Lightweight Multi-Object Detection for Electrical Equipment Based on the Improved YOLOv8s Algorithm

Li Li[1], Changjian Zeng[1], Qingbin Guo[1], Zhiyan Zhang[2*], Qiyuan Huang[1], He Li[2]

[1] Fujian Electric Power Co., Ltd. Quanzhou Electric Power Skills Research Institute, Quanzhou 362000, China
[2] School of Electrical and Information Engineering, Zhengzhou University of Light Industry, Zhengzhou 450002, China

Corresponding Author Email: 2004074@zzuli.edu.cn

ABSTRACT

To tackle the issues of small object detection difficulty, and complexity and high computational demands of multi-object detection models for electrical equipment images, a lightweight multi-object detection algorithm based on the improved YOLOv8s algorithm is proposed. Firstly, by integrating Ghostnet lightweight network structure, Convolutional Block Attention Module (CBAM), and Spatial Pyramid Pooling Fast Cross Stage Partial Channel (SPPFCSPC) convolution topology, the model complexity and computational cost are reduced absent a degradation in its discriminative prowess. Secondly, the GELU activation function and Wise-IoU penalty term are introduced to improve the model's robustness and extrapolation prowess. Finally, an experimental dataset that comprises 4,000 images featuring eight types of electrical equipment is constructed using online images and open-source datasets enhanced through data augmentation algorithms. This dataset is employed to train and evaluate the improved YOLOv8 algorithm. The experimental results demonstrate that compared to the original YOLOv8 algorithm, the improved YOLOv8 achieves a 5.2% increase in mAP@50%-95% (mean Average Precision) while increasing the model size by 1MB and computational load by 3.2 GFLOPs. Additionally, all the other metrics of the model also show the improvements. At the same time, a comparative analysis of ablation experiments is conducted between the improved YOLOv8 algorithm and three variant algorithms, namely YOLOv8s, YOLOv8s+CBAM+Ghostnet, and YOLOv8s+CBAM+Ghostnet+SPPFCSPC. The significant improvements in model size, recognition accuracy for small objects, and computational speed of the improved YOLOv8s algorithm are verified by the analysis, which provides an effective solution for lightweight multi-object electrical equipment recognition and detection.

1. INTRODUCTION

As a linchpin of the power grid, the functional integrity of electrical equipment is inextricably linked to the steadiness and trustworthiness of electricity delivery [1]. Throughout the protracted service life of such equipment, a gradual deterioration in performance alongside an emergence of assorted malfunctions is prone to take place [2]. The traditional state detection and fault diagnosis of electrical equipment mainly rely on manual inspection and experience judgment, which is inefficient and easy to miss [3]. Additionally, the technical competence and practical experience of operators also constitute critical factors limiting detection accuracy, thereby posing a formidable obstacle to the full satisfaction of the testing requirements for modern complex electrical equipment. Against this backdrop, sophisticated automation of operational status surveillance and failure etiology analysis for electrical equipment, encompassing processing, analysis, and diagnosis, holds significant engineering value in reducing the limitations of human judgment.

In recent years, deep learning methodologies, notably convolutional neural networks (CNN), have gained widespread application within the purview of object detection. This technology has not only effectively improved the accuracy of conventional recognition tasks but also demonstrated outstanding performance in complex scene recognition and processing [4, 5]. The common algorithms for deep learning-based CNN object detection include R-CNN, Faster R-CNN, Single Shot Multibox Detector (SSD) and YOLO [6]. Because YOLO algorithm balances detection speed and accuracy while also supporting a variety of computer vision tasks, it has become a preferred basic algorithm for object detection tasks, and has been updated to the Yolov8 version with better performance [7-9]. Li et al. [10] developed an improved version of the YOLOv8s algorithm named LW-YOLO. By introducing hierarchical convolutions, multi-depth feature extraction, and multi-scale feature fusion strategies, this algorithm effectively enhances small object detection accuracy. However, its detection rate decreases compared to the original YOLOv8s, and the computational complexity increases accordingly. Cheng et al. [11] integrated an exponential moving average (EMA) attention mechanism and C2F-Faster module into the YOLOv8s backbone network,

constructing a dual YOLOv8 feature pyramid architecture. It further optimizes the loss function by introducing a dynamic non-monotonic focus mechanism, forming a detection algorithm suitable for drone aerial images. The empirical evidence demonstrates that the feature extraction capabilities are strong and the small object recognition performance is good. However, in terms of lightweight design, there is still room for optimization to improve balance between detection speed and accuracy. Yang et al. [12], a strawberry recognition and location method based on an improved YOLOv8 algorithm and pose key point detection technology is proposed. By introducing BiFPN and Generalized Attention Module (GAM) to optimize YOLOv8, the bidirectional information flow and feature weight distribution of the model are strengthened, and the feature extraction of small objects and covered areas is focused, which improved the accuracy of strawberry recognition and stem picking point location. However, the increased complexity of the model structure also leads to longer training cycles and imposes higher demands on hardware configurations, which to some extent limits the model's application in scenarios with constrained computational resources. Wang et al. [13] proposed a Rep-YOLOv8 algorithm, which employs the RepECA architecture as its backbone network. This model replaces standard convolutional layers with RepVGG modules and assimilates an Efficient Channel Attention (ECA) mechanism to augment its representational efficacy. Additionally, an Effective Squeeze Excitation (ESE) self-attention module is embedded within the C2F module to mitigate channel information loss. The experiments demonstrate that Rep-YOLOv8 outperforms the original YOLOv8 algorithm in both inference speed and detection segmentation accuracy. Nevertheless, its practical application still relies on substantial computational power support. On resource-constrained edge devices, the algorithm's demand for computing resources is more obvious. Lou et al. [14], a new network structure Dense Convolution (DC)-YOLOv8 is constructed by using Modified Downsampling Convolution (MDC) module and DC module. The network can effectively downsample, enhance model learning ability and retain more context information. Although the parallel architecture of the DC module enhances feature extraction

efficiency, it incurs higher computational overhead. In contrast, the MDC module further increases the complexity of the network structure, thereby exacerbating demands on computational resources and training conditions. The lightweight multi-object detection algorithm YOLOv8 achieves practical performance standards in detection accuracy and speed, but its substantial parameter count, high storage overhead, and costly computational demands limit its deployment on resource-constrained embedded or edge devices [15].

In order to solve the above problems, reduce the performance requirements of hardware equipment and improve the accuracy of small object recognition, a lightweight multi-object detection algorithm for electrical equipment based on improved YOLOv8s algorithm is proposed. This algorithm integrates the lightweight Ghostnet architecture, the Convolutional Block Attention Module (CBAM), and the Spatial Pyramid Pooling Fast Cross Stage Partial Channel (SPPFCSPC) convolutional module. It significantly reduces model complexity and computational overhead, thereby achieving high-precision detection of electrical equipment. Furthermore, the introduction of the GELU activation function and the Wise-IoU penalty term further amplifies the algorithm's robustness and extrapolation prowess. The resultant data demonstrate that the improved YOLOv8s algorithm fruitfully reduces model complexity and computational burden while maintaining recognition accuracy.

2. YOLOv8 ALGORITHM

YOLO is a low-latency object detection algorithm built upon a deep CNN architecture. This model relies on training with large-scale image datasets to achieve its detection capabilities. Through operations such as normalization, convolution, pooling, and activation, it learns to extract image features and performs object detection by calculating the Intersection over Union (IoU). Compared with traditional object recognition algorithm, YOLO algorithm has faster detection speed and higher detection accuracy [16], and can handle more object objects.
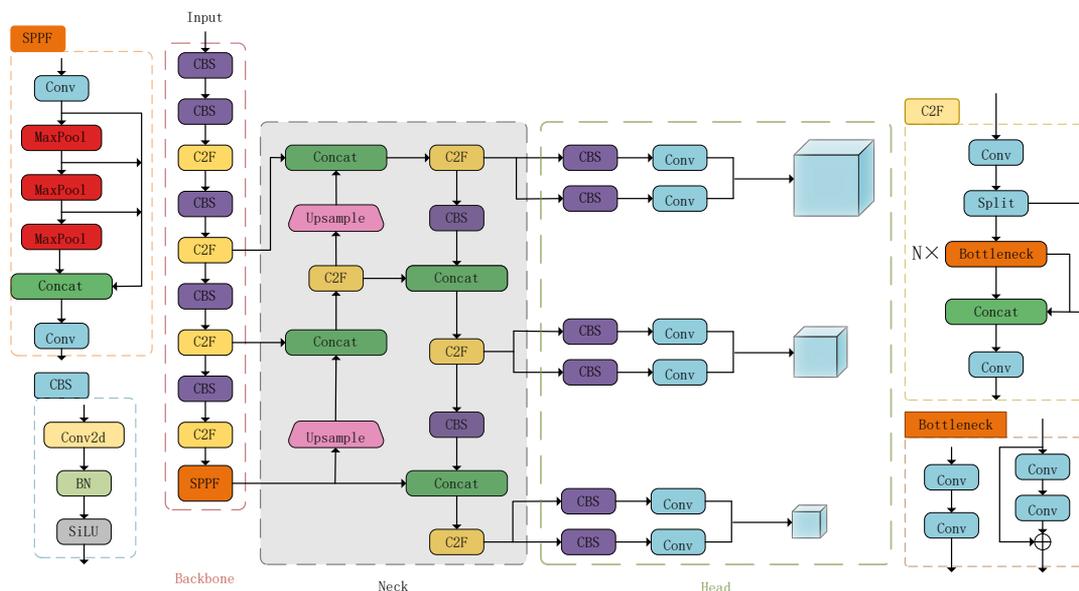

Figure 1. Structure of YOLOv8 object detection algorithm

YOLOv8 is a single-stage object detection algorithm. Compared to its predecessors, this model effects a marked diminution in both parametric volume and computational overhead. It can navigate a disparate array of object recognition tasks and ensure the stability of recognition [17]. YOLOv8 algorithm is mainly composed of four network layers, that is, Input, Backbone, Neck and Head. At the same time, YOLOv8 algorithm solves the problem of accuracy and small object recognition to a certain extent because it adds C2F and SPPF optimization algorithm. Its structure is demonstrated in Figure 1.

3. IMPROVED YOLOv8s ALGORITHM

The direct use of YOLOv8 algorithm for multi-object detection of electrical equipment still has some problems, such as too complex model training calculation, slow training speed, unfavorable to low-end equipment operations. At the same time, the application scenario in the paper is power system, which contains many types of electrical equipment in power system, each type having numerous models. The actual operating condition background is considerably complex, and various equipment is arranged in a staggered manner. After the image is taken, many types of object equipment overlap with each other, and some of this equipment located at a high

distance is far away, resulting in a dark image of the object equipment and a small image area.

Based on this, a comprehensive evaluation of the model size, mean Average Precision (mAP), and training speed across different YOLOv8 versions are performed. The YOLOv8s algorithm, which has a moderate model size and recognition accuracy, is selected as the foundational algorithm for multi-object detection, and the improved YOLOv8s algorithm is hereby proposed. The Conv and C2F modules in the backbone and head networks of the original model are replaced with GhostConv and C3Ghost modules, respectively, to achieve Ghostnet lightweight optimization. CBAM and SPPFCSPC modules are added to the backbone network, with layer counts adjusted to optimize attention mechanisms and small object detection while reducing SPPFCSPC training computational load. The CIOU loss function is replaced with Wise-IoU to enhance learning capability for low-quality annotated boxes, overall performance, robustness, and flexibility. Finally, the activation function is changed from ReLU to GELU to mitigate neuron death and gradient truncation issues, resulting in the improved YOLOv8 algorithm. These enhancements address model size, computational load, small object detection, lightweight architecture, low-quality datasets, and recognition accuracy of the YOLOv8 algorithm. The improved YOLOv8s algorithm model structure is shown in Figure 2.
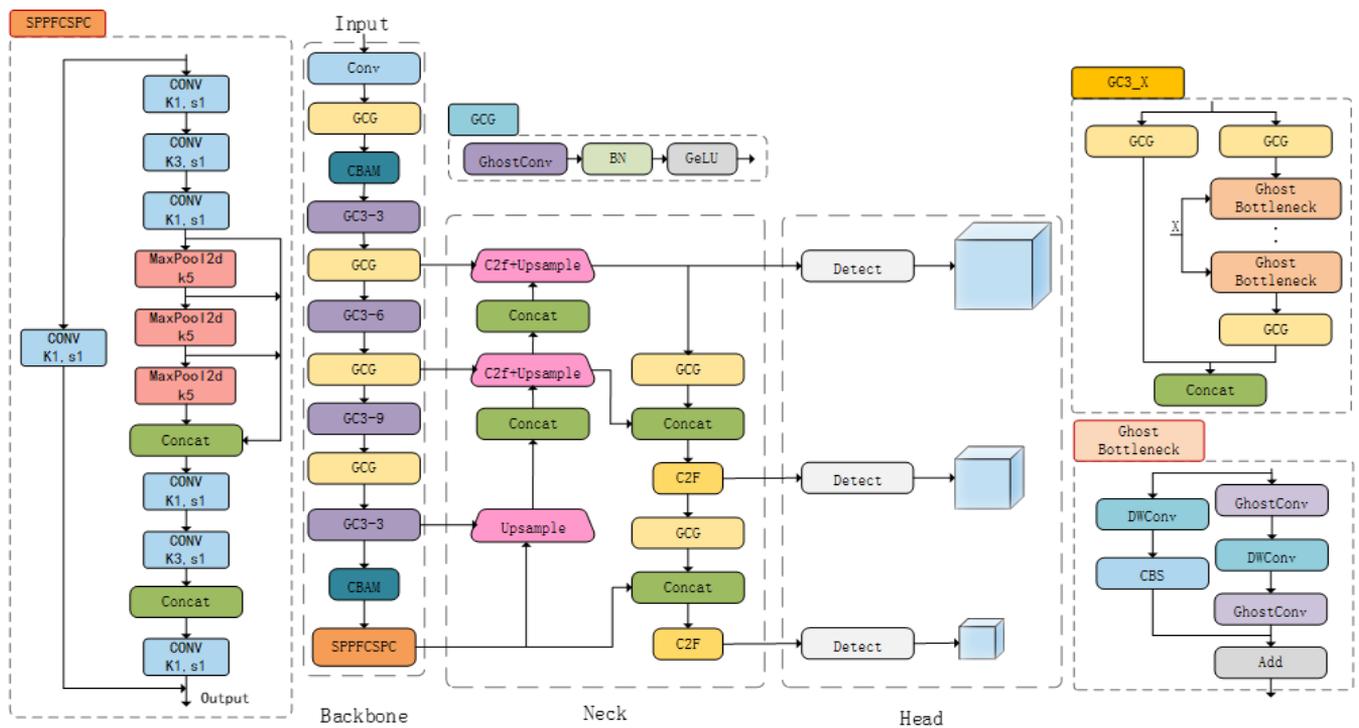

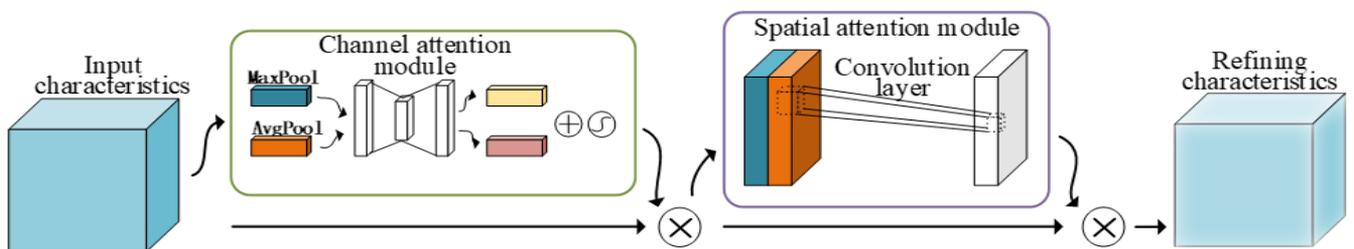**Figure 2.** Structure of the improved YOLOv8 algorithm model

Figure 3. Convolutional Block Attention Module (CBAM) structure diagram

3.1 Convolutional Block Attention Module introduction

This paper enhances the original YOLOv8 model by incorporating a CBAM after the C2F and SPPF modules within its backbone network. This addition equips the YOLOv8 model with an attention mechanism that improves the accuracy of the model. CBAM is a streamlined attention module put forth by Yan [18] that integrates both channel and spatial attention mechanisms. By reasonably allocating feature weights, the model adaptively refines the input image features and extracts informative attentional cues. This mechanism enables it to autonomously learn and attend to key salient features, thereby attenuating the parameter amount and algorithmic complexity [19]. The specific topology of the CBAM module is demonstrated in Figure 3.

As shown in Figure 3, CBAM is capable of extracting the input image and deriving refined features, while effectively reducing redundant information and computational complexity. Specifically, in the channel attention module, CBAM concurrently applies max-pooling and average-pooling to the ingress activation tensor. The two resulting vectors are then processed through a fully connected layer, followed by element-wise addition and an activation function, ultimately producing a channel attention feature that is highly responsive to the location of target objects. Subsequently, in the spatial attention module, the outputs from max-pooling and average-pooling are concatenated, and then processed through a convolutional layer and a Sigmoid activation function, resulting in a spatial weight map. Finally, by multiplying the spatial attention feature and the channel attention feature, an optimized feature representation is obtained that incorporates both spatial and channel dimensions. Integrating CBAM into existing network architectures can significantly diminish redundant information and model size during image processing, while also enhancing computational efficiency.

3.2 Construction of GhostNet lightweight network structure

The YOLO algorithm involves intensive convolutional operations, resulting in high computational demands when training the model on hardware devices. Notably, YOLOv8s algorithm incorporates the C2F convolutional module with richer gradient flow, significantly enhancing detection accuracy. However, it is difficult to achieve good reasoning speed and accuracy due to the inclusion of multiple convolution layers [20]. To attenuate the model's parametric footprint and computational overhead while balancing equilibrating localization fidelity and inference speed on resource-constrained devices, the improved YOLOv8s algorithm integrates the lightweight GhostNet architecture into its backbone and head networks.

The GhostNet network employs multiple convolutional modules based on GhostConv. Among them, the GhostConv convolutional module uses 1×1 convolution kernels to perform simple linear transformations, aggregating information features between the channels. It then applies grouped convolution to generate new Ghost feature maps. While preserving the correlation of the original feature channels and avoiding the loss of feature information, it further adds parallel linear channels, which enhances the operational width of the network. Simultaneously, by replacing the original BottleNeck module with the GhostBottleNeck module, the C3Ghost module is further

constructed. Building upon this foundation, the YOLOv8s algorithm also replaces the original Conv module and C2F module with the GhostConv module and C3Ghost module, respectively. This effectively enhances the lightweight characteristics of the GhostNet model and significantly reduces its overall computational complexity.

3.3 Improvement of small object recognition ability

In order to improve small object recognition of electrical equipment in power system, the SPPF module of YOLOv8s algorithm is improved to SPPFCSPC module, and the perception of diminutive targets is improved, so that the model has stronger recognition ability of small objects than YOLOv8 algorithm.
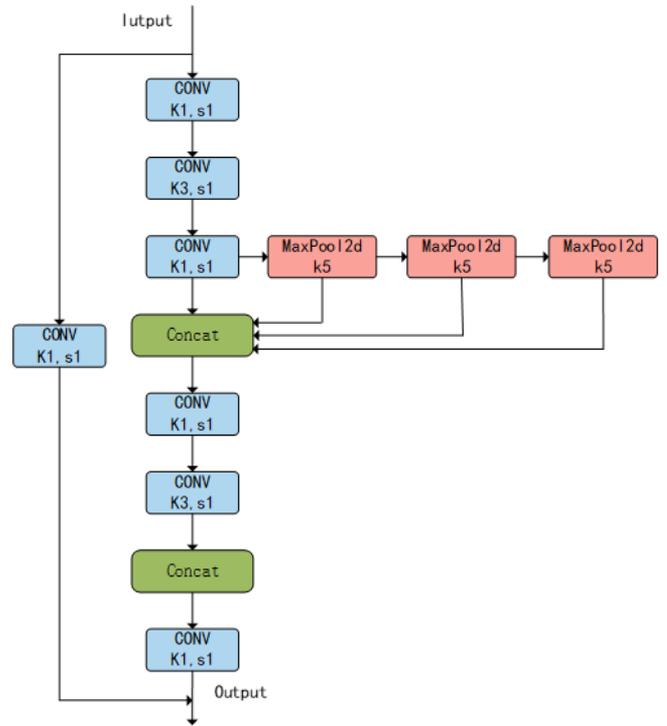

Figure 4. Spatial Pyramid Pooling Fast Cross Stage Partial Channel (SPPFCSPC) spatial convolution structure

The SPPFCSPC convolutional structure is shown in the Figure 4, which is composed of two modules: SPPF and CSPC. Firstly, the original image is partitioned via CSPC convolution. The CSPC convolutional structure integrates 1×1 and 3×3 grouped convolution blocks, along with three identical serial pooling layers. In this paper, the grouped convolution blocks of the CSPC divide the input features into four groups, with each group undergoing independent convolution. Subsequently, the SPPF operation is performed on the partitioned image. The SPPF structure uses smaller kernel structures for pooling, which increases the speed of a single pooling layer. Additionally, it optimizes the original parallel pooling layers of SPP into a serial structure, enabling multiple pooling operations to achieve more abundant feature fusion, while reducing computational complexity and increasing model speed. Finally, the pooled results are concatenated and convolved again to obtain features. The SPPFCSPC module can effectively improve the model's operational efficiency, ensure the recognition accuracy of small objects, and reduce computational complexity.

3.4 Improvement of activation function and loss function

3.4.1 GELU activation function introduction

Activation function commonly used in deep learning models is RELU. This activation function is linear, which is essentially a maximum function. It is simple to calculate, converges rapidly, and avoids vanishing gradient and exploding gradient problems [21]. However, the linear activation function will cause problems such as neuron death and gradient truncation. In order to solve this problem, the improved YOLOv8 algorithm introduces GELU nonlinear activation function with Gaussian error unit [22]. The graphs of these two activation functions and their derivative functions are shown in Figure 5.
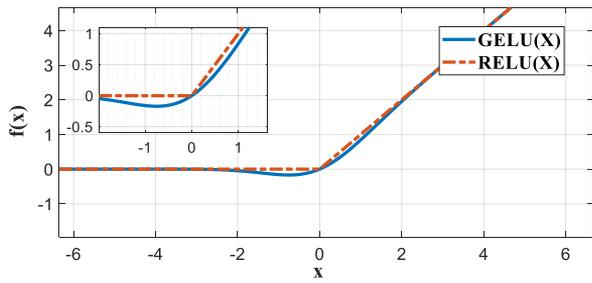
(a) Activation function
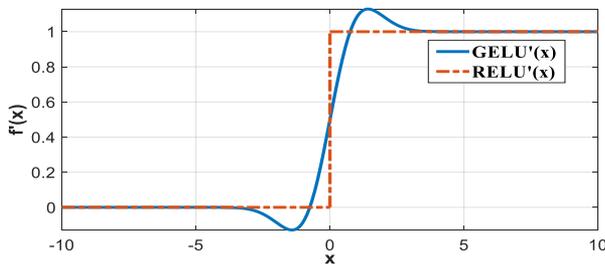
(b) Derivative of activation function

Figure 5. Activation function and their derivative images

It can be seen from Figure 5 that the response of the RELU activation function in the negative interval is 0, which will cause some neurons in the negative interval to be set to zero, resulting in neuron death. At the same time, it can be observed from its derivative function graph that the activation function exhibits truncation near zero; thus it causes gradient truncation to a certain extent. Compared with the RELU activation function, the GELU activation function is smoother near 0 and has a certain gradient value in the negative region, which minimizes the problem of neuronal death. Furthermore, its derivative is smoother and continuous, facilitating backpropagation. Therefore, the GELU activation function not only effectively accelerates convergence during model training but also significantly enhances overall model performance across a multitude of domains such as natural language processing and computer vision. Additionally, the nonlinearity of the activation function can make deep learning model produce more complex mapping.

3.4.2 Wise-IoU loss function introduction

When training neural network model, loss function has a great influence on training effect of the model and performance of the object detection algorithm. When the data presents class imbalance or sample imbalance, selecting appropriate loss function can enhance generalization ability of

the model and reduce negative effects of low-quality samples on the training results, such as bounding box loss and accuracy reduction [23]. However, excessive enhancement of low-quality data can also affect the improvement of model accuracy. To tackle the issue of imbalanced training sample quality in datasets, the improved YOLOv8s algorithm introduces the Wise-IoU non-monotonic penalty term predicated upon an adaptive focusing apparatus, superseding the conventional CIoU loss function.

The Wise-IoU loss function uses dynamic outlier to evaluate the detection box, and the dynamic outlier formula is:

$$\beta = \frac{\mathcal{L}^*_{IoU}}{\bar{\mathcal{L}}_{IoU}} \in [0, +\infty) \tag{1}$$

where, $\mathcal{L}^*_{IoU}$ is monotone focusing coefficient; $\bar{\mathcal{L}}_{IoU}$ is a sliding average value with dynamic change characteristics, so the gradient gain allocation strategy can be dynamically adjusted according to the quality of training samples.

The definition of the Wise-IoU loss function is as follows:

$$\mathcal{L}_{WIoUv3} = r R_{WIoU} \mathcal{L}_{IoU} \tag{2}$$

where, $\mathcal{L}_{IoU}$ is IoU bounding box loss function value; R_{WIoU} is distance between the prediction box and the real box; r is a non-monotone focusing coefficient defined by using the outlier degree β, which is dynamically adjusted according to the quality of the training samples.

After replacing the original CIoU loss function in the YOLOv8s algorithm with the Wise-IoU loss function, samples of high, medium, and low quality all achieve more balanced gradient gains, so this effectively adjusts the contribution of samples of varying quality during training. This methodology not only augments the model's learning capability for low-quality annotated boxes but also improves its overall performance, robustness, and adaptability to different data distributions.

4. DATASET FORMATION AND MODEL EVALUATION INDEX

4.1 Experimental dataset formation

The objects to be detected in the dataset include low-voltage distribution cabinet (dypdg), power transformer (dlbyq), insulator (jyz), transformer bushing (byqtg), safety helmet (helmet), transformer heat dissipation sink (byqsr), transformer oil pillow (byqyz) and drop fuse (dlrdq). A total of 800 images of 8 kind electrical equipment are included. These images encompass diverse background elements such as farmland, buildings, fences, and peripheral equipment, thereby enabling the model to maintain high prediction accuracy even when encountering targets within other complex scenes. After the dataset is formed, the Labelme dataset annotation program is used to annotate the above eight objects.

Because the high definition images of power system electrical equipment are mostly confidential, the high definition image data published on the network is limited, and some image background information in the open source dataset is simple. To mitigate potential underfitting issues in the dataset, data augmentation algorithms are applied to expand the original dataset, thereby reducing sample

distribution imbalance. The data augmentation algorithm in this paper performs the following operations on the original images: adding the salt-and-pepper noise with a noise ratio of 0.3, increasing the Gaussian noise with a mean of 0 and a standard deviation of 0.3, adjusting the brightness by adding or subtracting 70 to each channel of the image, and inverting the original images. The original image and dataset augmentation images are shown in Figure 6.

It can be seen from Figure 6 that each image is amplified into 5 enhancement images by the data augmentation algorithm. After the original 800 images are enhanced, 4000 images are obtained as the dataset, and the augmented dataset is partitioned into training, validation, and testing subsets following a 7:1:2 proportional allocation.
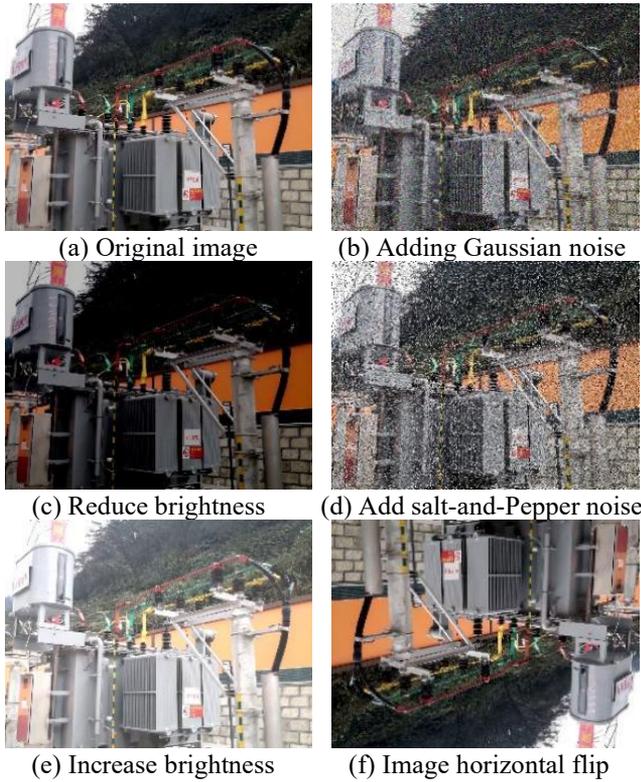

(a) Original image	(b) Adding Gaussian noise
(c) Reduce brightness	(d) Add salt-and-Pepper noise
(e) Increase brightness	(f) Image horizontal flip

Figure 6. Dataset augmentation algorithms

4.2 Model evaluation metric

Following the training phase, the model's performance is weighed via confusion matrices, precision, recall, F1 curves, and mAP. Additionally, the number of parameters (Params) and Giga Floating-point Operations Per Second (GFLOPs) generated during model training serve as metrics for assessing training efficiency. Based on these seven metrics, the proposed model's performance and effectiveness in electrical equipment detection tasks are comprehensively evaluated.

(1) Confusion matrix

YOLO algorithm provides a confusion matrix normalization table for multi-classification problems. Through this table, the classification accuracy between categories can be intuitively understood, and the categories with high accuracy of the training model can be obtained. The confusion matrix of YOLO algorithm is demonstrated in Table 1.

As demonstrated in Table 1, True Positive (TP) represents cases where both the predicted value and the actual value are 1; False Negative (FN) refers to samples where the actual value is 1 but the predicted value is 0; False Positive (FP)

corresponds to instances where the actual value is 0 but the predicted value is 1; True Negative (TN) corresponds to samples where both the predicted value and the actual value are 0.

Table 1. Confusion matrix

Confusion Matrix	True_Label=1	True_Label=0
Predict_Label=1	TP	FP
Predict_Label=0	FN	TN

(2) Precision rate

In the YOLO, the precision rate is ratio of samples with prediction of model value as 1, and actual value of sample is also 1. Calculation by the following method:

$$P = \frac{TP}{TP+FP} \tag{3}$$

where, TP is the number of samples where the predicted value and the true value are both equal to 1; FP is the number of samples where the predicted value is 1 and the true value is 0.

(3) Recall rate

Recall rate indicates the sample probability that the real value is 1, and the predicted value is also 1. The calculation method is shown as follows:

$$R = \frac{TP}{TP+FN} \tag{4}$$

(4) F1 fraction curve

The abscissa of the F1 fraction curve is the confidence, and the ordinate is the F1 fraction. The F1 score serves as a comprehensive performance metric that simultaneously considers both the precision and recall of a model. The magnitude of the quantitative measure is directly correlated with the model's fidelity in object localization tasks. And the missed detection rate and false alarm rate are the smallest. The F1 fraction calculation formula is shown as follows:

$$F1 = \frac{2 \times (P \times R)}{P+R} \tag{5}$$

(5) Average Precision

Average Precision (AP) represents the average of the precision values across all categories. In binary classification tasks, AP measures the average accuracy for a single class; in multi-class scenarios, mAP is typically used as the overall performance metric. Generally, a higher mAP value indicates superior model training effectiveness. The mathematical formulations governing AP and mAP are delineated as follows:

$$AP = \sum_{k=0}^{k=n-1} \big(R(k) - R(k+1) \big) * P(k) \tag{6}$$

$$mAP = \frac{1}{nc} \sum_{k=1}^{k=nc} AP_k \infty \tag{7}$$

where, n is the threshold size, nc is the number of categories, and AP_k is the AP value of the k-th category.

(6) Params and GFLOPs

The number of parameters reflects the spatial scale of the model, representing the total number of parameters that need to be learned during network training. FLOPS denotes the number of floating-point operations a hardware unit can execute per second, while GFLOPs corresponds to the scale of one billion floating-point operations per second. Therefore,

GFLOPs serves as a prevalent benchmark for assessing a model's computational efficiency on the same hardware platform, and together with parameter scale, it collectively reflects the model's overall complexity.

5. EXPERIMENTAL VERIFICATION AND RESULT ANALYSIS

5.1 Experimental environment and training parameter settings

Using PyCharm software as the development environment, based on Python 3.9.19 version, running in the PyTorch deep learning framework, using NVIDIA graphics card and graphics accelerator CUDA to train the improved YOLOv8s algorithm. The PyTorch memory monitoring tool is employed to ensure memory does not overflow. The precise experimental setup is delineated in Table 2.

Table 2. Experimental environment configuration

Name	Specific Configuration
Operating system	Windows 10
CPU processor	Intel i7-12700K
Memory	32GB
Graphics card	NVIDIA GeForce RTX 3050
Video memory	8GB
Graphics accelerator	CUDA 12.6.0
Deep learning framework	PyTorch 2.4.0
Development environment	PyCharm 2023.3.4 Community Edition
Programming environment	Python 3.9.19

The setting of training parameters has a great influence on training results and effects. By configuring parameters in the enhanced YOLOv8s model configuration file, the model is optimized to orient itself towards multifarious operational contexts and functional imperatives. In this case, the iteration count is set to 500 rounds, the input image size is adjusted to 640×640 pixels, and the batch size is set to 16.

To ensure consistency and the uniformity of all the hyperparameters during each model training in this paper, the hyperparameter configuration recommended by the Ultralytics official website is adopted. Both the initial learning rate and final learning rate are set to 0.01. The cosine annealing (cos) algorithm is used to make the model's learning rate dynamically decrease, preventing the model from falling into local optimum. Additionally, the model's training regularization weight decay parameter is set to 0.0005, which penalizes and decays larger weights to prevent the model from overfitting. The detailed parameter configurations are demonstrated in Table 3.

Table 3. Model training parameter settings

Parameter Name	Parameter Value
Image input size	640*640 pixels
Number of training batches	16
Number of training iterations	500
Initial learning rate	0.01
Final learning rate	0.01
Learning rate decline way	Cos
Weight attenuation	0.0005

5.2 Analysis of experimental evaluation metric

To corroborate that the proposed model improves the performance of the YOLOv8s algorithm, the Precision rate, recall rate, mAP@0.5, mAP@0.5:0.95, and F1 Score of both the original and the improved YOLOv8 algorithms are calculated and compared. The training results of YOLOv8s and the improved YOLOv8 algorithms under the same dataset are demonstrated in Table 4.

As shown in Table 4, all the evaluation metrics of the improved YOLOv8s algorithm have been improved, of which 50%-95% mAP has increased by 5.2% and R has increased by 4.1%, indicating that the improvement of the original YOLOv8s algorithm is effective.

Table 4. Evaluation metrics for the training results of two algorithms

Model	P (%)	R (%)	mAP@50%	mAP@50-95%	F_1(%)
YOLOv8s	83.8	75.2	82.6	50.2	76.33
Improved YOLOv8s	86.8	79.3	86.2	55.4	83.02

5.3 Case recognition result comparison

The transformer images in two natural environments are selected from the open source dataset for recognition, which are close to the images containing the dataset labels taken by the inspector's mobile phone. The improved YOLOv8s algorithm is employed for prediction, with the recognition results showed in Figure 7.

(a) Case 1: YOLOv8s algorithm recognition results

(b) Case 1: Improved YOLOv8s algorithm recognition results

(c) Case 2: YOLOv8s algorithm recognition results

(d) Case 2: Improved YOLOv8s algorithm recognition results

Figure 7. Two case recognition results of two algorithms

As demonstrated in Figure 7, the improved YOLOv8s algorithm achieves high-precision recognition of the images taken by mobile phones or with low definition on the network. It can be seen from Figure 7 (a) that the YOLOv8 s recognition algorithm only recognizes the safety helmet (helmet) and transformer bushing (byqtg) with large area display, and the recognition accuracy is relatively low. As demonstrated in Figure 7 (b), the recognition results of the improved YOLOv8s algorithm reach more than 90% when the objects such as power transformer (dlbyq), safety helmet (helmet) and transformer heat dissipation sink (byqsr) are covered.

It can be seen from Figure 7 (c) that in the case 2 of complex background objects, most of the object using YOLOv8s algorithm cannot be effectively identified. In addition, due to

the high image resolution, some small objects cannot be identified, resulting in false recognition errors. As demonstrated in Figure 7 (d), the improved YOLOv8s algorithm can effectively identify most of the objects, and the accuracy of small object recognition is above 75%. The recognition results of case 1 and case 2 show that the improved YOLOv8s algorithm has better recognition effect for both multiple objects and small objects in electrical equipment images, demonstrating its feasibility.

5.4 Ablation experimental analysis

To ascertain the potency of the improved YOLOv8s algorithm, the same dataset, experimental environment platform and training parameters are used to compare the ablation experiments of YOLOv8s, YOLOv8s+CBAM+ Ghostnet, YOLOv8s+CBAM+Ghostnet +SPPFCSPC and the improved YOLOv8s algorithm, and each type results of improvement are verified. The experimental results are shown in Table 5.

As shown in Table 5, compared to the original YOLOv8s algorithm, the YOLOv8s + CBAM + Ghostnet algorithm reduces the number of model parameters by 5 million due to the introduction of the lightweight Ghostnet network architecture. At the same time, R, mAP and F1 do not change much, but the calculational cost increases by 1.9 GFLOPs due to the inception of the CBAM attention mechanism. Compared with the YOLOv8s algorithm, the YOLOv8s+CBAM+ PPFCSPC+Ghostnet algorithm sees increases of 5.5% in P, 5.7% in the F1 score, and 2% in mAP, due to the addition of the SPPFCSPC convolutional structure. However, this comes at the cost of an increase of 1M parameters and a 3.2 GFLOPs rise in computational cost. By incorporating the GELU activation function and Wise-IoU loss function, the improved YOLOv8s algorithm achieves a 3% reduction in accuracy rate compared to the ablation study version, yet demonstrates a 3% improvement over the original YOLOv8s algorithm. Concurrently, the improvements are observed in R-value, mAP, and F1 score. Furthermore, the enhanced algorithm demonstrates strengthened average precision and overall recognition performance, while maintaining comparable model parameter counts and GFLOPs computational requirements to the original YOLOv8s algorithm. Therefore, according to the analysis of ablation experiments, the improved YOLOv8s algorithm can effectively identify small objects and reduce the volume of YOLOv8s algorithm, which meets the needs of multi-object recognition and lightweight detection algorithms for electrical equipment.

Table 5. Ablation experiment results comparison

Algorithm	P (%)	R (%)	mAP@50%	F_1(%)	Model Size (MB)	Computation (GFLOPs)
YOLOv8s	83.8	75.2	82.6	76.33	17 M	28.6
YOLOv8s+CBAM+Ghostnet	84.8	73.8	82.8	75.89	12 M	30.5
YOLOv8s+CBAM+SPPFCSPC+Ghostnet	89.3	73.6	84.6	82.03	18 M	31.8
Improved YOLOv8 algorithm	86.8	79.3	86.2	83.02	18 M	31.8

6. CONCLUSION

Based on the analysis of YOLOv8 algorithm, the lightweight multi-object detection of electrical equipment is realized by improving the lightweight network structure, attention mechanism, convolution structure, loss function and activation function. The enhanced dataset is used to compare

the case identification and ablation experiments. It is verified that the improved YOLOv8s algorithm achieves the expected accuracy while ensuring that the model volume is as small as possible. The following conclusions are obtained:

(1) Compared with YOLOv8s algorithm, the recognition accuracy of the improved YOLOv8s algorithm is enhanced due to the introduction of CBAM and the use of Ghostnet

lightweight network. In parallel, the parametric count remains unchanged and the amount of calculation only increases by 3.2 GFLOPs.

(2) Compared with YOLOv8s algorithm, the improved YOLOv8s algorithm has a great improvement in volume, recognition accuracy, calculation amount and parameter quantity. The mAP and F1 fraction are increased by 3.6% and 6.69% respectively, and the improved YOLOv8s algorithm is more obvious.

(3) The results of the image recognition cases show that the improved YOLOv8s algorithm is ideal for the scenery with a large number of equipment and a small object area. At the same time, it also has a certain recognition accuracy for objects covered by a large area.

(4) The GELU activation function has better performance in dealing with large amounts of data and deep neural networks while effectively preventing overfitting. Furthermore, the dynamic bounding box mechanism in Wise-IoU contributes to enhanced detection by reducing localization loss and improving object positioning accuracy. These improvements collectively indicate that the modified YOLOv8s algorithm is well-suited for lightweight multi-object detection in electrical equipment.

The recognition accuracy and effect of the improved YOLOv8s algorithm all meet the needs of lightweight electrical equipment detection and recognition, but the model volume and training time still have compression space. In addition, the training and verification of the improved YOLOv8s algorithm are performed using a desktop computer. The next step is to transplant the model and program to mobile terminals such as mobile phones or embedded systems, and run the model at any time for image recognition, providing a more convenient way for power grid operation and maintenance and electrical equipment recognition.

REFERENCES

[1] Long, L. (2022). Research on status information monitoring of power equipment based on Internet of Things. Energy Reports, 8: 281-286. https://doi.org/10.1016/j.egyr.2022.01.018

[2] An, Y., Yin, K., Huang, T., Hu, Y., Ma, C., Yang, M., Chen, D. (2022). Study on the insulation performance of SF6 gas under different environmental factors. Frontiers in Physics, 10: 820036. https://doi.org/10.1016/j.egyr.2022.01.018

[3] Lu, F., Niu, R., Zhang, Z., Guo, L., Chen, J. (2022). A generative adversarial network-based fault detection approach for photovoltaic panel. Applied Sciences, 12(4): 1789. https://doi.org/10.3390/app12041789

[4] Zhang, Y.T., Huang, D.Q., Wang, D.W., He, J.J. (2023). A review of research and application of object detection algorithms based on deep learning. Computer Engineering and Applications, 59(18): 1-13.

[5] Amjoud, A.B., Amrouch, M. (2023). Object detection using deep learning, CNNs and vision transformers: A review. IEEE Access, 11: 35479-35516. https://doi.org/10.1109/ACCESS.2023.3266093

[6] Lin, S.Y., Wu, Y.Q. (2024). A review of vision-based defect detection methods for LCD/OLED screens. Journal of Image and Graphics, 29(5): 1321-1345.

[7] Mi, Z., Lian, Z. (2024). A review of YOLO methods for general object detection. Computer Engineering and Applications, 60(21): 38-54. https://doi.org/10.3778/j.issn.1002-8331.2404-0130

[8] Chen, B., Liu, Z. (2024). Lightweight target detection algorithm based on partial convolution. Journal of Electronic Imaging, 33(2): 023049-023049. https://doi.org/10.1117/1.JEI.33.2.023049

[9] Zhou, F., Guo, D.D., Wang, Y., Wang, Q.Q., Qin, Y., Yang, Z.M., He, H.J. (2024). Traffic surveillance vehicle detection algorithm based on improved YOLOv8. Computer Engineering and Applications, 60(6): 110-120. https://doi.org/10.3778/j.issn.1002-8331.2310-0101

[10] Li, S., Li, S.Y., Liu, G.Q. (2024). Research on small object detection algorithm based on YOLOv8 for UAV aerial images. Journal of Chinese Computer Systems, 45(9): 2165-2174. https://doi.org/10.1109/JSTARS.2023.3339235

[11] Cheng, H.X., Qiao, Q.Y., Luo, X.L., Yu, S.J. (2024). object detection algorithm for UAV aerial images based on improved YOLOv8. Radio Engineering, 54(4): 871-881. https://doi.org/10.3969/j.issn.1003-3106.2024.04.010

[12] Yang, Z.Y., Wang, X.C., Qi, Z.H., Wang, D.Z. (2024). Strawberry recognition and stem picking KEYPOINT detection based on improved YOLO v8. Transactions of the Chinese Society of Agricultural Engineering, 40(18): 167-175. https://doi.org/10.11975/j.issn.1002-6819.202405044

[13] Wang, Y.S., Hua, H.B., Kong, M., Liang, X.Y. (2024). Rep-YOLOv8 vehicle and pedestrian detection and segmentation algorithm. Modern Electronics Technique, 47(9): 143-149. https://doi.org/10.16652/j.issn.1004-373x.2024.09.026

[14] Lou, H., Duan, X., Guo, J., Liu, H., Gu, J., Bi, L., Chen, H. (2023). DC-YOLOv8: Small-size object detection algorithm based on camera sensor. Electronics, 12(10): 2323. https://doi.org/10.3390/electronics12102323

[15] Guo, X.T., Xie, X.S., Lang, X. (2023). Pruning feature maps for efficient convolutional neural networks. Optik, 281: 170809. https://doi.org/10.1016/j.ijleo.2023.170809

[16] Lin, P., Li, D., Jia, Y., Chen, Y., Huang, G., Elkhouchlaa, H., Lu, H. (2022). A novel approach for estimating the flowering rate of litchi based on deep learning and UAV images. Frontiers in Plant Science, 13: 966639. https://doi.org/10.3389/fpls.2022.966639

[17] Shan, F., Li, H., Sun, H., Nie, S.G., Shen, Z.H. (2024). Pavement distress identification method based on improved simAM-YOLOv8. Journal of Jilin University (Engineering and Technology Edition), 1-15.

[18] Yan, J.J. (2022). Research on damage detection algorithm for catenary insulators based on deep learning. East China Jiaotong University.

[19] Wang, Z., Qu, S.J. (2024). Research progress and challenges in real-time semantic segmentation based on

deep learning. Journal of Image and Graphics, 29(5): 1188-1220. https://doi.org/10.11834/jig.230605

[20] Wang, C.Y., Bochkovskiy, A., Liao, H.Y.M. (2023). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 7464-7475. https://doi.org/10.48550/arXiv.2207.02696

[21] Yang, Y., Yang, S., Yan, M., Hu, C.L., Pei, S.T. (2024). Edge-end insulator discharge severity assessment algorithm based on UDD-YOLO. Journal of Electronic Measurement and Instrumentation, 38(1): 219-227.

[22] Hendrycks, D., Gimpel, K. (2016). Gaussian error linear units (Gelus). arXiv preprint arXiv:1606.08415. https://doi.org/10.48550/arXiv.1606.08415

[23] Ma, C.F., Li, X., Wang, X.X., Chen, X. (2024). Lightweight non-motor vehicle target detection based on re-parameterization. Computer Engineering and Applications, 16(19): 190-198. https://doi.org/10.3778/j.issn.1002-8331.2403-0292