., ... & Wei, S. (2024). Automated Equipment Defect Knowledge Graph Construction for Power Grid Regulation. Electronics, 13(22), 4430.

[19] Li, K., Wang, Y., Zhu, T., Li, J., Ren, J., & Chen, Z. (2024). Knowledge graph‐based multimodal neural networks for smart‐grid defect detection. Engineering Reports, 6(9), e12824.

[20] Zhu, L., Tian, N., Li, W., & Yang, J. (2022). A Text Classification Algorithm for Power Equipment Defects Based on Random Forest. International Journal of Reliability, Quality and Safety Engineering, 29(05), 2240001.

[21] Zhao, Z., Chen, Z., Liu, J., Huang, Y., Gao, X., Di, F., ... & Ji, X. (2019, August). Chinese named entity recognition in power domain based on Bi-LSTM-CRF. In Proceedings of the 2nd International Conference on Artificial Intelligence and Pattern Recognition (pp. 176-180).

[22] Xizi Yu,Shuang Li & Yu Zhang. (2024). Incorporating convolutional and transformer architectures to enhance semantic segmentation of fine-resolution urban images. European Journal of Remote Sensing(1),

[23] Aitong Liu,Yelin Shi,Ajay Kumar,Xin Chen & Manhua Liu. (2025). A hybrid graph transformer network for fingerprint and palmprint minutiae matching of young children. Neurocomputing129180-129180.

[24] Guolian Hou,Qingwei Li & Congzhi Huang. (2025). Spatiotemporal forecasting using multi-graph neural network assisted dual domain transformer for wind power. Energy Conversion and Management119393-119393.

[25] Tasarruf Bashir,Huifang Wang,Mustafa Tahir & Yixiang Zhang. (2025). Wind and solar power forecasting based on hybrid CNN-ABiLSTM, CNN-transformer-MLP models. Renewable Energy122055-122055.

[26] Fei Li,Chu Yang Luo,Ying Zi Wen,Sheng Lv,Feng Cheng,Guo Qiang Zeng... & Bing Hai Li. (2024). A nuclide identification method of γ spectrum and model building based on the transformer. Nuclear Science and Techniques(1),7-7.