International Journal of Science and Engineering Applications
Volume 13-Issue 10, 118 – 123, 2024, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1310.1024

Insulator Defect Detection Model based on Improved YOLOv7

Shang Yuhao
College of Electronic Information and Electrical Engineering
Yangtze University
Jingzhou City, Hubei Province
China

Abstract: Insulators are important devices for circuit insulation and fixing wires in transmission lines, and their surfaces are often damaged by lightning discharge, rainwater, and other harsh natural environments. To address the issues of low recognition accuracy, complex recognition background, and small targets caused by outdoor lighting and complex background, this paper proposes a transmission line insulator defect detection method based on improved YOLOv7. The model introduces the CNeB module, which uses convolutional kernels of different sizes and depths to capture features at different scales, thereby helping to capture the diversity and complexity of the target. Secondly, the GAM attention mechanism is introduced in the head part to reduce information dispersion while enhancing the important features of the target in both spatial and channel dimensions. Finally, the Wise IOU loss function is used to improve the accuracy and generalization ability of the model. Through testing on the public dataset CPLID, the average accuracy value of this model is 98.3%, which is 5% higher than the Yolov7 model, fully proving that this model can meet the requirements of insulator defect detection.

Keywords: yolov7; insulator defect detection; target detection; attention mechanism; loss function

1. INTRODUCTION

Electricity production is the nerve center of contemporary socio-economic development. Electricity supply promotes the development of industrial production, commercial operations, and service industries. It provides reliable electricity supply for enterprises and lays the foundation for various services to society, laying the foundation for sustainable development and future social progress. As the neural network in power production, transmission lines are key facilities for transmitting the electrical energy generated by power plants from power plants to users[1], playing a crucial role in energy transmission, distribution, grid stability, and transmission loss control[2]. Insulators, as an important component of high-voltage transmission lines, are devices used for electrical insulation and mechanical fixation of electrical equipment or conductors with different potentials. They play an important role in electrical insulation and mechanical support.

High voltage insulators not only have to withstand the working voltage during operation, but also suffer from overvoltage voltage; Not only should it be able to withstand mechanical forces, but it should also be affected by temperature changes and surrounding environmental conditions. Therefore, insulators are prone to damage during operation, but this damage accumulates over time. When damaged to a certain extent, insulators will lose their required insulation or the ability to support and fix current conductors, ultimately causing faults in the high-voltage power grid system[3]. According to statistics, accidents caused by various types of insulation faults account for over 50% of power grid accidents[4].

The insulator string is located near the tower of the transmission line, which passes through various terrains such as forests, towns, rivers, lakes, deserts, etc., resulting in complex backgrounds in aerial images captured by drones. Therefore, these captured images of aviation insulators often suffer from inevitable image quality difficulties such as occlusion, blurring, and object interference. In addition, insulators are made of different materials (glass, ceramics, and composite materials) and have different shapes due to different voltage levels. The different appearance characteristics of insulators also make it difficult to detect insulator defects in practical scenarios.

In recent years, with the continuous development of deep learning technology, more and more object detection algorithms have been applied in the field of insulator defect detection. At present, most object detection algorithms are mainly divided into single step detection method and two-step detection method. The commonly used two-step detection methods such as CNN, Fast R-CNN, Faster R-CNN, etc. first generate candidate regions through some simple candidate region generation methods (such as Selective Search), and then classify and locate these candidate regions. These algorithms have high accuracy and can handle complex scenes and multi-scale targets, but have high computational complexity and slow speed. Based on this, various improvement methods have been proposed to solve the problems encountered in two-step object detection for insulator defect detection tasks. Antwi Beko[6] et al. used Convolutional Neural Networks (CNN) to detect and classify defective insulators in transmission line images, achieving high-precision defect target recognition and localization. Z.[5] We used Faster R-CNN to search for insulators from the original image, and then used U-Net to segment insulators and cut out local insulator images for further object segmentation. Li et al. [7]proposed a method for identifying defects using a cascaded detection and segmentation network, using Faster R-CNN to capture defects and insulators throughout the entire image, and then feeding the detection results to the U-Net network for further detection, thereby reducing missed detections.

Common single step detection algorithms such as YOLO (You Only Look Once) and SSD (Single Shot Multi Box Detector) can directly predict the position and category of targets from images without the need for candidate region generation. These algorithms have fast computation speed and are suitable for real-time applications, but their accuracy is relatively low, and their processing ability for complex scenes and multi-scale targets is limited. In order to further improve the recognition

ability of neural networks for visual defects in transmission lines, Yuan et al.[8] introduced attention mechanism and small object detection layer in YOLOv5 neural network; To accelerate the convergence speed of the model, Han et al. [9]proposed the SIoU loss function to reduce the degrees of freedom in regression, which accelerates the convergence of the network and successfully improves the recognition accuracy of the network model; Huang et al.[10] integrated the parameter free attention module SimAM into the YOLOv5 network to enhance its feature extraction capability, and introduced the Wise IoU loss function to evaluate the quality of anchor boxes and allocate various gradient gains to them, thereby improving network performance and generalization ability; Yan et al.[11] attempted to replace the ELAN module with ConvNeXt's CNeB module, using the normalization layer and LN layer of the CNeB module to reduce the model's dependence on initialization parameters and further improve accuracy. Finally, a linear classifier was used to output feature maps to improve detection accuracy and address false positives in complex environments.

At present, the detection of insulator defects in transmission lines is mainly based on unmanned aerial vehicle (UAV) technology cruise photography. Due to the complex detection environment such as small detection targets, low visible light, and many obstructions, previous algorithm models have led to serious problems such as decreased detection accuracy, incorrect target detection, and missed detection in UAV inspection tasks. To address this phenomenon, this paper proposes an insulator defect detection model based on improved YOLOv7 (as shown in Figure 1), which introduces the addition of GAM (Global Attention Mechanism)[12] attention mechanism module and improves the original loss function in the original YOLOv7 model. Relevant experiments have shown that the improved model can effectively improve the detection accuracy of insulator defects. The improvements of this article are as follows:

(1) The improvement of the backbone network and head network replaces the two sets of E-Elan modules in the backbone network with a lighter CNeB[13] module, which is used to capture different scale features of the target, extract richer target information, and improve the speed and accuracy of model detection.

(2) Introducing the global GAM attention mechanism with excellent performance into the YOLOv7 head network MPConv module effectively focuses on the key feature information of small and occluded targets, improving the detection accuracy of small and occluded targets.

(3) A new regression loss function Wise IOU has been introduced to enhance the accuracy of detection box localization and model convergence speed.

2. YOLOv7 Network Architecture

As a representative of single step detection algorithms, YOLOv7 has been widely used in the field of object detection since its proposal by Alexey Bochkovskiy et al.[14] in 2022. The detection results on the COCO dataset show that YOLOv7 has greatly improved both accuracy and speed compared to the previously proposed YOLO algorithm. In the range of 5FPS to 120FPS, it outperforms previous YOLO algorithms in both speed and accuracy. At the same time, it also has high accuracy in real-time object detection algorithms. Therefore, this paper chooses the YOLOv7 algorithm model to detect insulator defects. The YOLOv7 algorithm model consists of four parts: Input layer, Backbone network, Head network, and Detect layer.

2.1 Input

Mosaic data augmentation, adaptive anchor box calculation, and adaptive image scaling were used in the input layer to enrich the detection dataset, adaptively calculate the best anchor box values from different training sets, and uniformly scale the original images to a standard size of 640x640 × 3.

2.2 Backbone

The main function of Backbone backbone network is to extract features from input images and generate feature maps with rich semantic information. The Backbone network consists of multiple convolutional layers, E-ELAN layers, and MPConv layers, where the CBS convolutional layer is composed of CBS convolutional layer+BN layer+activation function, and the activation function is SiLU. Firstly, the input image is processed through four convolutional layers (CBS), named C1, C2, C3, and C4. Each convolutional layer is followed by a normalization layer (BN) and an activation function (SiLU). Through these processes, the size of the feature map is changed from the original input image size to 160 * 160 * 128. Then, the feature map will go through the ELAN module proposed in the paper, which is composed of multiple CBSs. The input and output feature sizes remain unchanged, and the number of channels will change in the first two CBSs. The subsequent input channels are consistent with the output channels. After the ELAN module, there are three combined outputs of MPConv and ELAN, corresponding to the outputs of C2/C3/C4, with sizes of 80 * 80 * 512, 40 * 40 * 1024, and 20 * 20 * 1024, respectively.

2.3 Head

The head network is the core part of the model, whose main function is to further process the information features extracted by the backbone network to generate the final object detection output. Firstly, the head network adopts the SPPCSPC network structure, which extracts feature information of different scales such as the position, size, and category of the target by using pooling kernels of different sizes. Then, these feature maps are concatenated together to form a feature map with a larger receptive field. Then, the head network will upsample the feature map through UPSample and fuse it with the original feature map. This operation can help the model better understand the contextual information in the image, reduce computational complexity, achieve speed improvement, and enhance the accuracy of object detection.

After upsampling and feature fusion, the head network will divide the feature map into different regions. Each region will predict a certain number of targets and their location information. This is achieved through the use of grid attention mechanism, which can help the model focus on important regions in the image.

2.4 Detect

The detection layer of YOLOv7 mainly consists of a Head detection head, a loss function part, and optimization strategies on the Head side. Among them, the Head detection head is the last layer of the model, responsible for further processing and predicting the features extracted from the head to generate the final object detection result. The YOLOv7 algorithm has three detection heads, which are used to detect and output the predicted category probability, confidence, and predicted localization information of the target object. The detection head outputs three feature scales: small target 20 × 20, medium target 40 × 40, and large target 80 × 80.

International Journal of Science and Engineering Applications
Volume 13-Issue 10, 118 – 123, 2024, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1310.1024

3. Improved YOLOv7 detection model

3.1 Improvement of Backbone Network

ConvNeXt[15] is a pure convolutional neural network proposed by Facebook AI Research and UC Berkeley. Through a series of experimental comparisons, under the same FLOPs, ConvNeXt has faster inference speed and higher accuracy compared to Swin Transformer. ConvNeXt achieved an accuracy of 87.8% on ImageNet 22K. The ConvNeXt structure is a convolutional neural network architecture used for image classification tasks, which enhances feature representation ability by concatenating feature maps from multiple paths. It has the characteristics of high accuracy, high efficiency, strong scalability, and very simple design.

ConvNeXt is an improvement based on ResNet50, which, like Swin Transformer, has 4 stages; The difference is that ConvNeXt has changed the ratio of the number of blocks in each stage from 3:4:6:3 to the same 1:1:3:1 as Swin Transformer. In addition, in terms of feature map downsampling, ConvNeXt adopts a convolution kernel with a stride of 4 and a size of 4 × 4, which is consistent with Swin Transformer. The ConvNeXt structure diagram is shown in Figure 1.

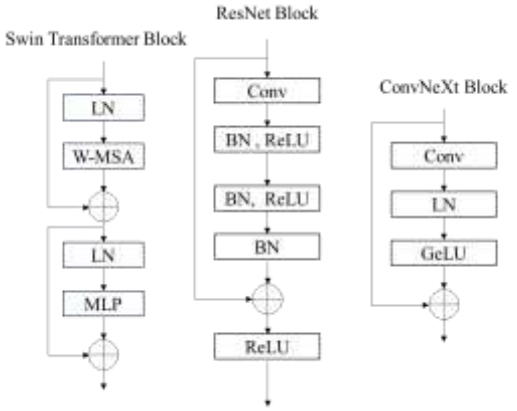

Figure.1 Structural diagram of Swin Transformer, ResNet, and ConvNeX Block

The CNeB (ConvNeXt Block) block consists of a set of parallel convolutional paths, each consisting of a series of convolutional layers and activation functions, used to extract features at different levels of abstraction, as shown in Figure 3. Specifically, we can use convolution kernels of different sizes and depths to capture features at different scales. Therefore, in order to enhance the feature extraction capability of the backbone network and improve recognition accuracy, we applied the basic unit of CNeB to the backbone network in YOLOv7, replacing the original ELAN network module, as shown in Figure 2.

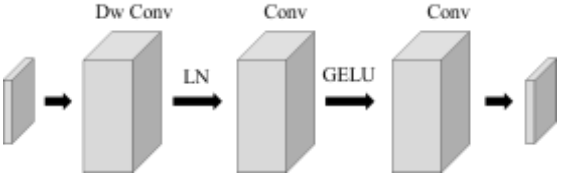

Figure.2 ConvNeX Block Structure Diagram

3.2 Improve attention mechanism

In the detection of insulator defects in transmission lines, transmission lines pass through various terrains (forests, towns, rivers, lakes, deserts, etc.), and insulators are often obstructed by various objects, which increases the difficulty of identification. In 2021, Liu et al. proposed the Global Attention Mechanism (GAM), which solves the problem of insulator occlusion during the detection of small or occluded objects. By using GAM, more key features that are beneficial for identifying the region can be extracted.

GAM adopts the channel attention and spatial attention module structure of CBAM, but designs internal sub modules as shown in Figure 3.

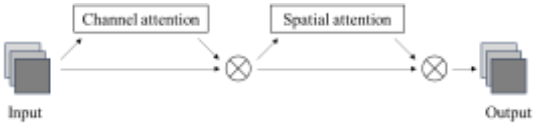

Figure.3 Module diagram of GAM attention mechanism

The channel attention submodule uses a three-dimensional arrangement to preserve information in three dimensions. It amplifies cross dimensional channel spatial dependencies using a two-layer MLP (Multi Layer Perceptron) while establishing nonlinear relationships between channels (MLP is an encoder decoder structure with a reduction rate r, the same as BAM). The channel attention submodule does not include pooling operations, so the shape of its output tensor will remain consistent with the shape of the input tensor. The construction of the channel attention submodule can capture the relationships between channels and enhance information exchange, thereby helping to improve the classification performance of the model.

Figure.4 GAM attention mechanism channel sub attention module diagram

The spatial attention submodule is mainly composed of two 7x7 convolutional layers, with the main goal of focusing on spatial information. These two convolutional layers perform the task of spatial information fusion, which helps enhance the model's attention to various regions in the input image. At the same time, pooling operations were also removed in the spatial sub attention module to further preserve feature mapping.

In CBAM, the spatial attention submodule also uses the same reduction ratio r as BAM from the channel attention submodule to better focus on the relationships between channels and suppress unimportant information. Due to the potential reduction in information utilization caused by max pooling operation, it was not used in the GAM attention mechanism. In this way, the GAM attention mechanism can effectively capture important areas of the image and suppress unnecessary information, thereby improving the performance of deep neural networks in processing image classification tasks.

Figure.5 GAM attention mechanism spatial sub attention module diagram

3.3 Improve the loss function

In object detection, the bounding box loss function, as a key step in determining the performance of target localization, can effectively improve the detection accuracy and convergence

speed of the model. Figure 6 shows the real box and predicted box regions of the prior box, with an overlapping area of $S_u = wh + w_{gt}h_{gt} - W_iH_i$。

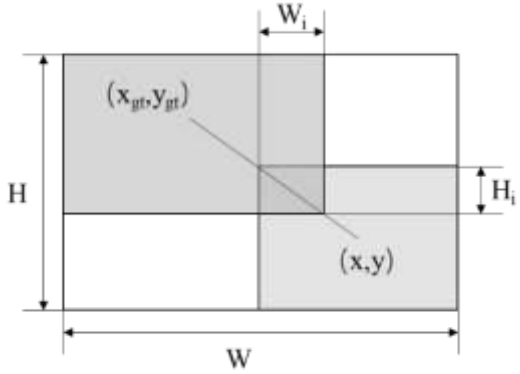

Figure.6 Calculation parameter diagram of the real box and predicted box of the prior box

In the YOLOv7 model, the CIoU loss function[16] is used as the loss function for bounding box regression, which is obtained by adding aspect contrast to DIoU. CIoU solves the problem of IoU gradient vanishing, while considering the overlapping areas of the bounding boxes and the distance between the center points, making the aspect ratio of the bounding boxes consistent. The formula is as follows:

$$IoU = \frac{|B \cap B^{gt}|}{|B \cup B^{gt}|} \quad (1)$$

$$L_{IoU} = 1 - IoU = 1 - \frac{W_iH_i}{S_u} \quad (2)$$

$$L_{DIoU} = 1 - IOU + \frac{\rho^2(B, B^{qt})}{c^2} \quad (3)$$

$$L_{CIoU} = 1 - IoU + \frac{\rho^2(B, B^{gt})}{c^2} + \frac{v}{(1 - I_CU) + v} \times v \quad (4)$$

$$v = \frac{4}{\pi^2}(arctan\frac{W^{gt}}{h^{gt}} - arctan\frac{W}{h})^2 \quad (5)$$

Among them, $B^{gt} = (x^{gt}, y^{gt}, w^{gt}, h^{gt})$ represents the predicted box, and B=(x, y, w, h) represents the real box, representing the Euclidean distance between the center points of the predicted box and the real box. The diagonal distance of the smallest closed area that can contain both predicted and real boxes, and represent the width and height of the real box, and represent the width and height of the predicted box.

CIoU considers the issues of bounding box overlap distance, center point distance, and aspect ratio. However, when two predicted boxes completely overlap, it cannot reflect the most realistic situation, which reduces the generalization performance of the model. In addition, CIoU calculates the minimum bounding box for each predicted box and the true box, which limits the calculation and convergence speed of the loss function. Therefore, in the improved network, we choose to use the Wise IOU loss function[17].

Wise IoU, as a bounding box regression loss, includes a dynamic non monotonic mechanism and designs a reasonable gradient gain allocation strategy that reduces the occurrence of large or harmful gradients in extreme samples. This loss method focuses more on ordinary quality samples, thereby improving the generalization ability and overall performance of the network model.

Based on distance measurement, distance attention is constructed, and Wise IoU_V1 with two-layer attention mechanism is obtained on the basis of Wise IoU. The calculation is shown in formulas (6) and (7):

$$L_{\mathrm{WIoU_{v1}}} = R_{\mathrm{WIoU}} \times L_{IoU} L_{\mathrm{WIoU_{v1}}} = R_{\mathrm{WIoU}} \times L_{IoU} \quad (6)$$

$$\mathrm{R_{WIoU}} = \exp\left(\frac{(\mathrm{x} - \mathrm{x_{gt}})^2 + (\mathrm{y} - \mathrm{y_{gt}})^2}{(\mathrm{W_g^2} + \mathrm{H_g^2})^*}\right) \quad (7)$$

In order to effectively reduce the contribution of simple examples to the loss value, Wise IoU_V2 constructed a monotonic focusing mechanism for cross entropy similar to Focal Loss, which enables the model to focus on complex targets and achieve improved classification performance. The formula is as follows (8):

$$L_{\mathrm{WIoU_{v2}}} = \left(L_{IoU}^* / \overline{L_{IoU}}\right)^{\gamma} \times L_{\mathrm{WIoU_{v1}}} \quad (8)$$

$\overline{L_{IoU}}$is the mean of L_{IoU}, which serves as the normalization factor; $(L_{IoU}^* / \overline{L_{IoU}})^{\gamma}$ is the gradient gain, and the normalization factor is dynamically updated to keep the gradient gain at a high level, thereby solving the problem of slow training speed in the later stage.

Wise IoU_V3 redefines the outlier degree to describe the quality of anchor boxes, and uses this outlier degree to construct non monotonic focusing coefficients, allowing the gradient gain to exhibit non monotonic changes with increasing loss values. Wise IoU_V3 formulas such as (9) and (10):

$$\beta = L_{IoU}^* / \overline{L_{IoU}} \in [0, +\infty) \quad (9)$$

$$L_{\mathrm{WIoU_V}} = \mathrm{r} L_{\mathrm{WIoU_{v1}}}, \mathrm{r} = \frac{\beta}{\delta\alpha^{\beta-\delta}} \quad (10)$$

β is the outlier degree, and a small outlier degree means high anchor box quality. Assigning smaller gradient gains to anchor boxes with higher outlier degrees will effectively prevent low-quality examples from generating large harmful gradients. R is the non monotonic focusing coefficient. Due to the dynamic nature of $\overline{(L\_IoU)}$, the quality division criteria for anchor boxes are also dynamic, which enables Wise IoU_V3 to make the most suitable gradient gain allocation strategy for the current situation at every moment.

For object detection tasks in drone aerial scenes, the high proportion of small objects increases the difficulty of detection. In this paper, Wise IoU_V3 is used to dynamically optimize the weighting of lost small objects to improve the detection performance of the model.

4. experimental design

4.1 Experimental Environment Construction

The experimental operating platform for this article is Windows 11 system, using Pytorch 1.10.1 deep learning framework, CUDA 11.3 and CUDNN 8.2. The CPU model is Inter 13th generation 13400f, the graphics card (GPU) model is NVIDIA GeForce RTX 3070, the graphics card memory is 12GB, and the programming language is Python. The uniform size of the training image is 640 * 640 px; The number of samples selected for each training round, batch_2, is set to 16, and the training rounds are 200. Use SGD optimizer during the training phase to minimize the loss function while accelerating training speed and improving accuracy.

4.2 Dataset

At present, there are a large number of samples in China, and the most widely used publicly available aerial insulator dataset

is CPLID, which consists of normal insulator images captured by drones and synthesized defective insulator images. The statistical dataset has 848 samples, including normal insulators and defective insulators. Among them, the number of positive insulator samples (normal insulator images) is 600, and the number of negative insulator samples (defective insulator images) is 248. Although this dataset only contains two types of targets, the samples have the characteristics of rich scenes, diverse backgrounds, and variable lighting conditions, which are suitable for training and performance testing of simple insulator target detection models to help us choose suitable basic target detection algorithms for transmission line components.

Meanwhile, in order to enhance the robustness of the network and ensure considerable accuracy in complex backgrounds, a method for automatically expanding the dataset is proposed in this paper. This method automatically enhances the dataset by flipping each image in the x-direction, y-direction, adding salt and pepper noise, adding Gaussian noise, adjusting image contrast, and automatically expanding the corresponding data labels without the need for manual calibration of the expanded images. After data augmentation, the dataset was reduced from 848 images to 3392 images. In this paper, 80% of them were selected for model training, and 20% were used as test samples for algorithm accuracy and real-time performance.

5. Experimental Results and Analysis

5.1 Evaluation indicators

The experimental evaluation indicators include Pr (accuracy), AP (average accuracy of each classification), mAP (mean accuracy), and Re (recall rate). The solution for accuracy and recall is shown in equations (11) and (12).

$$R_c = \frac{T_p}{T_p + F_N} \tag{11}$$

$$P_r = \frac{T_p}{T_p + F_p} \tag{12}$$

In equation (11), Tp is the number of positive samples correctly identified as positive samples, Fp is the number of negative samples incorrectly identified as positive samples, and Fn is the number of positive samples incorrectly identified as negative samples. The recall rate represents the ability of the algorithm to retrieve positive samples from the dataset, while the accuracy is used to measure the accuracy of the algorithm in finding positive samples from the dataset.

When using different preset IOU values, the recall rate Re and accuracy Pr of the detection target are also different. The P-R curve represents the curve formed by the precision Pr and recall Re with different IOU values. AP represents the area enclosed by the P-R curve and the coordinate axis. The larger the AP value for a category, the better the detection performance of the model for that category. MAP represents the average AP of all category targets, as shown in equation (13).

$$mAP = \frac{\sum_{i=1}^{n} AP_i}{n} \tag{13}$$

In equation (13), APi represents the AP value of the i-th type target, and n represents the number of targets that need to be detected. The larger the mAP, the better the overall performance of the algorithm.

5.2 Experimental Comparison Results

This study selected Fast R-CNN, SSD, YOLOv3, YOLOv4, YOLOv5, and YOLOv7 algorithms as comparative algorithms. The experimental comparison results between the improved YOLOv7 and non YOLO series models FastR-CNN and SSD in this article are shown in Table 1:

Table1. Comparison of Accuracy of Various Algorithm Models

Algorithms	AP(%)		mAP(%)
	Insulator	Defect	
SSD	93.81	75.03	84.42
YOLOv3	89.00	87.33	88.17
YOLOv4	96.14	96.43	96.29
YOLOv5s	98.54	96.07	97.31
YOLOv7	98.58	96.50	97.35
Ours	98.70	98.00	98.30

From the table, it can be seen that SSD performs well in detecting normal insulators with larger targets, but its accuracy for detecting defective insulators with smaller targets is only 75.03%. In contrast, the YOLOv7 algorithm performs better in algorithm detection accuracy, and improving the detection performance of the YOLOv7 algorithm greatly outperforms the detection performance of YOLOv3 and SSD. The convergence detection results of its model are shown in Figure 7. It can be seen that YOLOv7 can stably detect targets of different sizes.

Figure.7 Improved Yolov7 Model Detection Results

5.3 Comparison results of ablation experiments

This article mainly improves the YOLOv7 algorithm and obtains the YOLOv7 improved algorithm. In order to compare the YOLOv7 algorithm with the improved algorithm proposed in this paper and ensure the reliability of the experiment, all training and testing data were trained in the same training environment, and the epochs of each training were guaranteed to be 200. After training, use the optimal model weights for testing.

To verify the effectiveness of various improvement strategies, ablation experiments were conducted on the CPLID dataset, and mAP and Recall were selected as performance evaluation metrics. Based on the original YOLOv7 algorithm, attention module, CNeB module, and Wise IoU loss function are introduced in sequence. On the test set, the evaluation metrics of the improved YOLOv7 algorithm are shown in Table 2.

Table 2. Comparison of Precision, Recall, and Accuracy of Ablation Experiments

Algorithms	Improve module			P	R	mAP@0.5	mAP@0.5：0.95
	GAM	CNeB	Wise IoU				
A	×	×	×	0.949	0.965	0.973	0.785

B	√	×	×	0.975	0.972	0.975	0.785
C	√	√	×	0.983	0.976	0.980	0.788
D	√	√	√	**0.987**	**0.98**	**0.983**	**0.791**

Based on the above experimental results, compared with the original YOLOv7 model, Yolov7-GAM, Yolov7-GAM_CNeB, and our model have higher detection accuracy. By comparing the precision, recall, and mAP@0.5 The introduction of CNeB module and GAM attention mechanism effectively enhances mAP@0.5 Improved by 0.2% and 0.7% respectively on the original YOLOv7 model, and further improved the loss function, mAP@0.5 Reached 98.3%, an increase of nearly 1%. The experimental results show that introducing the CNeB module and GAM attention mechanism can enhance the accuracy of the model, improve the accuracy of the transmission line defect detection model, and enhance the reliability of target localization.

6. Conclusion

Insulators are important components in transmission systems. This paper proposes an insulator missing detection method based on improved YOLOv7. The ELAN module in the backbone network is replaced in the original YOLOv7 model, and the GAM (Global Attention Mechanism) attention mechanism module is added and the original loss function is improved. The powerful feature extraction ability of the improved algorithm is utilized to detect insulators from samples, determine whether the insulators are normal, and mark their positions. The experimental results show that the model of this method performs well on the test set mAP@0.5 (also known as Map50) has reached 98.30%, with an accuracy improvement of 1% compared to the original YOLOv7 algorithm, and can effectively distinguish whether insulators are missing. In addition, this method still has some shortcomings, such as the small proportion of insulators in the image, which will affect recognition. When collecting images, factors such as external weather conditions, floating dust in the environment, and shooting angles can all cause noise in the obtained images. This noise can have a negative impact on image quality, such as reducing contrast, blurring image features, and thus affecting recognition accuracy.

7. REFERENCES

[1] AHMED MD F, MOHANTA J C, SANYAL A. Inspection and identification of transmission line insulator breakdown based on deep learning using aerial images[J/OL]. Electric Power Systems Research, 2022: 108199.

[2] DIAN S, ZHONG X, ZHONG Y. Faster R-Transformer: An efficient method for insulator detection in complex aerial environments[J/OL]. Measurement, 2022, 199: 111238.

[3] Xiao Y , Li Z , Zhang D ,et al.Detection of Pin Defects in Aerial Images Based on Cascaded Convolutional Neural Network[J].IEEE Access, 2021, PP(99):1-1.DOI:10.1109/ACCESS.2021.3079172.

[4] Wen Q, Luo Z, Chen R, et al. Deep learning approaches on defect detection in high resolution aerial images of insulators[J]. Sensors, 2021, 21(4): 1033.

[5] Haiyan C , Yongjie Z , Rui C .Faster R-CNN based recognition of insulators in aerial images[J].Modern Electronics Technique, 2019.

[6] Antwi-Bekoe E, Zhan Q, Xie X, et al. Insulator recognition and fault detection using deep learning approach[C]//Journal of Physics: Conference Series. IOP Publishing, 2020, 1454(1): 012011.

[7] Li X, Su H, Liu G. Insulator defect recognition based on global detection and local segmentation[J]. IEEE Access, 2020, 8: 59934-59946.

[8] Yuan J, Zheng X, Peng L, et al. Identification method of typical defects in transmission lines based on YOLOv5 object detection algorithm[J]. Energy Reports, 2023, 9: 323-332.

[9] Han G, Li T, Li Q, et al. Improved algorithm for insulator and its defect detection based on YOLOX[J]. Sensors, 2022, 22(16): 6186.

[10] Huang H, Lan G, Wei J, et al. TLI-YOLOv5: A Lightweight Object Detection Framework for Transmission Line Inspection by Unmanned Aerial Vehicle[J]. Electronics, 2023, 12(15): 3340.

[11] Yan G, Guo J, Zhu D, et al. A Flame Detection Algorithm Based on Improved YOLOv7[J]. Applied Sciences, 2023, 13(16): 9236.

[12] Liu Y, Shao Z, Hoffmann N. Global attention mechanism: Retain information to enhance channel-spatial interactions[J]. arXiv preprint arXiv:2112.05561, 2021.

[13] Liu Z, Mao H, Wu C Y, et al. A convnet for the 2020s[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 11976-11986.

[14] 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.

[15] Tong Z, Chen Y, Xu Z, et al. Wise-IoU: Bounding Box Regression Loss with Dynamic Focusing Mechanism[J]. arXiv preprint arXiv:2301.10051, 2023.

[16] S.-J. J , Q.-H. L , Han F .An improved algorithm for small object detection based on YOLO v4 and multi-scale contextual information[J].Computers and Electrical Engineering, 2023.

[17] Tong Z, Chen Y, Xu Z, et al. Wise-IoU: Bounding Box Regression Loss with Dynamic Focusing Mechanism[J]. arXiv preprint arXiv:2301.10051, 2023.