

Research on Improved YOLOv8 Algorithm for Insulator Defect Detection

Lin Zhang

University of Science and Technology Liaoning

Boqun Li (✉ Lbqhylyxab@163.com)

University of Science and Technology Liaoning

Yang Cui

University of Science and Technology Liaoning

Yushan Lai

University of Science and Technology Liaoning

Jing Gao

University of Science and Technology Liaoning

Research Article

Keywords: Object detection, Insulator defect, Attention mechanism, YOLOv8, SIoU

Posted Date: September 15th, 2023

DOI: <https://doi.org/10.21203/rs.3.rs-3337929/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Journal of Real-Time Image Processing on January 13th, 2024. See the published version at <https://doi.org/10.1007/s11554-023-01401-9>.

Research on Improved YOLOv8 Algorithm for Insulator Defect Detection

Lin Zhang¹, Boqun Li¹, Yang Cui¹, Yushan Lai¹, Jing Gao²

Abstract

With the rapid advancement of artificial intelligence technologies, drone aerial photography has gradually become the mainstream method for defect detection of transmission line insulators. To address the issues of slow recognition speed and low accuracy in existing detection methods, this paper proposes an insulator defect detection algorithm based on an improved YOLOv8s model. Initially, a Multi-scale Large Kernel Attention (MLKA) module is introduced to enhance the model's focus on features of different scales as well as low-level feature maps. Additionally, by employing lightweight GSCov convolution and constructing the GSC_C2f module, the computational process is simplified and memory burden is reduced, thereby effectively improving the performance of insulator defect detection. Finally, an improved loss function using SIOU is adopted to optimize the model's detection performance and enhance its feature extraction capability for insulator defects. Experimental results demonstrate that the improved model exhibits excellent performance in drone aerial photography for insulator defect detection, achieving an mAP of 99.22% and an FPS of 55.73 frames per second. Compared to the original YOLOv8s and YOLOv5s, the improved model's mAP increased by 2.18% and 2.91%, respectively, and the model size is only 30.18MB, meeting the requirements for real-time operation and accuracy.

Keywords: Object detection, Insulator defect, Attention mechanism, YOLOv8, SIOU

1. Introduction

Insulators play an indispensable and crucial role in power transmission lines, responsible for securing conductors and preventing short circuits between lines and towers. However, due to their sheer number and susceptibility to natural environmental factors and wear over time, insulators may suffer from issues such as detachment, self-explosion, and contamination. These issues can lead to the loss of their normal function, causing failures and unnecessary losses. To ensure the safe and stable operation of the electrical system, it is imperative to explore methods capable of identifying and detecting insulator defects.

Currently, drone inspections are gradually replacing traditional manual inspections due to their efficiency and convenience. Accordingly, insulator defect detection methods based on aerial images have emerged. Traditional algorithms for insulator defect detection first process features such as color, texture, and edges, followed by identification through edge detection algorithms, HOG algorithms[1], and SIFT algorithms[2]. These algorithms heavily rely on high-quality images and

appropriate shooting angles, resulting in relatively weak robustness.

In recent years, deep learning algorithms have been widely applied in the field of image defect detection both domestically and internationally. Given the limitations of traditional algorithms, researchers have turned to deep learning. Methods using deep learning for insulator defect detection can be broadly divided into two categories. The first category includes single-stage detectors with high real-time detection performance. For instance, Xu et al.[3] proposed an insulator detection method based on SSD[4], which achieved high-precision detection in electrical systems but required high computational complexity and cost. Liu et al. [5] improved the YOLOv3 algorithm[6] by adopting the SPP feature pyramid pooling module[7] and multi-scale prediction network structure, enhancing the feature representation ability of insulator fault locations. Moreover, Qiu et al.[8] used a lightweight MobileNet[9] convolutional neural network as the feature extraction network for YOLOv4[10], solving the problems of excessive model parameters and slow detection speed. Han et al.[11] proposed an enhanced YOLOv5 model[12] that integrates the ECA-Net attention mechanism[13] and incorporates a bidirectional feature pyramid network in the feature fusion layer, effectively improving the accuracy of insulator defect detection. However, the model is parameter-intensive, potentially leading to computational overhead.

The second category involves two-stage detectors. For instance, Zhang et al.[14] put forth a hybrid approach for insulator defect image detection that combines morphological operations with deep learning techniques. Specifically, they utilized Faster R-CNN[15] for fine-grained localization, and augmented it with rotation algorithm preprocessing, sliding window techniques, and

Boqun Li
Lbqhylyxab@163.com

Lin Zhang
Z1578136361@163.com

1 College of Electronics and Information Engineering, University of Science and Technology Liaoning, China
2 College of Business and Administration, University of Science and Technology Liaoning, China

object image segmentation to achieve high-accuracy defect detection in transmission lines. Despite its remarkable detection precision, the method suffers from relatively slow inference speed. In a parallel development, Zhou et al. [16] integrated attention mechanisms into the backbone network of the Mask R-CNN model [17], thereby enhancing the model's focus on smaller objects and improving localization accuracy. Furthermore, they incorporated rotational mechanisms into the loss function, allowing for precise defect localization through the consideration of various rotational angles. While these innovative methods demonstrate exceptional performance in terms of detection accuracy, the issue of inference speed remains a significant challenge.

In light of the existing models' issues with low detection accuracy, large network structure parameters, and difficulty in achieving real-time accurate detection of insulator defects, this paper selects the YOLOv8s[18] lightweight deep learning detection model with smaller network depth and high detection efficiency and improves it to propose the YOLOv8-GSC model.

The main contributions are as follows:

1. First, an MLKA module[19] is added to the backbone detection network. This module uses a multi-scale LKA [20] strategy to process features of different scales and fuses their outputs, effectively capturing the local and non-local contextual information of the input feature map, thereby improving the model's accuracy in detecting insulators and their defects in complex backgrounds.

2. Furthermore, by introducing lightweight convolution GSConv[21] to construct an efficient GSC_C2f module to replace the original C2f module, the aim is to enhance target feature extraction efficiency and enrich deep semantic information. Due to its streamlined computational process, the GSC_C2f module can reduce the number of model parameters and complexity, thereby improving detection efficiency.
3. Lastly, SIoU[22] is adopted as the loss function. SIoU ingeniously combines angle loss, distance loss, and shape loss, providing a more comprehensive and accurate assessment of the similarity of target bounding boxes. Compared with the CIoU loss function[23], SIoU shows significant advantages in accelerating model convergence speed and improving small target detection accuracy.

2. YOLOv8 Model Structure

YOLOv8, released by Ultralytics in early 2023, is the latest object detection model. Compared to previous YOLO models, YOLOv8 demonstrates significant performance improvements across various key tasks, including object detection, semantic segmentation, and image classification. The YOLOv8 model consists of four main components: the input layer, the backbone network, the neck network, and the prediction network, as detailed in Figure 1.

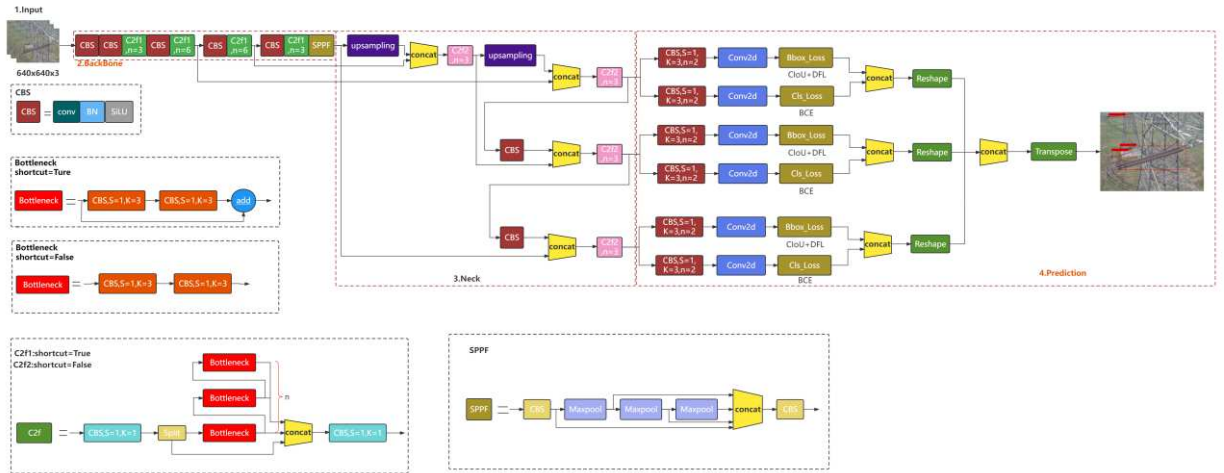


Fig. 1 YOLOv8 Structure

The input layer of YOLOv8 integrates adaptive anchoring, adaptive image scaling, as well as Mosaic and Mix Up data augmentation techniques. Adaptive anchoring and image scaling ensure the generation of prediction boxes and the standardization of input image dimensions. Specifically, Mosaic and Mix Up serve as data augmentation methods that enhance the diversity of training samples and the model's ability to recognize small targets. They achieve this by integrating and fusing randomly scaled, cropped, and arranged images, as illustrated in Figure 2. In complex tasks such as insulator detection, these two techniques have been particularly effective in improving the model's ability to recognize and segment overlapping targets, thereby ensuring

YOLOv8's performance under diverse conditions.



(a) Mosaic data augmentation

(b) Mix Up data augmentation

Fig. 2 (a) shows the effect of mosaic data augmentation, and(b) shows the effect of mix up data augmentation.

In the design of YOLOv8, the backbone detection network still adopts the CSPNet architecture [24]. The input image size is adjusted to $640 \times 640 \times 3$, and the image passes through the backbone network to generate multi-scale feature maps of three different dimensions: $20 \times 20 \times 256$, $40 \times 40 \times 128$ and $80 \times 80 \times 64$. Drawing upon the concept of ELAN in YOLOv7 [25], an effective C2f structure is introduced, as shown in Figure 4.3. Compared to traditional C3 and CSPLayer modules, the C2f module has denser residual connections. In each residual structure computation, the output is preserved and added to the output of subsequent layers, enhancing the gradient flow and making the network easier to train. To reduce computational complexity and memory requirements, the C2f module employs a smaller expansion ratio to decrease the number of channels in intermediate layers. Additionally, the bottleneck module in the C2f structure allows the use of different kernel sizes, increasing the network's flexibility in feature extraction.

In the neck structure of YOLOv8, a path aggregation network and feature pyramid structure are employed. Through a series of upsampling and downsampling operations, the three feature tensors from the backbone network are fused and further enhanced through convolutional operations. The original C3 structure is replaced by the C2f structure, improving the model's flexibility and efficiency.

Significant changes can be observed in the predictive

network structure of YOLOv8 as compared to that of YOLOv5. It adopts the current mainstream decoupled network structure, comprising two independent prediction networks specifically designed for classification and regression tasks, allowing the model to flexibly adapt to various task requirements. Notably, the decoupled network demonstrates excellent performance in handling class imbalance and object scale variation, enhancing the robustness of object detection. Furthermore, YOLOv8 transitions from anchor-based to anchor-free, successfully avoiding the complex calculations and hyperparameter settings associated with anchor boxes, significantly impacting the model's performance.

3. Improved Network Structure of YOLOv8

To enhance the detection performance of insulator defects in drone images, we propose the YOLO-GSC model, the structure of which is illustrated in Figure 3. YOLO-GSC incorporates an MLKA module into its backbone network, aiming to improve the detection accuracy of insulators and their defects, which vary in size and shape, in complex backgrounds. In line with the requirements for a lightweight model, the introduction of the GSC_C2f module in both the backbone and neck networks leads to a reduction in the number of model parameters while enhancing the feature extraction capabilities for insulator defects. Furthermore, the adoption of SIOU as the bounding box loss function aids in accelerating model convergence and significantly boosts the detection accuracy for small-target insulator defects.

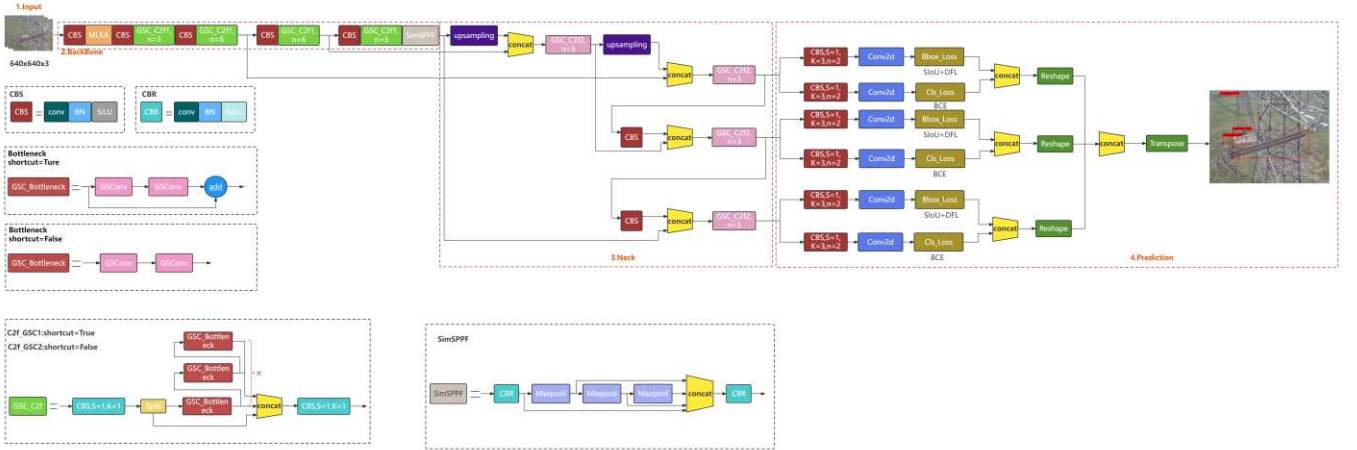


Fig. 3 Structure of YOLOv8-GSC

3.1 Large Kernel Attention

In the domain of computer vision tasks, there are generally two main approaches for establishing relationships between different feature regions within an image. The first approach utilizes self-

attention mechanisms[26] to capture long-range dependencies. However, this method has significant limitations in the realm of computer vision. Treating the image as a one-dimensional sequence leads to the neglect of the image's two-dimensional structure, resulting in a loss of spatial relationship information.

Additionally, the quadratic complexity of self-attention mechanisms in processing high-resolution images escalates computational costs, potentially limiting its applicability to large-scale images. More importantly, while self-attention mechanisms can capture adaptability in the spatial dimension, they overlook adaptability in the channel dimension. In visual tasks, different channels typically represent different features; therefore, ignoring channel adaptability could adversely affect task performance. The second approach involves using large convolutional kernels[27] to construct correlations and generate attention maps. Although this method is capable of capturing spatial relationships, it too has its limitations, such as high computational costs and a large number of parameters.

To overcome these limitations and take advantage of both self-attention and large kernel convolutions, a new method is introduced that captures long-range relationships by decomposing large kernel convolutional operations. As illustrated in Figure 4, large kernel convolution can be decomposed into three components: spatial local convolution (depthwise convolution), spatial long-range convolution (depthwise dilated convolution), and channel-wise convolution (1×1 convolution). Specifically, a convolution of size $k \times k$ can be broken down into a $(2d-1) \times (2d-1)$ depthwise convolution, a $\left[\frac{k}{d}\right] \times \left[\frac{k}{d}\right]$ depthwise dilated convolution, and a 1×1 convolution.

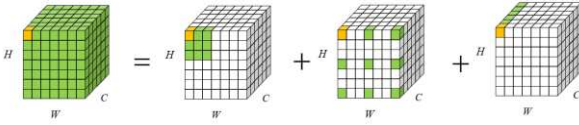


Fig. 4 Decomposition diagram of large kernel convolution

Through the aforementioned decomposition, long-range relationships can be captured with reduced computational cost and fewer parameters. After obtaining these long-range relationships, the importance of each point can be estimated to generate attention maps. The structure of the LKA module is illustrated in Figure 5. This can be expressed as:

$$Attention = Conv_{1 \times 1}(DW - D - Conv(DW - Conv(F))), \quad (1)$$

$$Output = Attention \otimes F. \quad (2)$$

The formula includes input feature $F \in R^{C \times H \times W}$ and $Attention \in R^{C \times H \times W}$ denotes attention map, which represents the importance of each feature. The symbol \otimes denotes element-wise multiplication.

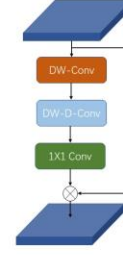


Fig. 5 Large Kernel Attention (LKA)

3.2 Multi-scale Large Kernel Attention

The MLKA module combines large-kernel decomposition with multi-scale learning. It essentially consists of three core components: large-kernel attention for establishing interdependencies, a multi-scale mechanism for capturing correlations across heterogeneous scales, and gated aggregation for dynamic calibration, as illustrated in Figure 6.

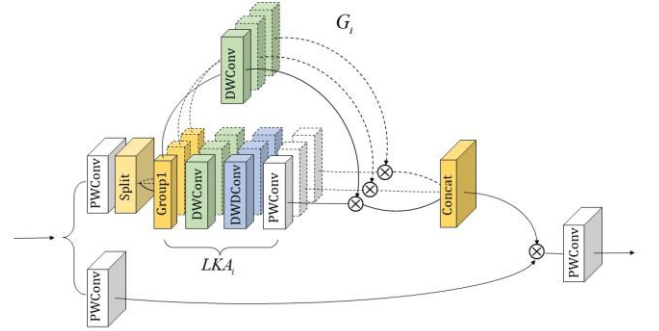


Fig. 6 The structure of Multi-scale Large Kernel Attention(MLKA) block

The input feature maps $X \in R^{C \times H \times W}$, the module initially divides the input feature map into n partitions, denoted as X_1, X_2, \dots, X_n , each with a size of $\left[\frac{C}{n}\right] \times H \times W$. Using the

LKA decomposed by $\{k_i, d_i\}$, a homogeneous scale of attention map LKA_i is generated. The multiscale LKA branch enables the module to capture features at various scales and spatial contexts, thereby obtaining richer feature representations.

Although larger LKA values capture broader pixel responses, blocking artifacts appear in the attention maps generated with larger LKA . To address the issue of blocking artifacts and to capture more local information, a spatial gating mechanism can be dynamically employed to adapt $LKA_i(\cdot)$ to $MLKA_i(\cdot)$. The gating function is utilized to control the weights of each $LKA_i(\cdot)$, resulting in a weighted sum for the final attention map. By introducing this gating mechanism, $MLKA_i(\cdot)$

can effectively capture both global and local information, thereby enhancing the capability to capture fine-grained features. The specific expression is as follows:

$$MLKA_i(X_i) = G_i(X_i) \otimes LKA_i(X_i), \quad (3)$$

Where $G_i(\cdot)$ is the i -th gating function generated by $a_i \times a_i$ depth separable convolution for gated aggregation. At the same time, $LKA_i(\cdot)$ is LKA decomposed by $a_i - b_i - 1$.

3.3 C2f_GSC Structure

In practical scenarios where drones are used for power line inspections, the trade-off between detection speed and accuracy is of critical importance. While some large-scale models like ResNet[28] and Vision Transformer are capable of achieving high detection accuracy, the time consumption of their detection processes is too long to meet real-time requirements. On the other hand, some lightweight networks such as Xception[29], MobileNets[30-32], and ShuffleNets[33-34] significantly improve detection speed by adopting depthwise separable convolutions, but they compromise on detection accuracy. For power line inspection tasks requiring high precision, the applicability of these models is relatively poor.

To address this issue, a novel lightweight convolutional structure, GSConv, is introduced. As illustrated in Figure 7, the GSConv structure first processes the input feature map through a module composed of a 2D convolution layer, Batch Normalization (BN), and SiLU activation function. The resulting feature map has a channel count that is half of the final output channel count. Subsequently, this feature map is processed through a DWConv module and is stacked with the original feature map along the channel dimension. Finally, a shuffle operation is performed to produce the final output feature map.

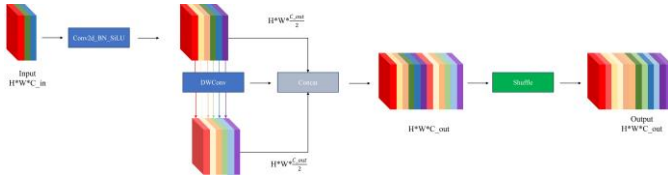


Fig. 7 Structure of GSConv

This paper incorporates the lightweight GSConv convolution into the C2f architecture. By performing grouping and shuffling along the channel dimension, GSConv not only enhances the model's feature representational power but also lowers computational complexity, thus boosting the model's computational efficiency. More crucially, the lightweight design of GSConv reduces the model's parameter count, effectively mitigating overfitting and enhancing the model's generalizability. Concurrently, the C2f module in YOLOv8s is replaced with the GSC_C2f module. The GSC_C2f module continues to employ the CSP approach and consists of two CBS modules and n bottleneck modules, as depicted in Figure 8.

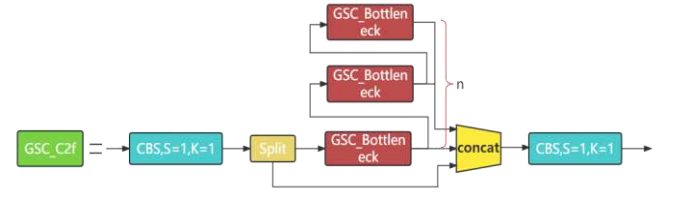


Fig. 8 GSC_C2 Module

3.4 SIOU Loss Function

The IoU (Intersection over Union) loss, also known as the Jaccard index, is a metric used to measure the degree of overlap between two bounding boxes. In object detection tasks, IoU loss offers advantages such as intuitiveness and robustness. However, it also presents disadvantages, including poor smoothness, sensitivity to thresholds, and higher computational complexity. The IoU loss formula is defined as the ratio of the intersection area to the union area of the predicted and ground truth bounding boxes, as illustrated in Figure 9.

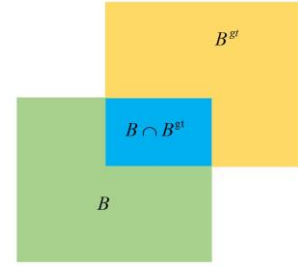


Fig. 9 The diagram illustrates the principle of IoU, the green box represents the predicted box, the yellow box represents the ground truth box, and the blue part denotes the intersection between the predicted box and the ground truth box.

In the original YOLOv8 model, CIOU loss was chosen as the bounding box loss function. The CIOU loss accounts for the overlap of bounding boxes, the distance between their central points, and the aspect ratio. This consideration offers advantages in solving gradient smoothness issues and positively impacts the gradient descent optimization algorithm. However, CIOU loss has certain limitations as it does not consider the orientation between the ground-truth and predicted bounding boxes, which to some extent limits the convergence speed during model training. To address this shortcoming of CIOU loss and further accelerate the model's convergence speed, this study introduces the SIOU loss function.

The SIOU loss function ingeniously integrates angle loss, distance loss, and shape loss. Compared to CIOU loss, SIOU loss exhibits significant superiority in improving the performance of object detection algorithms, particularly in terms of convergence speed and accuracy. This superiority makes SIOU loss an ideal choice for the loss function. The principle behind it is illustrated in Figure 10. Here, w and h represent the width and height of the

predicted bounding box, while w^{gt} and h^{gt} represent the width and height of the ground-truth bounding box. α denotes the angle between the central points of the predicted and ground-truth bounding boxes. Additionally, c^w and c^h correspond to the width and height of the rectangle formed by the two central points, and C^w and C^h represent the width and height of the minimum bounding rectangle encompassing both boxes.

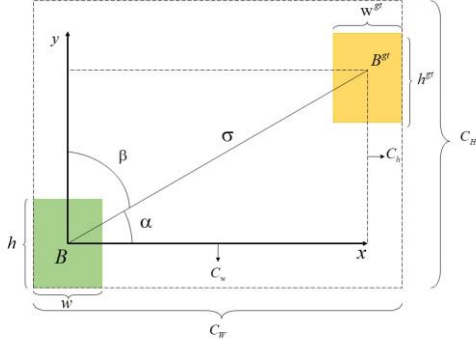


Fig. 10 SIOU Principal Diagram

The SIOU loss function incorporates angle cost, distance cost, shape cost, and IoU cost, with the underlying principle as follows:

(1) Angle Loss

$$\Delta = 1 - 2 * \sin^2 \left(\arcsin(x) - \frac{\pi}{4} \right) \quad (4)$$

$$x = \frac{c_h}{\sigma} = \sin(\alpha) \quad (5)$$

$$c_h = \max(b_{cy}^{gt}, b_{cy}) - \min(b_{cy}^{gt}, b_{cy}) \quad (6)$$

$$\sigma = \sqrt{(b_{cx}^{gt} - b_{cx})^2 + (b_{cy}^{gt} - b_{cy})^2} \quad (7)$$

(2) Distance Loss

$$\Delta = 2 - e^{(\Delta-2) * \left(\frac{c_w}{C^w}\right)^2} - e^{(\Delta-2) * \left(\frac{c_h}{C^h}\right)^2} \quad (8)$$

(3) Shape Loss

$$\Omega = \left(1 - e^{-\frac{|w-w^{gt}|}{\max(w, w^{gt})}} \right)^0 + \left(1 - e^{-\frac{|h-h^{gt}|}{\max(h, h^{gt})}} \right)^0 \quad (9)$$

(4) IoU Loss

$$IoU = \frac{|B \cap B^{gt}|}{|B \cup B^{gt}|} \quad (10)$$

By integrating these four types of loss functions, the final SIOU loss function is calculated using the following equation:

$$L_{SIOU} = 1 - IoU - \frac{\Delta + \Omega}{2} \quad (11)$$

4. Experiment and analysis

4.1 Experimental Environment

The experimental environment was configured on Windows 10,

with an AMD EPYC 7601 CPU and an NVIDIA GeForce RTX 3090 GPU. Programming was done using Python 3.7 and the model architecture was built on the PyTorch framework. The model was trained for 300 epochs, divided into two phases. During the first 50 epochs, the backbone of the model was frozen, the initial learning rate was set to 1×10^{-2} , and the batch size was 32, focusing only on fine-tuning the network. In the second phase, the "unfreeze" training method was used, training the model for 250 epochs with a batch size of 16. At this point, the backbone network was not frozen, thus all network parameters were updated. A momentum value of 0.937 and a weight decay of 5×10^{-4} were used with the SGD optimizer. To further enhance the effectiveness of the training process, techniques such as Mosaic data augmentation and cosine annealing learning rate strategy were employed. This approach leverages the powerful optimization capabilities of the SGD optimizer, combined with fine-grained techniques like data augmentation and custom learning rate strategies, to improve the model's performance during training. Table 1 shows the experimental environment and configuration parameters.

Table 1 Basic Configuration of Local Computer

Computer configuration	Specific parameters/versions
Operating System	Windows10
CPU	AMD EPYC 7601
Python	3.7.13
Pytorch	1.7.1
Cudnn	11.0
Opencv	4.5.3

4.2 Dataset

The experimental dataset partially comes from the Chinese Power Line Insulator dataset, which includes 600 images of normal insulators and 248 images of defective insulators. The remaining data were collected from real-world scenarios, totaling 2348 images. To address the issue of overfitting due to the small dataset size, which could affect the detection performance of both insulators and their defects, data augmentation techniques were employed to expand the original dataset. By using a variety of data augmentation methods, including random rectangular occlusion, horizontal flipping, random pixel zeroing, random cropping, and padding, the dataset was expanded to 5322 images. One advantage of this approach is that the expanded dataset enhances the model's robustness and generalizability, enabling it to handle detection tasks in different scenarios while avoiding overfitting, thereby improving the accuracy and reliability of detection. Subsequently, the expanded dataset was randomly divided into training, validation, and test sets. The training set was used for training the network, the validation set for checking for overfitting and assessing network convergence, and the test set for evaluating the model's performance on new data. The division ratio was 8:1:1 to ensure the dataset is fully utilized and effectively prevent overfitting. Some results of image data augmentation are shown in Fig. 11.

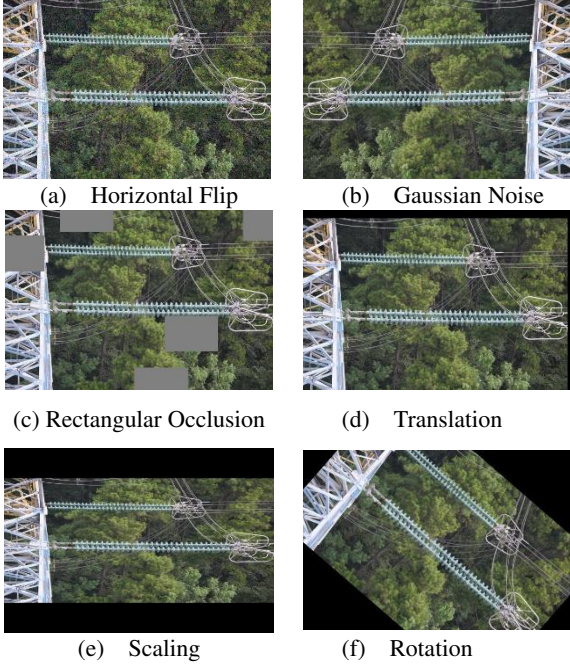


Fig. 11 Example of data expansion

4.3 Evaluation Metrics

This paper employs evaluation metrics such as mAP (Mean Average Precision) and FPS (Frames Per Second) for a comprehensive assessment of the insulator detection model. mAP serves as an integrated evaluation metric that assesses the overall performance of an object detection model by calculating the mean of the AP (Average Precision) across all detection categories. The AP value gauges the model's performance in a specific category and reflects the model's detection capabilities for various classes. By calculating the mAP value, the model's composite performance across all categories can be assessed. Equations (12) and (13) outline the methods for calculating AP and mAP, while equations (14) and (15) explain the recall and precision required to compute the AP for different detection categories. In these equations, recall measures the model's ability to detect positive samples, while precision evaluates the model's predictive accuracy against actual results. By calculating the number of True Positives (TP), False Positives (FP), and False Negatives (FN), the model's precision and recall can be obtained, thereby enabling the calculation of average precision and mean average precision.

$$AP = \int_0^1 P(R) dR \quad (12)$$

$$mAP = \frac{\sum PA}{N_c} \quad (13)$$

$$P = \frac{TP}{TP + FP} \times 100\% \quad (14)$$

$$R = \frac{TP}{TP + FN} \times 100\% \quad (15)$$

4.4 Experimental Results Analysis

As illustrated in Figure 12, a detailed comparative analysis of the

training loss curves between the improved model and YOLOv8s is presented. By analyzing the dynamic changes in the loss curves, it can be observed that as the number of training iterations increases, the loss curves of both models gradually transition to a stable state. This phenomenon indicates that the models are continuously extracting information from the training data and adjusting their internal parameters to more accurately adapt to and fit the data.

During the initial 20 epochs of training, both models show a significant downward trend in loss, symbolizing that the models have already begun to learn at a faster rate in the early stages. However, the loss in the initial stages of training is slightly higher for the improved model than for the YOLOv8s model. This is primarily because the improved model integrates the MLKA module and GSC_C2f structure into its backbone network. The initial adjustments and optimizations of these new modules may temporarily increase the loss. Despite this situation in the early stages, the improved model begins to show a more rapid decrease in loss and superior convergence performance relative to the YOLOv8s model after approximately 30 epochs.

In the later stages of the training process, the loss values of both models show a gradually decreasing trend. However, the loss value of the improved model consistently remains lower than that of the YOLOv8s model. This phenomenon indicates that the improved model not only converges faster but also exhibits better convergence results, further substantiating the significant effect of the SIOU loss function on model optimization and performance enhancement.

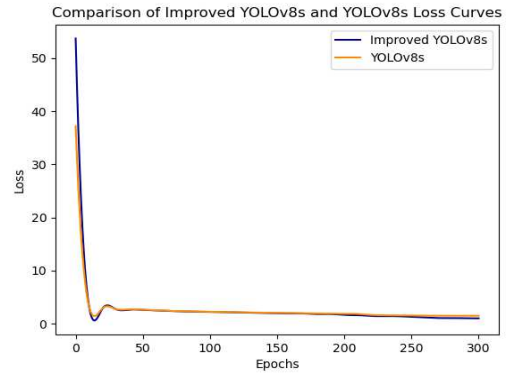


Fig. 12 Comparison of Loss between YOLOv8s and Our Model

4.4.1 Comparative Experiment on the Embedding of Different Attention Modules

To validate the impact of attention modules on object detection performance, four different attention mechanisms were embedded into the backbone detection network: CA[35], CBAM, ECA, and MLKA. This resulted in the improved models YOLOv8s+MLKA, YOLOv8s+CA, YOLOv8s+CBAM, and YOLOv8s+ECA, with no other modifications made to other parts. These models were tested and compared on the insulator dataset, and the results are shown in Table 2.

Table 2 Comparison of Different Attention Modules in the Experiment

Model	Insulator	Defect	mAP50	FPS
-------	-----------	--------	-------	-----

	AP%	AP%		
YOLOv8s	97.80%	96.27%	97.04%	50.89
YOLOv8s+SE	98.40%	96.33%	97.36%	47.50
YOLOv8s+ECA	98.49%	96.38%	97.43%	46.65
YOLOv8s+CBAM	98.72%	96.09%	97.40%	47.81
YOLOv8s+CA	98.98%	96.57%	97.76%	46.23
YOLOv8+MLKA	98.93%	96.89%	97.91%	46.73

The experimental results indicate that all five embedded attention modules, including YOLOv8s+MLKA, YOLOv8s+CA, YOLOv8s+CBAM, YOLOv8s+ECA, and YOLOv8s+SE, can effectively enhance the model's detection performance. Despite a slight decrease in detection speed due to the increased size of all models, the model with the MLKA module exhibited a less significant reduction in speed and a marked improvement in detection accuracy. Specifically, the YOLOv8s+MLKA model achieved the best detection results on the insulator dataset, with an mAP value 0.87 percentage points higher than the original YOLOv8s model. The YOLOv8s+CA model also showed a relatively large performance increase, adding 0.72 percentage points. In contrast, the performance gains of YOLOv8s+SE, YOLOv8s+CBAM, and YOLOv8s+ECA were relatively smaller, with increases of 0.32, 0.36, and 0.39 percentage points, respectively. Based on these experimental results, adding the MLKA module is the most effective for improving object detection performance, better enhancing the model's detection capabilities.

4.4.2 Ablation Experiment

To comprehensively evaluate the impact of the three improvement strategies on the performance of the YOLOv8s model, a series of systematic ablation experiments were meticulously designed and conducted under unified parameter settings, environmental configurations, and dataset conditions. Table 3 provides detailed information on the implementation of each improvement strategy and its specific contribution to model performance.

Table 3 Experimental results of different improvement methods

Model	Insulator AP%	Defect AP%	mAP50	FPS
YOLOv8s	97.80%	96.27%	97.04%	50.89
YOLOv8s+GSC_C2f	98.66%	96.45%	97.56%	56.57
YOLOv8s+MLKA	98.93%	96.89%	97.91%	46.73
YOLOv8s+SIOU	98.11%	97.30%	97.71%	50.52
YOLOv8-GSC	99.43%	99.01%	99.22%	5.73

Based on the aforementioned data analysis, by replacing the C2f module in the backbone detection network of the original YOLOv8s model with the GSC_C2f module, the model's mAP increased by 0.32%, and the FPS increased by 5.14 frames. This improvement not only enhances the model's feature representation capabilities but also effectively reduces computational complexity, further improving the model's computational efficiency. Adding the MLKA module to the backbone detection network of the model increased the mAP by 0.87%. Figure 13(a) depicts the detection heat map of the original model, while Figure 13(b) shows the detection heat map after adding the MLKA module. In

the heat maps, areas with higher brightness indicate where the model's attention is more focused. These images clearly demonstrate the improved accuracy of the modified model in insulator defect detection, further confirming that the addition of the MLKA module significantly enhances the detection precision for insulator defects.

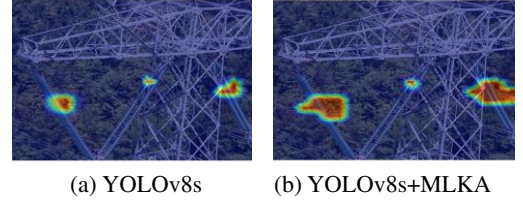


Fig. 13 Comparison of the Improved Model's Heatmap with the YOLOv8s Model Heatmap

Adopting the SIOU loss function in place of the CIoU loss function yields superior performance in guiding model convergence and enhancing the regression accuracy of the prediction network. The AP for the insulator defect category in the improved model increased by 1.03%. Overall, these three improvement strategies have made significant progress in enhancing the model's detection accuracy and speed, resulting in a total increase of 2.18% in mAP. The experimental results strongly validate the practicality and effectiveness of the improvement strategies proposed in this paper.

4.4.3 Comparative experiment of different algorithms

To deeply evaluate the performance of the improved model, a series of carefully designed comparative experiments were conducted. These include Faster R-CNN, SSD, CenterNet, YOLOv3, YOLOv4, YOLOv5s, and the YOLOv8 series. During this process, special attention was paid to three key performance metrics of each detection model: GFLOPS (Giga Floating Point Operations Per Second), Parameters, and model size. These metrics are summarized in Table 4.

Table 4 The parameter comparisons for several algorithms are summarized.

Model	GFLOPS	Parameter	Model Size
Faster R-CNN	370.210G	137.099M	521.51MB
SSD	62.747G	26.285M	100.27MB
CenterNet	70.217G	32.665M	124.61MB
YOLOv3	66.171G	61.949M	236.32MB
YOLOv4	59.960G	63.943M	243.92MB
YOLOv5s	17.156G	7.277M	27.76MB
YOLOv8s	28.817G	11.167M	42.70MB
YOLOv8-GSC	21.517G	7.912M	30.18MB

Through an in-depth analysis and comprehensive comparison of the data in Table 4, it can be observed that the size of the improved model is only 30.18MB, making it the smallest among the tested models, except for YOLOv5s. Additionally, the model

shows significant advantages in terms of GFLOPS and the number of parameters. These results collectively validate the superior performance of the YOLOv8-GSC algorithm in computational efficiency and resource utilization.

The evaluation metrics for each model are summarized in Table 5. The corresponding Precision-recall curves are plotted and

visualized in Figure 14.

Table 5 Comparison of several methods

Model	Insulator AP%	Defect AP%	mAP50	FPS
Faster R-CNN	93.43%	71.28%	82.35%	16.10
SSD	93.94%	83.53%	88.74%	45.98
CenterNet	96.16%	93.89%	95.02%	27.41
YOLOv3	96.97%	93.71%	95.34%	44.16
YOLOv4	96.42%	95.60%	96.01%	46.37
YOLOv5s	96.90%	95.71%	96.31%	53.59
YOLOv8s	97.80%	96.27%	97.04%	50.89
YOLOv8-GSC	99.43%	99.01%	99.22%	55.73

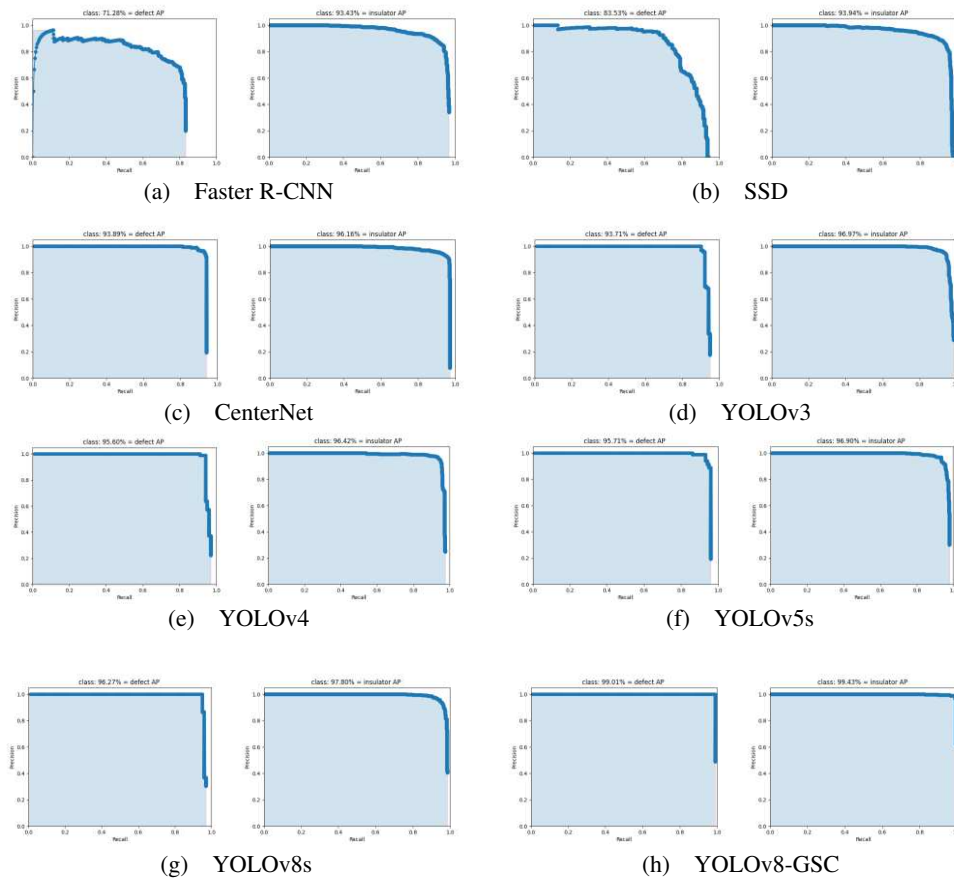


Fig. 14 Precision-recall curves of different models

After a thorough analysis of the data in Table 5 and Figure 13, it is evident that the YOLOv8-GSC algorithm shows significant improvements in all evaluation metrics, outperforming other comparison models. Its mAP value reached 99.22%, which is an increase of 16.87%, 10.48%, 4.20%, 3.87%, 3.21%, 2.91%, and 2.18% compared to Faster R-CNN, SSD, CenterNet, YOLOv3, YOLOv4, YOLOv5s, and the original YOLOv8s, respectively. These results strongly validate the effectiveness of the three

improvement strategies proposed in this paper for enhancing the model's performance.

In terms of processing speed, YOLOv8-GSC achieves 55.73 frames per second, making it the fastest among all tested models. This performance ensures that the improved model can meet the real-time dynamic detection requirements of drones and other embedded devices with limited computational resources.

4.4.4 Comparison of Detection Results Across Different Models

To verify the superior performance of the improved model, eight different detection models were tested on a dataset of insulator images with diverse sizes and shapes against complex backgrounds, as shown in Figure 15. The results indicate that, compared to the other seven models, the YOLOv8-GSC algorithm

exhibits higher confidence in identifying insulator defects without any false negatives or false positives. These empirical findings confirm that the improved model has made significant progress in enhancing the accuracy and efficiency of insulator defect detection tasks, making it more adaptable to the demands of real-world detection applications.

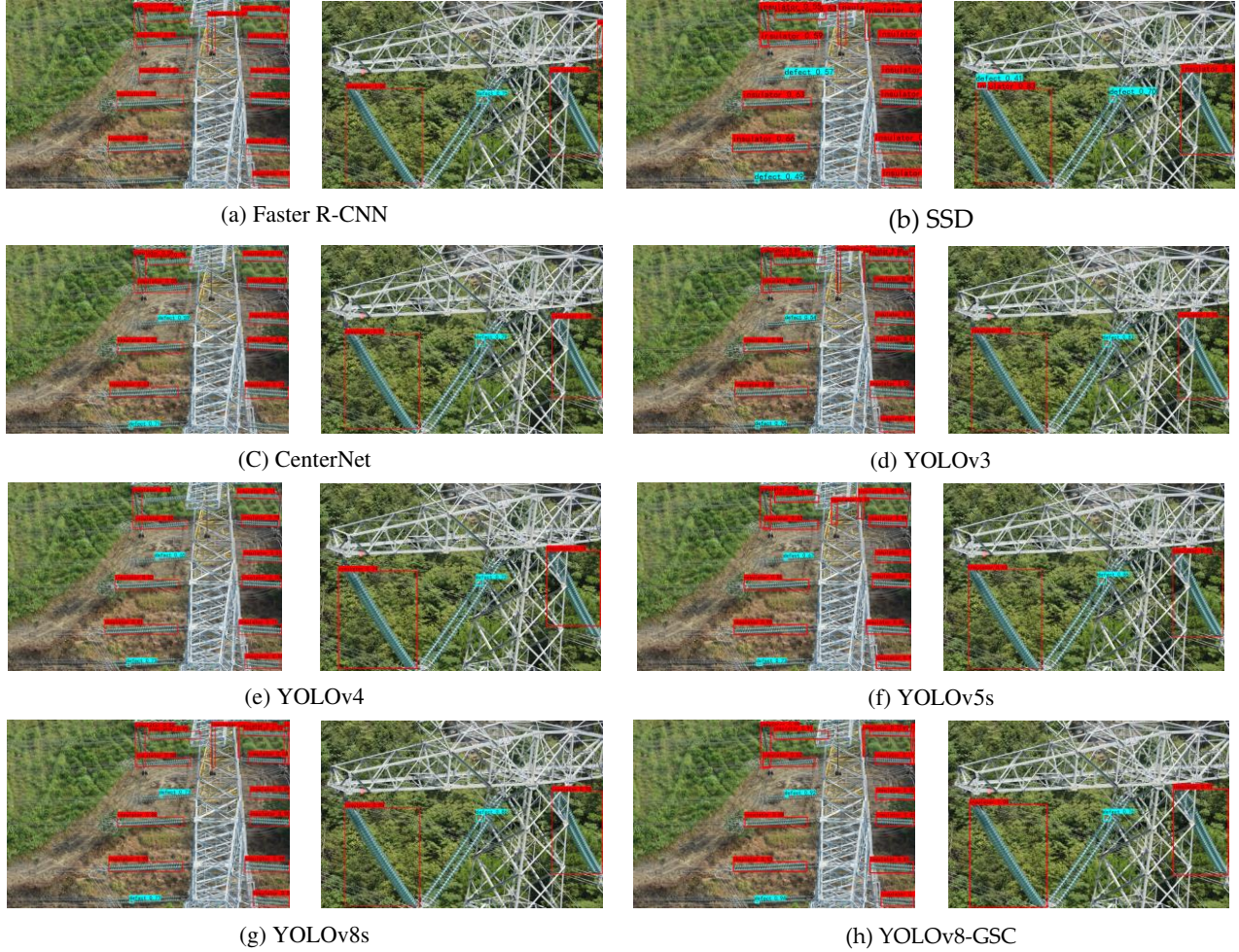


Fig. 15 Display of different model detection effects

4.4.5 Robustness Analysis

To validate the robustness of the improved model, experimental tests were conducted on insulators with and without defects under challenging environmental conditions. The results are illustrated in Figure 16. A comparison between parts (a) and (b) of Figure 16 reveals a noticeable improvement in the detection accuracy of the enhanced model. By contrasting the first and third images in part (a), it's evident that the original YOLOv8s algorithm misses

detections when the background is complex. In contrast, the improved model, which incorporates the MLKA attention mechanism into the backbone detection network and utilizes the SIOU loss function, effectively enhances the feature extraction capability and defect localization accuracy. These improvements significantly increase the accuracy of target identification and substantially reduce the rate of missed detections. Therefore, the experiments confirm that the improved algorithm possesses a certain level of robustness in insulator defect detection.

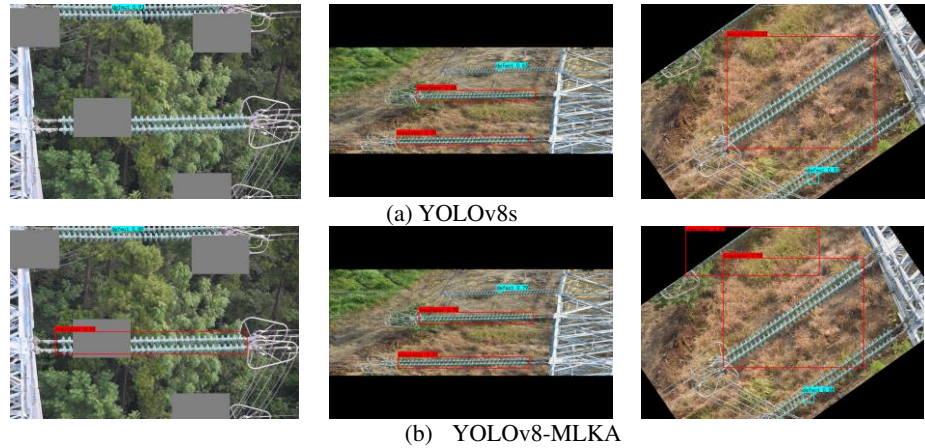


Fig. 16 Comparison of Detection Results in Different Scenarios

5. Conclusions

The primary aim of this paper is to delve deeply into the methods for detecting defects in transmission line insulators within aerial images. We chose the lightweight YOLOv8s network as the foundational model and made improvements to enhance its identification accuracy. Considering the relatively small size of existing datasets, innovative data augmentation techniques were employed to effectively expand the dataset and enhance the model's robustness.

In comparative analysis experiments, the improved model was benchmarked against several advanced models, including Faster R-CNN, SSD, CenterNet, YOLOv3, MobileNetv3-YOLOv4, YOLOv5s, and YOLOv8s, under consistent experimental conditions. The results demonstrate that the enhanced model exhibits significant advantages in detection performance, with a mAP value considerably surpassing other comparison models. The AP value for the normal insulator category increased by 1.63%, and the defect category saw a 2.74% rise, both of which significantly outperformed the YOLOv8s model. Moreover, aside from YOLOv5s, the enhanced model is the smallest in size and has the fewest parameters among all compared models, making it highly valuable for real-time applications and resource-constrained environments.

Furthermore, robustness tests were carried out to assess the model's detection capabilities in complex environmental backgrounds and at various angles. The results further validate the excellent performance of the improved model. It can accurately detect all targets, and its stability and accuracy in complex and changing environments are noticeably better than the original YOLOv8s algorithm.

The primary direction for future work is to deploy the YOLOv8-GSC algorithm into embedded devices. This will better apply object detection algorithms to real-world insulator defect detection projects, improving both the accuracy and efficiency of insulator defect detection, and thus providing strong support for the continuous development of the electrical industry.

Acknowledgements The link to the insulator experimental dataset is as follows:

<http://github.com/InsulatorData/InsulatorDataSet>.

Funding This research is supported by the Youth Foundation of the National Natural Science Foundation of China and the General Program of the Liaoning Provincial Department of Education, with numbers (61903169) and (LJKMZ20220655).

Declarations Conflict of interest The authors declare that there is no conflict of interest

References

1. Martinez C, Sampedro C, Chauhan A, et al. Towards autonomous detection and tracking of electric towers for aerial power line inspection[C]//2014 international conference on unmanned aircraft systems (ICUAS). IEEE, 2014: 284-295.
2. Jinsha Y, Kebin C, Baoshu L. Identification and location of insulator video images based on ASIFT algorithm [J][J]. Electrical Measurement & Instrumentation, 2015, 52(7): 106-112.
3. Xu C, Bo B, Liu Y, et al. Detection method of insulator based on single shot multibox detector[C]//Journal of Physics: Conference Series. IOP Publishing, 2018, 1069(1): 012183.
4. Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector[C]//Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14. Springer International Publishing, 2016: 21-37.
5. Liu J, Liu C, Wu Y, et al. An improved method based on deep learning for insulator fault detection in diverse aerial images[J]. Energies, 2021, 14(14): 4365.
6. Redmon J, Farhadi A. Yolov3: An incremental improvement[J]. arXiv preprint arXiv:1804.02767, 2018.
7. He K, Zhang X, Ren S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE transactions on pattern analysis and machine intelligence, 2015, 37(9): 1904-1916.
8. Qiu Z, Zhu X, Liao C, et al. Detection of transmission line insulator defects based on an improved lightweight YOLOv4 model[J]. Applied Sciences, 2022, 12(3): 1207.

9. Howard A G, Zhu M, Chen B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications[J]. arXiv preprint arXiv:1704.04861, 2017.
 10. Bochkovskiy A, Wang C Y, Liao H Y M. Yolov4: Optimal speed and accuracy of object detection[J]. arXiv preprint arXiv:2004.10934, 2020.
 11. Han G, He M, Gao M, et al. Insulator breakage detection based on improved YOLOv5[J]. Sustainability, 2022, 14(10): 6066.
 12. Cristi, F. Pre-commit-ci[bot] and Glenn-jocher. Ultralytics, Yolov5. 2020. Available online: <https://github.com/ultralytics/yolov5/> (accessed on 11 April 2022).
 13. Wang Q, Wu B, Zhu P, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 11534-11542.
 14. Zhang Z, Huang S, Li Y, et al. Image detection of insulator defects based on morphological processing and deep learning[J]. Energies, 2022, 15(7): 2465.
 15. Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[J]. Advances in neural information processing systems, 2015, 28.
 16. Zhou M, Wang J, Li B. ARG-Mask RCNN: An Infrared Insulator Fault-Detection Network Based on Improved Mask RCNN[J]. Sensors, 2022, 22(13): 4720.
 17. He K, Gkioxari G, Dollár P, et al. Mask r-cnn[C]//Proceedings of the IEEE international conference on computer vision. 2017: 2961-2969.
 18. Reis D, Kupec J, Hong J, et al. Real-Time Flying Object Detection with YOLOv8[J]. arXiv preprint arXiv:2305.09972, 2023.
 19. Wang Y, Li Y, Wang G, et al. Multi-scale attention network for single image super-resolution[J]. arXiv preprint arXiv:2209.14145, 2022.
 20. Guo M H, Lu C Z, Liu Z N, et al. Visual attention network[J]. Computational Visual Media, 2023: 1-20.
 21. Li H, Li J, Wei H, et al. Slim-neck by GSConv: A better design paradigm of detector architectures for autonomous vehicles[J]. arXiv preprint arXiv:2206.02424, 2022.
 22. Gevorgyan Z. SiO loss: More powerful learning for bounding box regression[J]. arXiv preprint arXiv:2205.12740, 2022.
 23. Zheng Z, Wang P, Ren D, et al. Enhancing geometric factors in model learning and inference for object detection and instance segmentation[J]. IEEE transactions on cybernetics, 2021, 52(8): 8574-8586.
 24. Wang C Y, Liao H Y M, Wu Y H, et al. CSPNet: A new backbone that can enhance learning capability of CNN[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. 2020: 390-391.
 25. Wang C Y, Bochkovskiy A, Liao H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 7464-7475.
 26. Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[J]. arXiv preprint arXiv:2010.11929, 2020.
 27. Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 3-19.
 28. He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.
 29. Chollet F. Xception: Deep learning with depthwise separable convolutions[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 1251-1258.
 30. Howard A G, Zhu M, Chen B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications[J]. arXiv preprint arXiv:1704.04861, 2017.
 31. Sandler M, Howard A, Zhu M, et al. Mobilenetv2: Inverted residuals and linear bottlenecks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 4510-4520.
 32. Howard A, Sandler M, Chu G, et al. Searching for mobilenetv3[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2019: 1314-1324.
 33. Zhang X, Zhou X, Lin M, et al. Shufflenet: An extremely efficient convolutional neural network for mobile devices[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 6848-6856.
 34. Ma N, Zhang X, Zheng H T, et al. Shufflenet v2: Practical guidelines for efficient cnn architecture design[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 116-131.
 35. Hou Q, Zhou D, Feng J. Coordinate attention for efficient mobile network design[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 13713-13722.
- Zhang Lin** a graduate student at the School of Electronic and Information Engineering, Liaoning University of Science and Technology. In 2021, obtained a bachelor's degree in software engineering from Liaoning University of Science and Technology. The main research directions include deep learning and image processing.
- Boqun Li** a Professor at Liaoning University of Science and Technology. He obtained his Doctorate in Engineering from Beijing University of Science and Technology in 2004. His main areas of research are control and optimization, as well as deep learning.
- Yushan Lai** a graduate student at the School of Electronic and Information Engineering, Liaoning University of Science and Technology. He received his bachelor's degree in management from Beijing University of Architecture in 2018. His primary area of research is computer vision.
- Yang Cui** is an Associate Professor and Master's supervisor at Liaoning University of Technology. He graduated from Northeast University in 2018, where he specialized in Control Theory and Control Engineering and earned his Ph.D. in Engineering. From September 2019 to August 2020, he worked as a postdoctoral researcher at Lakehead University in Canada. From March 2021 to February 2022, he was a visiting scholar at Harbin Institute of Technology.

Jing Gao a graduate student at the School of Business Management, Liaoning University of Science and Technology.

She obtained her bachelor's degree in management from Hebei University of Technology in 2021. Her main research direction is data prediction analysis.