# Recognition Method of Aerial Insulator Defects Based on Deep Learning

# Recognition Method of Aerial Insulator Defects Based on Deep Learning

**J F Wang**[1,*]

[1] School of Electrical Engineering, Shanghai Dianji University, Pudong, Shanghai, China

[*]451661315@qq.com

**Abstract.** Aiming at the problem of low accuracy of aerial photography insulator defect identification during UAV patrol, a method of insulator defect identification and positioning based on improved YOLOv3 target detection model is proposed. This method uses asymmetric convolutional block (ACB) as the structural block of the convolutional neural network model in the training phase. On the basis of not increasing the inference time, the depth of neural network learning is improved. At the same time, the feature map combination of the pyramid multi-scale detection model is improved, and the distance index parameters of the K-means ++ algorithm are optimized, which is used for target prior frame clustering. The experimental results show that the improved method proposed in this paper improves the recognition accuracy of small defective targets. Without reducing the speed of inference, the average accuracy is increased from 84.5% to 91.9%, which meets the requirements of intelligent inspection of UAVs for overhead transmission lines.

## 1. Introduction

Insulators are necessary equipment for electrical insulation and mechanical fixation in high-voltage power transmission lines[1]. From the statistical point of view, the transmission system faults caused by the self-detonation faults of insulators account for more than 50% of the power system faults, and the defect detection has gradually become an important part of the line inspection task[2]. Overhead lines need to pass through rivers, mountains and rivers, and the difficulty of traditional manual line inspections continues to increase. In order to improve the inspection efficiency, State Grid Corporation has gradually established a mode combining drones and manual inspections[3]. Although the UAV aerial line patrol greatly reduces the field work intensity of patrol inspectors, it still adopts the work method of manpower observing images and marking defect information in data processing[4]. This work method is not only inefficient, but also the accuracy rate is limited by the visual observation skill level of the inspectors. At the same time, there is also a hidden danger of visual fatigue leading to an increase in the rate of missed inspections.

As a research hotspot in recent years, deep learning occupies a dominant position in the field of target detection and image segmentation. It does not require manual design to extract the technical advantages of image features, and can better solve the problem of complex background and diversified target features of power inspection images[5]. Tao et al. Proposed an insulator defect recognition method based on convolutional neural networks, and constructed an insulator positive and negative sample data set[6]. Wang et al.[7] was the first to apply the Faster R-CNN[8] algorithm to target detection in power inspection images. The detection accuracy is higher but the speed is slow, which

cannot meet the real-time detection requirements. Prasad et al. [9] used the LBP-HF feature training SVM classifier to detect insulator damage defects, with an accuracy rate of 93.33%. The detection speed is faster, but the detection accuracy of the defective block cannot meet the actual needs of the project.

The purpose of this article is to propose a target detection method that can solve the low accuracy of insulator defect recognition, and finally achieve real-time positioning and accurately identify the insulator and the position of the insulator defect in the high-definition aerial image of the drone Inspection provides technical ideas.The organization of the paper is as follows. In Sect. 2, YOLOv3 target detection model and asymmetric convolution block are briefly outlined. In Sect. 3, model improvements are analyzed in detail. In Sect. 4, a brief analysis of the experimental results and data. In Sect. 5, presents the conclusions.

## 2. Basic theory

### *2.1. YOLOv3 target detection model*

YOLO (You Only Look Once) was quickly attracted a lot of attention as soon as it was proposed, and experienced three iterations of YOLO[10], YOLO9000[11], and YOLOv3[12], constantly optimizing itself and absorbing other detection models. advantage. YOLOv3 combines the two operations of traditional target detection classification and positioning in one step. It has the characteristics of multi-scale detection and multi-target classification. Compared with mainstream target detection models such as Faster R-CNN and SSD[13], it has the same resolution. For better accuracy and faster detection speed, the model structure is shown in Figure 1.

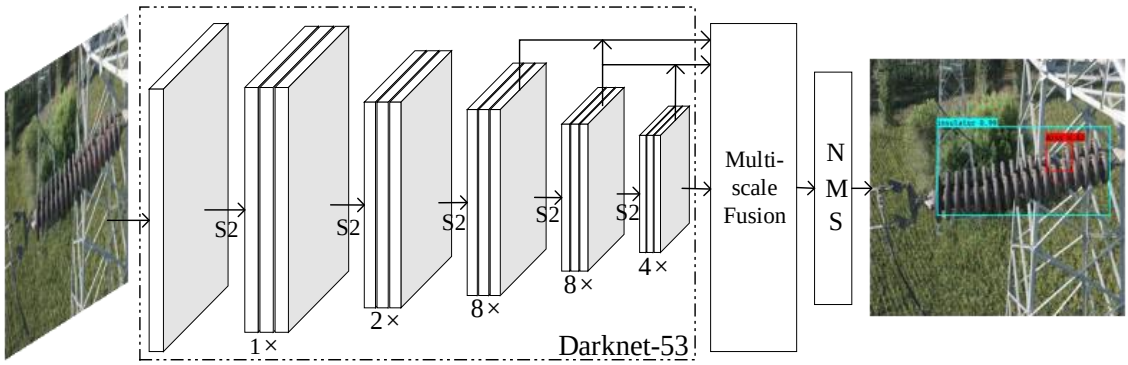


**Figure 1.** YOLOv3 model structure diagram

Based on the second generation Darknet-19, YOLOv3 proposed a deep residual neural network model Darknet-53 (with 53 layers of convolution neural network) by inheriting the advantages of the residual network ResNet[14]. The residual network module reduces the training complexity and achieves the effect of preventing gradient disappearance and gradient explosion during deep neural network training, and uses a convolution kernel with a step size of 2 to replace the pooling in Darknet-19.The layer is down-sampled twice, which greatly reduces the number of parameters in the network training process and improves the training speed at the same network depth. At the same time, combined with multi-scale features (2-fold up-sampling fusion) for detection, the recognition accuracy of targets of different sizes is improved. Finally, in order to improve the accuracy and readability of the detection results, a non-maximum suppression algorithm is used to eliminate the anchor frames with similar positions and repeated categories, thereby obtaining the final detection results.

### *2.2. Asymmetric Convolution Block*

Asymmetric Convolution Block (ACB) is a new type of convolution neural network building block proposed by Ding et al.[15]. During training, ACB mainly consists of three parallel $d \times d$ , $1 \times d$ and

$d\times1$ kernels to replace the original convolution kernel $d\times d$ .During deployment(inference), ACB uses the network information after skeleton weight fusion. We take the commonly used 3×3 size ACB convolution kernel as an example, the main structure is shown in Figure 2.

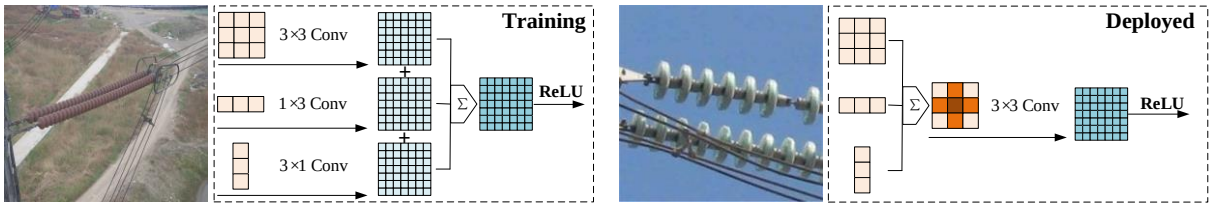


**Figure 2.** ACB structure diagram

In the training phase, ACB uses a parallel structure, using the additivity of convolution calculation (as shown in Equation 1), the three branches are batch normalized after each layer, and the output of the three branches is synthesized as ACB Output without adjusting any additional hyper-parameters. Thus, the skeleton information of the square convolution kernel is enhanced without increasing the number of network layers, and the learning depth of image features is improved.

$$I*K^{(1)}+I*K^{(2)}=I*(K^{(1)}\oplus K^{(2)}) \tag{1}$$

Where $I$ represents the matrix, $K^{(1)}$ and $K^{(2)}$ are two-dimensional convolution kernels with compatible sizes, $*$ is the convolution operation, and $\oplus$ is the sum operation of the corresponding positions.

In the inference stage, ACB uses the homogeneity of convolution, so that the batch normalization and linear transformation during inference can be equivalently fused into the convolutional layer with additional bias. The batch normalization calculation process is shown in Equation 2, and the offset value calculation is shown in Equation 3:

$$F'^{(j)}=\frac{\gamma_j}{\sigma_j}F^{(j)}\oplus\frac{\bar{\gamma}_j}{\bar{\sigma}_j}\bar{F}^{(j)}\oplus\frac{\hat{\gamma}_j}{\hat{\sigma}_j}\hat{F}^{(j)} \tag{2}$$

$$b_j=-\frac{\mu_j\gamma_j}{\sigma_j}-\frac{\bar{\mu}_j\bar{\gamma}_j}{\bar{\sigma}_j}-\frac{\hat{\mu}_j\hat{\gamma}_j}{\hat{\sigma}_j}+\beta_j+\bar{\beta}_j+\hat{\beta}_j \tag{3}$$

Among them, for the No. $j$ convolution kernel, $F'^{(j)}$ represents the convolution kernel after fusion, $b_j$ represents the offset, $\mu_j$ and $\sigma_j$ represent the batch-normalized channel average and standard deviation, $\gamma_j$ and $\beta_j$ represent the scaling factor and offset, $F^{(j)}$, $\bar{F}^{(j)}$ and $\hat{F}^{(j)}$ represent 3×3, 1×3,and 3×1 convolution kernels.

It is worth noting that although ACB can be equivalently converted to a standard conv layer, the equivalence is only established when inference, because the training depth is different, so it produces different performance. In addition, the number of network layers in the inference stage is the same as the original convolutional network, so it does not increase the time consumed in the inference stage.

## 3. Model improvement

### *3.1. Improvement of YOLOv3 target detection model combined with ACB*

Taking the input image size of 416 × 416 as an example, Darknet-53 combined with asymmetric convolution block originally modified the multi-scale index according to the characteristics of the data set, and the improved model is shown in Figure 3.

The convolution block marked by the rectangular box on the left in the figure is composed of a

convolution kernel, a batch normalization layer and a Relu activation function. In the training phase, the original 3×3 convolution kernel is replaced with an ACB-Conv convolution block. Specifically, 3×3, 1×3 and 3×1 convolution kernels are used in parallel convolution mode, and the final feature matrix is summed according to the weights. In the inference stage, the original Darknet-53 network model structure is adopted, in which the 3×3, 1×3 and 3×1 convolution kernels in the trained ACB-Conv are added as equations (2,3) as the original 3×3 convolution kernel for inference.

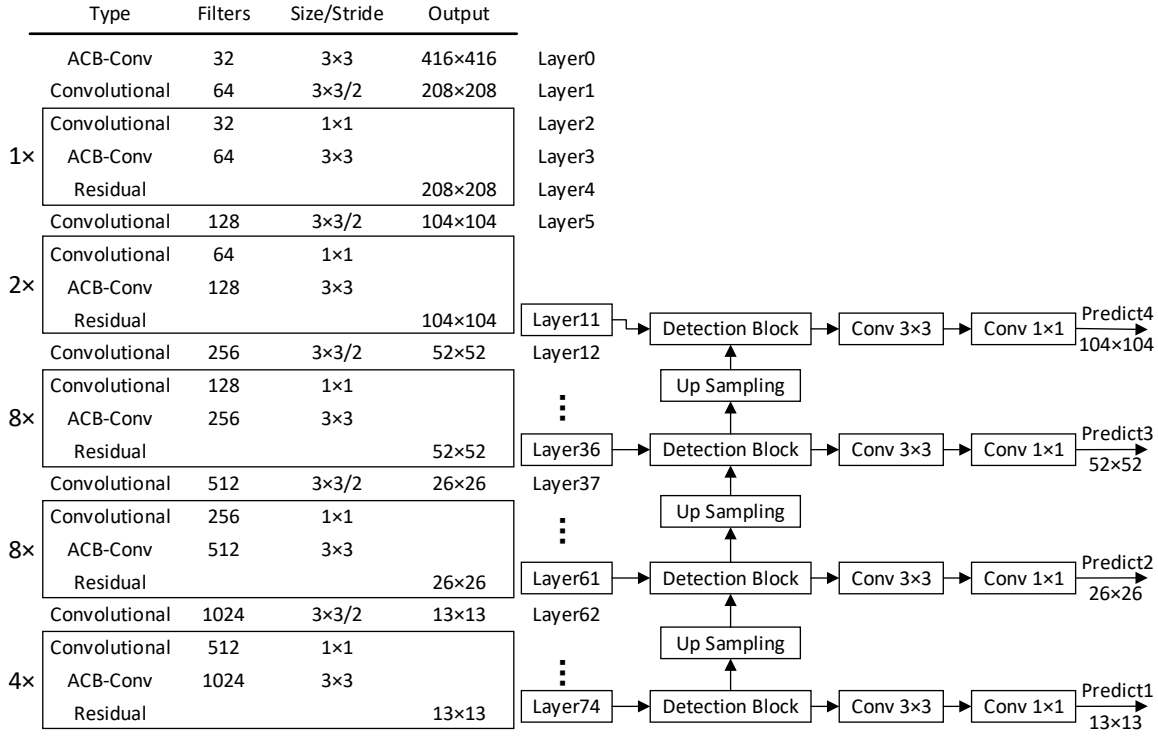


**Figure 3.** Improved network model diagram.

The detection features of the multi-scale pyramid model extracted by the original YOLOv3 algorithm through the Darknet-53 network are 13×13, 26×26, 52×52. Among them, the size of the feature map is proportional to the size of the receptive field, that is, the 13×13 feature map is responsible for detecting large targets, 26×26 is responsible for detecting the target, and 52×52 is responsible for detecting the smaller target. In the aerial image, the insulator defect is smaller.

Therefore, based on the original YOLOv3 multi-scale target detection model, this paper connects the multi-scale detection model to the output of Darknet-53 layer 11 and adds 104×104 scale. At the same time, up-sampling fusion of other scale feature maps is performed to improve the detection accuracy of the insulator defect.

### *3.2. K-means ++ algorithm for modifying distance index*

The YOLO series algorithm obtains a priori frame by clustering the target frame of the training set to guide network learning. The selection of the initial a priori frame will directly affect the accuracy and speed of the target detection by the network. The original YOLOv3 algorithm uses the K-means clustering algorithm to calculate the size of 9 prior frames, of which 1-IOU (IOU is the intersection ratio of the target frame and the clustering center, the calculation method is shown in equation (4)) It is a distance indicator.

$$IOU = \frac{Area(centroid \cap box)}{Area(centroid \cup box)} \quad 0 \tag{4}$$

The K-means algorithm needs to manually select the initial clustering center, and the initial clustering center selection is not suitable, the clustering result is difficult to meet the requirements. In order to solve this problem, this paper uses a more stable K-means ++ algorithm to increase the degree of discrete selection of the initial clustering center, thereby improving the accuracy of the clustering results. At the same time, it is considered that part of the image data containing insulator defect parts in the data set marked in this paper is obtained by data augmentation, which leads to a more concentrated distribution of the true value of the target frame scale. If the original cluster distance index is used, the result of the a priori frame with a good fitting state cannot be obtained. This paper proposes 1-IOU with a center point Euclidean distance penalty factor as the distance index of the clustering algorithm, as shown in equations (5) and (6):

$$D = 1 - IoU + \frac{d^2}{c^2} \tag{5}$$

$$d = \rho^2(obj, cen) \tag{6}$$

In the formula, $obj$ is the target frame, $cen$ is the cluster center, $\rho$ is the Euclidean distance between the center point of the target frame and the cluster center frame, and $c$ is the minimum diagonal distance of the circumscribed rectangular frame, as shown in Figure 4.

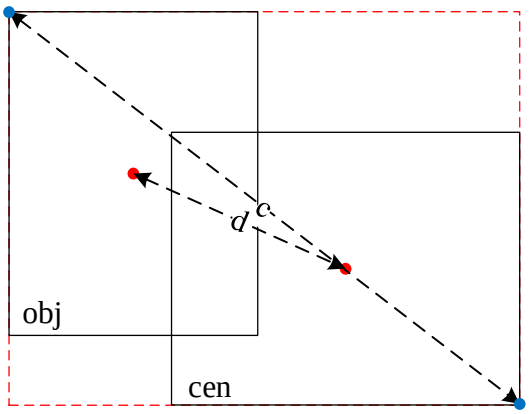


**Figure 4.** Penalty factor for distance

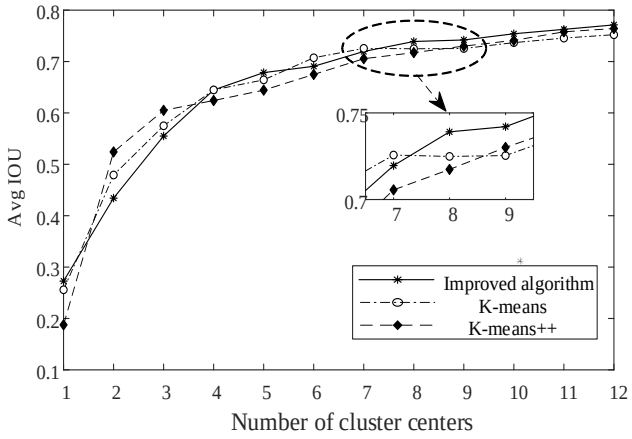


**Figure 5.** Comparison of clustering effects

This paper selects k = 1 ~ 12 to perform cluster analysis on the samples in the data set, and uses Avg IOU as the measurement of cluster analysis (see Equation 7). The relationship between k and Avg IOU is shown in Figure 5.

$$Avg\,IoU = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n_k} I_{IoU}(obj, cen)}{n} \tag{7}$$

In the formula, $n$ represents the number of samples, $k$ represents the number of cluster centers, $n_k$ represents the number of samples in the No. $k$ cluster, and $I_{IoU}(obj, cen)$ represents the intersection ratio of the cluster centers and the target frame.

As the value of k increases, the objective function changes more and more smoothly, and the inflection point of the change can be used as the optimal number of a priori frames. In this paper, when the inflection point of clustering is k = 8, the number of anchor frames is 8, which can accelerate the convergence of the loss function and eliminate the error caused by the candidate frame. The final eight prior frames are: (15,17), (26,60), (39,197), (80,73), (86,22), (118,34), (256,53), (262,80).

## 4. Experiment and result analysis

### *4.1. Performance evaluation index*

In order to verify the detection effect and stability of the model and the influence of the sample on it, F1-score, Precision, mean Average Precision(mAP) and detection speed (FPS) were used to compare the test results.

### *4.2. Comparison and analysis of results*

*4.2.1. Comparison of the effects of ACB-based convolutional neural network model.* Combined with the asymmetric convolution block to conduct comparative experiments in different convolutional neural network models, the experiments were carried out using YOLOv3 and Tiny-YOLO frameworks, respectively. The experimental results are shown in Table 1. Experimental results show that combining ACB can effectively improve model accuracy without increasing inference time.

**Table 1.** Combined with ACB network model comparison

| Testing framework | Network model | P/ Insulator | P/ Defect | mAP | Detection speed /FPS |
|---|---|---|---|---|---|
| YOLOv3 | Darknet-53 | 85.5 | 83.7 | 84.5 | 33.2 |
| | **ACBnet-53** | **86.7** | **85.4** | **86.1** | **33.9** |
| Tiny-YOLO | Tinynet | 79.3 | 77.2 | 78.3 | 51.2 |
| | **ACBtinynet** | **81.4** | **79.9** | **80.7** | **51.1** |

*4.2.2. Comparison of multi-scale detection effects.* In order to better adapt to the problem of small targets in the image of insulator defect areas, this paper proposes a method of adding 104×104 scale feature maps for up-sampling fusion. The experimental results are shown in Table 2. The results show that the method in this paper effectively improves the recognition accuracy of small targets with insulator defects.

**Table 2**. Combined with ACB network model comparison

| Detection algorithm | Network model | Number of scales | P/ Insulator | P/ Defect | mAP |
|---|---|---|---|---|---|
| YOLOv3 | Darknet-53 | 3 | 85.5 | 83.7 | 84.5 |
| | Darknet-53 | **4** | **86.9** | **90.6** | **88.7** |

*4.2.3. Analysis of the final experimental results.* In the training process, we used the GIoU proposed by Hamid et al. In CVPR2019 as the loss function of the bounding box regression[16], to improve the detection accuracy, in which the GIoU calculation is shown in Equation 8:

$$GIoU = \frac{A \cap B}{A \cup B} - \frac{C - (A \cup B)}{C} \tag{8}$$

In the formula, A and B represent any two boxes, C represents the smallest circumscribed rectangle of the two boxes A and B, and the area measurement is used in the calculation process.

The training process is shown in Figure 6(a). The Tiny-YOLO model converges fastest due to the low network depth. The improved algorithm in this paper uses an asymmetric convolution block to construct a convolutional neural network, which converges faster than the original Darknet-53 network model, which verifies the effectiveness of ACB. In addition, from the perspective of mAP@0.5 and F1 performance indicators (as shown in Figure 6(b), (c)), the algorithm of this paper is superior to the other two algorithms under the insulator data set. The final model test results are shown

in Table 3. The improved YOLOv3 target detection algorithm proposed in this paper effectively improves the detection accuracy of small targets with insulator defects while maintaining the original algorithm detection speed. In addition, the average accuracy and F1-score performance indicators have been improved. A part of the detection effect diagram is shown in Figure 7.

**Table 3.** Final experimental results

| Testing framework | Network model | Feature map size | Number of prior frames | Insulator | Miss | mAP@0.5 | F1-score | FPS |
|---|---|---|---|---|---|---|---|---|
| Tiny-yolo | Tinynet | 13*13、26*26 | 6 | 79.3 | 77.2 | 79.9 | 0.84 | 51.1 |
| YOLOv3 | Darknet53 | 13*13、26*26、52*52 | 9 | 85.5 | 83.7 | 84.5 | 0.89 | 33.2 |
| Improved YOLOv3 | ACBnet53 | **13*13、26*26、52*52、104*104** | **8** | **91.3** | **92.1** | **91.9** | **0.93** | **29.8** |

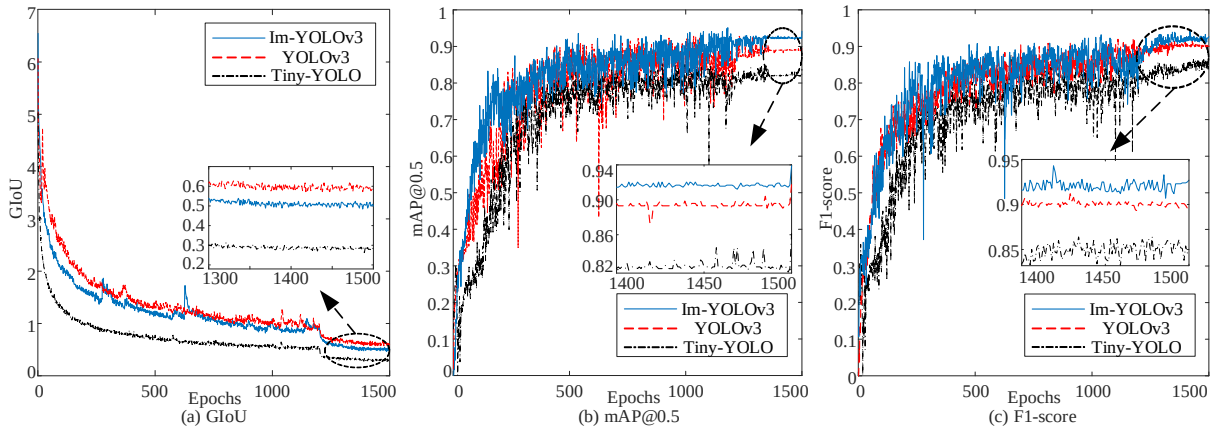


**Figure 6.** Experimental process comparison

*4.2.4. Analysis of the final experimental results.*

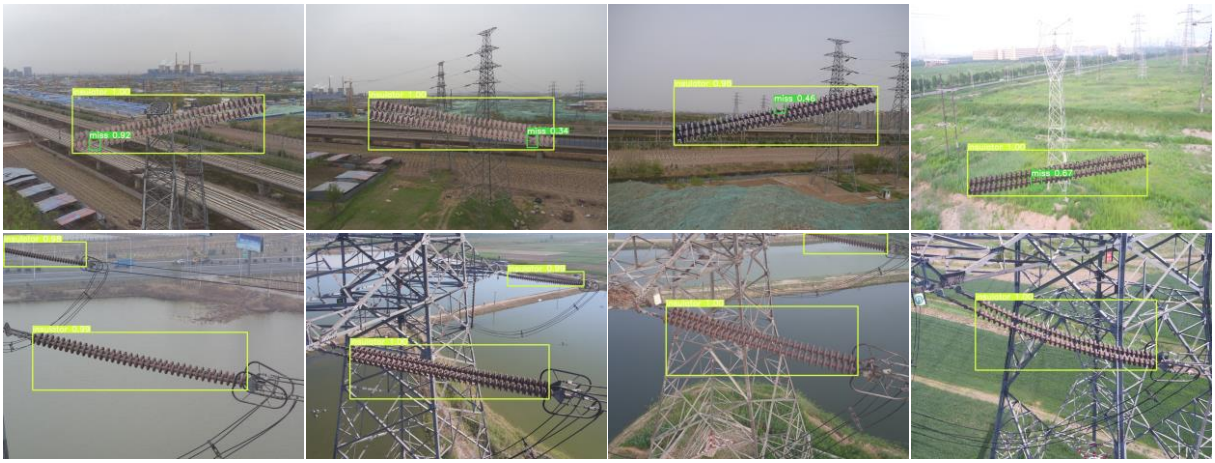

**Figure 7.** A part of the detection effect chart

## 5. Conclusion

Combining with the development status of deep learning, we proposed an ACBnet-53 convolutional neural network model using asymmetric convolutional blocks, improved the original YOLOv3 multi-scale pyramid feature fusion model, and added 104×104 scale feature maps for up-sampling fusion. In addition, in order to improve the fit between the model and the data set constructed in this paper, this paper uses the K-means++ algorithm to modify the distance index to perform a priori frame clustering, and modify the number of the original network a priori frame to 8.

On the basis of maintaining the detection speed of the original algorithm, the recognition accuracy of the area where the small defect target is located is effectively improved. The mean average precision reached 91.9%, and the recognition speed reached 29.8 FPS, which met the needs of real-time detection in the transmission line drone aerial inspection.

How to build a large-scale transmission line inspection image data set to further improve the detection efficiency, and build a cloud intelligent detection system in combination with 5G technology will be the focus of the next research.

**References**

[1] Park K C, Motai Y and Yoon J R 2017 Acoustic fault detection technique for high power insulators *IEEE Transactions on Industrial Electronics* **12**(64) 9699-9708
[2] Rui C, Wei Z and Bin C 2017 Analysis on Explosion of Toughened Glass Insulators Occurred in 500kV Tianping Double Transmission Lines *Smart Power* **45**(09) 74-79
[3] Van N N, Robert J and Davide R 2018 Automatic autonomous vision-based power line inspection: A review of current status and the potential role of deep learning *Electrical Power and Energy Systems* **99** 107-120
[4] Zhiying L, Xiren M and Jing C 2020 Review of Visible Image Intelligent Processing for Transmission Line Inspection *Power System Technology* 1-14
[5] Zhou G, Yuan J and Yen I 2016 Robust real-time UAV based power line detection and tracking *IEEE International Conf. on Image Proc.*(Phoenix) 744-748
[6] Tao X, Zhang D and Wang Z 2018 Detection of power line insulator defects using aerial images analyzed with convolutional neural networks *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 1-13
[7] Wang W X, Tian B, Liu Y, et al. 2017 Study on the electrical devices detection in UAV images based on region based convolutional neural networks *Journal of Geo-information Science* **19**(02) 256-263
[8] Ren S, He K, Girshick R, et al. 2017 Faster R-CNN: towards real-time object detection with region proposal networks *IEEE Transactions on Pattern Analysis & Machine Intelligence* **39**(6) 1137-1149
[9] Prasad P S, Rao B P. 2016 LBP-HF features and machine learning applied for automated monitoring of insulators for overhead power distribution lines *International Conf. on Wireless Communications*(Chennai, India) IEEE 808-812
[10] Redmon J, Divvala S, Girshick R, et al. 2016 You Only Look Once: unified, real-time object detection *Conf. on Computer Vision and Pattern Recognition (CVPR)* (San Francisco) IEEE 779-788
[11] Redmon J and Farhadi A 2017 Yolo 9000: Better, faster, stronger *Conf. on Computer Vision and Pattern Recognition(CVPR)* (San Francisco) IEEE 6517-6525
[12] Redmon J and Farhadi A 2018 YOLOv3: An Incremental Improvement *Conf. on Computer Vision and Pattern Recognition(CVPR)*
[13] Liu W, Anguelov D, Erhan D, et al. 2016 SSD: Single shot multiBox detector *European Conf. on Computer Vision* (Berlin: Springer-Verlag) 21-37
[14] He K, Zhang X, Ren S, et al. 2016 Deep residual learning for image recognition *Conf. on Computer Vision and Pattern Recognition (CVPR)* (San Francisco) IEEE 770-778
[15] Ding X, Guo Y, Ding G, et al. 2019ACNet: Strengthening the Kernel Skeletons for Powerful CNN via Asymmetric Convolution Blocks *IEEE International Conf. on Computer Vision(ICCV)*
[16] Rezatofighi H, Tsoi N, Gwak J, et al. 2019 Generalized Intersection Over Union: A Metric and a Loss for Bounding Box Regression *Conf. on Computer vision and pattern recognition* 658-666.