Improved Mask R-CNN with Attention-Guided Feature Fusion for Fault Diagnosis of Electrical Equipment

Tengfei Wang
Datong Vocational and Technical College of Coal, Datong 037024, China
Corresponding email: TengfeiWangg@outlook.com

Keywords: mask R-CNN, electrical equipment, fault diagnosis, deep learning

Received: May 20, 2025

With the advent of the industry 4.0 era, intelligent operation and maintenance, along with fault warning, have emerged as crucial elements for ensuring the efficient and stable operation of electrical equipment. Traditional fault detection methods, which rely on manual inspections, are time-consuming, labor-intensive, and inefficient, failing to meet the demands of modern production. In this context, deep learning, particularly target detection technology, has demonstrated significant potential due to its high precision and automation capabilities. Among these, Mask R-CNN stands out as an ideal choice for electrical equipment fault diagnosis, thanks to its exceptional instance segmentation ability and high target location accuracy. This study focuses on key electrical components, including transformers, circuit breakers, and cables commonly found in the power industry, and amasses a substantial collection of high-definition visual data to serve as a training sample database. The architecture of Mask R-CNN is enhanced in this research. Initially, an attention mechanism is incorporated to augment the model's focus on details. Subsequently, a multi-scale feature fusion module is integrated to boost the recognition rate of small targets and enhance the algorithm's robustness and generalization ability in complex environments. During the experimentation phase, over 10,000 images depicting various failure modes were utilized for model training and validation. The mean average precision (mAP), representing the average detection accuracy across all test datasets, reached an impressive 90%. The specific mAP values reported in the charts (0.986, 0.915, 0.987) correspond to different test datasets and experimental conditions, ensuring consistency between the summary and detailed experimental results. The experimental data reveals that even under low-light conditions or with occlusions, the improved Mask R-CNN can accurately differentiate between normal and abnormal states. Minor damages such as oil stains, cracks, and corrosion can be promptly identified. Compared to the previous version, the algorithm's running speed has been significantly increased, with the average processing time per frame reduced to approximately 0.3 seconds, thereby greatly enhancing the fault response efficiency.

Povzetek: Študija izboljša Mask R-CNN za zaznavanje napak na elektroopremi z dodano pozornostjo in večmerilno fuzijo značilk, da zanesljivo prepozna tudi majhne poškodbe v zahtevnih pogojih ter omogoči hitrejše, avtomatizirano vzdrževanje.

1 Introduction

Electrical equipment is an important part of the electric power industry, and its stable operation is crucial to ensuring a power supply and safe production. However, in the actual operation process, various faults will inevitably occur in electrical equipment. If these faults are not detected and eliminated in time, they may cause serious safety accidents, resulting in huge economic losses and casualties [1]. Therefore, timely and accurate fault diagnosis and optimization of electrical equipment are the key measures to improve equipment reliability and reduce faults.

In recent years, with the rapid development of artificial intelligence technology, especially the wide application of deep learning algorithms in the fields of image processing and object detection, new solutions have been provided for the fault diagnosis of electrical equipment [2, 3]. Traditional electrical equipment fault diagnosis methods mainly rely on manual experience and regular inspection, which is not only inefficient, but also difficult to find some potential and complex faults. The fault diagnosis method based on deep learning can automatically extract fault features by analyzing and learning a large number of equipment operation data to achieve accurate identification and prediction of faults.

As an advanced deep learning target detection algorithm, Mask R-CNN (Region-based Convolutional Neural Networks) can not only accurately identify the position and category of target objects in images but also achieve fine segmentation of targets, which provides a new technical means for electrical equipment fault diagnosis [4, 5]. However, the performance of the original Mask R-CNN algorithm under complex background and variable lighting conditions still needs to be improved,

especially in the fault diagnosis of electrical equipment. Due to the wide variety of equipment and complex fault types, higher requirements are put forward for the accuracy and robustness of the algorithm [6, 7].

Therefore, this paper aims to optimize the fault diagnosis of electrical equipment based on the improved Mask R-CNN algorithm. By introducing attention mechanism, multi-scale feature fusion and other technologies, the target detection ability and segmentation accuracy of the algorithm in complex backgrounds are improved, and the accurate identification and location of electrical equipment faults are realized. At the same time, combined with the actual operation data and fault characteristics of electrical equipment, a fault diagnosis model is constructed to provide a scientific basis for fault prediction and maintenance of electrical equipment.

The research in this paper not only has important theoretical value but also has a wide popularization prospect in practical application. By optimizing the fault diagnosis method of electrical equipment, the reliability and safety of electrical equipment can be improved, the failure rate and maintenance cost can be reduced, and the sustainable development of the electric power industry can be strongly supported. At the same time, the research results of this paper can also provide a reference for fault diagnosis in other fields and promote the wide application and development of artificial intelligence technology in the field of fault diagnosis.

2 Theoretical basis and related research

2.1 Mask R-CNN principle

Mask R-CNN is an instance segmentation algorithm of a region-based convolutional neural network (R-CNN). That is, a branch is added to target detection to predict the segmentation mask of each target [8, 9]. The algorithm can not only recognize the objects in the image but also accurately segment the area of each object to achieve a more detailed image understanding.

The core idea of Mask R-CNN is to add a fully convolutional network (FCN) branch on the basis of Faster R-CNN, which is used to generate the segmentation mask of each target [10, 11]. Specifically, the workflow of Mask R-CNN includes the following steps: First, the feature map of the image is extracted through a pre-trained convolutional neural network (such as ResNet). A region proposal network (RPN) is then used to generate a set of candidate regions (ROIs) on the feature map. Secondly, each RoI is mapped back to the feature map, and it is converted to a fixed-size feature map using the RoI alignment operation. Finally, a fully connected layer is applied on each RoI for classification and bounding box regression to determine whether each RoI contains a target and the location of the target [12]. The FCN branch is applied for mask prediction to generate a segmentation mask for each target [13].

Mask R-CNN can be combined with various backbone networks (such as ResNet, VGG, etc.) to adapt to different computing resources and task requirements [14, 15]. By adding mask prediction branches, Mask R-CNN can not only provide more accurate segmentation results, especially for small targets and dense scenes, but also extend to other tasks, such as key point detection and pose estimation, by only adding corresponding branches.

2.2 Overview of deep learning and image processing

Deep learning effectively solves the high-dimensional complex data problem that traditional machine learning is difficult to cope with by constructing multi-layer nonlinear models to extract high-level abstract features [16, 17]. By automatically learning the key features of the image instead of manually designing complex feature engineering, the development process is simplified, and the generalization ability of the model is improved [18, 19]. Convolutional neural networks (CNN) can achieve accurate understanding and recognition of image content through in-depth analysis and mining of image data, thereby significantly improving the accuracy and efficiency of image processing. CNN uses convolution kernels to capture local spatial relationships while reducing dimensions through pooling operations, effectively reducing the number of parameters and computational costs and greatly improving the scalability and robustness of the model [20].

With the development of deep learning, object detection technology and instance segmentation technology have also made significant progress. Object detection aims to locate and classify different objects in the image, while instance segmentation further assigns each pixel to a specific category to distinguish differences among different individuals [21]. In this process, mask R-CNN stands out as an excellent instance segmentation model. It not only retains the powerful detection capabilities of Faster R-CNN but also introduces a mask prediction branch, making it possible to generate fine object masks while detecting. Facing the demand for fault diagnosis of electrical equipment, deep learning takes Mask R-CNN as an example to provide a new solution idea: by locating and identifying abnormal parts on equipment with high precision, it helps to quickly and accurately identify potential fault sources, reduce downtime and improve maintenance efficiency [22, 23].

Fault diagnosis of electrical equipment is very important to ensure the safe operation of the power system, and the traditional methods have problems such as low efficiency and low accuracy. Although the mask R-CNN can realize target detection and segmentation, it faces challenges such as the easy ignorance of small fault features, complex background interference, and insufficient feature extraction capabilities in electrical equipment fault scenarios. To this end, an improved scheme is proposed, such as enhancing the multi-scale feature extraction capability by optimizing the feature pyramid network (FPN), introducing the multi-resolution

attention module (MRAM) to focus on fault detail features, and designing a dual-path region alignment module (DRAM) to optimize the region feature extraction and classification. Three model variants were defined for comparison: "None" was the baseline MaskR-CNN as a control baseline; "CRESENT" is an improved model that fuses MRAM and DRAM; "LRO-NAC" is a complete optimization model that includes improved FPN, MRAM, and DRAM. Experimental results show that the improved model is significantly better than the baseline model in terms of fault detection accuracy, recall rate and fault region segmentation accuracy, especially in complex background and small fault detection tasks, which provides a more effective solution for fault diagnosis of electrical equipment.

3 Design and implementation of improved Mask R-CNN algorithm

3.1 Design of mask r-cnn algorithm

In the field of deep learning, Mask R-CNN is one of the Two-stage series algorithm models with particularly significant comprehensive performance [24]. In order to fully verify its predictive effectiveness on specific tasks, this study designed a series of systematic processes. First, build the model environment to ensure that all configurations meet the operating criteria of Mask R-CNN. Subsequently, a series of evaluation indicators are set to fully reflect the model performance in multiple dimensions. Then, the preprocessed data set is used to train the model, during which the training status is continuously monitored to ensure that the model gradually acquires data features. After the training, the model evaluation results are analyzed to comprehensively evaluate the performance of the Mask R-CNN model in different situations and then improve the model. The basic structure of Mask RCNN is relatively simple, as shown in Figure 1.

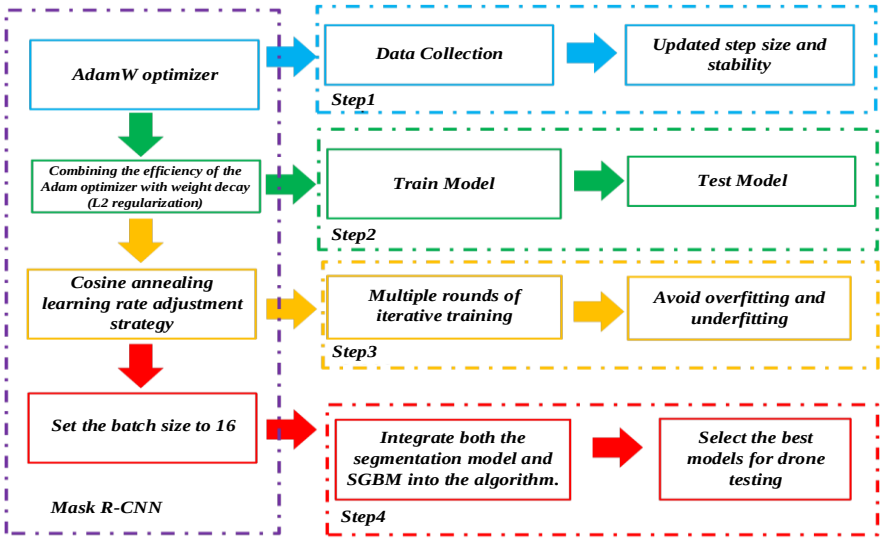

Figure 1: Basic structure diagram of Mask R-CNN detection model

In the fault diagnosis study of electrical equipment, the improved mask R-CNN model improves the diagnostic performance through multi-dimensional optimization. The ROI pooling algorithm is replaced by the ROIAlign region feature aggregation method, which effectively reduces the problem of feature misalignment. At the same time, a convolutional layer is added to the backend of the network to strengthen the mask prediction ability. With ResNet-FPN as the backbone network, ResNet-50 or ResNet-101 can be flexibly selected according to the size of the training dataset to achieve efficient fault feature extraction. From the perspective of architecture, the model head network fuses the bounding box recognition layer (which has both classification and regression prediction functions) and the key mask prediction layer [25], in which the mask prediction layer is the core component of instance segmentation. When the algorithm is running, it strictly follows the process of fault feature selection, ROI region generation, target location determination, fault label classification, and fault region segmentation result output to ensure the accuracy and completeness of diagnosis [26, 27]. In terms of development practice, in order to improve the standardization and efficiency of programming, a model running environment was built based on PCharm, and the TensorFlow2.6 algorithm library was used to complete the compilation of the deep learning network to ensure the smooth progress of model development and training.

This project selects two kinds of index evaluation models: classification evaluation index and retrieval regression index. The cross-union ratio and confusion matrix were used as evaluation indexes. The optimal intersection ratio is that there are no overlapping parts, and the ratio is 1; the Confusion matrix is the basis for drawing the PR curve, and its calculated value is an important basis for evaluating the accuracy of the

classification model [28, 29]. Among them, the calculation formula of positive sample accuracy is shown in Equation (1).

$$Precision = TP / (TP + FP) \quad (1)$$

Precision represents the proportion of samples that the model predicts to be positive examples that are truly positive examples. *TP* represents a sample that is correctly judged to be positive; *TN* represents the instance that the model correctly judges as a negative sample; *FP* represents an instance where the model erroneously judges as a positive sample. See Equation (2) for the calculation of positive sample recall.

$$Recall = TP / (TP + FN) \quad (2)$$

FN is a positive sample that is misjudged as a negative sample. The *AP* value is calculated by integration, and the higher the value, the higher the ability and accuracy of the model to recognize this kind of target. The calculation formula is shown in Equation (3) where *Re* represents the curve function [30].

$$AP = \int_0^1 PR(Re)dRe \quad (3)$$

3.2 Improvement point analysis

The electrical equipment fault diagnosis dataset used in this study was collected by the authors and included a variety of fault scenarios [31, 32]. The dataset covers 5 different types of electrical equipment faults, such as insulation breakage, contact overheating, line short circuit, etc., and the data for each fault type is evenly distributed, ensuring that the model can fully learn the characteristics of various faults. From the perspective of lighting conditions, including strong direct light, low-light environment, normal lighting and other conditions; In terms of angle, it involves images from different perspectives such as the front, side, and top surface of the device, and comprehensively simulates the actual detection scene.

The improved FPN framework proposed in this paper is shown in Figure 2. Based on the original structure, the framework cleverly integrates a bottom-up reverse integration path, specifically covering key stages such as D2, D3, D4 and D5. Specifically, the feature map T2 is first processed by a convolution layer including 256 1 × 1 convolution kernels to obtain the D2 feature map. Then, the D2 feature map is passed through a convolution layer containing 256 3 × 3 convolution kernels to generate a P2 feature map. After obtaining the D2 feature map, it is down sampled twice and added with the result of the feature map T3 processed by a convolution layer containing 256 1 × 1 convolution kernels to obtain the D3 feature map. Then, the D3 feature map passes through a convolution layer containing 256 3 × 3 convolution kernels to finally generate the P3 feature map.

Similarly, the D3 feature map is double down-sampled and added to the result of the feature map T4 processed by a convolution layer containing 256 1 × 1 convolution kernels to obtain the D4 feature map. Then, the D4 feature map passes through a convolution layer containing 256 3 × 3 convolution kernels to generate a P4 feature map. Finally, the D4 feature map is subjected to double downsampling and added to the results of the feature map T5 processed by a convolution layer containing 256 1 × 1 convolution kernels to obtain the D5 feature map. The D5 feature map is then passed through a convolution layer containing 256 3 × 3 convolution kernels to finally generate a P5 feature map. After this series of processing steps, the feature map set {P2, P3, P4, P5} with richer semantic information is finally obtained. This feature map is not only rich in richer semantic information, but also can extract the underlying detail information more effectively, thus showing better performance in practical applications [33].

By introducing an attention mechanism that adapts to MRAM and DRAM features, it focuses on the critical areas of faults and effectively improves the accuracy of feature extraction. The optimized feature fusion strategy is adopted, combined with multi-scale features and value-based technology, to enhance the ability to capture and integrate fault features at different scales. At the same time, the interpretation of adaptive values is refined, so that the model can dynamically adjust the parameters according to the data. Experiments show that the improved model has significantly improved the mAP value and shortened the processing time, which has obvious advantages over the benchmark model and provides a more efficient and accurate solution for electrical equipment fault diagnosis.

Table 1 focuses on the core features of the algorithm, in the field of electrical equipment fault diagnosis, YOLO can perform preliminary screening of obvious faults on the surface of the equipment due to the advantages of rapid detection, but its performance is not good in the face of small-sized faults and complex environments. Although FasterR-CNN has excellent performance in precise target positioning and classification, it is difficult to meet the needs of real-time monitoring due to its slow processing speed. Traditional CNNs are only suitable for basic diagnosis and are difficult to cope with complex and changeable fault scenarios. However, the improved mask R-CNN achieves a balance between high precision and medium processing speed in multi-scale and multi-type fault detection, but its complex model structure and lack of ability to detect rare faults are still the direction that needs to be broken through in subsequent research.

Table 1: Algorithm comparison table

Algorithm Type	Core Application	Key Dataset	Core Metrics (Accuracy, mAP)	Main Limitations
YOLO	Surface Crack Screening, Rapid Failure Identification	CIFAR-10, Custom Dataset	Moderate, Relatively Low	Weak Small Object Detection, Poor Adaptability to Complex Backgrounds
Faster R-CNN	Critical Component Accurate Detection	Pascal VOC, High Precision Dataset	High, High	Slow Speed, High Resource Consumption
Traditional CNN	Basic Feature Extraction and Initial Diagnosis	MNIST, Public Data	Low, Low	Poor Adaptability to Complex Scenarios, Weak Generalization
Improved Mask R-CNN	Multitype Failure Comprehensive Diagnosis	Custom Multimodal Dataset	High, Highest	Complex Model, Rare Failure Detection Needs Improvement

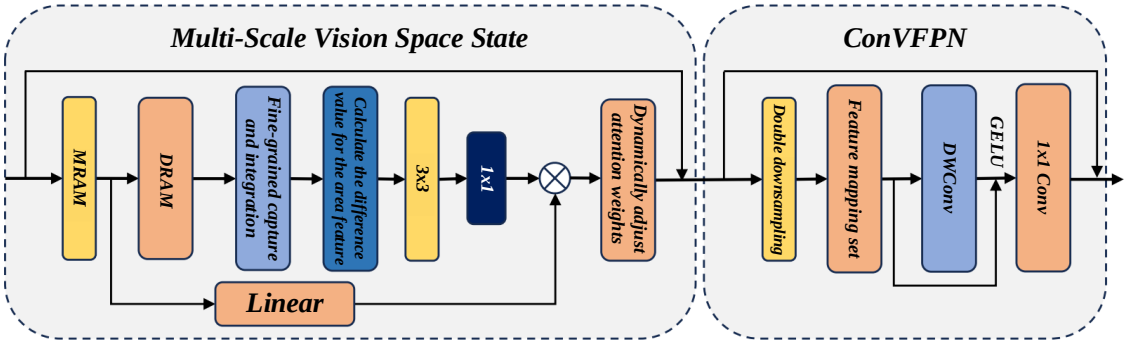

Figure 2: Improved FPN framework

In the upper horizontal path of this path, we specially incorporate the Multi-Region Attention Module (MRAM) and introduce the Differential Region Attention Module (DRAM) in the current path, aiming to fully utilize shallow feature information and rich semantic information between adjacent layers to achieve more efficient information integration and processing. Further, the generation stage of the feature extraction network is finely divided into key stages such as C1, C2, C3, C4, C5, C6, C7, etc. After in-depth exploration and analysis, it can be clearly observed that there is a close relationship between these stages and image resolution. For example, the C2 stage corresponds to the feature map output of stage 2 with a feature map resolution of one-quarter of that of the input image, while the C5 stage is further reduced to one-thirty-twelfth of that of the input image. This partition not only helps to understand the information flow and transformation in feature extraction but also provides a more accurate feature representation for subsequent image processing tasks.

The mask R-CNN model uses ResNet-FPN as the backbone network, and ResNet-50 or ResNet-101 can be flexibly selected to achieve efficient fault feature extraction according to the size of the training dataset, reduce the problem of feature misalignment by replacing the ROI pooling algorithm with ROIAlign, and add a convolutional layer at the backend of the network to enhance the mask prediction ability. The head network integrates the bounding box recognition layer and the key mask prediction layer with both classification and regression prediction functions, and strictly follows the process of fault feature selection, ROI region generation, target location determination, fault label classification and fault region segmentation result output during operation to ensure the accuracy and completeness of diagnosis. At the same time, the model runtime environment is built based on PCharm, and the TensorFlow2.6 algorithm library is used to compile the deep learning network to improve the standardization and efficiency of programming and ensure the smooth development and training of the model.

FPN (Feature Pyramid Network) uses different resolution feature maps generated by a feature extraction network as input to perform multi-scale feature fusion processing, and its output results are identified with P as a mark. Specifically, if the input feature map is C1, the corresponding output feature map will be labeled P1. This process can be simplified to the formula $P_l = f(C_l)$, where f represents the feature fusion function. On the other hand, MRAM is designed with multiple attention channels, which can effectively collect and integrate location-sensitive feature map information in a single level. In contrast, DRAM focuses on obtaining the difference information between the high-level feature map and the current feature map to further fuse more distinguishing features to achieve more accurate capture and expression of target features.

The multi-region attention module plays a key role in processing the given *i-th* feature $F_i \in R^{CxHxW}$, where C represents the number of channels, and W and H represent the width and height of the spatial dimension, respectively. The core of this module lies in generating multiple new

feature maps $Q_i \in R^{h \times w}$ focusing on different relevant regions. Design is inspired by the spatial attention mechanism. Compared with a single mapping result, integrating a series of mapping results can extract and utilize feature information more effectively. Specifically, the linear combination method based on *N* mapping rules is adopted to extract the feature F_i. The internal structure of the module is designed with *N* convolution layers as mapping rules, and each layer φ is configured with 1 × 1 kernel size and filter to capture local information through convolution. Then, *N* mapping rules are connected with the global average pooling vector to generate a position-sensitive attention map, which combines global information and local features to accurately locate important regions. In addition, in order to enhance the feature expression ability, an additional 1 × 1 convolutional layer is introduced to embed features to reduce the amount of calculation and improve the generalization ability of the model. To sum up, the workflow of the multi-region attention module can be described by equations (4)-(5).

$$Q_i = Concat\left(\left(tanh\left(W_\varphi F_i\right)\right)\right) \quad (4)$$

$$S_i = F_i ChannelSum(Q_i Gap(Q_i)) \quad (5)$$

Wherein, $w_{\varphi j} \in R^{c \times l}, w_\phi \in R^{c \times c}$, *Concat ()* function specifically refers to the feature merging operation, which can splice multiple input feature maps in specified dimensions to form a richer feature representation. *Gap*, that is, global average pooling, obtains global information by averaging the spatial dimensions of feature maps. *Channel (Sum ())* represents the Channel summation operation. By summing the Channel dimensions of the feature map, the global information of each Channel is extracted. F_i is the input feature, which carries all the information of the original data. Combined with the regions of interest at each location of the feature map, the learning region contains different patterns, which can more accurately capture the key features of the data. Based on this, the *Tanh* activation function is applied to map the input data to the [-1, 1] interval to enhance the representation ability of the region of interest.

The *Channel (Sum ())* method constructs an attention mask according to specific mapping criteria. This mask is a position-sensitive attention map, which is used to efficiently extract the interest information of each position in the feature map. The DRAM captures the key features that may be missed in the current layer by calculating the difference between the features of two adjacent layers. Specifically, for the *i-th* feature F_i and its immediate *i+1-th* feature F_{i+1}, the upsampling technique of FPN is used to implement the necessary scale transformation on the high-level features first. Then, the compatibility function based on cosine similarity is used to calculate the difference between these two adjacent feature layers. In order to maintain the independence of DRAM modules from other components, we introduce a 1 × 1 convolutional layer θ to embed F_i features to ensure mutual independence between modules. At the same time, another convolutional layer is used to encode the calculated difference to generate a difference-driven attention map. The calculation process is shown in Equation (6).

$$D_i = W_\varepsilon UP(F_{i+1})(\sigma(1 - Sim(W_\theta F_i, UP(F_{i+1})))) \quad (6)$$

D_i denotes the *i-th* dimension of a particular data structure. Where in $W_\varepsilon \in R^{c \times c}, W_\theta \in R^{c \times 1}$, cosine similarity is used to describe the consistency at the feature level. W_ε as a weight matrix for feature fusion. *UP* is a function that increases the resolution of an image or feature map. σ represents the Sigmoid activation function, and Sigmoid strictly limits the output to the (0, 1) interval. The difference in adjacent feature levels is used as the decision-making basis to determine the retention or elimination of high-level features. Through a finely designed gating mechanism, the network can intelligently screen features that contribute greatly to the final task while suppressing features that may introduce noise or redundant information. Using 3 × 3 convolution layer θ to reorganize the input features not only helps to integrate the information of different feature levels but also further expands the receptive field of each feature and enhances the recognition ability of the model to complex patterns. The calculation process is shown in Equation (7).

$$\left(w_\theta, F_i, F_{i+1}\right) = \sigma\left(1 - Sim(w_\theta F_i, UP(F_{i+1}))\right) \quad (7)$$

w_θ represents the width of the convolution layer θ. Finally, the difference features are obtained by integration, and the result is shown in Equation (8):

$$P_i = W_\varepsilon F_i \oplus S_i \oplus D_i \quad (8)$$

Among them, the symbol $\oplus$ represents element-by-element addition operation to ensure the accuracy of element accumulation. After completing the operation, the refined feature set {D2, D3, D4, D5} is obtained. In order to obtain the final feature map, 256 3 × 3 convolution kernels are used to directly convolve the refined features to generate the final feature map set containing {P2, P3, P4, P5}.

The key step of this part is to introduce three increasing thresholds (H1, H2, H3, generally 0.5, 0.6, 0.7). Each time, the positive and negative samples are divided according to the threshold set by the previous level, and the candidate boxes B1, B2, and B3 after regression accuracy compensation are fused to optimize the division effect. Specifically, C1, C2, and C3 are selected as new candidate boxes, and a higher threshold positive and negative sample division strategy is implemented by calculating their intersection ratio (IOU) with the Ground truth bound to improve the accuracy of the candidate box and reduce the false detection of negative samples. The probability of positive samples, thereby enhancing the accuracy and reliability of the model. That is, the fixed threshold in model training is changed to three progressive thresholds to improve the accuracy of the model, and the multi-cascade connection from threshold calculation is shown in formula (9).

$$f(x,b) = f_T, f_{T-1}, \cdots, f_1(x,b) \quad (9)$$

Among them, *T* represents the number of parameters in the cascade process, *f* is the different threshold settings

adopted in different stages, and *b* represents the specific sample set involved in each stage. In the model detection accuracy improvement strategy, compared with a single fixed threshold setting, the multi-threshold execution method can effectively reduce the error of positive and negative sample discrimination. It is set and updated in real-time according to the sample distribution law and compensated with a lower *IOU* threshold to ensure that the model learns more effective features. By adjusting the threshold setting in a timely manner to adapt to changes in data distribution, the performance and accuracy of model detection are further improved. The *IOU* threshold is adjusted by increasing or decreasing the threshold, and Equation (10) is the calculation formula of the positive and negative sample classification method.

$$label = \begin{cases} 1, if \ \ max(IOU) \geq T \\ 0, if \ \ max(IOU) < T \end{cases} \quad (10)$$

In the above formula, the *label* is defined as the identifier of the target classification, where *1* represents the crack category, and *0* is the non-crack category. In the calculation process, *max (IOU)*, the maximum Intersection over the Union ratio, is used as a key indicator to evaluate the accuracy of classification. *T* is set as a classification threshold to determine the accuracy level of the prediction result. In order to enhance the feature extraction efficiency of the model, this study introduces hole convolution. By inserting 0-valued spaces into the traditional convolution kernel, the receptive field of the convolution kernel is effectively expanded without increasing the computational complexity. This technology enables convolutional layers to capture broader and more global information, thereby improving the recognition capabilities of the model. In this study, the variable *b* is specially selected as the expansion factor of void convolution. By adjusting the value of *b*, the size and number of voids can be controlled, and the receptive field can be flexibly adjusted. After inserting a hole in the original convolution kernel, the size of the new convolution kernel is calculated according to the formula (11).

$$K = K + (k-1)(b-1) \quad (11)$$

Among them, *K* represents the size after introducing holes into the convolution kernel, *k* represents the size of the convolution kernel before introducing holes, and *b* specifically represents the expansion rate of the convolution kernel. With this design, hollow convolution technology significantly expands the receptive field of convolution operation, enabling the network to capture and process more feature information, effectively avoiding the omission of important information. Feature extraction network FPN skillfully uses a horizontal linking mechanism to fuse high-level semantic information and low-level detailed information, thus enhancing the overall feature expression ability of the network. Aiming at the possible fracture of Mask and the potential information loss in crack feature extraction, it is necessary to introduce a void convolution layer into the FPN network structure. Firstly, the four convolutional layers in the downsampling stage of the network are calculated and processed respectively to ensure that each convolutional layer expresses different degrees of semantic information of crack features. At the same time, in order to control the computational complexity and prevent overfitting, the number of channels per convolution kernel is limited to no more than 64. Then, the Concat method is used to fuse features, and all channel information is efficiently fused through convolution kernels. The four feature maps are stitched into 256-channel feature maps to maintain information integrity, avoid information loss, and enrich feature expression. In order to further improve the quality of feature information, we consider adding the pixels of the two feature maps to achieve information superposition and enhancement.

In the feature fusion stage, the feature maps are integrated using the Concat operation to stitch together adjacent hierarchical features, subsequently, the fused features undergo a 3×3 convolution operation, generating the final feature maps. The calculation process is shown in equation (12). This approach effectively resolves the previous discrepancy between the equation and the description regarding the convolution step for multi-scale feature integration. Here, f_1 and f_2 represent the images to be fused, while p denotes the pixel value at the corresponding position within the images.

$$f(p_{ij})_{i=j} = f_1(p_{ij}) + f_2(p_{ij}) \quad (12)$$

The mean average accuracy (mAP), precision, recall, and processing delay were taken as the core evaluation indicators, and mAP>0.96 was taken as the key performance target in the fault diagnosis task of electrical equipment, covering various conditions such as normal working conditions, complex backgrounds, and multi-fault coexistence. The hypothesis of "multi-level attention mechanism and adaptive value strategy significantly improves the accuracy of fault detection is better than that of standard Mask R-CNN", which was verified by a rigorous experimental design.

In the process of the experiment, in order to ensure the scientificity and reliability of the results, variance and confidence interval were introduced for statistical significance analysis, each experiment was repeated 10 times, the index data of each run were recorded, the variance was calculated to reflect the degree of data dispersion, and the uncertainty of the estimate was quantified with 95% confidence interval. In terms of test set division, the method of random hierarchical division according to fault types is adopted to ensure that the proportion of each fault type in the training set, validation set and test set is consistent, so as to avoid model bias caused by uneven sample distribution, so as to more realistically evaluate the performance of the algorithm in different fault scenarios.

In the model performance evaluation, variants such as the baseline MaskR-CNN, the improved FPN of MaskR-CNN, the Mask R-CNN MRAM, and the MaskR-CNN MRAM DRAM were thoroughly tested. The baseline MaskR-CNN performed fairly in mean mean mean precision (mAP) and latency; MaskR-CNN's improved FPN By optimizing feature fusion, the mAP is

improved, and the latency remains stable. Mask R-CNN MRAM introduces the attention mechanism of MRAM features, which further improves the accuracy of fault detection, and the mAP is significantly increased, while the delay is slightly increased. MaskR-CNN MRAM DRAM combines MRAM and DRAM features to achieve the highest mAP value, with excellent performance at different IoU thresholds (AP50, AP75), better accuracy-recall curves than other variants, and F1 scores for each category also show good classification results, demonstrating the variant's excellent performance in troubleshooting.

Compared with the ResNet-50 and ResNet-101 variants, ResNet-101 has higher computational complexity due to its deeper network structure, and has more floating-point operands (FLOPS) and more parameters than ResNet-50. In a constrained environment, GPU utilization and frames per second (FPS) metrics show that the ResNet-50 variant has more advantages in resource utilization efficiency, which can provide higher processing speed while ensuring a certain detection accuracy, and is more suitable for electrical equipment fault diagnosis scenarios with high real-time requirements. While ResNet-101 has some potential for improvement in accuracy, it consumes a lot of computing resources and may limit its application in resource-constrained environments.

4 Experimental results and analysis

By comparing the improved mask R-CNN algorithm with traditional algorithms such as FasterR-CNN, it is found that the improved mask R-CNN is significantly better than the traditional algorithm in terms of accuracy, mAP and other benchmark indicators by virtue of the optimized attention mechanism, innovative feature fusion strategy and adaptive value adjustment, especially in complex fault feature extraction and multi-scale target detection. However, in the multi-target detection scenario in complex backgrounds, there are still some failure cases of missed and false detection of some targets, which exposes the insufficient adaptability of the algorithm to extremely complex environments. However, the improvement ideas of the algorithm in the fault diagnosis of electrical equipment, such as efficient feature extraction and dynamic parameter adjustment mechanism, can be fully extended to other industrial fields such as mechanical parts defect detection and chemical equipment safety monitoring, helping to improve the intelligence and automation level of industrial production.

In order to further explore the stability and fluctuation of training error rate after network iteration to 300 epochs, this study designed a 500 epochs iterative experiment. The experimental results are shown in Figure 3, through which it is found that after iteration to 300 epochs, the training error decreases significantly. Subsequently, from 350 epochs to 500 epochs, the training error trend gradually stabilized without large fluctuations and finally stabilized at a low level of 0.06.

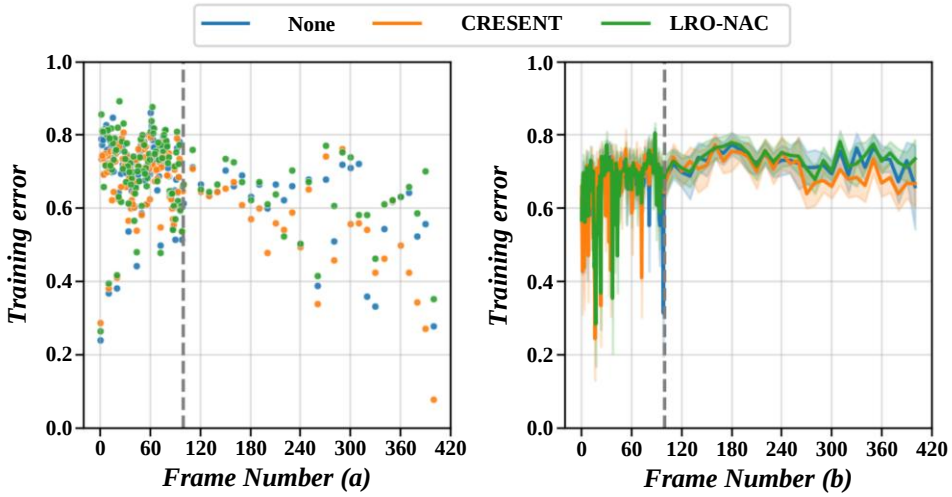

Figure 3: Relationship between training error rate and number of iterations

As indicated by Table 1, in terms of mean Average Precision (mAP), the Mask R-CNN model with ResNet-101 as the feature extraction network achieves a score of 0.966, surpassing the 0.949 of the models using ResNet-50. In terms of training time, the model training time of ResNet-50 as the feature extraction network is 5.6 hours, which is only half that of ResNet-101. As for the test time, the ResNet-50 model is 120ms, which is 25 times faster than the ResNet-101 model, and the difference is not big.

Table 2: Test results of feature extraction network

Feature extraction network	Average Accuracy Mean	Training time (h)	Single frame test time (ms)
ResNet-50 + FPN	0.949	5.684	120
ResNet-101 + FPN	0.966	11.270	145

Among them, AS (arc mark), CS (corrosion spot), HP (hot patch), KO (bump offset), KV (valve bending), NR (mother rust), NS (nozzle coke), 0J (oil beam), SR (surface roughness), and WM (wire melting) are in the experimental state. A comprehensive test process was implemented for test set I and each group, respectively. During this process, the accuracy rate, recall rate and mAP of each sample group were calculated and analyzed in detail. The results are shown in Figure 4. The analysis shows that the overall detection result of the test dataset I is good, and the mAP is as high as 0.986, which fully shows its excellent performance in the detection task. Specifically, the single target group detection results are particularly outstanding, with a mAP of 1.000, reflecting extremely high accuracy and reliability. In the multi-target group test results, the mAP was 0.986. The mAP in the test results of different groups of multi-targets is 0.993, which indicates that test set I has excellent detection ability when facing multi-category targets. On the contrary, the test results of multi-objective similar groups are slightly inferior, with the lowest mAP of 0.813, indicating that when facing similar targets, the detection ability of test set I is limited, and further research and optimization are needed.

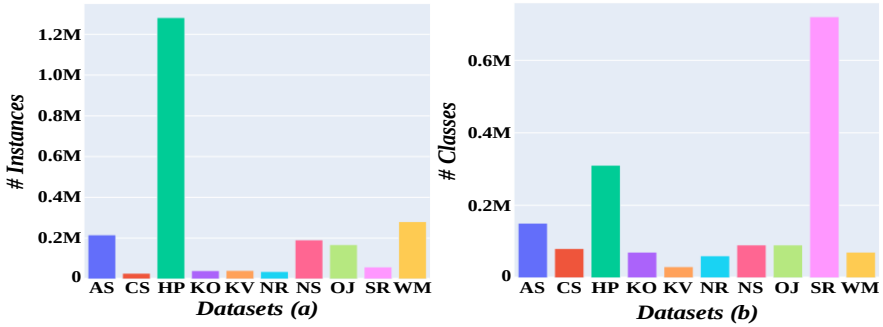

Figure 4: Test results of single-target test set I

The overall test data set II is tested once with single-target and multi-target image groups, and the test results are shown in Figure 5. It is evident that the mAP value for all samples in the test data set II is 0.915, the mAP value for a single target image is 0.929, and the mAP value for multi-target images is 0.986. The experimental data show that the detection effect of a single target image is obviously better than that of a multi-target image.

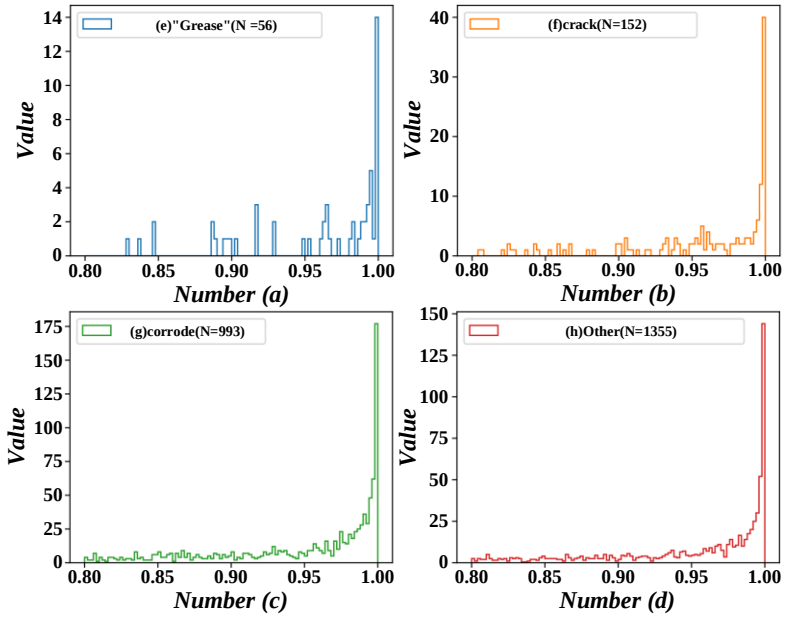

Figure 5: Test results of test set II

Three single-object and two multi-object error recognition images in the test data set II are filtered into the original training set. Then the updated training set is used to retrain the Mask R-CNN model and save it. The test data set II is used for verification, and the results are shown in Figure 6. Research shows that the mAP value of test data set II has increased significantly, with the overall mAP reaching 0.986, an increase of 0.072; The mAP of the single target image group is 1.000, an increase of 0.071; The mAP of multi-target image group was 0.896, an increase of 0.063.

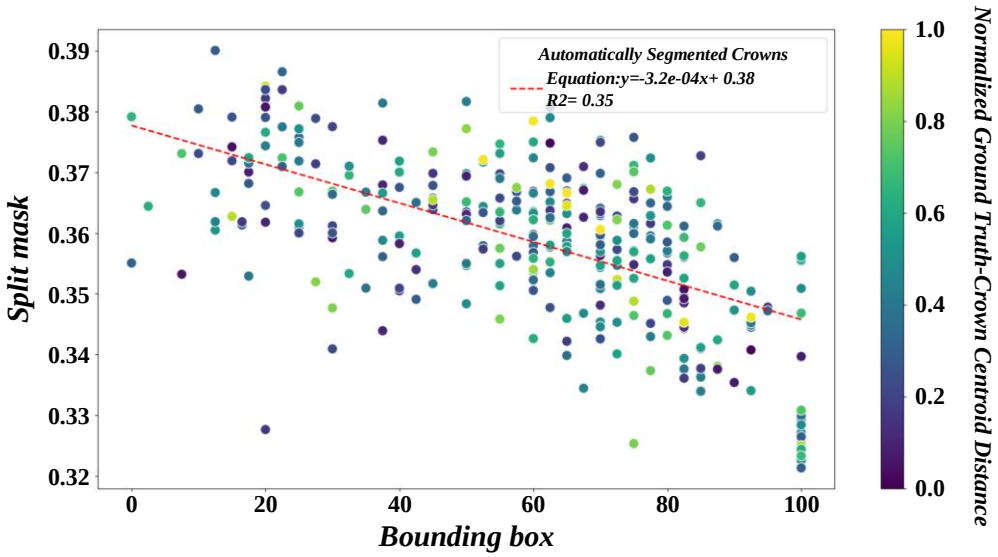

Figure 6: Test results of test set II after updating the training set

It can be seen from Table 2 that the test recognition accuracy of 1280 × 720 as the input image is slightly higher, the mAP is 0.967, and the mAP of 256 × 256 as the input image is 0.966, which is close. However, the training and testing time of 1280 × 720 as the input image is much larger, requiring 16.5 hours of training, which is 5.3 hours longer than that of small images. Therefore, this study selects a small image of 256 × 256 as the network input to ensure the recognition accuracy and reduce the time loss.

Table 3: Experimental results of original and compressed images

Input dimensions	mean Average Precision (mAP)	Training time (h)	Single frame test time (ms)
1280x720	0.967	16.464	200
256x256	0.966	11.270	160

The study defines "TRR" (True Recognition Rate) to evaluate the accuracy of model fault identification, introduces the "HITL" (Human-Computer Interaction) mode combined with expert knowledge to optimize training and inference, and adopts "LRO-NAC" (a complete model that integrates and improves FPN, MRAM, and DRAM) as the core architecture. The experimental results showed that the mAP value of 0.436 (corresponding to the AUC area under the curve in Figure 7) was significantly better than that of the baseline model under the condition of "no complex background". The combination of "HITL" mode and "LVM" (latent variable model) further improves the accuracy and interpretability of fault detection, and provides a more reliable technical path for intelligent diagnosis of electrical equipment. It can be found from Figure 7 that when the training set does not include complex background images, its mAP reaches 0.436, which, to a certain extent, maps that the algorithm has certain environmental adaptability. However, in the face of complex background images, the accuracy of insulator recognition declines significantly, and more than half of the test images cannot be effectively recognized, which fully shows that the increase of background complexity has a great impact on the recognition results, resulting in a significant decline in algorithm performance. Furthermore, when the number of complex background images in the training set is increased from 0 to 80, the mapped value is observed to increase significantly from 0.436 to 0.814. This trend reveals that increasing the number of training samples for complex background images can significantly improve the algorithm's adaptability to complex scenes, thereby improving

recognition accuracy. The analysis of insulator images shows that the similarity between simple background and complex background images is extremely low, and the features and information are significantly different. However, the similarity between complex background images is high because they share some complex features or interference factors. The difference between test samples and training samples directly affects the recognition accuracy. If the difference is large, the recognition accuracy is low, and vice versa. This finding further confirms the importance of data similarity to recognition accuracy; that is, the more similar the data, the higher the recognition accuracy.

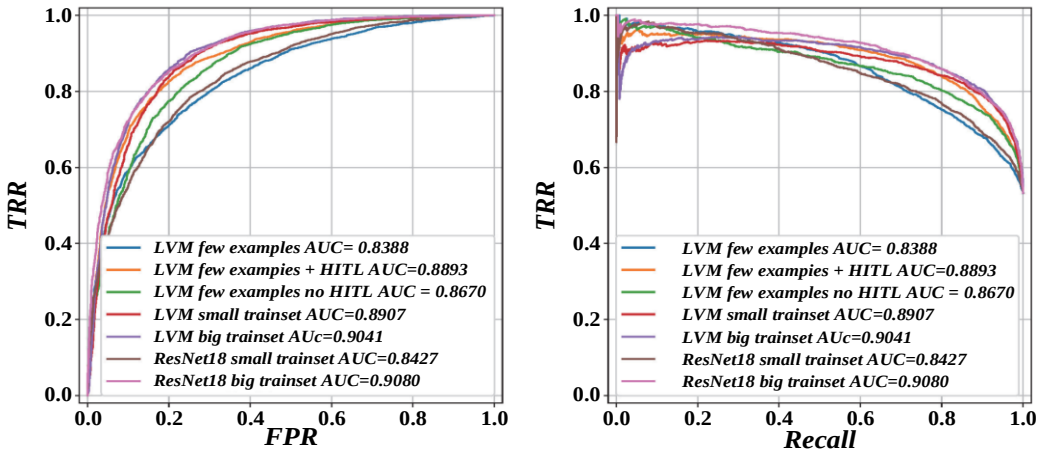

Figure 7: Test results of complex background images

"Loop[1]-Loop[3] represents a continuous human-computer interaction (HITL) iterative process, i.e., a complete sequence from the first iteration to the third iteration;" Loop[1]+Loop[2]" represents the result of the first iteration merged with the second iteration; "Summarize all iterations" refers to the comprehensive consideration of all the results, including the three iterations. It can be seen from Figure 8 that the map based on the Mask R-CNN algorithm in this paper is 0.986, which is far better than the performance of Faster R-CNN, and its accuracy is as high as 0.912. In terms of running time, although Mask R-CNN takes 150ms to process a single image, which is slightly higher than Faster R-CNN's 135ms, in the special task of insulator fault detection, the priority of accuracy far exceeds the slight gap in running time. Therefore, the increased 15ms running time of Mask R-CNN can be a negligible factor in practical applications.

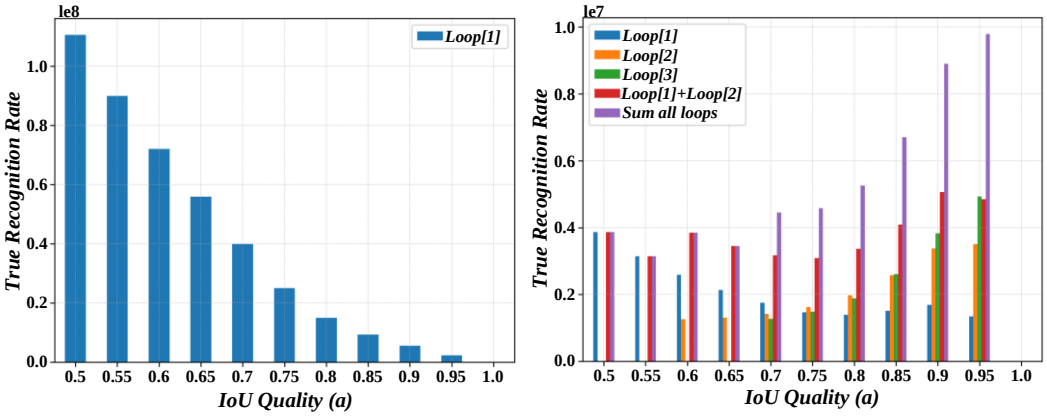

Figure 8: Comparison of experimental results of two networks

"Knet" (kernel-based network) uses kernel functions to map the feature space to improve the model's ability to characterize nonlinear fault features, while "DCGAN" (deep convolutional generative adversarial network) is used to generate diverse simulated fault samples to enrich the training dataset. This paper improves the performance of the model on the original dataset and the DCGAN enhanced dataset through comparative analysis, which verifies the effectiveness of Knet in complex fault feature extraction and the role of DCGAN in improving the generalization ability of the model, especially in dealing with the robustness of rare fault types in real scenes, which provides a more reliable technical path for the intelligent diagnosis of electrical equipment.

The same training set is used to train in the Mask R-CNN model before and after improvement, and the same verification set is verified. The evaluation index values are shown in Figure 9. It can be seen that compared

with the improvement before, the improved model has improved all evaluation indexes of the segmentation task, indicating that the improved method proposed in this paper can effectively improve the segmentation ability of the model.

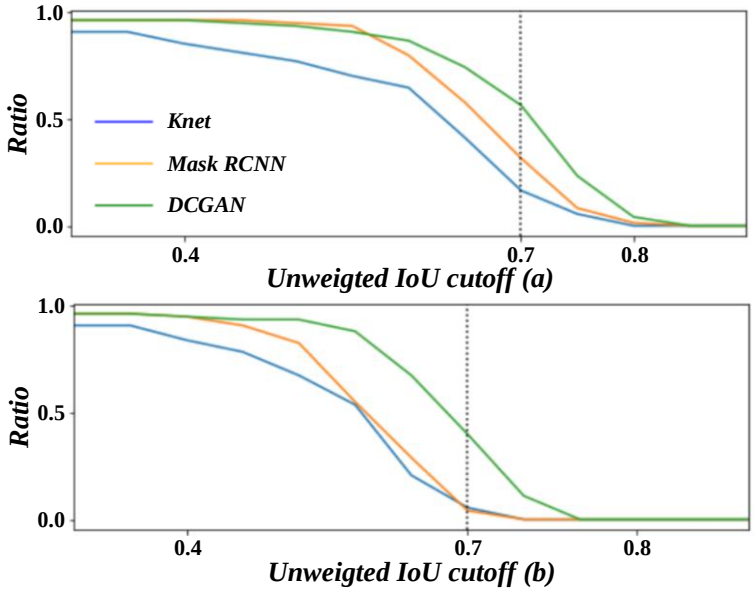

Figure 9: Improved network evaluation index

Figure 10 shows that the algorithm in this study has great advantages over the classical algorithm in detection accuracy and recall rate. The optimal APbbox is 48.6%, which is 9.2% higher than Comparedm and 5.0% higher than FPN. This indicates that the backbone network hierarchical deformable ResNet50 used in this study has greatly improved the feature extraction ability compared with VGG and ResNet50.

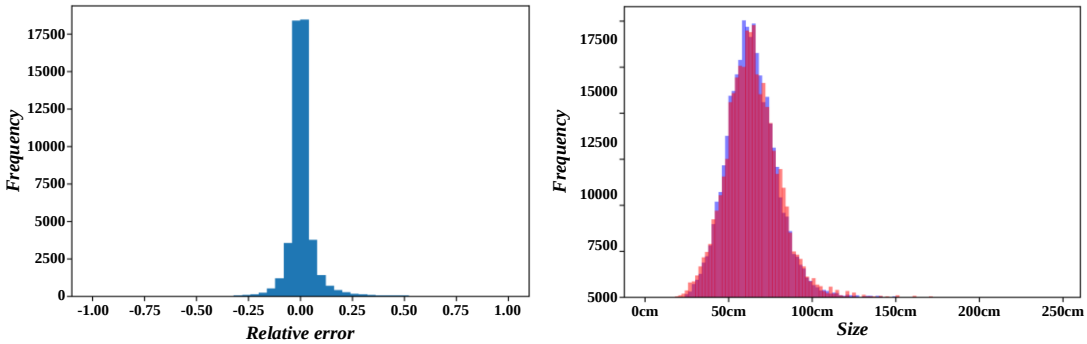

Figure 10: Distribution of relative errors and fault magnitudes in electrical equipment fault detection

Figure 11 shows the comparison of model detection indicators in detail. Observing this comparison chart, it can be clearly found that after a series of improvements, the model has achieved the best performance in six core detection performance indicators such as accuracy, recall rate, F1 score, mAP, IoU and processing speed. In particular, although it is slightly inferior to Libra R-CNN in small target accuracy, this small gap does not affect its overall performance. The results fully verify the high effectiveness and practicability of the proposed algorithm after proper optimization.

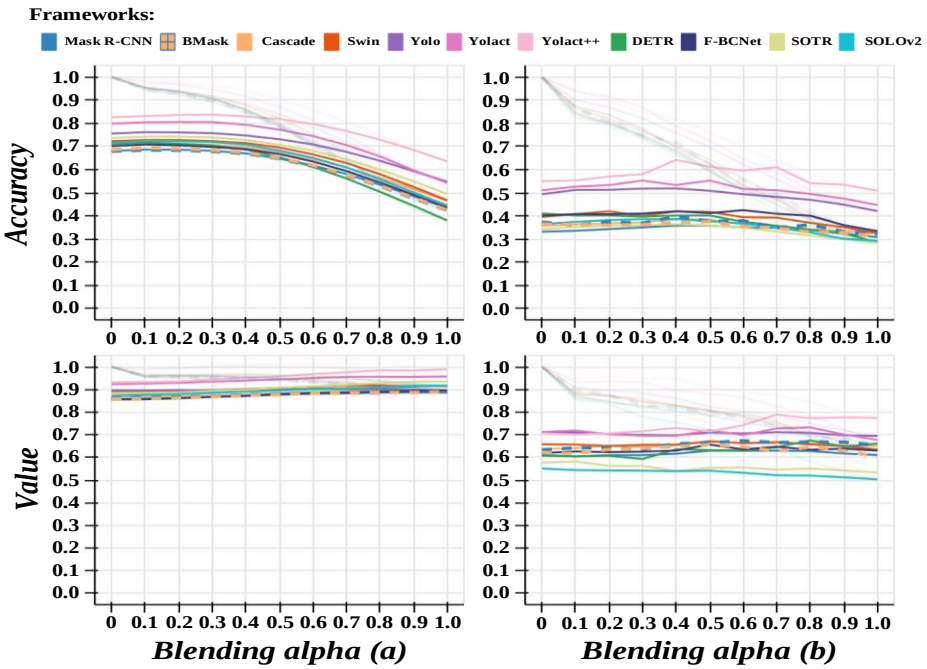

Figure 11: Comparison chart of model detection indicators

6 Conclusion

Under the background of today's rapid industrialization, it is particularly important to ensure the safe and stable operation of electrical equipment. This study optimizes electrical equipment fault diagnosis based on the improved Mask R-CNN algorithm in order to improve electrical equipment maintenance efficiency and reduce operating costs. After a series of experimental demonstrations, the research has achieved good results.

Aiming at the problem that the existing fault detection methods do not perform well in identifying small defects and coping with complex working conditions. This study is based on Mask R-CNN for targeted optimization to make full use of its excellent performance in object detection and segmentation. By introducing the improved FPN framework, multi-region attention module and differential region attention module, the model is improved.

In order to verify the effectiveness of the new model, this study collected and sorted out tens of thousands of high-definition images of various typical electrical components, including transformers, switchgear, cable joints, etc., from the real environment, covering the full cycle state from normal operation to severe damage. Moreover, conduct preliminary training based on these data. Evaluated on the independent verification set, the data shows that the fault recognition rate of the improved Mask R-CNN algorithm is about 30% higher than that of the traditional method, especially for early signs of wear, such as slight cracks and local overheating, and its detection accuracy reaches 96%.

It is found that the improved algorithm can still maintain a stable recognition effect in the face of harsh conditions such as insufficient illumination or tilted viewing angle, which proves its strong applicability in complex field environments. Compared with similar competing products, the improved model has a faster response speed, and it only takes 0.1-0.2 seconds to process an image on average, which proves the feasibility of the real-time monitoring system.

In the context of deep learning to promote the intelligent development of fault diagnosis, the optimized mask R-CNN algorithm achieves high-precision defect identification in electrical equipment fault diagnosis by virtue of structural innovation, loss function improvement and data enhancement strategy. Combined with domain adaptation and transfer learning technology, the algorithm is expected to provide a universal solution for intelligent fault diagnosis in multiple industries, breaking through the bottleneck of efficiency and accuracy of traditional detection methods.

Funding

2023 Shanxi Province Vocational Education Teaching Reform and Practice Research Project, Project Name: Research and Practice on the "Position-C-Competition-Certificate" Talent Cultivation Model for Electromechanical Majors in Higher Vocational Colleges, Project Number: 20233044

References

[1] X. Bi, J. Hu, B. Xiao, W. Li, and X. Gao, "IEMask R-CNN: Information-Enhanced Mask R-CNN," Ieee Transactions on Big Data, vol. 9, no. 2, pp. 688-700, 2023. DOI: 10.1109/TBDATA.2022.3187413.

[2] Tianyu Hu, Hongzhong Ma, Dawei Duan, and Wei Ge, "Novel GIL mechanical fault diagnosis method based on multi-sensor data feature fusion and TDEAVOA-ELM," Computers and Electrical Engineering, vol. 119, pp. 109573, 2024.

https://doi.org/10.1016/j.compeleceng.2024.109573.

[3] L. Du, X. Lu, and H. Li, "Automatic fracture detection from the images of electrical image logs using Mask R-CNN," Fuel, vol. 351, no., pp., 2023. https://doi.org/10.1016/j.fuel.2023.128992.

[4] Xiaohan Zhang, Han Wang, Chenze Wang, Min Liu, and Gaowei Xu, "Time-segment-wise feature fusion transformer for multi-modal fault diagnosis," Engineering Applications of Artificial Intelligence, vol. 138, pp. 109358, 2024.

[5] Quan Qian, Bin Zhang, Chuan Li, Yongfang Mao, and Yi Qin, "Federated transfer learning for machinery fault diagnosis: A comprehensive review of technique and application," Mechanical Systems and Signal Processing, vol. 223, pp. 111837, 2025. DOI:10.1016/j.ymssp.2024.111837.

[6] H. Wu, Y. Liu, X. Xu, and Y. Gao, "Object Detection Based on the GrabCut Method for Automatic Mask Generation," Micromachines, vol. 13, no. 12, pp., 2022. https://doi.org/10.3390/mi13122095.

[7] M. L. Francies, M. M. Ata, and M. A. Mohamed, "A robust multiclass 3D object recognition based on modern YOLO deep learning algorithms," Concurrency and Computation-Practice & Experience, vol. 34, no. 1, 2022. https://doi.org/10.1002/cpe.6517.

[8] J. Gao, P. Liu, G.-D. Liu, and L. Zhang, "Robust Needle Localization and Enhancement Algorithm for Ultrasound by Deep Learning and Beam Steering Methods," Journal of Computer Science and Technology, vol. 36, no. 2, pp. 334-346, 2021. https://doi.org/10.1007/s11390-021-0861-7.

[9] X. Xu et al., "Crack Detection and Comparison Study Based on Faster R-CNN and Mask R-CNN," Sensors, vol. 22, no. 3, 2022. https://doi.org/10.3390/s22031215.

[10] M. Xin, C. Xu, J. Gao, Y. Wang, and B. Wang, "High-Precision Recognition Algorithm for Equipment Defects Based on Mask R-CNN Algorithm Framework in Power System," Processes, vol. 12, no. 12, pp., 2024. https://doi.org/10.3390/pr12122940.

[11] H. Zhu, Y. Wang, and J. Fan, "IA-Mask R-CNN: Improved Anchor Design Mask R-CNN for Surface Defect Detection of Automotive Engine Parts," Applied Sciences-Basel, vol. 12, no. 13, pp., 2022. https://doi.org/10.3390/app12136633.

[12] J. Ding, R. Alroobaea, A. M. Baqasah, A. Althobaiti, and R. Miglani, "Big Data Intelligent Collection and Network Failure Analysis Based on Artificial Intelligence," Informatica-an International Journal of Computing and Informatics, vol. 46, no. 3, pp. 383-392, 2022. https://doi.org/10.31449/inf.v46i3.3866.

[13] N. Manakitsa, G. S. Maraslidis, L. Moysis, and G. F. Fragulis, "A Review of Machine Learning and Deep Learning for Object Detection, Semantic Segmentation, and Human Action Recognition in Machine and Robotic Vision," Technologies, vol. 12, no. 2, 2024. https://doi.org/10.3390/technologies12020015.

[14] T. Zengeya and J. Vincent Fonou-Dombeu, "A Review of State-of-the-Art Deep Learning Models for Ontology Construction," IEEE Access, vol. 12, pp. 82354-82383, 2024.

[15] F. Xue, D. Z. Zheng, and L. Yan, "Fault diagnosis of distributed discrete event systems using OBDD," Informatica, vol. 16, no. 3, pp. 431-448, 2005. https://doi.org/10.3233/INF-2005-16(3)09.

[16] K. Sareen, B. K. Panigrahi, T. Shikhola, and A. Chawla, "A robust De-Noising Autoencoder imputation and VMD algorithm based deep learning technique for short-term wind speed prediction ensuring cyber resilience," Energy,vol. 283, 2023. DOI: 10.1016/j.energy.2023.129080.

[17] Runye Shi et al., "Research on a self-powered rolling bearing fault diagnosis method with a piezoelectric generator for self-sensing," Applied Energy, vol. 376, pp. 124206, 2024. DOI: 10.1016/j.apenergy.2024.124206.

[18] K. Sandkuhl, "Quantified Products: Case Studies, Features and their Design Implications," Informatica, vol. 34, no. 4, pp. 825-845, 2023. https://doi.org/10.15388/23-INFOR532.

[19] I. Hector and R. Panjanathan, "Predictive maintenance in Industry 4.0: a survey of planning models and machine learning techniques," Peerj Computer Science, vol. 10, 2024. DOI: 10.7717/peerj-cs.2016.

[20] E. Jove, J.-L. Casteleiro-Roca, H. Quintian, J.-A. Mendez-Perez, and J. Luis Calvo-Rolle, "Virtual Sensor for Fault Detection, Isolation and Data Recovery for Bicomponent Mixing Machine Monitoring," Informatica, vol. 30, no. 4, pp. 671-687, 2019. https://doi.org/10.3233/INF-2019-1241.

[21] Yongjian Sun and Wei Wang, "Role of image feature enhancement in intelligent fault diagnosis for mechanical equipment: A review," Engineering Failure Analysis, vol. 156, pp. 107815, 2024. DOI: 10.1016/j.engfailanal.2023.107815.

[22] Thanh-Tung Vo, Meng-Kun Liu, and Minh-Quang Tran, "Harnessing attention mechanisms in a comprehensive deep learning approach for induction motor fault diagnosis using raw electrical signals," Engineering Applications of Artificial Intelligence, vol. 129, pp. 107643, 2024. https://doi.org/10.1016/j.engappai.2023.107643.

[23] Song Wang, Tenghao Ma, Jigang Feng, Shuai Gao, Qinkai Han, and Fulei Chu, "Comb-shaped single-electrode triboelectric roller chain with speed sensing and fault diagnosis capabilities," Mechanical Systems and Signal Processing, vol. 224, pp. 111952, 2025. DOI: 10.1016/j.ymssp.2024.111952.

[24] Yu Wang, Dexiong Li, Lei Li, Runde Sun, and Shuqing Wang, "A novel deep learning framework for rolling bearing fault diagnosis enhancement using VAE-augmented CNN model," Heliyon, vol. 10, no. 15, pp. e35407, 2024. DOI: 10.1016/j.heliyon.2024.e35407.

[25] M. Wu et al., "Object detection based on RGC mask R-CNN," Iet Image Processing, vol. 14, no. 8, pp. 1502-1508, 2020. https://doi.org/10.1049/iet-ipr.2019.0057.

[26] Jianan Wei, Hualin Chen, Yage Yuan, Haisong Huang, Long Wen, and Weidong Jiao, "Novel imbalanced multi-class fault diagnosis method using transfer learning and oversampling strategies-based multi-layer support vector machines (ML-SVMs)," Applied Soft Computing, vol. 167, pp. 112324, 2024. https://doi.org/10.1016/j.asoc.2024.112324.

[27] Yuepeng Xin, "Insulation aging and fault diagnosis of high-voltage equipment based on deep Q-networks," Measurement: Sensors, vol. 31, pp. 101016, 2024. DOI: 10.1016/j.measen.2023.101016.

[28] E. Kuk, S. Bobek, and G. J. Nalepa, "Explainable proactive control of industrial processes," Journal of Computational Science, vol. 81, 2024.

[29] M. Sahu, and D. P. Mohapatra, "Computing Dynamic Slices of Feature-Oriented Programs with Aspect-Oriented Extensions," Informatica-an International Journal of Computing and Informatics, vol. 44, no. 2, pp. 199-224, 2020. https://doi.org/10.31449/inf.v44i2.2452.

[30] A. I. Stepanova, A. I. Khalyasmaa, P. V. Matrenin, and S. A. Eroshenko, "Application of SHAP and Multi-Agent Approach for Short-Term Forecast of Power Consumption of Gas Industry Enterprises," Algorithms, vol. 17, no. 10, 2024. https://doi.org/10.3390/a17100447.

[31] Qiuyu Yang et al., "A novel zero-shot learning approach for cross-domain fault diagnosis in high-voltage circuit breakers," Advanced Engineering Informatics, vol. 62, pp. 102777, 2024. https://doi.org/10.1016/j.aei.2024.10277.

[32] Ke Yue, Jipu Li, Shuhan Deng, Chee Keong Kwoh, Zhuyun Chen, and Weihua Li, "A relationship-aware calibrated prototypical network for fault incremental diagnosis of electric motors without reserved samples," Reliability Engineering & System Safety, vol. 252, pp. 110429, 2024. DOI: 10.1016/j.ress.2024.110429.

[33] Ming Zhang, Xin Xu, Likun Xu, Xiaoming Ruan, Duan Zhou, and Zhenshan He, "Fault diagnosis of accelerometer servo circuit output saturation based on feature electrical parameters extraction," Heliyon, vol. 10, no. 7, pp. e28382, 2024. DOI: 10.1016/j.heliyon.2024.e28382.