IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 18 August 2023, accepted 24 August 2023, date of publication 29 August 2023, date of current version 5 September 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3309693

RESEARCH ARTICLE

Insulator and Defect Detection Model Based on Improved YOLO-S

WEIGUO YI[1], SIWEI MA[2], AND RONGHUA LI[2]
[1]School of Computer and Communication Engineering, Dalian Jiaotong University, Dalian 116028, China
[2]School of Mechanical Engineering, Dalian Jiaotong University, Dalian 116028, China

Corresponding author: Ronghua Li (lironghua705@163.com)

This work was supported in part by the Foundation of the Key Laboratory of Defense Science and Technology under Grant 2022-JCJQ-L8-015-020, in part by the Key Project of Scientific Research Project of Liaoning Provincial Education Department under Grant LJKZ0475, and in part by the Dalian High-Level Talent Innovation Support Plan under Grant 2022RJ03.

ABSTRACT A large number of insulators play an important role in insulating and supporting complex power grids, and they are constantly exposed to challenges such as lightning strikes and contamination from the external environment. Only by accurately detecting insulator damage can potential safety hazards and equipment damage due to damaged insulators be avoided in time. Aiming at the problems of low accuracy and large model computation of existing insulator and defect detection algorithms, this paper proposes the insulator and defect detection model YOLO-S (YOLO-Small). First, the PAN structure in Neck uses lightweight convolutional GSConv instead of the standard convolutional Conv, introduces GSbottleneck on the basis of GSConv, and uses the method of OSA to design the cross-level partial network GSCSP module (VoV-GSCSP). The use of VoV-GSCSP instead of the C3 module in Neck reduces the computational effort and maintains the accuracy. Secondly, a new attention module, MaECA (MainECA), is designed based on the ECA attention mechanism to enhance target perception. After that, the SILU function in the SPPF in YOLOv5s is replaced with the Mish function, HardSwish function, and ReLU function, respectively, and the results show that the replacement of the Mish function (MishSPPF) is more effective in preventing the distortion of the image caused by image cropping and scaling, and thus improving the accuracy. Finally, the SIoU loss function is used to replace the original loss function CIoU, which improves the number of transmitted frames per second and the detection accuracy of the pictures. The mAP of YOLO-S is 4.2% higher than that of YOLOv5s, the detection accuracy is improved by 2.1%, and the model parameter computation is reduced by 6.0%, which still gives YOLO-S a higher recognition accuracy when compared with existing detection algorithms under the same conditions.

INDEX TERMS YOLOv5s, insulator defect detection, GSConv, attentional mechanisms, activation function, SIoU.

I. INTRODUCTION

Insulators are indispensable components in power equipment, and their existence directly affects the stable operation of power equipment and the safety of the power grid. If the insulator is broken or aging, it will cause the insulator to be pierced, and even further lead to equipment failure and accidents. Therefore, the detection of insulators is of great significance.

Many researchers and scholars have used image processing and machine learning techniques to detect insulator defects. Li et al. [1] in 2012 proposed contour projection algorithm for locating insulators in aerial images. Wu and An [2] proposed a new active contour model for effective segmentation of insulator images with uneven overlapping textures. In 2016 Zao et al. [3] is the one that utilizes azimuthal detection, combined with the a priori knowledge of binary shapes, to realize the localization detection of heterogeneous insulators. Liao and An [4] on the other hand, proposed an algorithm with multi-scale representation and multi-featured descriptors, which determines that the matching strategy of

The associate editor coordinating the review of this manuscript and approving it for publication was Senthil Kumar.

the insulator region is gradual from coarse to fine, and it can eliminate the background noise. The design of these algorithms can provide references and lessons for insulator fault identification in practical engineering. Gao et al. [5] investigated how to combine deep learning and traditional image processing algorithms to identify insulator faults. Zuo et al. [6] used the theory of segmentation on images to identify insulator faults. On the contrary, Zhao et al. [7] proposed a method to classify insulator states based on mutant patch features in convolutional neural networks. All these algorithms have some application value in identifying insulator faults, which can provide references and lessons for engineering practice. Liu et al. [8] used the Faster R-CNN [9] algorithm to detect and localize insulators. This method automatically identifies and localizes insulators and provides a basis for further defect detection. Cheng et al. [10] use an edge detection operator to extract insulators, and then use the spatial properties of insulators to detect and localize spontaneous insulator defects, and then use the full convolutional neural network FCN algorithm to perform a segmentation operation of the insulator caps on individually extracted insulators, and finally by calculating the distance between neighboring insulator caps to determine their defect locations.

The above methods show that the application of machine vision in the field of defect detection is very feasible and solves the problems of manual inspection, such as difficult work, high danger, lack of flexibility and high labor cost, but their shortcomings are the low detection accuracy and slow processing speed of the model. In this paper, we make some improvements based on YOLOv5s model through deep learning theory, and propose a YOLO-S insulator and defect detection model. This new method has high flexibility and processing speed, which not only effectively solves the problems faced by traditional manual inspection, but also provides an in-depth study of insulator localization and defect diagnosis, improves the defect detection accuracy and reduces the calculation of model parameters, and also has a positive guiding and reference effect on the detection of other components such as anti-vibration hammers on transmission lines.

The direction of YOLO-S algorithm structure is basically the same as YOLOv5s, compared with YOLOv5s:

YOLO-S adopts MaECA module instead of C3 module in Backbone, which can make the model more focused on the target area or important features, and improve the target detection accuracy.

YOLO-S adopts the MishSPPF module instead of the SPPF module, which solves the problem of repeated feature extraction by convolutional neural networks for graph correlation and greatly improves the speed of generating candidate frames.

YOLO-S uses GSConv to replace the first two Convs in Neck's PAN structure, and VoV-GSCSP replaces the first two C3 modules, reducing the amount of unnecessary parameter calculations while improving the accuracy of detection.

YOLO-S replaces the original loss function CIoU of YOLOv5s with the SIoU loss function, which optimizes the regression of the target frame and improves the accuracy of the regression, thus enabling the model to locate the target more accurately.

The structure of YOLO-S incorporating the four improvement points is shown in Fig. 1.

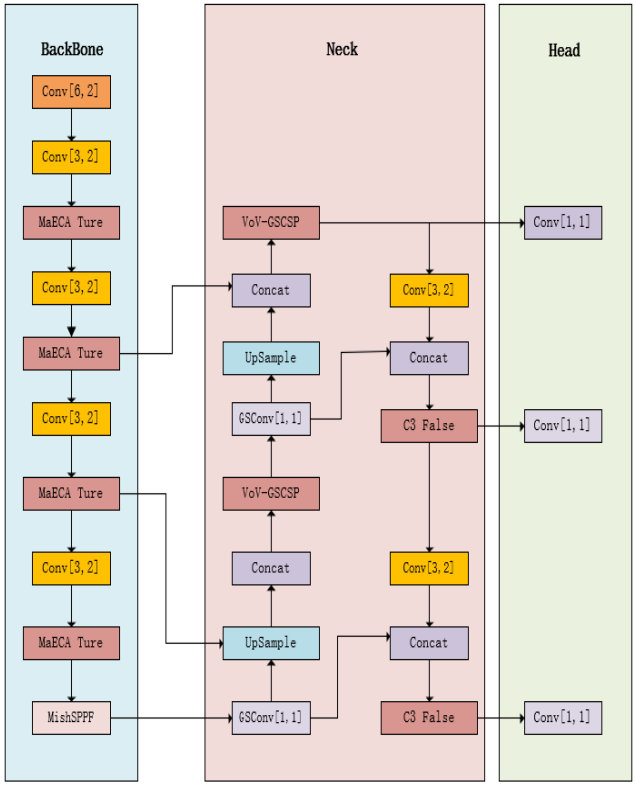


FIGURE 1. YOLO-S model diagram.

The specific innovations are as follows:

a) The PAN structure in Neck uses a lighter weight convolutional GSconv instead of the standard convolutional Conv, and the VoV-GSCSP module replaces the C3 module, which reduces the amount of model parameter computation and improves accuracy.

b) The MaECA attention module was designed. MaECA is embedded in the global maximum pooling in the attention module of ECA and fused with the channel attention feature map obtained from the global average pooling, and then multiplied with the original input feature map on a channel-by-channel basis. MaECA is embedded in the Bottleneck of the C3 module in Backbone to improve the perception of the target and achieve a good performance improvement.

c) MishSPPF module was designed. That is, the SILU [11] function in SPPF in YOLOv5s is replaced withMish function, HardSwish function, and ReLU function, respectively, and the experimental data comparison shows that replacing Mish function is better to solve the problem of low detection accuracy.

d) In order to improve the convergence speed and detection accuracy of the model, the SIoU loss function is used to replace the original loss function CioU [12].

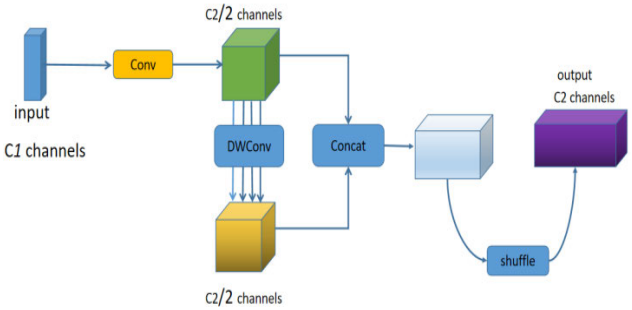


FIGURE 2. GSConv module.

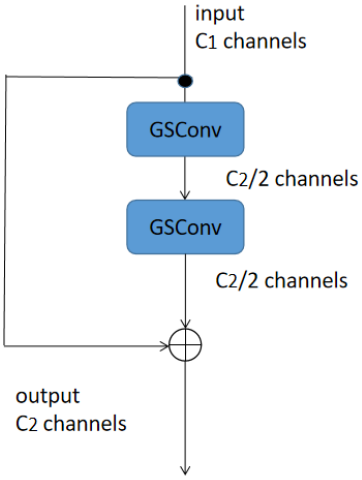


FIGURE 3. GSbottleneck module.

II. RELEVANT THEORETICAL KNOWLEDGE

A. GSCONV

Lightweight convolution GSConv [13] uses grouped convolution and depth-separable convolution to reduce the computation and number of parameters. The main advantage is that it can provide relatively high accuracy while having faster speed and smaller model size. Using GSConv in all phases of the model would result in a deeper network, increasing the resistance of the data flow and significantly increasing the inference time. Therefore, a better choice is to use GSConv only in the Neck, since at this stage the feature map has become slender and no longer needs to be transformed. The stitched feature maps are better processed using GSConv and are computationally less expensive, about 60% to 70% of the standard convolution. This avoids redundant information and complex computations, and improves the effectiveness of the attention module. the GSConv module diagram is shown in Figure 2.

The GSbottleneck module is further introduced on the basis of GSConv. The GSbottleneck module is shown in Figure 3.

An efficient architecture approach, called VoVNet, was designed by using the OSA (one-shot-aggregation) method for building cross-level partial network (GSCSP) modules (VoV-GSCSP). VoV-GSCSP is shown in Figure 4.

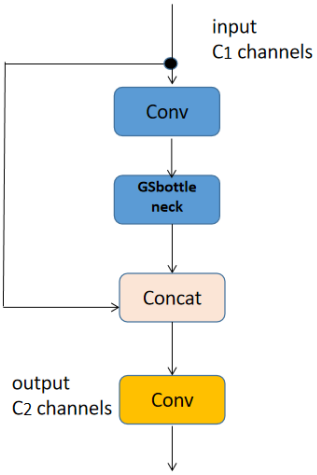


FIGURE 4. VoV-GSCSP module.

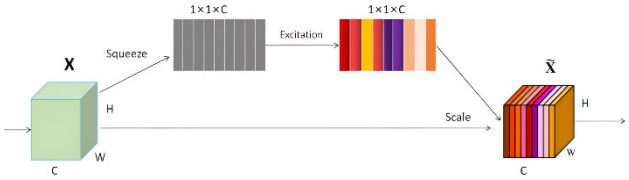


FIGURE 5. SE structure.

B. ATTENTIONAL MECHANISMS

1) SE

An implementation of the SE attention mechanism [14]:

Squeeze:Compress the 2D features (H*W) of each channel into 1 real number by global average pooling, and get the global features at channel level by transforming the feature map from [h, w, c] ==> [1,1,c].

Excitation: generates a weight value for each feature channel, constructs the correlation between channels through two fully connected layers, and outputs the same number of weight values as the number of channels in the input feature map. That is, [1,1,c] ==> [1,1,c].

Scale: the normalized weights obtained earlier are weighted to the features of each channel, multiplying the weight coefficients channel by channel. [h,w,c]*[1,1,c]==> [h,w,c].

The structure of the SE attention mechanism is shown in Figure 5.

2) CBAM

CBAM [15] consists of a traditional channel attention module (CAM) and a spatial attention module (SAM). Among them, CAM allows the model to selectively learn the importance of each channel, thus improving the robustness and differentiation of the features, while SAM allows the model to focus on the important parts of the object or the background region, thus improving the model's understanding of the image. The structure of CBAM is shown in Fig. 6.

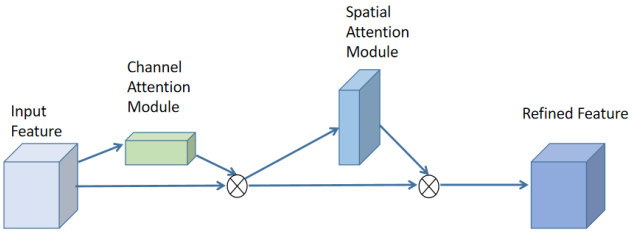


FIGURE 6. CBAM structure.

3) CA

Implementation of CA Attention Mechanism [16].

The input feature maps of C × H × W shape are average pooled channel by channel, and each channel is encoded using the pooling kernels of (H,1) and (1,W) in the X and Y axis directions, respectively, to produce feature maps of C × H × 1 and C × 1 × W shapes.

The generated feature maps are transformed and then concat operation is performed. The formula is as follows:

$$f = \sigma\left(F_1\left(\left[z^h, z^w\right]\right)\right) \tag{1}$$

The z^h, z^w is subjected to the concat operation with the F1 operation (dimensionality reduction using a 1×1 convolution kernel) and the activation operation to generate the feature map $f \in R^{C/r\times(H+W)\times 1}$.

Along the spatial dimension, f is then split into $f^h \in R^{C/r\times H\times 1}$ and $f^h \in R^{C/r\times 1\times W}$ by performing the split operation, and then the ascending dimension operation is performed using 1×1 convolution, respectively, and combined with the sigmoid activation function to obtain the final attention vectors $g^h \in R^{C\times H\times 1}$ and $g^h \in R^{C\times 1\times W}$.

Finally, the output equation of CA is shown in (2):

$$y_c(i,j) = x_c(i,j) \times g_c^h(i) \times g_c^w(j) \tag{2}$$

The CA structure is shown in 7.

4) ECA

The computational procedure of the ECA attention mechanism [17] is as follows:

Input the feature map whose dimension is C × H × W, where C is the number of channels, H is the height, and W is the width.

First, for each channel, compute the mean value of, The formula is shown in (3):

$$Avg(X_c) = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W} X_{cij} \tag{3}$$

The mean values are then mapped to a new representation space by a linear transformation, which helps the model to learn the relationship between channels,

The formula is shown in (4):

$$Avg_{proj}(X_c) = W_{proj} \cdot Avg(X_c) \tag{4}$$

W_{proj} is the weight matrix used for mapping.

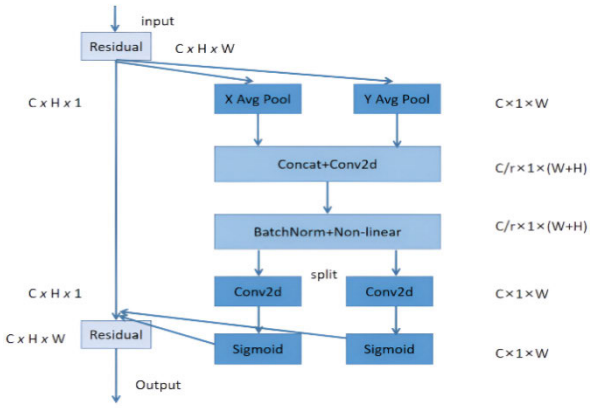


FIGURE 7. CA structure.

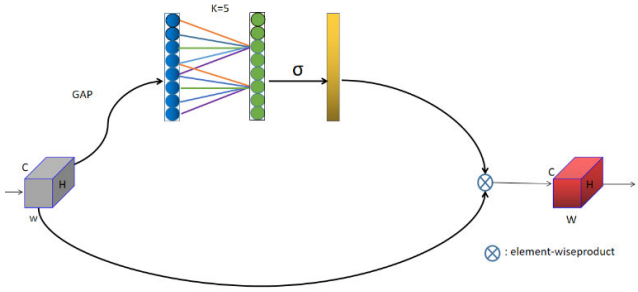


FIGURE 8. ECA structure.

A Sigmoid function is applied to the mapped values to obtain a channel weight factor indicating the importance of each channel, The formula is shown in (5):

$$ChannelWeights(X_c) = \sigma\left(Avg_{proj}(X_c)\right) \tag{5}$$

σ denotes the Sigmoid function.

The channel weight factors are multiplied element-by-element with the channel dimensions of the original input feature map to obtain a weighted feature map, The formula is shown in (6):

$$ECA(X_c) = ChannelWeight(X_c) \odot X_c \tag{6}$$

$\odot$ denotes element-by-element multiplication.

The ECA structure diagram is shown in Figure 8.

5) MaECA

The basic idea is to embed the global maximum pooling into the ECA module and fuse it with the channel attention feature maps obtained from the global average pooling, and then multiply it with the original input feature maps channel by channel.

The computational procedure of MaECA attention mechanism is as follows:

The input feature map dimension is C × H × W, where C is the number of channels, H is the height, and W is the width.

First, the feature maps for channel attention are computed using global average pooling, The formula is shown

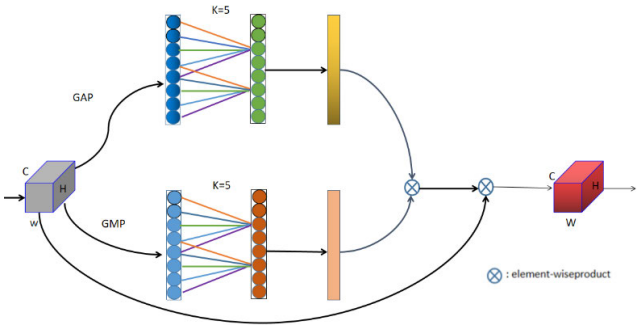


FIGURE 9. MaECA structure.

in (7), (8):

$$Avg\ (X_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} X_{cij} \quad (7)$$

$$ChannelWeightsAvg(X_c) = \sigma(W_{proj} \cdot Avg(X_c)) \quad (8)$$

σ denotes the Sigmoid function.

The feature map of channel attention is computed using global maximum pooling with the formula shown in (9), (10).

$$Max\ (X_c) = max_{i,j} X_{cij} \quad (9)$$

$$ChannelWeightsMax\ (X_c) = \sigma\left(W_{proj} \cdot Max\ (X_c)\right) \quad (10)$$

The channel attention feature map incorporating global average pooling and global maximum pooling is shown in (11) by the formula.

$$MergedChannelWeights\ (X_c) = \alpha \cdot A + (1 - \alpha)\, B \quad (11)$$

A for ChannelWeightsAvg(X_c), B for ChannelWeightsMax (X_c), α is the fusion weight, usually 0.5.

The fused channel attention feature maps are multiplied channel-by-channel with the original input feature maps to obtain the weighted feature maps as shown in (12) by the formula.

$$MaECA\ (X_c) = MergedChannelWeight\ (X_c) \odot X_c \quad (12)$$

The MaECA structure is shown in Figure.

C. ACTIVATION FUNCTION

1) ReLU

ReLU [18] takes all negative values as 0 and keeps the positive values unchanged, the formula is shown in (13).

$$F\ (x) = \max\ (0, x) = \begin{cases} x, & if\ x > 0 \\ 0, & if\ x \leq 0 \end{cases} \quad (13)$$

The ReLU activation function and its derivative curves are shown in Figure 10.

2) HARDSWISH

Hardswish activation function [19] has the advantages of good numerical stability and fast calculation. The formula is shown in (14).

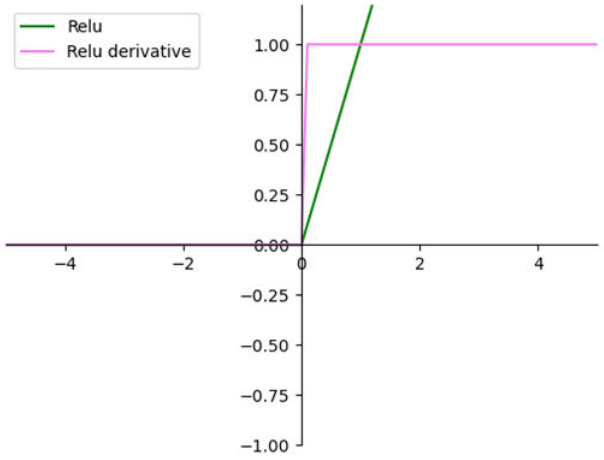


FIGURE 10. ReLU activation function.

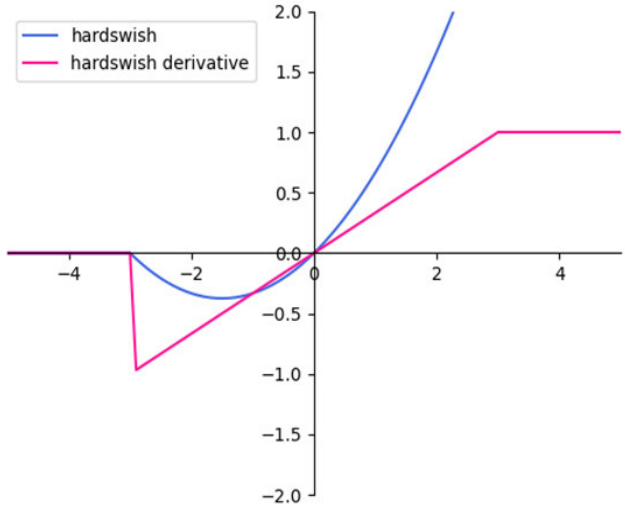


FIGURE 11. Hardswish activation function.

$$F\ (X) = \begin{cases} 0, & x \leq -3 \\ x, & x \geq 3 \\ x(x+3)/6, & otherwise \end{cases} \quad (14)$$

Figure 11 shows the hardswish activation function and its derivative curve.

3) MISH

The Mish activation function [20] is a smooth nonmonotonic activation function., The formula is shown in (15).

$$F\ (x) = xTanh(\ln(1 + e^x)) \quad (15)$$

Figure 12 shows the Mish activation function.

D. SPPF

Based on SPP [21] pooling technique, the authors of YOLOv5 propose a novel pooling method: spatial pyramid pooling SPPF (SPP with Fast pooling). This method, based on SPP pooling, can significantly reduce the computation time

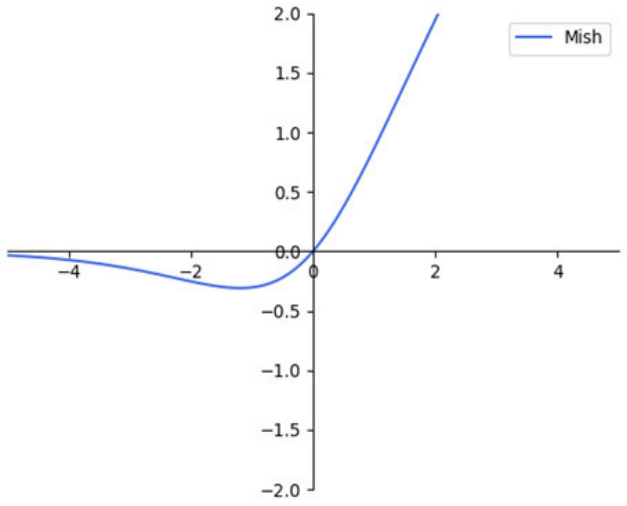


FIGURE 12. Mish activation function.

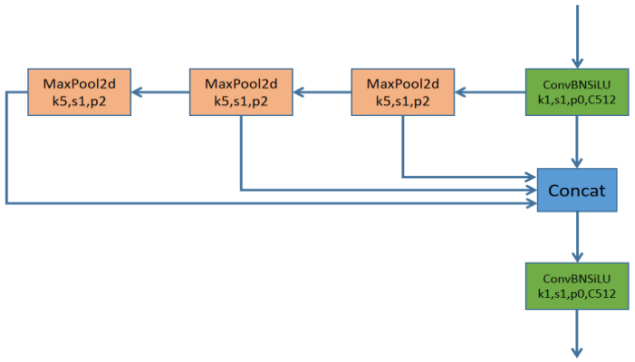


FIGURE 13. SPPF structure.

and memory consumption, and further improve the accuracy and efficiency of image recognition and target detection. The SPPF structure diagram is shown in Figure 13.

E. SIoU

Compared with the original CIoU [22]loss function, the main highlight of the SIoU loss function [23] is that the SIoU Loss integrally considers four factors: the angle loss, distance loss, the shape of the predicted frame and the real frame, and the interaction ratio, which can largely accelerate the convergence speed of the model and improve the accuracy of the model. It can well solve the problem that CIOU does not take into account the mismatch angle between the real frame and the predicted frame. The loss function is mainly composed of Angle cost, Distance cost, Shape cost and IoU cost, and the specific formula is as follows:

The Angle cost section is defined in the following form:

$$A = 1 - 2sin^2(arcsin(\frac{c_h}{\sigma}) - \frac{\pi}{4}) \tag{16}$$

where is the height difference between the center point of the real frame and the predicted frame, and is the distance between the center point of the real frame and the predicted frame.

The Distance cost component is defined in the following form:

$$\Delta = \sum_{t=x,y}(1 - e^{-\gamma\rho_t}) = 2 - e^{-\gamma\rho_x} - e^{-\gamma\rho_y} \tag{17}$$

$$\rho_x = (\frac{b^{gt}_{c_x} - b_{c_x}}{c_w})^2, \quad \rho_y = (\frac{b^{gt}_{c_y} - b_{c_y}}{c_h})^2 \tag{18}$$

$$\gamma = 2 - A \tag{19}$$

where, is the center coordinate of the real frame, is the center coordinate of the prediction frame,, is the width and height of the smallest outer rectangle of the real frame and the prediction frame.

The Shape cost section is defined in the following form:

$$\Omega = \sum_{t=w,h}\left(1 - e^{-w_t}\right)^\theta = (1 - e^{-w_w})^\theta + (1 - e^{-w_h})^\theta \tag{20}$$

$$w_w = \frac{|w - w^{gt}|}{max(w, w^{gt})}, \quad w_h = \frac{|h - h^{gt}|}{max(h, h^{gt})} \tag{21}$$

where w represents the width of the prediction box, h represents the height of the prediction box, represents the width of the real box, and represents the height of the real box for controlling the attention to Ldis.

The intersection and merger ratio loss (IoU cost) component is defined in the following form:

$$L_{IoU\_Cost} = 1 - IoU \tag{22}$$

In summary, the entire SIoU loss function is composed of the following formulas:

$$SIoU = 1 - IoU + \frac{\Delta + \Omega}{2} \tag{23}$$

F. YOLOV5

Yolov5 [24] is an efficient target detection algorithm, Backbone structure is the first important component in Yolov5, which is responsible for extracting featureinformation from the original image and passing it to the subsequent processing layers. Neck structure is the second important component in Yolov5. YOLOv5-6.1 version uses a network structure called PAN plus FPN as an intermediate part for fusing feature maps of different resolutions to further improve the accuracy of detection. Head is the third important component in Yolov5. The Head layer is the last layer in the whole Yolov5 model, which is used to transform the feature map after Backbone and Neck processing into the output of prediction frames and category probabilities. The structure diagram of YOLOv5-6.1 is shown in Figure 14.

III. DATASET

The insulator dataset in this paper was collected in many different ways with 1700 original images because the sample size of the dataset was too small to avoid overfitting due to insufficient number of samples, which in turn had an impact on the detection of insulators and defective parts. Therefore, the initial dataset was rotated, panned, scaled, mirror-flipped,

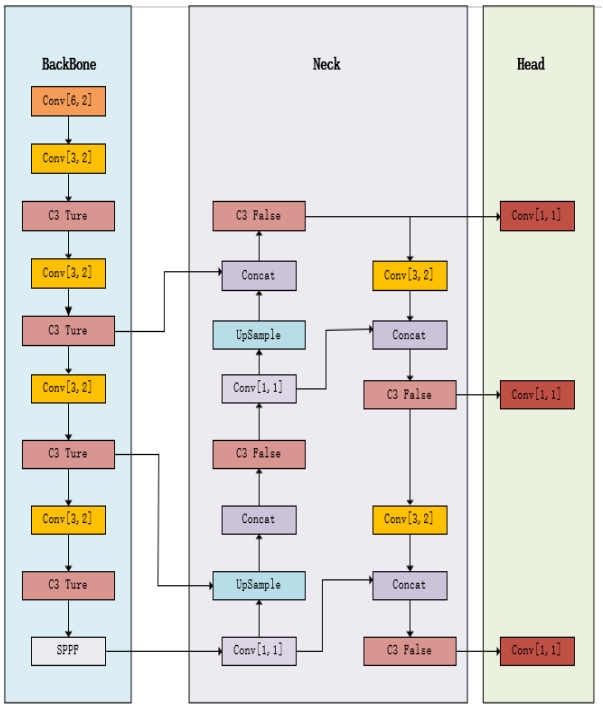


FIGURE 14. YOLOv5 model diagram.

FIGURE 15. Dataset images.

cropped, noise added, and Cutout [25] transformed using the data enhancement tool to effectively increase the size of the training set and thus improve the generalizability of the model. After image enhancement, a total of 5180 images were added to the dataset, and the ratio of training set, validation set, and test set was randomly distributed as 8:1:1. The dataset was annotated by MakeSense software to obtain txt tag files that could be used for YOLOv5s training. A partial image of the dataset is shown in Figure 15.

IV. EXPERIMENT AND ANALYSIS

A. EVALUATION METRICS

The experiments in this paper use evaluation metrics commonly used in deep learning, including Precision (Precision), mean average precision (mAP), GFLOPs and FPS.

TABLE 1. Performance comparison after incorporating lightweight modules.

Module	Precision/%	Reacall/%	mAP@0.5/%	mAP@0.5:0.95/%	GFLOPs	FPS
YOLOv5s	86.4	83.8	83.9	55.6	15.8	116
+Method 1(PAN+FPN)	86.7	84.8	84.5	55.8	13.7	101
+Method 1(PAN)	86.9	85.3	85.1	56.3	14.9	98

TABLE 2. Performance comparison of four attention mechanisms added in the C3 module.

Module	Precision/%	Recall/%	mAP@0.5/%	mAP@0.5:0.95/%	GFLOPs	FPS
YOLOv5s	86.4	83.8	83.9	55.6	15.8	116
+SE	85.5	85.0	84.5	57.5	15.8	108
+CBAM	85.8	84.1	84.2	55.7	15.8	104
+CA	86.5	83.7	84.9	56.9	15.8	100
+ECA	85.6	84.4	85.1	56.2	15.8	98

TABLE 3. Comparative Evaluation of the Effects of Enhanced ECA.

Module	Precision/%	Recall/%	mAP@0.5/%	mAP@0.5:0.95/%	GFLOPs	FPS
YOLOv5s	86.4	83.8	83.9	55.6	15.8	116
+ECA	85.6	84.4	85.1	56.2	15.8	98
+MaECA	86.8	84.8	85.3	55.8	15.8	109

Among them, TP (true positives) predicts positive samples as positive and correct predictions as the correct number of insulator detections, FN (false negatives) predicts positive samples as negative and missed insulators, and FP (false positives) predicts negative samples as positive, i.e., false detections.

Precision: The ratio of the number of samples determined to be positive by the classifier to the total number of samples determined to be positive. The precision is calculated as an indication of how accurate the classifier is in predicting the positive class, i.e., how many of the samples that the model predicts as positive are actually positive. The higher the accuracy rate, the lower the probability that the classifier will predict negative samples as positive samples, and the better the performance of the classifier. The formula is shown below:

$$P = \frac{TP}{TP + FP} \tag{24}$$

Recall: how many of the samples that were actually positive were predicted to be positive, the higher the recall the better. The formula is shown below:

$$P = \frac{TP}{TP + FN} \tag{25}$$

TABLE 4. Comparison of the results of replacing the SPPF function.

Module	Precision/%	Recall/%	mAP@0.5/%	mAP@0.5:0.95/%	GFLOPs	FPS
YOLOv5s	86.4	83.8	83.9	55.6	15.8	116
+REL U	86.7	83.5	84.5	56.0	15.8	109
+Hard Swish	86.5	84.3	84.4	56.2	15.8	111
+Mish	86.1	85.2	85.5	56.5	15.8	115

TABLE 5. Comparison chart of the results of replacing the SIoU loss function.

Module	Precision/%	Recall/%	mAP@0.5/%	mAP@0.5:0.95/%	GFLOPs	FPS
YOLOv5s	86.4	83.8	83.9	55.6	15.8	116
+SIoU	85.8	84.7	85.2	56.7	15.8	122

Mean Average Precision (mAP), abbreviated as mAP. is derived by calculating the average precision for each category, where the precision for each category is the proportion of positive samples correctly predicted by the algorithm over all positive samples predicted at different confidence thresholds. Mean Average Precision (mAP), or mAP for short, is derived by calculating the average precision of each category, where the precision of each category is the proportion of positive samples correctly predicted by the algorithm to all positive samples predicted at different confidence thresholds. Higher mAP values indicate more accurate predictions.

Among them, mAP@0.5 Represents the average mAP at an IoU threshold of 0.5, mAP@0.5:0.95 represents the average mAP at different IoU thresholds (0.5 to 0.95, step size 0.05). The formula is shown below:

$$mAP = \int_0^1 P(R)\, dR \tag{26}$$

GFLOPs are the number of floating point operations that can be performed per second. In deep learning, GFLOPs are usually used to measure the computational complexity and speed of a neural network. A higher value of GFLOPs usually means that the neural network requires more computational resources and longer training time.

FPS is the definition in the field of graphics and refers to the number of frames per second that a picture is transmitted, or in layman's terms, the number of frames in an animation or video. fps is a measure of the amount of information used to save and display moving video. The more frames per second, the smoother the action displayed will be.

B. EXPERIMENTS AND RESULTS

YOLOv5 has four versions: YOLOv5s, m, l, and x. YOLOv5s has the smallest network structure and computation volume, and adopts a lighter and more efficient network structure, which can achieve faster operation speed without reducing accuracy; YOLOv5m has a larger network structure and computation volume, with higher accuracy than YOLOv5s, and is suitable for use in application scenarios requiring higher accuracy; YOLOv5l has a larger network structure and computational volume with higher accuracy than YOLOv5m; and YOLOv5x has the largest network structure and computational volume with the highest accuracy. Given that the requirements of insulator detection include not only the accuracy but also the calculation amount of model parameters, etc., the YOLOv5s model with the fastest detection speed is used as the basis of this paper.

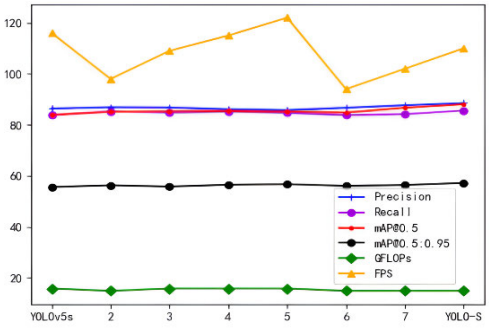


FIGURE 16. Visualization of ablation experiment results.

Experimental parameter configuration, Learning_ Rate is 0.01, Batch_ Size set to 64, Epoch set to 200, Iou_ Threshold is set to 0.5.

1) IMPROVEMENTS TO THE YOLOV5S MODEL BASED ON GSCONV

The GSConv module is used to replace the Conv module and the VoV-GSCSP to replace the C3 module in the PAN structure, PAN+FPN structure in Neck, respectively, referring to this method as Method 1, The results are shown in Table 1. The P, R, mAP@0.5, mAP@0.5: 0.95 values after replacement in PAN structure are the highest, GFLOPs are reduced by 6% compared to YOLOv5s, and FPS is slightly lower; replacement in PAN+FPN structure, only GFLOPs and FPS due to the replacement in PAN structure, and the others are inferior to PAN structure. It shows that the improvement point GSConv+VoV-GSCSP replacement in PAN structure in Neck is effective to improve the accuracy of the model and reduce the computational overhead.

2) IMPROVEMENTS TO THE YOLOV5S MODEL BASED ON ECA

In order to improve the accuracy and sensitivity of target detection, localize objects or regions more precisely, and effectively deal with complex scenes, four different types of attention mechanisms are embedded into the Bottleneck of the C3 module in the Backbone of the YOLOv5s network structure for experiments.

The effects of the four different types of attention mechanisms in YOLOv5s are comprehensively analyzed and

TABLE 6. Comparison of ablation experimental results.

Number	GSConv+VoV-GSCSP(PAN)	MaECA	Mish	SIoU	Precision/%	Recall/%	mAP@0.5/%	mAP@0.5:0.95/%	GFLOPs	FPS
YOLOv5s	×	×	×	×	86.4	83.8	83.9	55.6	15.8	116
2	√	×	×	×	86.9	85.3	85.1	56.3	14.9	98
3	×	√	×	×	86.8	84.8	85.3	55.8	15.8	109
4	×	×	√	×	86.1	85.2	85.5	56.5	15.8	115
5	×	×	×	√	85.8	84.7	85.2	56.7	15.8	122
6	√	√	×	×	86.7	83.8	84.8	56.1	14.9	94
7	√	√	√	×	87.7	84.2	86.7	56.4	14.9	102
YOLO-S	√	√	√	√	88.5	85.6	88.1	57.2	14.9	110

TABLE 7. Comparison results on coco128 dataset.

Module	Precision/%	Recall/%	mAP@0.5/%	mAP@0.5:0.95/%	GFLOPs	FPS
YOLOv5s	66.6.	54.2	62.3	43.2	16.4	72
YOLO-S	88.1	77.2	86.7	63.4	15.7	84

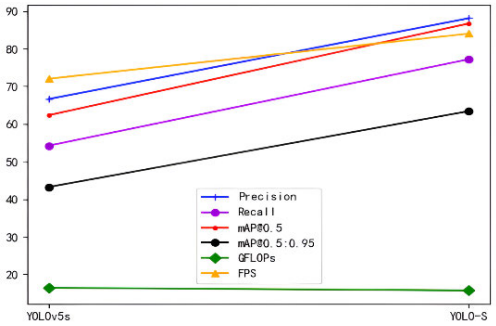


FIGURE 17. Comparative visualization in coco128 dataset.

compared, and the optimal attention mechanism is derived. The results of this analysis are displayed in Table 2 below. Comparing the effects of the four different C3 attention mechanisms in the YOLOv5s network structure, it can be concluded that the addition of the ECA attention mechanism has the highest mAP of 85.1%, indicating a better performance in the target detection task. The differences between the four attention mechanisms in other indicators are not significant, and, mAP as an evaluation indicator more accurately reflects the accuracy and application value of the target detection algorithm, and is therefore more practical in practical applications. In summary, the effect of adding the ECA attention mechanism in the C3 module is more excellent relative to other methods, which can provide a reference for improving the accuracy of the target detection algorithm.

The experiments concluded that the ECA attention mechanism is more effective, so we choose to improve the ECA attention mechanism to further improve the performance of the model The improved ECA is named MaECA (MainECA) and added to the C3 module of Backbone, and the training results are shown in Table 3. Compared with the ECA, adding the improved MaECA improves the accuracy by 1.2 percentage points, and the mAP is 0.2% higher; compared with the original YOLOv5s, adding MaECA improves the accuracy by 0.4 percentage points, and the mAP is 1.4% higher than that of the original YOLOv5s, and the FPS is better than that of the YOLOv5s. This shows that the MaECA attention mechanism is an effective strategy that can be efficiently applied.

3) IMPROVEMENTS TO THE YOLOV5S MODEL BASED ON SPPF

In order to better solve the problem of repeated feature extraction for images as well as to improve the accuracy of model detection, the SILU function in SPPF is chosen to be replaced by RELU function, Mish function, and HardSwish function, respectively. Among them, RELU function can make the neural network converge faster; Mish has better convergence and smaller accuracydegradation; HardSwish function can improve the training effect of deep neural network well. The training results are shown in Table 4. With other values similar, the SPPF of replacing the Mish function has the highest mAP, so this improved method is named MishSPPF.

4) IMPROVING THE YOLOv5s MODEL BASED ON SIoU

In order to improve the training speed and inference accuracy, the original CIoU function was replaced by the SIoU loss function. According to the table below, the mAP of YOLOv5s with the SIoU function replaced was 1.3% higher than that of the original YOLOv5s, and the accuracy was slightly lower. The training results are shown in Table5.

5) ABLATION EXPERIMENTS

Using GSConv instead of standard convolutional Conv in the PAN structure in Neck, replacing C3 with VoV-GSCSP, adding MaECA attention mechanism to C3 in Backbone, replacing SPPF with MishSPPF, and replacing the original CIoU function with SIoU loss function, all of which are aimed at increasing the attention and reducing the computational effort to the features in the important areas and enhancing the feature extraction capability for different scales and sizes of targets. All of them are to increase the attention

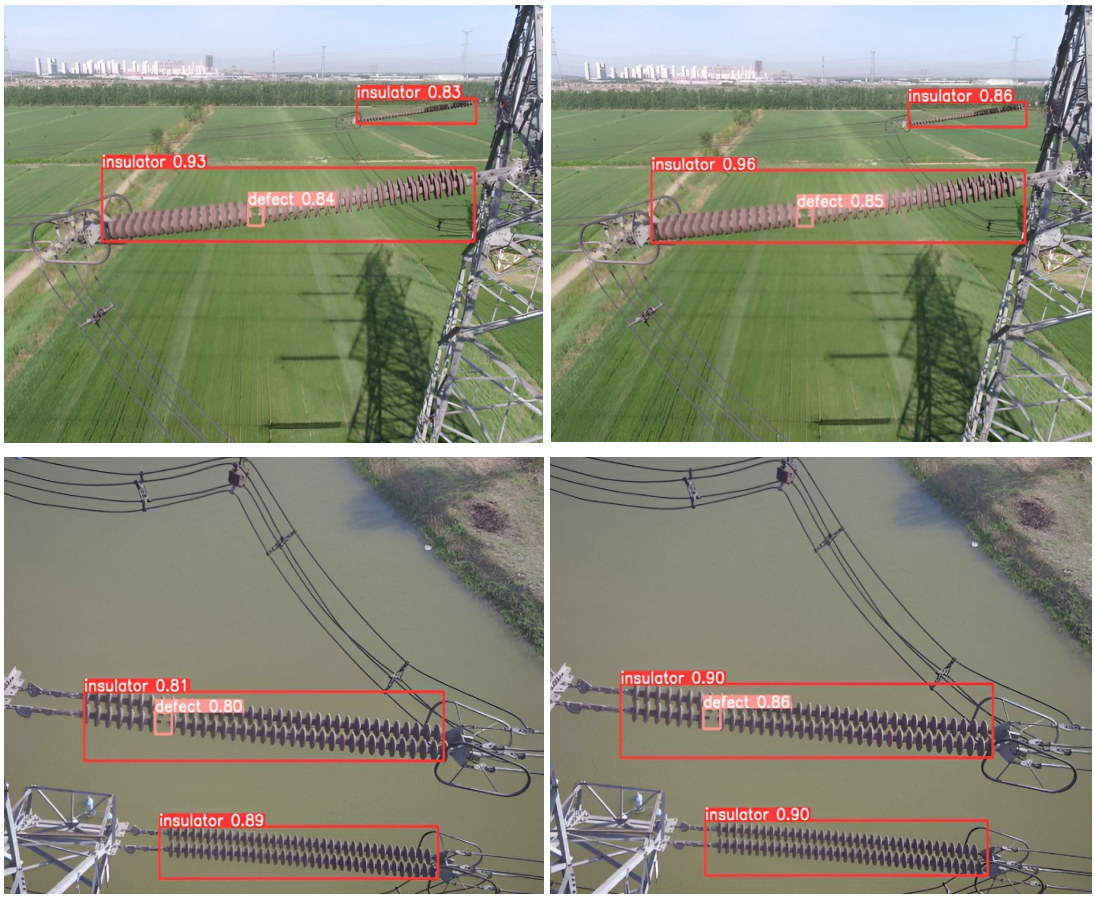


FIGURE 18. Test Comparison chart.

TABLE 8. Comparison with other object detection algorithms.

Module	Precision /%	Recall /%	mAP@0.5/%	mAP@0.5:0.5/%	GFLOPs	FPS
YOLOv3	87.0	86.0	85.8	58.2	154.6	86
YOLOv3-spp	86.8	85.4	85.5	58.0	155.4	83
YOLOv4	86.8	83.4	85.0	56.8	20.6	78
YOLOv5s	86.4	83.8	83.9	55.6	15.8	116
YOLOv6	85.1	83.2	83.6	55.2	24.8	102
YOLOv5-GSConv	84.9	84.8	84.4	55.4	15.2	109
YOLO-S	88.5	85.6	88.1	57.2	14.9	110

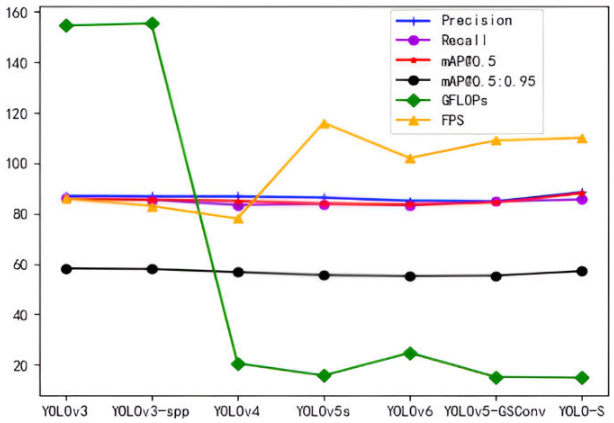


FIGURE 19. Visualization of the results compared to other detection algorithms.

to the important area features and reduce the computation amount as well as to enhance the feature extraction ability of the model for different scales and sizes of targets. In order to verify the positive impact of these improvement mechanisms on the model performance, ablation experiments are conducted to gradually add these improvement mechanisms and train the model, which are designed to demonstrate the positive effect of adding the improvement points. The specific results of the ablation experiments are shown in Table 5. The GSConv+VoV-GSCSP module, MaECA attention mechanism, MishSPPF module and SIoU loss function, were introduced to improve the performance of the model, and the mAP was improved by 1.2%, 1.4%, 1.6%, and 1.3%, respectively. The experimental results show that all these improvement modules have a positive impact on the network to improve the detection accuracy. The simultaneous fusion of these improvement modules, named YOLO-S (YOLO-Small), reduces the amount of parameter computation compared to the original YOLOv5s network, while keeping the FPS from dropping too much. The accuracy is improved by 2.1% and mAP by 4.2%, and the detection

accuracy can be up to 88.5% and mAP can be up to 88.1%, and the computation is reduced by 6%. In the complex environment, the model can identify and detect normal insulators and defective insulators better. The ablation experiment results are shown in Table 6.

The larger the values of Precision and Recall, the higher the precision rate, and the larger the value of mAP, the higher the average precision of detection and the higher the accuracy. the lower the value of GFLOPs, the lower the parameter computation. the larger the value of FPS, the faster the detection speed. So from Fig. 16, it can be visualized that each innovation point added is superior to the original YOLOv5s network in terms of mAP while keeping the values of the other metrics about the same, all of them have a positive impact. (YOLOv5s represents the result without adding any improvement point, serial number 2, 3, 4, 5 represents each kind of improvement point, serial number 6 represents the fusion of serial number 2, 3 improvement points, serial number 7 represents the fusion of serial number 2, 3, 4 improvement points, and YOLO-S represents the fusion of 2, 3, 4, 5, and four improvement points.)

In order to further validate the feasibility of the improved model YOLO-S in this paper, it was chosen to be validated on the public dataset coco128, and the results after 1000 rounds of iterations are shown in Table 7 below.

From Figure 17, it can be seen that YOLO-S outperforms YOLOv5s in various evaluation indicators, mAP@0.5 It has also increased by 24.1%, so YOLO-S has certain applicability.

6) ANALYSIS OF DETECTION RESULTS IMAGE

In order to test the performance of the improved algorithm, a selection of normal and defective insulators were detected under different background conditions and some of the detection results are shown. The left figure shows the detection graph of the original YOLOv5s algorithm, while the right figure shows the detection graph of the improved YOLO-S algorithm. The comparison reveals that the original YOLOv5s algorithm suffers from leakage detection and has a lower value of confidence. In contrast, the YOLO-S algorithm proposed in this paper incorporates four improvement points, which significantly improves the feature extraction ability of the model, which in turn can accurately detect the target and effectively reduces the leakage detection rate. The detection comparison graph is shown in Fig. 18.

7) COMPARISON EXPERIMENT

The YOLO-S algorithm proposed in this article is compared with algorithms widely used in recent years, such as YOLOv3 [26], YOLOv4 [27], YOLOv6 [28], etc. The comparison results are shown in Table 8. Compared with YOLOv3 and YOLOv3 spp [29], YOLO-S has improved mAP indicators by 2.3% and 2.6%, respectively, while the accuracy has increased by 1.5% and 1.7%, greatly reducing computational complexity by more than 10 times. FPS hasalso increased by approximately 1.3 times, and there has been no significant change in recall rate. Compared with YOLOv4, YOLO-S has increased mAP by 3.1%, recall rate by 2.2%, and accuracy has also been improved. The computational complexity has been reduced by 27.7% compared to YOLOv4, and FPS has been improved by about 1.4 times. Compared to YOLOv6, YOLO-S has a 4.5% improvement in mAP, a 3.4% improvement in accuracy, and a 40.0% reduction in computational complexity. Additionally, FPS has also slightly improved. In addition, compared to YOLOv5-GSConv, YOLO-S has improved mAP by 3.7%, accuracy by 3.6%, computational complexity by 2.0%, and FPS has also slightly improved compared to YOLOv5-GSConv.

The performance comparison between YOLO-S and other mainstream models can be more intuitively seen in Fig. 19, which shows that YOLO-S's Precision, mAP, and GFLOPs are better than those of other models, and Recall and FPS also have good performance.

V. CONCLUSION

This study proposes a YOLO-S model for detecting insulators, which addresses issues such as high parameter calculation and low detection accuracy in existing models. Experimental results show that the improved YOLO-S model maintains high detection accuracy and low parameter calculation. However, given the variability of insulator scenes, there may still be cases of missed detection in more complex environments. Future work will focus on improving detection accuracy and addressing missed detection in various settings.

REFERENCES

[1] B. Li, D. Wu, Y. Cong, Y. Xia, and Y. Tang, "A method of insulator detection from video sequence," in *Proc. 4th Int. Symp. Inf. Sci. Eng.*, Dec. 2012, vol. 8330, no. 1, pp. 386–389.

[2] Q. Wu and J. An, "An active contour model based on texture distribution for extracting inhomogeneous insulators from aerial images," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3613–3626, Jun. 2014.

[3] Z. Zhao, N. Liu, and L. Wang, "Localization of multiple insulators by orientation angle detection and binary shape prior knowledge," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 22, no. 6, pp. 3421–3428, Dec. 2015.

[4] S. Liao and J. An, "A robust insulator detection algorithm based on local features and spatial orders for aerial images," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 5, pp. 963–967, May 2015.

[5] F. Gao, J. Wang, Z. Kong, J. Wu, N. Feng, S. Wang, P. Hu, Z. Li, H. Huang, and J. Li, "Recognition of insulator explosion based on deep learning," in *Proc. 14th Int. Comput. Conf. Wavelet Act. Media Technol. Inf. Process. (ICCWAMTIP)*, Dec. 2017, pp. 79–82.

[6] D. Zuo, H. Hu, R. Qian, and Z. Liu, "An insulator defect detection algorithm based on computer vision," in *Proc. IEEE Int. Conf. Inf. Autom. (ICIA)*, Jul. 2017, pp. 361–365.

[7] Z. Zhao, G. Xu, Y. Qi, N. Liu, and T. Zhang, "Multi-patch deep features for power line insulator status classification from aerial images," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2016, pp. 3187–3194.

[8] X. Liu, H. Jiang, J. Chen, J. Chen, S. Zhuang, and X. Miao, "Insulator detection in aerial images based on faster regions with convolutional neural network," in *Proc. IEEE 14th Int. Conf. Control Autom. (ICCA)*, Jun. 2018, pp. 1082–1086.

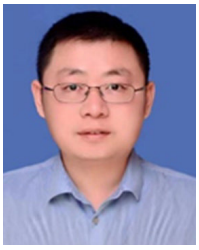
[9] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.

[10] H. Cheng, R. Chen, J. Wang, X. Liu, M. Zhang, and Y. Zhai, "Study on insulator recognition method based on simulated samples expansion," in *Proc. Chin. Control Decis. Conf. (CCDC)*, Jun. 2018, pp. 2569–2573.

[11] G. Jocher et al., "ultralytics/yolov5: v4. 0-nn.SiLU() activations, Weights & Biases logging, PyTorch Hub integration," Zenodo, Honolulu, HI, USA, Tech. Rep., 2021.
[12] X. Lang, Z. Ren, D. Wan, Y. Zhang, and S. Shu, "MR-YOLO: An improved YOLOv5 network for detecting magnetic ring surface defects," *Sensors*, vol. 22, no. 24, p. 9897, Dec. 2022.
[13] H. Li, J. Li, H. Wei, Z. Liu, Z. Zhan, and Q. Ren, "Slim-neck by GSConv: A better design paradigm of detector architectures for autonomous vehicles," 2022, *arXiv:2206.02424*.
[14] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.
[15] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Oct. 2018, pp. 3–19.
[16] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 13708–13717.
[17] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient channel attention for deep convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11531–11539.
[18] A. F. Agarap, "Deep learning using rectified linear units (ReLU)," 2018, *arXiv:1803.08375*.
[19] R. Avenash and P. Viswanath, "Semantic segmentation of satellite images using a modified CNN with hard-swish activation function," in *Proc. 14th Int. Joint Conf. Comput. Vis., Imag. Comput. Graph. Theory Appl.*, 2019, pp. 1–8.
[20] D. Misra, "Mish: A self regularized non-monotonic activation function," 2019, *arXiv:1908.08681*.
[21] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Sep. 2015.
[22] S. Du, B. Zhang, P. Zhang, and P. Xiang, "An improved bounding box regression loss function based on CIOU loss for multi-scale object detection," in *Proc. IEEE 2nd Int. Conf. Pattern Recognit. Mach. Learn. (PRML)*, Jul. 2021, pp. 92–98.
[23] Z. Gevorgyan, "SIoU loss: More powerful learning for bounding box regression," 2022, *arXiv:2205.12740*.
[24] G. Jocher. (2020). *Yolov5*. Journal of Code Repository. [Online]. Available: https:/github.com/ultralytics/yolov5
[25] T. DeVries and G. W. Taylor, "Improved regularization of convolutional neural networks with cutout," 2017, *arXiv:1708.04552*.
[26] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.
[27] A. Bochkovskiy, C.-Y. Wang, and H.-Y. Mark Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.
[28] C. Li, L. Li, H. Jiang, K. Weng, Y. Geng, L. Li, Z. Ke, Q. Li, M. Cheng, W. Nie, Y. Li, B. Zhang, Y. Liang, L. Zhou, X. Xu, X. Chu, X. Wei, and X. Wei, "YOLOv6: A single-stage object detection framework for industrial applications," 2022, *arXiv:2209.02976*.
[29] P. Shaotong, L. Dewang, M. Ziru, L. Yunpeng, and L. Yonglin, "Location and identification of insulator and bushing based on YOLOv3-spp algorithm," in *Proc. IEEE Int. Conf. Electr. Eng. Mechatronics Technol. (ICEEMT)*, Jul. 2021, pp. 791–794.

WEIGUO YI received the bachelor's and master's degrees from Northeast Normal University, in 2002 and 2005, respectively, and the Ph.D. degree from Dalian Maritime University, in 2012. He is currently an Associate Professor with Dalian Jiaotong University. He is also a Master's Degree Evaluation Expert of the Ministry of Education of China. His research interests include artificial intelligence, data mining, and deep learning.

SIWEI MA is currently pursuing the Ph.D. degree with Dalian Jiaotong University. His research interests include machine vision, deep learning, and object detection.



RONGHUA LI received the bachelor's degree from Dalian Jiaotong University, in 2005, and the Ph.D. degree from the Dalian University of Technology, in 2011. He is currently a Professor, the Associate Dean of scientific research, and a Ph.D. Supervisor of Dalian Jiaotong University, mainly engaged in configuration reconstruction and measurement technology research of space non-cooperative targets.