Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Monitoring Algorithm for Substation Equipment Based on OPGW Communications

Zibin Yu

Yuzbb743@163.com

Shenzhen Power Supply Bureau

Huan Zhao

Shenzhen Power Supply Bureau

Hao Yang

China Southern Power Grid (China)

Xuefeng Wang

Shenzhen Power Supply Bureau

Jinlong Tao

Shenzhen Power Supply Bureau

Mingzhang Fang

Shenzhen Power Supply Bureau

Shaowei Zhang

Shenzhen Power Supply Bureau

Yashan Tian

Shenzhen Power Supply Bureau

Feihong Diao

Shenzhen Power Supply Bureau

Research Article

Keywords: Substation equipment inspection, small-scale targets, multiple self-attention, equipment monitoring, OPGW communications

Posted Date: March 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8986909/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

A Monitoring Algorithm for Substation Equipment Based on OPGW Communications

Zibin Yu[1,*], Huan Zhao[1], Hao Yang[2], Xuefeng Wang[1], Jinlong Tao[1], Mingzhang Fang[1], Shaowei Zhang[1], Yashan Tian[1], Feihong Diao[1]

1. Shenzhen Power Supply Bureau Co. Ltd., Shenzhen 518038, China.

2. China Southern Power Grid Co., Ltd., Guangzhou 510450, China.

*e-mail: Yuzbb743@163.com

Abstract: As the electricity demand continues to grow in a digital society, substations, as a key part of the grid infrastructure, face many challenges to the reliability and stability of their equipment. Traditional substation equipment monitoring methods such as manual and helicopter inspection suffer from slow speed, high cost, and safety hazards, making it difficult to meet real-time monitoring and efficient detection demand. In recent years, machine learning-based target detection techniques have shown great potential for automatic classification and identification of substation equipment. However, the existing methods still have the problems of insufficient feature extraction and low positioning accuracy in small-scale target detection, especially under the influence of interfering factors in complex environments, such as occlusion, light change, and scale change, the detection effect is further degraded. Aiming at the above problems, this paper proposes an improved YOLOv8 target detection model, which is dedicated to the detection of substation equipment in complex transmission and distribution line scenarios. Specifically, a multi-head self-attention (MHSA) module is introduced in front of the backbone network of YOLOv8 to enhance the feature extraction capability of the network and improve the detection accuracy of small-scale targets. Secondly, the P2 detection layer is designed and integrated to specifically address the detection needs of small-scale targets, which significantly improves the detection rate and positioning accuracy of substation equipment. Through experimental validation in the actual complex transmission and distribution line environment, the proposed detection model shows superior detection performance and robustness, effectively improves the fault detection capability of substation equipment, and ensures the safe and stable operation of the power system. The results show that the proposed model has significant advantages in substation equipment monitoring and has a wide range of application prospects.

Keywords: Substation equipment inspection; small-scale targets; multiple self-attention; equipment monitoring; OPGW communications

Research funding declaration: This work was supported by grants from the Technology Project Foundation of China Southern Power Grid (grant number: SZKJXM20222171).

Ethics and Consent to Participate declarations: not applicable.

1. Introduction

As the demand for electricity in our increasingly digital society continues to grow, substations, which are critical components of power grid infrastructure, face escalating challenges in terms of reliability and stability. As key nodes in the power system, substations play a crucial role in the transmission and distribution of electricity, ensuring efficient power supply. Therefore, maintaining the normal operation and high reliability of substation equipment is essential for the safety of power supply [1]. However, the diverse characteristics of lightning in nature exert complex influences on the performance of substation equipment. These characteristics contribute to cumulative effects on energy absorption and temperature control, accelerating the aging process of internal components and potentially leading to thermal runaway, grounding faults, and even equipment damage or explosions. In extreme cases, busbar short circuits can severely threaten the safety and stability of the power system. Thus, real-time monitoring of substation equipment operations is of paramount importance. To promptly alert on-site operators and maintenance personnel and implement appropriate maintenance actions, it is necessary to enhance traditional leakage current monitoring technologies and incorporate external condition assessment methods for substation equipment.

In the context of monitoring power distribution lines based on the Optical Fiber Composite Overhead Ground Wire (OPGW) communication, OPGW not only facilitates power transmission but also integrates optical fiber communication capabilities, enabling efficient data transmission [2-3]. Within transmission lines, substation equipment is interconnected through methods such as welding, fulfilling critical roles in electrical isolation and mechanical support, ensuring reliable line operation. However, substation equipment is often exposed to outdoor environments for extended periods, making it susceptible to electrical, mechanical, and thermal stresses. These stresses can lead to increased surface resistance, reduced flashover voltage, and decreased puncture strength, resulting in cracks at the connection points of the equipment. Such cracks not only compromise the operational stability of the power system but also pose substantial economic risks. Therefore, timely detection and replacement of defective substation equipment are crucial for ensuring the safe operation of the power system. However, conventional monitoring methods such as manual inspections and helicopter surveys suffer from slow speed, high costs, and safety risks, relying heavily on human visual inspection, which limits detection efficiency and accuracy. Additionally, monitoring with climbing robots may damage power lines due to the equipment's weight. Hence, there is an urgent need to develop more efficient and safer monitoring methods to enhance the protection levels of substation equipment in transmission systems.

To address the shortcomings of traditional monitoring methods, machine learning technologies have been extensively applied in recent years for the automatic classification and recognition of substation equipment [4-6]. Machine learning algorithms achieve classification by extracting features from images. However, traditional machine learning approaches

often rely on manually designed features, which limit the stability and generalization of detection results. The introduction of Convolutional Neural Networks (CNNs) has been pivotal in addressing this issue, as CNNs can automatically learn and extract features such as textures, shapes, and patterns from images, significantly improving the accuracy of equipment recognition [7-9]. Additionally, image preprocessing techniques enhance CNN model performance by improving image quality and reducing noise interference. Simultaneously, single-stage object detection models like the YOLO series have revolutionized the field of object detection by significantly enhancing detection speed and efficiency [10-12]. With the deep integration of artificial intelligence and embedded systems, the real-time detection and monitoring capabilities of transmission equipment have greatly improved, particularly in the automatic detection of high-voltage substation equipment in remote areas, significantly reducing the need for manual inspections and enhancing the overall reliability of distribution networks.

Despite significant advancements in the field of object detection, small-scale object detection remains challenging [13-15]. First, small-scale obiects occupy limited areas in images, have low resolution, and contain less effective feature information, making it difficult for existing detection models to extract highly discriminative features. Moreover, small-scale obiects require high localization accuracy, and further complicate small-scale object detection [16-18].

To address these challenges, this paper proposes an improved model for detecting substation equipment in complex transmission and distribution line scenarios. Leveraging the advantages of the YOLOv8 detection model for small-scale object detection tasks, we have made specific optimizations. Specifically, we introduce a Multi-Head Self-Attention (MHSA) module into the backbone network of YOLOv8 to enhance feature extraction capabilities while maintaining detection accuracy. Additionally, we introduce a P2 detection layer tailored for small-scale objects and integrate it into the YOLOv8 network to improve the detection rate of substation equipment targets. These enhancements enable the proposed detection model to exhibit superior detection performance and robustness in complex transmission and distribution environments, effectively improving fault detection capabilities for substation equipment and ensuring the safe and stable operation of power systems. The main contributions of this paper are summarized as follows:

- Improvement of YOLOv8 Architecture: By ntroducing the MHSA attention module into the backbone network of YOLOv8, we ignificantly enhance feature extraction capabilities and improve detection accuracy for small-scale objects.
- Introduction of P2 Detection Layer: A P2 detection layer is designed and integrated to address the small-scale object detection problem, further improving the detection rate and localization accuracy of substation equipment.
- Comprehensive Performance Evaluation: The proposed detection model is comprehensively evaluated in complex transmission and distribution environments, validating its efficiency and robustness in practical applications.

The remainder of this paper is organized as follows. Section 2 describes the proposed improved model in detail; Section 3 presents the experimental setup and evaluation metrics; Section 4 concludes the paper with a summary of the main findings.

2. Elaboration of methodology

As shown in Fig. 1, the YOLO-MHSA substation detection model proposed in this paper consists of four parts: input, backbone network (Backbone), neck network (Neck), and output (Head). The input side is enhanced by using data augmentation to increase the complexity of the dataset and thus reduce the risk of overfitting.

2.l Core network

The backbone network is responsible for extracting multi-scale features from the input mage.YOLOv8 employs an improved CSPDarknet rchitecture, in which the CSP (Cross Stage Partial) module effectively enhances the gradient flow and feature representation by segmenting and fusing the feature maps.

2.1.1 Convolution operation

The basic building block of the backbone network is the convolutional layer. For the input feature map $F_{in} \in {}^{C_{in} \times H \times W}$, the convolution operation computes the output feature map F_{out}, through the filter $K \in {}^{C_{out} \times C_{in} \times k \times k}$, The computation process is shown in Equation (1).

$$F_{out}(c,i,j) = \sum_{c'}^{C_{in}} \sum_{m=1}^{k} \sum_{n=1}^{k} K(c,c',m,n) F_{in}(c', i+m, j+n) \quad (1)$$

where c denotes the output channel, i and j denote the spatial locations, and k×k is the size of the convolution kernel.

2.1.2 CSP module

The core idea of the CSP module is to segment the feature map into two parts, which are subjected to independent convolutional processing before fusion. Let the input feature map be F. F_1 and F_2 are obtained after segmentation. Where F_1 is processed through a series of convolutional layers to obtain F1 ,which is then spliced with F_2 .The computational process is shown in Equation (2). This mechanism of segmentation and fusion effectively reduces the repeated gradient information and improves the learning efficiency of the model.

$$F' = Concat(F_1', F_2) \quad (2)$$

2.1.3 BottleneckCSP lump

In CSPDarknet, the BottleneckCSP block is the main feature extraction unit. Each BottleneckCSP block consists of multiple Bottleneck layers and a CSP connection. Specifically, assume that the input is F and F ' is obtained after one Bottleneck layer. Then we fuse with the original feature F through the CSP connection, which is calculated as shown in Equation (3).

$$F_{out} = Concat(Bottleneck(F), F) \quad (3)$$

2.1.4 Feature extraction process

The backbone network progressively extracts high-level semantic features through multilayer BottleneckCSP blocks. Let the input image size be H×W. After network processing, the output multi-scale feature map $F_1 \times F_2 \times F_3$, corresponding to different spatial resolutions （H/8) ×W/8), （H/16) ×W/16), and （H/32)×W/32). These feature maps capture the target information at different scales, which helps to detect whether the current state of substation equipment of different sizes is running well.

2.2 Neck network

The neck network is responsible for fusing the multi-scale

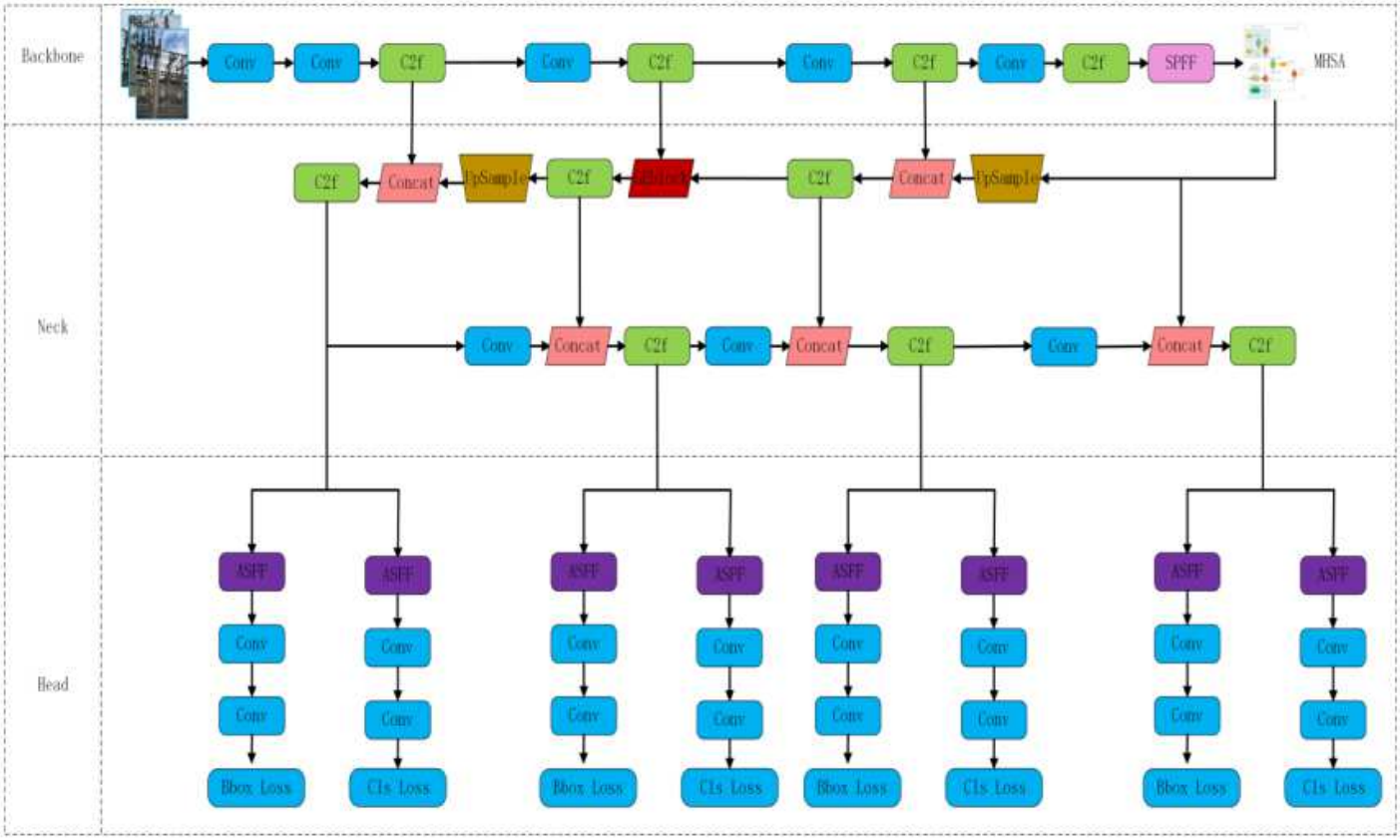


Fig. 1 YOLO-MHSA architecture diagram.

features extracted from the backbone network to further enhance the richness and semantic information of the feature representation. YOLOv8 adopts a combined structure of Feature Pyramid Network (FPN) and Path Aggregation Network (PANet) to realize the efficient fusion of features at different scales.

2.2.1 Characteristic feature pyramid network

FPN gradually fuses high-level semantic features into low-level spatial features through top-down paths and lateral connections. Let the high-level feature be F_3, and $\hat{F}_3$ is obtained through the up-sampling operation, which is shown in Equation (4).

$$\hat{F}_3 = Upsample(F_3) \tag{4}$$

Based on splicing with the low-level feature F_2 , as shown in Equation (5)

$$F_2^{'} = Concat\left(\hat{F}_3, F_2\right) \tag{5}$$

After further processing by the convolutional layer, the fused feature $F_2^{'}$ is obtained. Similarly, the fusion of $F_2^{'}$ with F_1 is continued to obtain the final FPN output.

2.2.2 Path aggregation network

PANet adds bottom-up paths to FPN to further enhance the semantic information of features through gradual down-sampling. Let the output of the FPN be F_{detect} . After a series of down-sampling and convolution operations, a further enhanced feature map is obtained as shown in Equation (6).

$$F_{pan} = PANet\left(F_{detect}\right) \tag{6}$$

Ultimately, the fused feature map F_{fusion} output from the neck network can be calculated using Equation (7).

$$F_{fusion} = PANet\left(FPN\left(F_1, F_2, F_3\right)\right) \tag{7}$$

2.3 Detector head

The detection head is responsible for target classification and bounding box regression on the fused feature maps. The detection head of YOLOv8 employs an optimized convolutional layer design that predicts the class probabilities and bounding box parameters of multiple anchors for each spatial location.

2.3.1 Anchor mechanism

For each feature map location (i,j) , the model predicts multiple a priori boxes (anchor boxes). Each anchor prediction contains the classification probability P_c and the bounding box regression parameter (t_x, t_y, t_w, t_h).

2.3.2 Bounding box decoding

The bounding box regression parameters are converted to actual bounding box coordinates (x ,y ,w ,h) . Let the step size of the feature map be S, the center coordinate of the current grid cell be (C_x ,C_y), and the width and height of the a priori box be (p_w ,p_h). Then the decoding process is shown in Equation (8).

$$\begin{cases} x = \sigma(t_x) \cdot S + C_x \\ y = \sigma(t_y) \cdot S + C_y \\ w = p_w \cdot e^{t_w} \\ h = p_h \cdot e^{t_h} \end{cases} \tag{8}$$

where σ is the Sigmoid activation function, ensuring that the values of x and y are within reasonable limits.

2.3.3 Multi-scale detection

YOLOv8 performs detection independently on feature maps at different scales. It is responsible for detecting large-size

targets and focuses on small-size targets when it is responsible for medium-size targets. The detection head at each scale outputs the corresponding bounding box and classification probability, and finally the results are fused and filtered by a non-maximal suppression (NMS) algorithm.

2.3.4 Model Optimization

Gradient propagation: The CSP module reduces the problem of gradient vanishing by feature segmentation and fusion, which ensures the effective training of the deep network. The gradient propagation is calculated as shown in Equation (9).

$$\frac{\partial L}{\partial \theta}=\sum_{i}\frac{\partial L}{\partial F_i}\frac{\partial F_i}{\partial \theta} \quad (9)$$

Boundary box regression accuracy: Through the design of the CIoU loss function, the model not only optimizes the overlap rate of the bounding box but also considers the centroid distance and width-to-height ratio to improve the positioning accuracy. Specifically, the calculation process of the CIoU loss function is shown in Equation (10).

$$\frac{\partial CIoU}{\partial x}=\frac{\partial IoU}{\partial x}+\frac{\partial}{\partial x}\left(\frac{\rho^2\left(b_p,b_g\right)}{c^2}\right)+\frac{\partial(\alpha v)}{\partial x} \quad (10)$$

Where b_p is the Predicted Bounding Box (PBF), which indicates the position and size of the detected target; bg is the Ground Truth Bounding Box (GDB), which serves to label the true position and size of the target in the data; α is the weighting factor, which is used to balance the contribution of the width-height ratio consistency loss v in the overall loss to ensure a reasonable trade-off between different loss terms. The calculation of v is shown in Equation (11).

$$v=\frac{4}{\pi^2}\left(\arctan\left(\frac{w_g}{h_g}\right)-\arctan\left(\frac{w_p}{h_p}\right)\right)^2 \quad (11)$$

where $\rho^2(b_p,b_g)$ is the Euclidean distance between the predicted bounding box and the center point of the real bounding box, which is calculated as in Equation (12).

$$\rho\left(b_p,b_g\right)=\sqrt{\left(x_p-x_g\right)^2+\left(y_p-y_g\right)^2} \quad (12)$$

where (x_p, y_p) and (x_g, y_g) are the center coordinates of the prediction frame and the real frame, respectively. c is the diagonal length of the minimum closed frame of the predicted and real bounding boxes, which is calculated as shown in Equation (13).

$$c=\sqrt{\left(x_{\max}-x_{\min}\right)^2+\left(y_{\max}-y_{\min}\right)^2} \quad (13)$$

where (x_{min}, y_{min}) and (x_{max}, y_{max}) are the coordinates of the lower-left and upper-right corners of the minimal closure box, respectively.

2.4 Improved YOLOv8 model

2.4.1 Attention Module

With the continuous progress of deep learning technology, the attention mechanism has gradually become an important component in neural networks [19]. The introduction of the attention mechanism is inspired by the working principle of the human visual system, i.e., when observing things, human beings will pay priority attention to the part of information that is crucial for judgment and recognition, while ignoring the unimportant information, to amplify the important features and suppress the irrelevant features.

To improve the monitoring accuracy of substation equipment, this paper optimizes the YOLOv8 algorithm and proposes the YOLO-MHSA algorithm, whose core module is shown in Fig. 2. The core of the self-attention mechanism is to generate a new representation by calculating the correlation between the elements in the input sequence. The Multi-Head Self-Attention (MHSA) mechanism captures richer feature representations by enabling the model to learn information from different subspaces through multiple attention heads in parallel [20].

The self-attention mechanism is usually a single attention head. Its computation can be formalized as Equation (14)

$$Attention(Q,K,V)=soft\max\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (14)$$

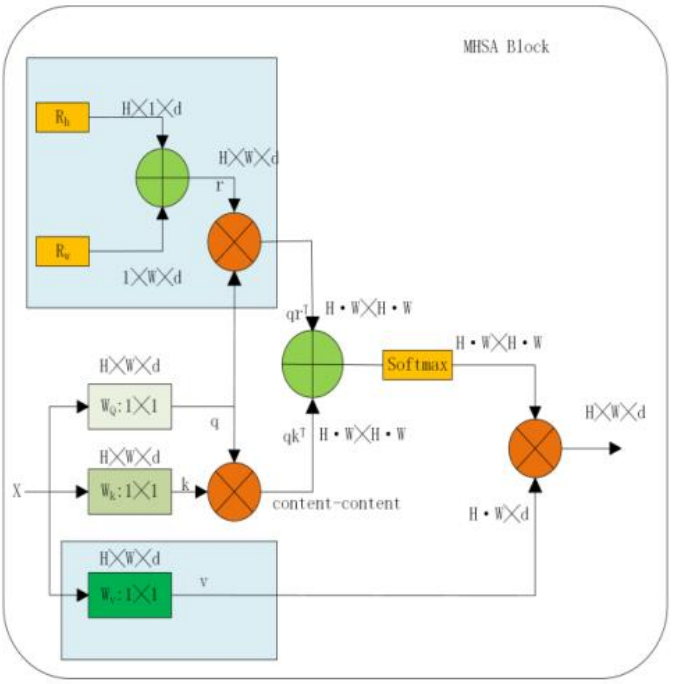


Fig. 2 Schematic diagram of MHSA module.

Algorithm 1. YOLO-MHSA

Input: Substation Images
Output: Test results

1 Core network
2 Initial Convolutional Layer
3 Repeat the following modules to extract multi-scale features
4 CSP module
5 Downsampling
6 Output multi-scale feature maps: $F_1\times F_2\times F_3$
7 Neck network
8 Characteristic pyramid networks: $F_{pan}=PANet\left(F_{dectect}\right)$
9 Path aggregation network: $F_{fusion}=PANet\left(FPN\left(F_1,F_2,F_3\right)\right)$
10 Detector head
11 Perform the following operations for each fused feature map F_{fusion}
12 Classification projections
13 Boundary box regression
14 Decode the bounding box
15 reprocess
16 Combining test results from all scales
17 Applying non-maximal suppression (NMS)
18 Return: Final test results

As shown in Fig. 2 [20], MHSA (Multi-Head Self-Attention) has multiple attention heads, and the purpose of the multi-head mechanism is to allow the model to learn information from different subspaces. Assuming that there are h attention heads, the MHSA computation steps are as follows.

Parallel computation of attention scores for multiple heads:

$$head_i=Attention(Q_i,K_i,V_i)i=1,2,\ldots,h \quad (15)$$

where each header has its corresponding weight.

Splicing the output of all heads is as follows:

$$MultiHead = Concat(head_1, head_2, \cdots, head_h) \quad (16)$$

As shown in Fig. 2, YOLO-MHSA aims to enhance the contextual information of the feature representation after introducing MHSA into the SPPF layer, so that the model can better understand the object relationships and global features in the image [20]. Combining the multi-scale features of SPPF with the global contextual information of MHSA can significantly improve the accuracy and robustness of target detection.

As shown in Algorithm 1, the steps of the YOLO-MHSA algorithm are given.

2.4.2 Introducing a small-scale target detection layer

Small-scale target detection is particularly important in practical applications, but its detection accuracy often depends on the effective fusion of high-level detail information and low-level semantic information. The traditional YOLOv8 network backbone loses detail information after multiple up-sampling, limiting the effectiveness of small-scale target detection. For this reason, this paper introduces a P2 small-scale target detection layer to enhance the multi-scale receptive field and extract global and local features of the image, to improve the detection rate of small equipment targets present in the substation [21].

3. Experimental and Results

3.1 experimental data

The substation equipment data set used in this paper covers 15 target objects, and the specific categories are shown in Table 1. The dataset contains 7539 substation images and 213566 annotated objects. Images come from a variety of sources, including mobile phone cameras, panoramic aerial cameras, stereo infrared thermal cameras, and cameras mounted on automated guided vehicles [22].

Table 1 Category and quantity of substation equipment data set

form	quantities
Open blade disconnect switch	1117
Closed blade disconnect switch	26068
Open tandem disconnect switch	4263
Closed tandem disconnect switch	4803
Fuse disconnect switch	1925
Glass disc insulator	13803
Porcelain pin insulator	114578
Muffle	8128
Lightning arrester	8788
Recloser	8059
Power transformer	2391
Current transformer	9293
Potential transformer	2904
Tripolar disconnect switch	2044

All the images in the data set come from a power distribution station, which took two years to collect. These images cover different periods, different weather conditions, and different seasons of the day, ensuring that the captured target objects exhibit different characteristics in a variety of light environments. All images are organized by the electrical engineering expert group to ensure that the shooting angle and distance of the target object in the image are suitable for automatic detection tasks.

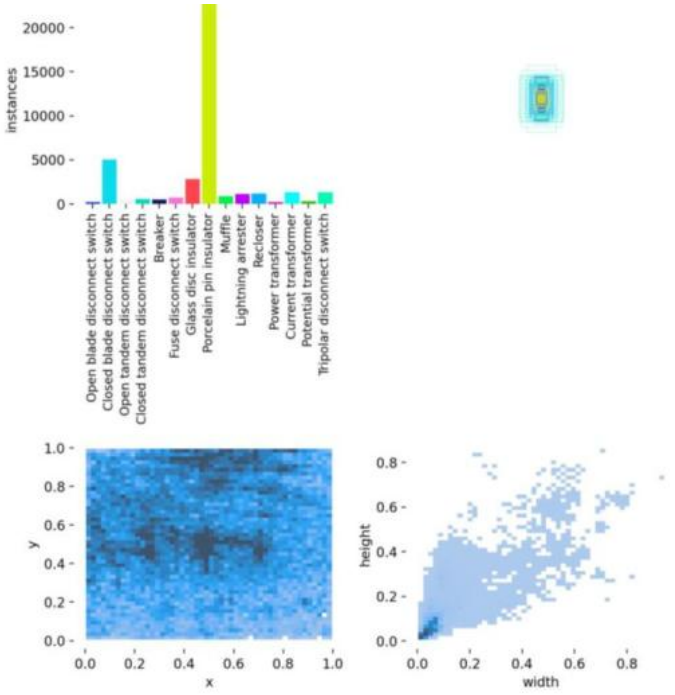


Fig. 3 Visual distribution of dataset labels.

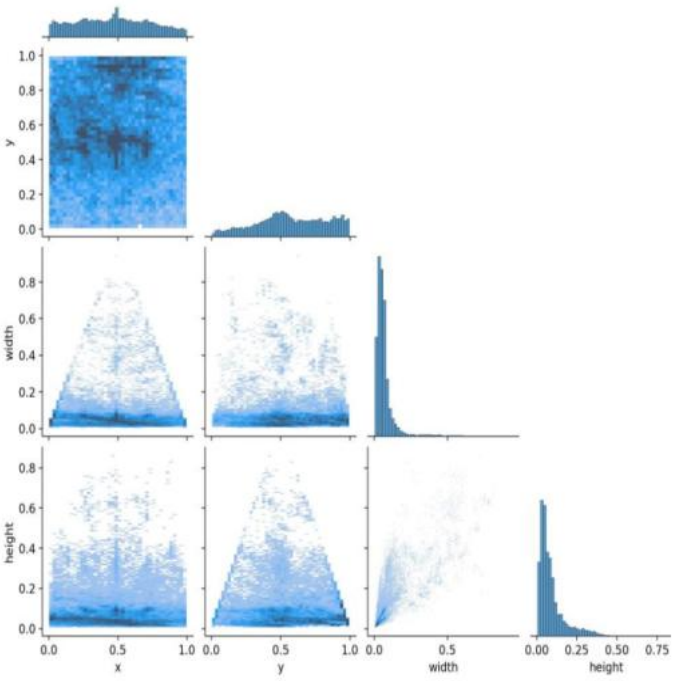


Fig. 4 Distribution of data sets used in substation equipment detection training process.

Fig. 3 shows the visual distribution map of the label of the data set, which intuitively describes the distribution of different types of samples in the data set. The upper left corner presents the number of samples for each target category in the dataset, provides an overview of the overall data distribution, and highlights the sample balance between categories. The lower left corner depicts the distribution of the center point of the object, emphasizing the focus and position information of the sample. The bounding boxes in the data set are mainly concentrated in the lower left corner, indicating that most of the substation's small equipment.

In addition, this paper analyzes all the images in the substation equipment data set in detail, mainly focusing on the location and size distribution of the annotation bounding boxes

of various types of equipment. Fig. 4 shows the distribution of data sets used in the training process of substation equipment detection. The coordinate value is normalized, where the coordinate value 0 corresponds to the upper left corner of the image, and the coordinate value 1 corresponds to the lower right corner of the image.

From Fig. 4, it can be seen that the x-coordinate distribution of the equipment target box is relatively uniform, while the y-coordinate is mainly concentrated in the central area of the image, and the width and height of the target box are normally distributed. In addition, there are more dense points in the area with smaller width and height in the image, which indicates that there are specific commonalities in the size or position of the equipment target in the data set. This aggregation feature helps the model to learn the position and size distribution of the target more accurately, thereby improving the detection accuracy.

3.2 Experimental Environment and Parameter Setting

This experiment is carried out under the Windows 10 operating system. The hardware environment is configured as AMD Ryzen 7 1800 X eight-core processor and NVIDIA GeForce GT 710 GPU. The running memory is 16 GB, the system type is the 64-bit operating system, and the processor is based on x64 architecture. The software environment uses Python 3.8, PyTorch 1.12.1, CUDA 11.3, and cuDNN acceleration library 8.3.0. The parameter settings of model training are shown in Table 2.

Table 2 The model training parameter setting

parameters	numerical value
Initial learning rate	0.01
momentum	0.931
weight decay	0.0004
round	200
batch size	8
IoU threshold	0.7
Bounding box loss weight (box)	0.05
Categorized loss weights (cls)	0.5
Distribution focus loss weight (dfl)	0.1

3.3 Evaluating Indicator

To comprehensively evaluate the performance of the proposed algorithm, this paper uses a variety of evaluation indicators, including Mean Average Precision (mAP), Confusion Matrix, F1-score, Precision, and Recall. These indicators can reflect the classification effect and stability of the model from different perspectives.

mAP: mAP is one of the most commonly used evaluation indicators in the field of target detection, which is used to measure the overall performance of the model in the detection task. Considering the accuracy and recall rate of the model, mAP is the key standard to evaluate the effect of the target detection model.

Confusion matrix: Confusion matrix is a basic tool for evaluating the performance of classification models. By displaying the prediction of the model on various categories, it can intuitively reflect the correctness and error of the model. The confusion matrix is displayed through the heat map, and the prediction performance of the model on different categories can be visually observed. The depth of the color in the heat map usually represents the number of different samples. The deeper the color, the more the number of corresponding samples, which is convenient to identify the advantages and disadvantages of the model.

F1-score, Precision, and Recall: These indicators are important indicators for comprehensive evaluation of the performance of classification models, especially in the case of category imbalance. F1-score combines precision and recall, reflecting the ability of the model to balance the two. Precision measures the proportion of the actual positive class in the samples predicted by the model, and Recall measures the proportion of the actual positive class samples that the model can detect. These indicators can help to understand the performance of the model on different categories, as well as the advantages and disadvantages of the model in dealing with unbalanced data. The evaluation formula is as follows:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{17}$$

$$Recall = \frac{TP}{TP + FN} \tag{18}$$

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{19}$$

3.4 Experimental Results and Analysis

3.4.1 Comparative Analysis

To assess the effectiveness of the introduced MHSA mechanism, we conducted comparative experiments between the YOLO-MHSA model and other mainstream YOLO baseline models. The results, as detailed in Table 3, highlight the superior performance of the YOLO-MHSA model. Specifically, the model achieved a precision of 93.3% and a recall of 85.8%, culminating in a mAP of 91.1%. This performance surpasses the baseline YOLOv9t by 0.5% in precision and mAP and provides improvements of 1% and 1.3% over YOLOv9t and YOLOv11 respectively. Despite the YOLOv10th model having the highest parameter count, its mAP was lower than that of YOLOv9t. These findings underscore the efficacy of incorporating the MHSA mechanism within YOLOv8s, effectively leveraging channel and spatial features to enhance model performance.

Table 3 Comparative experiments with different baseline models

model	Precision/% (Average of 15 categories)	Recall/% (Average of 15 categories)	mAP/% (Average of 15 categories)	Parameters
YOLO-MHSA(our)	93.3	85.8	91.1	3211149
YOLOv9t[23]	92.8	84.9	90.6	2008333
YOLOv10n[24]	91.8	83.7	90.1	2712890
YOLOv11[25]	90.6	82.9	89.8	2592765

3.4.2 Quantitative Analysis

The confusion matrix in Fig. 5 provides a detailed overview of the YOLO-MHSA model' s performance on the substation detection dataset. Categories such as "Closed Blade Disconnect Switch" (0.78), "Open Tandem Disconnect Switch" (0.80), and "Lightning Arrester" (0.84) exhibit high prediction accuracy,

demonstrating the model's robust recognition capabilities for these categories. However, the confusion between "Glass Disc Insulator" and "Porcelain Pin Insulator" (actual 0.77 and 0.87 respectively) suggests morphological and color similarities that warrant further attention during data collection and model training to bolster power inspection operations.

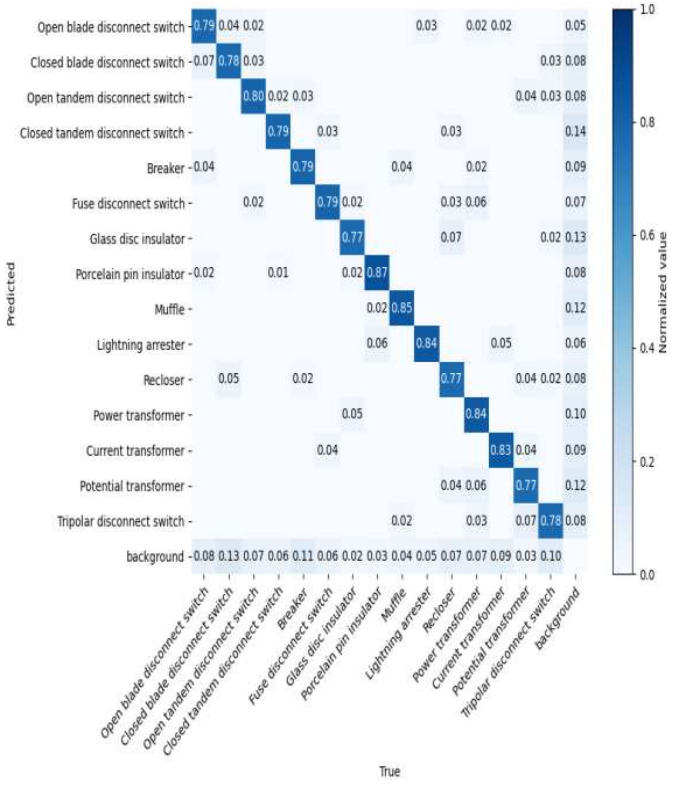


Fig. 5 Confusion matrix.

Fig. 6 provides a comprehensive analysis of the YOLO-MHSA model's performance across a dataset comprising 7,539 images and 213,566 annotated objects spanning 15 categories. We employed four evaluation metrics-F1-Confidlence Curve, Precision-Confidence Curve, Precision-Recall Curve, and Recall-Confidence Curve-to thoroughly exhibit the model's capabilities.

A. F1-Confidence Curve

Fig. 6(a) illustrates the F1 scores across various confidence thresholds, a crucial metric when evaluating classification models, particularly under imbalanced category distributions. The curve reveals that the overall F1 score, represented by the blue line, peaks at 0.897 with a confidence threshold of approximately 0.392. This balance between precision and recall suggests an optimal threshold for practical applications. Notably, categories such as "Fuse" and "Circuit Breaker" achieve near-perfect F1 scores, indicating robust predictive performance. Conversely, the lower F1 score for the "Glass Insulator" category suggests challenges in recognizing these instances, potentially due to limited training samples or category complexity. This highlights the YOLO-MHSA model's capability to balance precision and recall effectively, enabling accurate identification of crucial substation components. It is noteworthy that as confidence increases, the F1-score for certain categories (e.g., "Glass Insulator") slightly declines, reflecting potential under-detection issues at higher confidence levels.

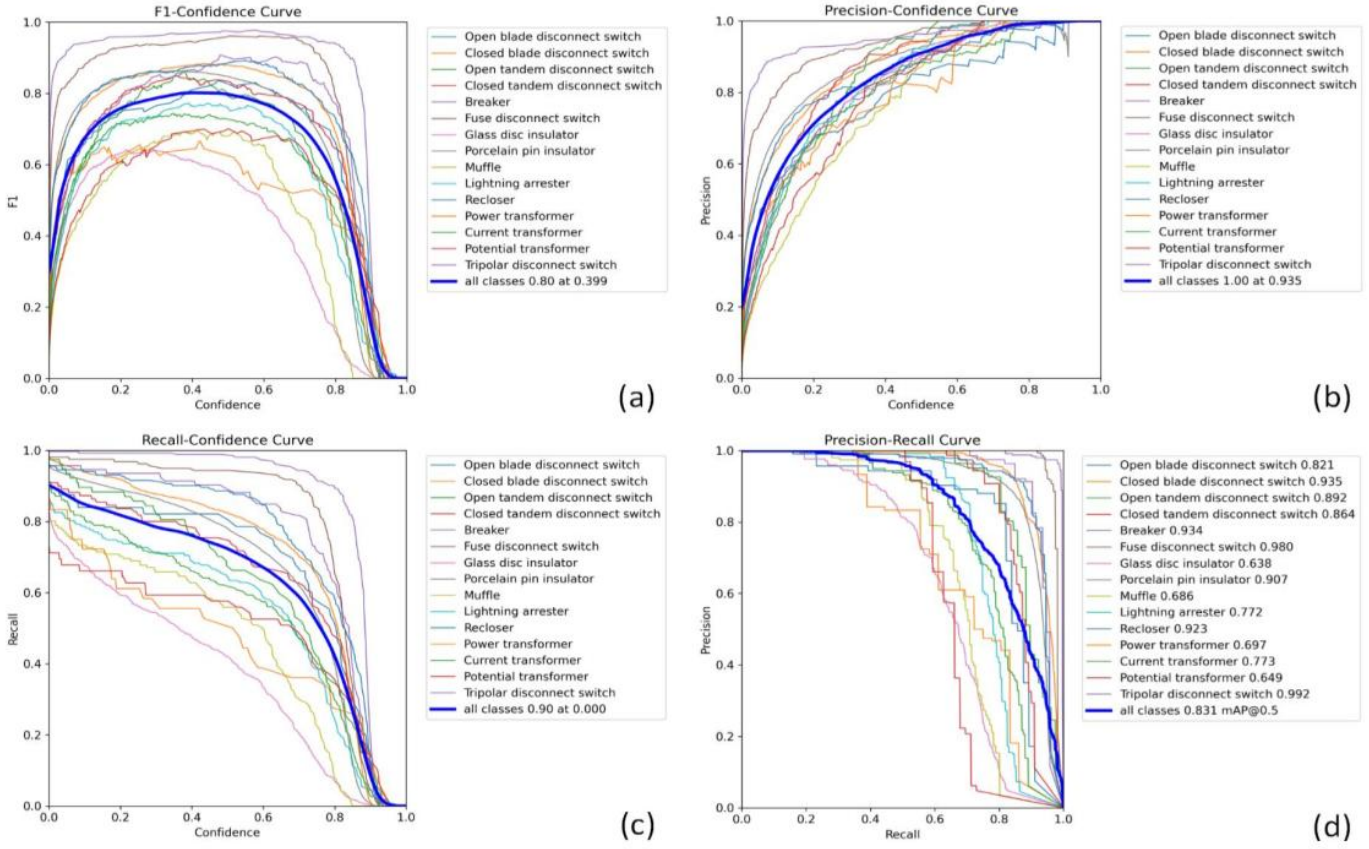


Fig. 6 Performance Evaluation Curve of YOLO-MHSA Model on Substation Inspection Dataset.

B. Precision-Confidence Curve

Fig. 6(b) delineates the variation in precision across different confidence thresholds for each category. Precision, representing the ratio of true positive predictions to all positive predictions, is particularly critical in applications where false positives carry high costs. The YOLO-MHSA model demonstrates outstanding performance in high-confidence regions, achieving a precision of 1.00 across all categories at a confidence level of 0.935. The consistently high precision for the "Fuse" category underscores the model's accuracy in this context. These results indicate that the model can significantly minimize false positives at high confidence levels, which is essential for automated substation detection systems, reducing maintenance costs associated with false alarms.

C. Recall-Confidence Curve

Fig. 6(c) visualizes the recall variation with different confidence levels, offering insights into the model's ability to identify positive instances. Recall, defined as the ratio of true positives to all actual positives, is vital in scenarios with high costs associated with missing positive instances. The curve reveals a recall of up to 0.90 at low confidence levels, indicating the model's capability to capture most positive instances. This is crucial for substation safety monitoring, where missed detections could lead to equipment failures or safety hazards. Although recall decreases at higher confidence levels, overall performance remains satisfactory, showcasing the model's adaptability to complex scenarios.

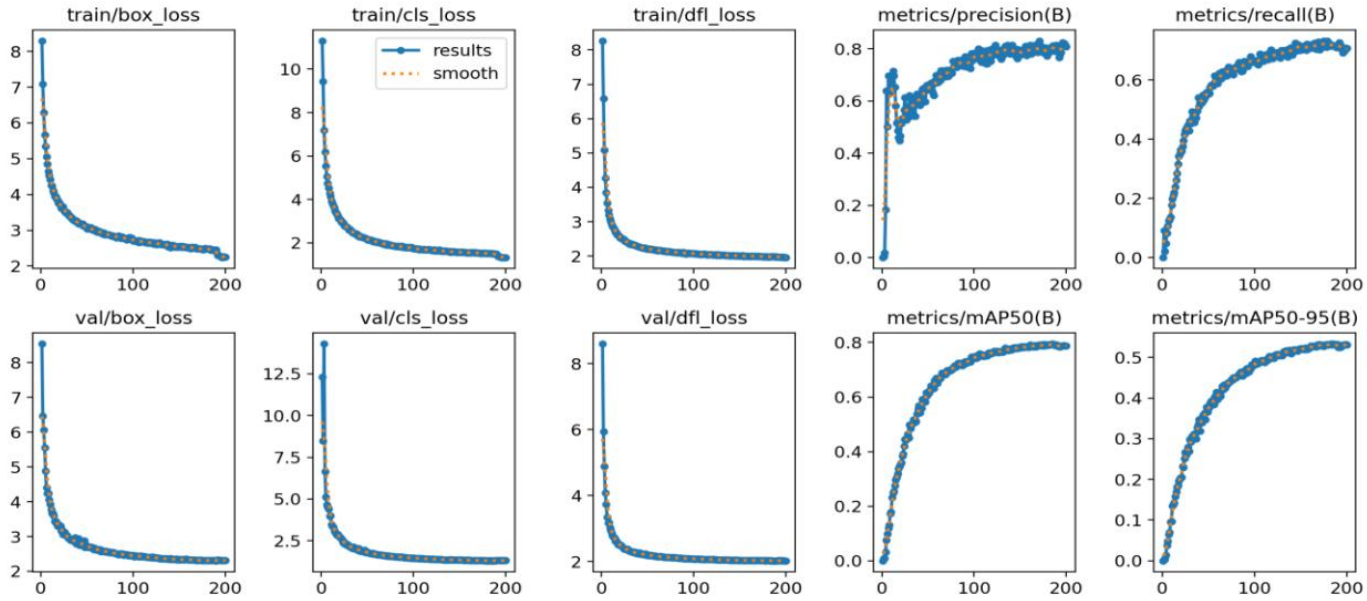


Fig. 7 Training loss analysis.

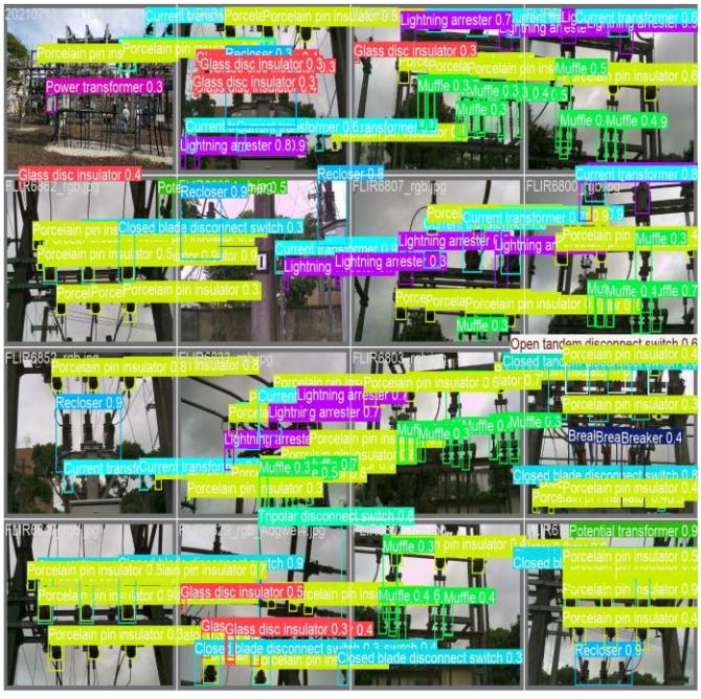


Fig. 8 Visualization of detection results of YOLO-MHSA model in substation scenario

D. Precision-Recall Curve

Fig. 6(d) outlines the trade-off between precision and recall across different categories, vital for evaluating model performance on imbalanced datasets. The blue line indicates an overall mAP of 0.831 at a recall threshold of 0.5, suggesting relatively good performance across categories, albeit with notable disparities. For instance, the "Fuse" category achieves a precision of 0.980, highlighting excellent recognition, whereas the "Glass Insulator" category's precision is only 0.638, underscoring the difficulty in recognizing these categories. The curve's shape illustrates the common trade-off where increasing recall often results in decreased precision, emphasizing the need for a balanced approach when tuning the model to prevent compromised precision or recall.

3.4.3 Loss curve

As shown in Fig. 7, the frames for the training and validation phases show the frame loss (box loss), classification loss (cls loss), and decision loss (dfl loss) respectively. As can be seen from the figure, the loss values during training are higher at the beginning and decrease significantly as the number of training sessions increases, indicating that the model is constantly optimizing and gradually adapting to the dataset. The smoothing curve during this process shows that the model achieves effective convergence with a low risk of overfitting. In terms of metrics evaluation, the graphs of precision and recall show the ability of the model to detect target objects. The

precision gradually increases while the recall enhances, indicating that the model can recognize more objects in the varactor category. Meanwhile, the improvement of the mAP metric further proves the accuracy and stability of the model, especially under different matching thresholds (0.5 to 0.95), the model still shows a robust performance.

3.4.4 Qualitative Analysis

Fig. 8 showcases the detection results of the proposed YOLO-MHSA model on selected test images. These qualitative results demonstrate the model's ability to accurately detect substation equipment under varying time-of-day, weather, and seasonal conditions. Notably, the model maintains robust detection performance in complex weather conditions such as cloudy or rainy days, indicating the YOLO-MHSA model's resilience to visual disturbances caused by weather variations.

4. Conclusions

Aiming at the real-time monitoring requirements of substation equipment in complex transmission and distribution line environments, this study proposes an improved target detection algorithm based on YOLOv8. By introducing the MHSA module in YOLOv8, the feature extraction ability is significantly enhanced, especially the detection accuracy of small-scale targets is improved. In addition, the designed and integrated P2 detection layer is further optimized for small-scale targets, thereby improving the detection rate and positioning accuracy of substation equipment. The comprehensive evaluation shows that the proposed detection model is superior to the traditional methods in terms of detection speed, accuracy, and robustness, which significantly enhances the early identification ability of substation equipment faults, and thus effectively guarantees the safe and stable operation of the power system. The research results not only make up for the shortcomings of the existing monitoring methods in efficiency and safety but also provide an efficient and reliable technical means for the automatic monitoring of substation equipment. Future research will focus on further optimizing the model structure and combining multi-source data fusion technology to improve the adaptability of the model in a changing environment, to promote the sustainable development and application of power equipment monitoring technology.

Acknowledgements

This work was supported in part by the funding project (No. SZKJXM20222171). The authors wish to thank the reviewers for their helpful comments.

On behalf of all authors, the corresponding author states that there are no conflicts of interest among all the contributing authors.

Data Availability Statement

All data generated or analysed during this study are included in this published article.

References

[1] He J, Zeng R, Gao Y, et al. Seasonal influences on safety of substation grounding system, IEEE Transactions on Power Delivery, 2003, 18(3): 788-795.

[2] Li Z, Fan M. OPGW insulation design for energy efficiency improvement and ground wire ice melting, in Proc. IEEE International Conference on Power Science and Technology (ICPST'24), 2024, pp. 380-384.

[3] Zhu Z, Li Y, Wang L, et al. Study on power frequency and switching overvoltage of typical transmission lines in pumped storage power substation, in Proc. the Asia Conference on Power and Electrical Engineering (ACPEE'24), 2024, pp. 1690-1697.

[4] Zhu W, Xie Q. Machine learning chain models for multi-response prediction of electrical equipment in substation subjected to earthquakes, Engineering Structures, 2024, 319(2): 118815-118830.

[5] Sourav S, Biswas P, Mohanraj V, et al. Machine learning assisted bad data detection for high-throughput substation communication, in Proc. IEEE International Conference on Communications (ICC'23), 2023, pp. 4131-4137.

[6] Gaspar J, Cruz T, Lam C, et al. Smart substation communications and cybersecurity: A comprehensive survey, IEEE Communications Surveys and Tutorials, 2023, 25(4): 2456-2493.

[7] Kuyumani E, Hasan A, Shongwe T. A hybrid model based on CNN-LSTM to detect and forecast harmonics: a case study of an Eskom substation in South Africa, Electric Power Components and Systems, 2023, 51(8): 746-760.

[8] Aksan F, Li Y, Suresh V, et al. CNN-LSTM vs. LSTM-CNN to predict power flow direction: A case study of the high-voltage subnet of northeast Germany, Sensors, 2023, 23(2): 901-913.

[9] Zhang N, Yang G, Wang D, et al. A defect detection method for substation equipment based on image data generation and deep learning, IEEE Access, 2024, 212(2024): 1-14.

[10] Wu T, Zhou Z, Liu J, et al. Ise-yolo: A real-time infrared detection model for substation equipment, IEEE Transactions on Power Delivery, 2024, 33(3): 1-13.

[11] Dong H, Yuan M, Wang S, et al. PHAM-YOLO: A parallel hybrid attention mechanism network for defect detection of meter in substation, Sensors, 2023, 23(13): 6052-6065.

[12] Zhang N, Yang G, Wang D, et al. A defect detection method for substation equipment based on image data generation and deep learning, IEEE Access, 2024, 213(3): 1-15.

[13] Zhang Y, Ye M, Zhu G, et al. FFCA-YOLO for small object detection in remote sensing images, IEEE Transactions on Geoscience and Remote Sensing, 2024, 62(3): 1-15.

[14] Ji S, Ling Q, Han F. An improved algorithm for small object detection based on YOLO v4 and multi-scale contextual information, Computers and Electrical Engineering, 2023, 105(3): 108490-108503.

[15] Ji C, Yu T, Gao P, et al. YOLO-TLA: An efficient and lightweight small object detection model based on YOLOv5, Journal of Real-Time Image Processing, 2024, 21(4): 141-153.

[16] Tang S, Zhang S, Fang Y. HIC-YOLOv5: Improved YOLOv5 for small object detection, in Proc. IEEE International Conference on Robotics and Automation (ICRA'24), 2024, pp. 6614-6619.

[17] Cheng G, Yuan X, Yao X, et al. Towards large-scale small object detection: Survey and benchmarks, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(11): 13467-13488.

[18] Mahaur B, Mishra K. Small-object detection based on YOLOv5 in autonomous driving systems, Pattern Recognition Letters, 2023, 168(2): 115-122.

[19] Zhang K, Cui T, Wu W, et al. Large kernel separable mixed convnet for remote sensing scene classification, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2024, 17(3): 4294-4303.

[20] Tan H, Liu X, Yin B, et al. MHSA-Net: Multihead self-attention network for occluded person re-identification, IEEE Transactions on Neural Networks and Learning Systems, 2022, 34(11): 8210-8224.

[21] Li J, Wu W, Zhang D, et al. Multi-pedestrian tracking based on KC-YOLO detection and identity validity discrimination module, Applied Sciences, 2023, 13(22): 12228-12230.

[22] Gomes A, Ganacim F, Magrin F, et al. A semantically annotated 15-class ground truth dataset for substation equipment to train semantic segmentation models, Data, 2023, 8(7): 118-131.

[23] Wang C, Yeh I, Mark L. Yolov9: Learning what you want to learn using programmable gradient information, in Proc. the European Conference on Computer Vision (ECCV'24), 2024, pp. 1-21.

[24] Wang A, Chen H, Liu L, et al. Yolov10: Real-time end-to-end object detection, Advances in Neural Information Processing Systems, 2024, 37(2): 107984-108011.

[25] Khanam R, Hussain M. Yolov11: An overview of the key architectural enhancements, arXiv, 2024, https://arxiv.org/html/2410.17725v1.

References

[1] He J, Zeng R, Gao Y, et al. Seasonal influences on safety of substation grounding system, IEEE Transactions on Power Delivery, 2003, 18(3): 788-795.

[2] Li Z, Fan M. OPGW insulation design for energy efficiency improvement and ground wire ice melting, in Proc. IEEE International Conference on Power Science and Technology (ICPST'24), 2024, pp. 380-384.

[3] Zhu Z, Li Y, Wang L, et al. Study on power frequency and switching overvoltage of typical transmission lines in pumped storage power substation, in Proc. the Asia Conference on Power and Electrical Engineering (ACPEE'24), 2024, pp. 1690-1697.

[4] Zhu W, Xie Q. Machine learning chain models for multi-response prediction of electrical equipment in substation subjected to earthquakes, Engineering Structures, 2024, 319(2): 118815-118830.

[5] Sourav S, Biswas P, Mohanraj V, et al. Machine learning assisted bad data detection for high-throughput substation communication, in Proc. IEEE International Conference on Communications (ICC'23), 2023, pp. 4131-4137.

[6] Gaspar J, Cruz T, Lam C, et al. Smart substation communications and cybersecurity: A comprehensive survey, IEEE Communications Surveys and Tutorials, 2023, 25(4): 2456-2493.

[7] Kuyumani E, Hasan A, Shongwe T. A hybrid model based on CNN-LSTM to detect and forecast harmonics: a case study of an Eskom substation in South Africa, Electric Power Components and Systems, 2023, 51(8): 746-760.

[8] Aksan F, Li Y, Suresh V, et al. CNN-LSTM vs. LSTM-CNN to predict power flow direction: A case study of the high-voltage subnet of northeast Germany, Sensors, 2023, 23(2): 901-913.

[9] Zhang N, Yang G, Wang D, et al. A defect detection method for substation equipment based on image data generation and deep learning, IEEE Access, 2024, 212(2024): 1-14.

[10] Wu T, Zhou Z, Liu J, et al. Ise-yolo: A real-time infrared detection model for substation equipment, IEEE Transactions on Power Delivery, 2024, 33(3): 1-13.

[11] Dong H, Yuan M, Wang S, et al. PHAM-YOLO: A parallel hybrid attention mechanism network for defect detection of meter in substation, Sensors, 2023, 23(13): 6052-6065.

[12] Zhang N, Yang G, Wang D, et al. A defect detection method for substation equipment based on image data generation and deep learning, IEEE Access, 2024, 213(3): 1-15.

[13] Zhang Y, Ye M, Zhu G, et al. FFCA-YOLO for small object detection in remote sensing images, IEEE Transactions on Geoscience and Remote Sensing, 2024, 62(3): 1-15.

[14] Ji S, Ling Q, Han F. An improved algorithm for small object detection based on YOLO v4 and multi-scale contextual information, Computers and Electrical Engineering, 2023, 105(3): 108490-108503.

[15] Ji C, Yu T, Gao P, et al. YOLO-TLA: An efficient and lightweight small object detection model based on YOLOv5, Journal of Real-Time Image Processing, 2024, 21(4): 141-153.

[16] Tang S, Zhang S, Fang Y. HIC-YOLOv5: Improved YOLOv5 for small object detection, in Proc. IEEE International Conference on Robotics and Automation (ICRA'24), 2024, pp. 6614-6619.

[17] Cheng G, Yuan X, Yao X, et al. Towards large-scale small object detection: Survey and benchmarks, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(11): 13467-13488.

[18] Mahaur B, Mishra K. Small-object detection based on YOLOv5 in autonomous driving systems, Pattern Recognition Letters, 2023, 168(2): 115-122.

[19] Zhang K, Cui T, Wu W, et al. Large kernel separable mixed convnet for remote sensing scene classification, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2024, 17(3): 4294-4303.

[20] Tan H, Liu X, Yin B, et al. MHSA-Net: Multihead self-attention network for occluded person re-identification, IEEE Transactions on Neural Networks and Learning Systems, 2022, 34(11): 8210-8224.

[21] Li J, Wu W, Zhang D, et al. Multi-pedestrian tracking based on KC-YOLO detection and identity validity discrimination module, Applied Sciences, 2023, 13(22): 12228-12230.

[22] Gomes A, Ganacim F, Magrin F, et al. A semantically annotated 15-class ground truth dataset for substation equipment to train semantic segmentation models, Data, 2023, 8(7): 118-131.

[23] Wang C, Yeh I, Mark L. Yolov9: Learning what you want to learn using programmable gradient information, in Proc. the European Conference on Computer Vision (ECCV'24), 2024, pp. 1-21.

[24] Wang A, Chen H, Liu L, et al. Yolov10: Real-time end-to-end object detection, Advances in Neural Information Processing Systems, 2024, 37(2): 107984-108011.

[25] Khanam R, Hussain M. Yolov11: An overview of the key architectural enhancements, arXiv, 2024, https://arxiv.org/html/2410.17725v1.