Insulator defect detection algorithm based on a lightweight network

Ying LAN 1*, Wenxiang Xu 1

1 Logistics Engineering College, Shanghai Maritime University, Shanghai, 201306, China

*lanying@shmtu.edu.cn, 202030210243@stu.shmtu.edu.cn

Abstract. Insulators are the key components of transmission lines. The identification and detection of insulator defects are directly related to the stable operation of transmission lines. In order to improve the efficiency of the insulator and its defect location, a faster defect detection algorithm based on YOLOv5 is proposed. Firstly, a lightweight Ghost module was introduced in the YOLOv5 backbone network, which significantly improved the detection speed with ensuring accuracy. Secondly, Secondly, CBAM is introduced into YOLOv5 Neck network to further improve the detection accuracy. The experimental results show that the model of the improved post-network is smaller compared to the YOLOv5 original network, and the detection speed improves greatly while ensuring the detection accuracy. It is of great significance to power grid operation and maintenance.

1. Introduction

Insulator is the key component of the transmission line, its main role is to support the wire and insulate the wire from the rod and tower, with the dual functions of electrical insulation and mechanical support[1]. Insulator state detection is an important guarantee to ensure the stable operation of electric power transmission lines. Because the insulator is in the outdoor environment, the weather and other factors are easy to cause damage to the insulator string and produce defects[2]. It has had a great impact on the safe and stable operation of the power grid, and also poses a great threat to the social economy[3]. Therefore, the defect detection of the insulator is of important significance for the power grid operation and maintenance.

With the rapid development of deep learning, various target detection algorithms based on deep learning have been proposed and have been widely used in insulator detection[4]. In literature [5], a convolution layer in the original YOLOv3 network was replaced with dense-Net, which realized multi-layer feature fusion of insulator images and improved detection accuracy. In literature [6], a new residual block with 4-fold reduction sampling friendly to small targets was added between the second and third residual blocks of YOLOv3 in feature extraction network Darknet-53 to improve the detection accuracy of small targets. In view of the difficulty in small target detection, literature [7] combined non-maximum suppression algorithm, sliding window and YOLOv4 neural network to improve the accuracy of small target detection in insulator defect area. Literature [8] proposed an intelligent identification algorithm of insulator defects in HSR catenary based on deep learning technology, efficientnet-B3 network was used to discriminate the insulator status, and then YOLOv5 model was used to accurately detect the information of insulator defects, which improved the accuracy of intelligent identification of insulator defects in catenary.
On the basis of the above research, the YOLOv5 algorithm with smaller model parameters and higher speed and accuracy is selected for insulator detection in this study. On this basis, the structure is improved. Experimental results show that compared with the original YOLOv5 network, the improved network model is smaller, and the detection accuracy is guaranteed while the detection speed is greatly improved.

2. Realization process of insulator defect detection

In this paper, YOLOv5 is used as an insulator defect detection algorithm. The realization process of insulator defect detection based on YOLOv5 is shown in Figure 1, and the specific process is as follows:

YOLOv5 network firstly performs feature extraction and feature fusion operations on the input image, and then divides the input image into an equal grid of $S \times S$. If the center position of the target to be measured falls into a grid, the target will be processed and detected by the grid.

Each grid will generate 3 prediction boxes, and each prediction box will output 5 elements $x$, $y$, $w$, $h$, $c$. Where $x$ and $y$ are the central coordinates of the prediction frame, $w$ and $h$ are the width and height of the prediction frame. $C$ is the confidence degree, reflecting the prediction of the model for this grid, which is defined as follows: If there is no object in this grid, the confidence degree should be 0. If there is an object in this grid, the confidence is the intersection ratio (IOU) between the prediction box and the real boundary box of the object. In addition, each grid will output D conditional category probabilities, that is, the probability that a grid contains an object belonging to a certain class. In summary, by using YOLOv5 detection network, the input image can be processed by the detection network to obtain a tensor of size $S \times S \times 3 \times (5 + D)$.

In the prediction stage, the detection network will use the tensor of $S \times S \times 3 \times (5 + D)$ obtained by NMS (non-maximum suppression) to further process, and the invalid prediction box is eliminated and the best prediction box is retained. Finally, the detection network successfully realized the location of insulators and their defect positions and output categories.

3. Improved insulator detection model based on YOLOv5.

3.1 The network structure of YOLOv5

The network structure of YOLOv5 consists of four parts: Input, Backbone, Neck, and Head. The network structure of YOLOv5 is shown in Figure 2 below.
Figure 2. Network structure of the YOLOv5

The Input network is used for pre-processing of images, including adaptive anchor frame calculations, adaptive image filling, and Mosaic Data augmentation in three parts. The Input network may process the original image as a uniform size image of 608×608×3 to facilitate further processing of the image by a subsequent network. The backbone network was used to feature extract the images, and the backbone network contained the Focus structure and the C3 structure. The Focus structure provides the slice manipulation. The input image of 608×608×3 was first sectioned to obtain a feature map of 304×304×12. Then 32 convolution checks were used to calculate the features of 304×304×12, resulting in a new feature map of 304×304×32. The role of C3 structure is to locally cross layers, using the feature information of different layers to obtain richer feature maps. The Neck network is used to generate feature pyramids, which contains FPN and PAN. FPN integrates strong semantic features at the high level from top to bottom. PAN is a complement to FPN and conveys strong localization features. Both are used simultaneously, strengthening the feature fusion capability of the network. The Head network was used for the final detection, including the loss function of the prediction box and the NMS (Non-Maximum suppression) Algorithm. The YOLOv5 uses the GIoU Loss as a loss function of the prediction box, the problem of overlapping boundary boxes can be effectively solved, and the speed and accuracy of the prediction box regression can be improved. The NMS is used to screen the target detection prediction stage detection box.

3.2 Improved insulator detection model

Currently, in practical insulator detection application environments, detection models are mostly deployed on mobile or embedded devices that fail to provide sufficient computing power for large network models. Due to the excessive network parameters of YOLOv5, and the feature extraction network of YOLOv5, there are a lot of redundant feature diagrams in the process of convolution operation, which will greatly consume the computing resources of hardware equipment and make the detection time of images longer. In order to further improve the detection speed of the model, the network structure of YOLOv5 is improved.

First, CBAM is introduced into the Neck network, and the structure diagram is shown in Figure 3 below. The CBAM module contains two submodules, the Channel Attention Module and the Spatial Attention Module.
Figure 3. Schematic diagram of the CBAM structure

The channel attention module focuses its attention on the channel to determine which channels have the main characteristics of the target, and find meaningful parts in the input image. The basic principle of the channel attention module is as follows: First, the input feature $F$ is fed into the global average pooling layer and the global maximum pooling layer compresses the feature map on the spatial dimension, and the compressed features, $F_1, F_2$. Then, $F_1, F_2$. Input Multi-Layer Sensor (MLP) for convolutional operation to obtain feature $MLP(F_1), MLP(F_2)$. Finally, $MLP(F_1), MLP(F_2)$ Add up and use the sigmoid activation function to get a feature diagram of the channel's attention weight, $M_c(F)$, $M_c(F)$ is calculated as Formula (1), where the $\sigma$ is the sigmoid activation function.

$$M_c(F) = \sigma(MLP(F_1), MLP(F_2))$$ (1)

The Spatial Attention module focuses on the spatial position and determines which locations contain the main information about the target. The basic principles of the spatial attention module are as follows: First, the input feature, $F'$, will obtain the spatial feature through the average pooling layer and the maximum pooling layer, $F_3, F_4$. Then, stitch together, $F_3, F_4$, to form a new spatial feature map with a channel number of 2. Finally, the new spatial feature map is convoluted and the sigmoid activation function is used to obtain the feature map of the spatial attention weight, $M_s(F)$. $M_s(F)$ is calculated as Formula (2), where the $\sigma$ is the sigmoid activation function, and $f^{7x7}$ represents the convolution layer of the 7x7 convolution core.

$$M_s(F) = \sigma(f^{7x7}([F_3, F_4]))$$ (2)

To sum up, the output characteristic is calculated by the input feature $F$, and the formula for $F''$ is shown in the following formula (3), where $M_c(F)$ is the channel attention weight feature, $M_s(F)$ is the spatial attention weight feature.

$$\begin{cases} F' = M_c(F) \ast F \\ F'' = M_s(F') \ast F' \end{cases}$$ (3)

The Conv module and C3 module in the backbone network of YOLOv5 are then replaced with the Ghost module[16] and the Ghost Bottleneck module. Through the appeal operation, the calculation amount of the model in the feature extraction stage can be reduced and the detection speed can be improved significantly under the condition that the detection accuracy of the model is unchanged. The Ghost module is a lightweight, efficient convolution module that integrates traditional volumes into two steps, first using fewer traditional convolutions to generate the underlying feature map, and then using the underlying feature diagram to generate a new feature map using a smaller linear convolution. Finally, the two sets of feature diagrams are stitched together to get the final output. The structure of the Ghost module is shown in Figure 4 (a) below.
To further reduce the need for hardware resources, the C3 module in YOLOv5 can be replaced with an efficient Ghost Bottleneck module, as shown in Figure 4 (b). There are currently two main types of Ghost Bottleneck: where Ghost Bottleneck 1 is the simple stacking of two Ghost modules, and then the input and output of the two Ghost modules are connected using shortcut. Ghost Bottleneck 2 increases the number of channels used to reduce feature mapping compared to Ghost Bottleneck 1, and then connects the inputs and outputs of the two Ghost modules using shortcut.

Finally, the improved model is named YOLOv5_CG, where CG is an abbreviation for the Ghost module and the CBAM initials.

4. Analysis of the experimental process and the results

4.1 The introduction of datasets

The datasets of defective insulators used in this study comes from the Internet. The pixel size of insulator images is $1152 \times 864$, and there are 240 images in total. To prevent too few experimental samples, resulting in over-fitting network training. The data set was enhanced by randomly adding noise and randomly changing image brightness, and a total of 2400 images were generated. Then labelImg was used to label the image: the insulator was marked as "insulator" and its defects as "defect". After annotation, the data set was divided into 90% training set and 10% test set.

4.2 Experimental evaluation indicators

MAP (mean accuracy), FPS (number of images processed per second) and Model Size (Model Size) are used as evaluation indexes in this study. The specific calculation process is as follows:

MAP is the mean of average accuracy for all categories in the dataset. The calculation of mAP with N detection categories is shown in Formula (4) below. Where $P$ and $R$ represent the accuracy rate and recall rate respectively, as shown in Equations (5) and (6). $TP$ is the number of samples correctly classified into a certain category; $TN$ is the true number of negative samples as negative sample prediction; $FP$ represents the number of samples incorrectly assigned to a category; $FN$ represents the number of misidentified targets. FPS represents the number of images processed per second, and the indicator is related not only to the calculation of the algorithm model, but also to the hardware performance during the experiment. Model Size represents the size of the system memory space that the algorithm model takes up, typically in MB.

\[
mAP = \frac{\sum_{i=1}^{N} P(R) dR}{TP} \tag{4}
\]
\[
P = \frac{TP}{TP + FP} \tag{5}
\]
\[
R = \frac{TP}{TP + FN} \tag{6}
\]
4.3 Experimental platform introduction and training parameter setting
This paper builds a model based on the Pytorch deep learning framework, the system is configured as NVDIA TITAN RTX ×2 GPU (32G), i7-10700k× 2 CPU, operating system Windows10, CUDA 11.2, Python 3.8. After the experimental environment is built, enter the training stage of the model, the main parameters are set as follows: enter the image size 640×640, the number of pictures per training session is 16, and the number of training times is 200.

4.4 Analysis of experimental results
Table 1 below compares the metrics on the dataset for the two algorithms. Experimental results show that the modified model size is only 66.04%, and the speed of running on the GPU is increased by 29.18 percent, and the mAP is increased by 1.4%. Therefore, the improved model can improve the detection speed and reduce the model parameters under the premise of ensuring the detection accuracy.

Table 1. Test results and algorithm comparison.

| Models     | FPS(f/s) | Model size(MB) | mAP(%) |
|------------|----------|----------------|--------|
| YOLOv5     | 107.14   | 13.7           | 98.2   |
| YOLOv5_CG  | 138.41   | 9.06           | 99.6   |

In order to compare the detection effects of YOLOv5 and YOLOv5_CG on insulator images, this paper randomly selects 4 images from the test set and uses YOLOv5 and YOLOv5_CG images for detection respectively. The detection and comparison results are shown in Figure 4

![Test image 1](image1)
![Test image 1](image2)
![Test image 2](image3)
![Test image 2](image4)
![Test image 3](image5)
![Test image 3](image6)
![Test image 4](image7)
![Test image 4](image8)

Figure 5. The test results of YOLOv5_CG
Table 2 shows the comparison of detection confidence between YOLOv5 and YOLOv5_CG on the insulators in Figure 5. By comparison with Table 2, it can be seen that, compared with the original YOLOv5 network, YOLOv5_CG improves the average reliability of the detection of insulators and their defect locations. Therefore, the effectiveness of the improvement work has been proved successfully.

| Image       | YOLOv5 insulator | YOLOv5_CG insulator | YOLOv5 defect | YOLOv5_CG defect |
|-------------|------------------|---------------------|---------------|------------------|
| Test image 1| 0.98             | 0.94                | 0.97          | 0.97             |
| Test image 2| 0.97             | 0.91                | 0.97          | 0.95             |
| Test image 3| 0.97             | 0.95                | 0.97          | 0.95             |
| Test image 4| 0.97             | 0.93                | 0.97          | 0.95             |

5. Conclusion
In the actual insulator detection environment, devices such as drones can not provide sufficient computing power for large network models. In this paper, we design and implement the lightweight YOLOv5_CG network model. The size of this model is only 66.04% of that of YOLOv5, and it is 29.18% faster on GPU and 1.4% faster on mAP. Therefore, it greatly reduces the dependence on the hardware environment, and can meet the actual application needs.

This algorithm also has some limitations, and the next step will continue to increase the diversity and complexity of the data, considering other related factors such as weather. At the same time, the network structure will be further improved, and the subsequent work will also solve how to efficiently deploy the mobile models, and verify and improve the proposed model in practical applications.

References
[1] Shang, J.P., Li, C.X., Chen, L. (2017) Visual-based insulator localization and self-explosion defect detection. Journal of Electronic Measurement and Instruments., 31 (06): 844-849.
[2] Zhou, Z.B., Gao, J., Zhang, W., Wang, X.J., Zhang J. (2018) visible light image target detection of transmission lines based on a deep convolutional neural network. LCD & Display, 33 (04): 317-325.
[3] Lv, Z.N. 2020. Review of common fault analysis and detection methods of power transmission lines . Automation and instrumentation., (01): 161-164 + 168.
[4] Lin, M.S., Lin, R., Hong, L.Y., Xiao, T.H. (2020) Study on Insulator Defects in Electric Power Inspection. Mathematical modeling and its Application., 9 (04): 85-93.
[5] Zhang, H.K., Li, J.Y., Zhang, B. (2020) Insulator foreign body detection method based on improved YOLO v3. China Power., 53 (02): 49-55.
[6] Chen, Z.Y., Min, F. (2020) Based on the improved YOLO V3 contact net insulator detection method, Journal of Wuhan University of Engineering,42(04):462-466.
[7] Lin, M.S., Lin, R., Hong, L.Y., Xiao, T.H. (2020) Study on Insulator Defects in Electric Power Inspection. Mathematical modeling and its Application., 9 (04): 85-93.
[8] He, E.L.(2021)Intelligent network insulation defect identification method Based on Deep Convolutional Network. China's high-tech. (03): 80-81.
[9] Qi, M., Zhu, D., Zou, W.X. Image processing technology review. Digital technology and application., 38 (02): 57 + 59.
[10] Zhong, Z.F., Xia, Y.F., Zhou, D.P., Yan, Y.T. (2021). Lightweight objective detection algorithm based on improved YOLOv4. Computer application,,1-8
[11] Li, J.Y., Yang, J., Kong, B., Wang, C., Zhang, L. (2021) Multi-scale vehicle and pedestrian detection algorithm based on the attention mechanism. Optical precision engineering, 29 (06): 1448-1458.
[12] Chen, P.C., Qi, B.R., Kong, E.Y. (2021) Application of an improved non-maximum suppression algorithm in the YOLOv3 model. Digital technology and application.,39 (01): 121-123.
[13] Yu, J.R., Huang, D.Q., Zeng, R., Zhao, H.H., (2021), Research on the model detection algorithm based on improved YOLOv4.Laser magazine.,1-8
[14] Xu, C.J., Wang, X.F., Yang Y.D. (2019) Attention-yolo: A YOLO detection algorithm based on attentional mechanism. Computer engineering and applications., 55(06):13-23+125.

[15] Yang, J., Zhang, C.I., Li, Z.X., Tang, Y.P. (2021) Integrated spatial attention and attitude estimation of blocking pedestrian reidentification. computer research and development., 1-11

[16] Guo, F. (2021) An improved YOLOv3 lightweight network model based on the traffic night vision scene. Electronic Technology and Software engineering., (10): 171-172