Improved YOLOX Transmission Line Insulator Identification

Zhongqi Zhao[1(✉)], Qing He[2], Sixuan Dai[2], and Qiongshuang Tang[2]

[1] Haibei Power Supply Company, State Grid Qinghai Power Company, Haiyan 812200, China
254739027@qq.com

[2] School of Electrical and Information Engineering, Changsha University of Science and Technology, Changsha 410114, China

Abstract. Aiming at the problem of low recognition accuracy of insulators in power system transmission lines and the recognition results contain many backgrounds, this paper proposes a high-performance detection model by combining the improved YOLOX target detection algorithm with the rotating frame detection algorithm. Firstly, the backbone network of YOLOX is replaced with ConvNext with a larger receptive field to improve the feature learning ability of the model for insulators. Secondly, the fusion between the output features of the feature pyramid pooling module is enhanced using the channel disorder operation. Finally, the angle classification of the detection frame is added to the network to realize the rotation frame detection and reduce the background interference in the recognition result. The model is trained and tested with manually marked aerial photography data. The test results show that the method has high accuracy in insulator identification and meets the high-performance detection requirements.

Keywords: Insulator identification · Target detection · Rotating box detection · Feature pyramid pooling

1 First Section

1.1 A Subsection Sample

In the power system, the safety of transmission lines is the primary factor for the failure-free operation of the power grid, so the grid needs to be inspected regularly. The traditional inspection mode requires employees to go to the site for inspection, which needs a lot of human and material resources and is inefficient.

Because of the rapid development of unmanned aerial vehicle (UVA) technology, using UVA instead of manual operation can simplify the work pattern and reduce the risk of aerial operation [1]. In order to accelerate the transformation of traditional inspection work mode to automatic operation mode, the power grid company has adopted a combination of helicopter and laser point cloud scanning to achieve full coverage of transmission line images. At present, the "machine patrol as the main, human patrol as a supplement" of "machine patrol + human patrol" operation and maintenance mode

H. Yang et al. (Eds.): SmartGIFT 2022, LNICST 483, pp. 186–199, 2023.
https://doi.org/10.1007/978-3-031-31733-0_17

has basically formed, transmission line operation and maintenance level continues to improve [2]. In order to be safe and stable, power transmission conductors are generally erected at high altitudes through poles and towers. The function of insulators is to ensure that the conductor is insulated from the tower and the earth, with the characteristics of high voltage resistance and mechanical stress resistance. However, the complex outdoor environment causes insulators to be damaged, partially dislodged, dirty, etc., which can easily lead to power line operation failure, it is necessary to replace the detected damaged insulator [3], and insulator detection provides significant data support for it.

In the traditional insulator identification algorithm, researchers use the edge features, histogram features, and color features of insulators as the basis for the identification of insulators. Reference [4] uses Euclidean distance to match insulator features extracted by Canny edge detection to intelligently identify insulators. Reference [5] uses the Dirichlet distribution combined with relative entropy to obtain the semantic relevance of features and improve the recognition accuracy. Traditional algorithms rely on human a priori knowledge, which requires manual design of classification features and increases the complexity of the algorithm.

The development of Graphics Processing Unit (GPU) technology has facilitated the widespread application of deep learning technology. In the insulator identification task, the deep learning model is constructed so that the model can learn the contextual semantic information of insulators autonomously and complete the automatic insulator identification work, which simplifies the complexity of the identification process and improves the work efficiency. The target detection algorithm can complete the insulator recognition and get the insulator location information through logistic regression. It can be mainly divided into two categories, single-stage detection algorithms including SSD [6], YOLO [7], RetinaNet [8], etc., and two-stage detection algorithms including R-CNN [9], Fast R-CNN [10], Faster R-CNN [11], etc.

Reference [12] applied the Faster R-CNN detection algorithm to the insulator identification task of transmission lines, completed the identification of different types of insulators, and proved the effectiveness of the algorithm. Based on Faster RCNN, the reference [13] uses an attention mechanism to improve each feature extraction module, forcing the network to learn channel feature correlation, suppressing background interference, and improving model accuracy. Reference [14] uses the YOLOV2 detection algorithm to identify insulators and combines fault diagnosis technology to complete insulator defect discrimination. Reference [15] uses multi-feature fusion technology to enhance the detailed information of small objects and then combines spatial attention mechanisms to optimize model performance. To address the problem of little semantic information, reference [16] uses multi-scale feature fusion to enhance the semantic features of targets and optimize the scale scaling module. The model is made more suitable for insulator identification with a large-scale span. To achieve real-time detection requirements, reference [17, 18] replaces the traditional convolution operation with depth-separable convolution. It can reduce the computational complexity of the model and improve the efficiency of the algorithm under the premise of ensuring accuracy.

Deep learning techniques have significantly improved the automation of transmission line inspection, and the accuracy of the current single-stage algorithm in the target detection network is improved to a high level of accuracy. Based on this, this paper proposes an insulator identification model with an improved YOLOX algorithm to improve the detection accuracy of insulator identification.

2 Detection Network

2.1 YOLO Network

In 2015, Redmon et al. proposed the yolov1 detection algorithm, which is a single-stage target detection algorithm and an anchors-free target detection algorithm.

Subsequent authors proposed V2 and V3 versions, in which V2 introduced the idea of anchors and proposed to use of a stronger feature extraction network, V3 version proposed a stronger backbone feature extraction network DarkNet53 and introduced the idea of FPN to detect multi-scale targets, among which the V3 algorithm is one of the most widely used algorithms. Later, some scholars proposed V4 and V5 versions based on the V3 version, which improved the speed and accuracy of the model. The V4 and V5 models have similar structures, but the V5 version is very friendly to engineering deployment. In July 2021, Megvii Technology released the YOLOX target detection algorithm.

2.2 YOLO Network

YOLOX is an anchors-free algorithm, YOLOV4 and YOLOV5 use the idea of hyperparameter evolution to update their hyperparameters during training, which will lead the algorithm to over-optimization on anchors, so the authors do not improve on V4 and V5 structure, but on the model of YOLOV3-SPP. The model structure of YOLOX is shown in Fig. 1. Conv in the figure represents convolution, BN represents batch normalization, and Sibu represents activation function.

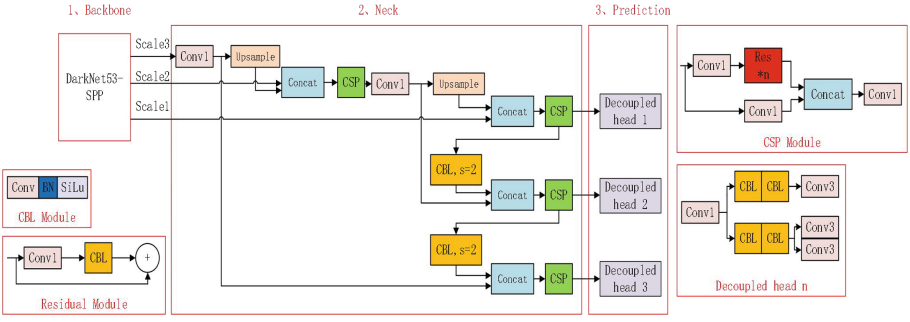

Fig. 1. YOLOX structure diagram.

In the backbone network, the feature extraction module is consistent with that in the YOLOV3-SPP algorithm, and both use DarkNet53 to extract higher-order semantic features and the SPP module to extract multi-scale features. The feature fusion module introduces the PAN structure based on the FPN structure. The shallow feature map contains more local detail information, and the local detail information is beneficial to the regression task, and the deep feature map contains more higher-order semantic features, and the higher-order semantic information is beneficial to the classification task.

The fusion of shallow and deep features can obtain a feature mapping of the two attributes, thus improving the detection performance of the model. In the prediction module, a Decoupled Head (DH) strategy is used to decouple the prediction branches, where the output branches of classification and regression share a single convolutional layer in YOLOV3. Since the classification branch requires more higher-order semantic information and the regression branch requires more local detail information. On this basis, the output of the classification branch is separated from the branch of the regression output and use two-way convolution to perform the regression task and the regression task separately. The addition of DH improves the detection accuracy of the model and speeds up the convergence of the network.

2.3 Optimization Strategy

It is well known that the detection algorithm based on anchors based has the problems of difficult adjustment of hyperparameters and poor versatility. To obtain better detection performance, the V3 algorithm sets three a priori boxes of different sizes to improve the detection accuracy of the model, but it reduces the detection speed of the model to a certain extent.

A large number of anchors free algorithms have emerged in recent years, which eliminate a large number of tuning links, reduce the complexity of the algorithm, and are no longer limited by the setting of anchors for extreme size targets, which can improve the detection performance of the model. The YOLOV3 algorithm is combined with the detectors of anchors free to extract the target detection algorithm of anchors free. The detection algorithm based on anchors free changes the original prediction of 3 different sizes of boxes in one location into 1 box, which reduces the computation of the model in the detection head link and improves the detection speed of the model. The performance of the target detection algorithm depends to some extent on the label assignment of positive and negative samples. The assignment of positive and negative samples is according to the Intersection over Union (IoU) ratio of the ground truth and the prior frame. However, the assignment of positive and negative samples is different in different situations (obscured), and at this point, the label assignment is regarded as an optimal transmission problem, and the simOTA label assignment method is proposed. The detection accuracy of the model is improved by proposing a cost matrix to select the a priori frame corresponding to the position with the smallest cost as the positive sample to match the ground truth.

3 Model Improvement

3.1 ConvNext

Reference [19] proposes a neural network model with a larger sensory field with a Base Unit (BU) as shown in the a-plot in Fig. 2b. Firstly, the input is goes through a deep separable convolution of size 7×7 for feature extraction, and the larger the convolution kernel, the wider the feature area learned. Then the features are up-dimensioned and down-dimensioned using two 1×1 ordinary convolutional operations to obtain more abstract semantic information, and finally the features are summed and fused with those of the residual mapping branch. The downsampling operation can reduce the input feature resolution and the computational complexity of the model. This paper shows the Down Sample Uint (DSU) in Fig. 2b, a 3×3 depth-separable convolution is added to the residual mapping, while the step size of the 7×7 depth-separable convolution is 2, which is applied to double the feature resolution and double the number of channels.

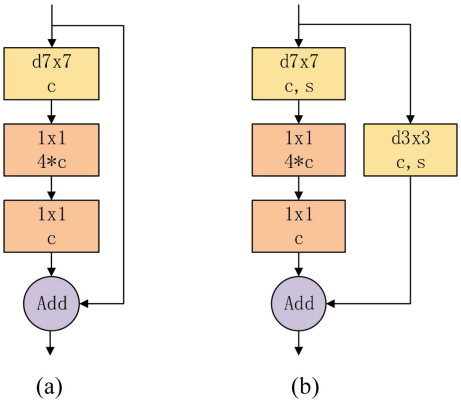

Fig. 2. ConvNext module. (a) Base Unit (BU); (b) Down Sample Uint (DSU).

3.2 Spatial Pyramid Pooling Module

In this paper, SPP [20] has been modified adaptively, and its structure is shown in Fig. 3. Firstly, four sets of feature compression vectors are obtained by using four sets of maximum pooling operations with different sizes for the input features. To reduce the number of parameters of the module and to ensure that the feature dimension of each group of features is consistent with the input when stitching, so compress its channel dimension by a factor of 4 using 1×1 convolution operation. Then the feature resolution is restored using upsampling, and the four groups of features are combined in the channel dimension, and the combined results are summed and fused with the original input features. Finally, to increase the transferability of information between features, a Channel Shuffle operation is used to reorder the final results in the channel dimension.

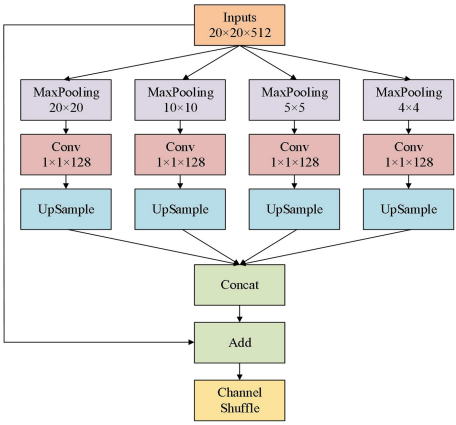

Fig. 3. ConvNext module. (a) Base Unit (BU); (b) Down Sample Uint (DSU).

Table 1. Backbone network

name	size	layer
Inputs	640 × 640 × 3	-
Conv1	320 × 320 × 32	BU × 1 DSU × 1
Conv2_x	160 × 160 × 64	BU × 2 DSU × 1
Conv3_x	80 × 80 × 128	BU × 5 DSU × 1
Conv4_x	40 × 40 × 256	BU × 2 DSU × 1
Conv5_x	20 × 20 × 512	BU × 1 DSU × 1
	20 × 20 × 512	SPP × 1

3.3 Backbone

To improve the model feature learning capability, the original backbone network in YOLOX is replaced using the network structure in Table 1. The network input size is 640 × 640 × 3. In the first level of features, 1 BU module and 1 DSU module are used to learn shallow features, and change the resolution of the features to half and double the number of channel dimensions. Similarly, in the second level of features, 2 BU modules and 1 DSU module are used to learn the features. In the third level features, 5 BU modules and 1 DSU module are used to learn the features, in the fourth level feature, 2 BU modules and 1 DSU module learning features are used, and in the fifth level, 1 BU module and 1 DSU module are used to learn the features. Finally, the improved SPP module is used to acquire multi-scale features.

3.4 Rotation Detection

YOLOX is a horizontal rectangular box based target detection algorithm, using a horizontal rectangular box for tilted target detection will contain a lot of background information. The best practice is to detect targets using rotating rectangular boxes, so this paper uses the YOLOX target detection algorithm to achieve the detection of rotating targets, with the following improvements.

Circular Smooth Label Coding. A branch is added to the output port of the network's detection head in order to get the rotation angle, which is necessary for rotating box detection in addition to the horizontal rectangular box's *x*, *y*, *w*, and h dimensions. This paper uses longside format to define the rotated rectangular frame, which is the format of *x*, *y*, *w*, *h*, and θ ($0 \leq \theta < 180°$). The acquisition of θ does not use the regression task, because the periodicity of the angle will interfere with the results of the regression, but the classification task is used to obtain the angle θ. However, due to the periodicity of the angle, a simple one-hot encoding cannot be used. Instead, the encoding method of the Circular Smooth Label (*CSL*) proposed in reference [21] is used to classify the angle of the box. Its expression The formula is shown in formula 1:

$$CSL(x) = \begin{cases} g(x), \ \theta - r < x < \theta + r \\ 0, \ else \end{cases} \tag{1}$$

where θ denotes the current rotation angle of the real frame, r denotes the window radius, whose default value is 6, and *g(x)* denotes the window function.

The window function expression is shown in formula 2, and its related expressions in formula 3 and 4 are respectively shown. The window function meets periodicity and symmetry.

$$g(x) = e^{-\frac{(x-\mu)^2}{2\delta^2}} \tag{2}$$

$$g(x) = g(x + KT), \ K \in N, T = 180/\omega \tag{3}$$

$$0 \leq g(\theta + \varepsilon) = g(\theta - \varepsilon) \leq 1, \ |\varepsilon| < r \tag{4}$$

where the mean $\mu = 0$, the variance $\delta = 0$, and N is the set of natural numbers.

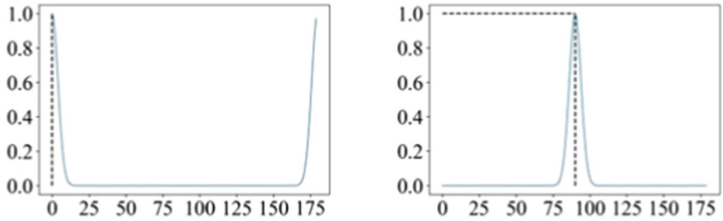

Fig. 4. *CSL* encoded label score map. (a) $\theta = 0°$; (b) $\theta = 90°$.

When the true θ value is 0° or 90°, the corresponding *CSL* coded label scores are shown in the left and right plots in Fig. 4, respectively, where the horizontal axis indicates the angle value and the vertical axis indicates the coded label score.

Loss Function Optimization. There are two places in YOLOX where the loss function needs to be calculated, the first time the loss is found for constructing the cost matrix for label assignment with the loss function shown in in formula 5, and this time the loss is calculated simply for screening positive and negative samples.

$$\begin{aligned} \text{Cos}t &= L^{cls} + \lambda L^{reg} + L_1 \\ L_1 &= \begin{cases} -1000, \ if \ in\, center\, or\, r \leq 2.5 \\ 0, \text{else} \end{cases} \end{aligned} \tag{5}$$

where the classification loss (L^{cls}) uses a binary cross-entropy loss function, the regression loss (L^{reg}) uses an iou loss function, λ is used to control the ratio of the two loss weights, and λ defaults to 3. The L_1 in formula is equivalent to adding more prior information to the equivalent, increasing the degree of matching of the high-quality prior (the center of the grid is within the range of the true box or 2.5 grid).

A second calculation of the loss is used to optimize the model with the loss function shown in formula 6.

$$Loss = L^{cls} + \lambda L^{reg} + L^{conf} \tag{6}$$

The equation contains classification loss, confidence loss and regression loss, where the binary cross-entropy loss function is used for classification loss, and confidence loss, and λ is used to control the regression loss weight ratio, which is 5 by default. This paper also uses the approach proposed in reference [22] to swap the true labels of confidence with the true labels of classification, which is used to solve the inconsistency between the training session and the testing session of the network due to the absence of supervised signals in the confidence output of some boxes during training. Since the output branch of the improved algorithm has an additional branch for rotation angle classification, the loss function in YOLOX needs to be improved. The improved cost function at label assignment is given in formula 7.

$$\text{Cos}t = L^{cls} + \lambda L^{reg} + L_1 + L^{\theta - cls} \tag{7}$$

where (θ-*cls)* is used to calculate the angular classification loss in the form of sigmoid combined with binary cross-entropy to calculate the angular loss. The improvement of the loss function during model training is shown in formula 8.

$$Loss = L^{cls} + \lambda L^{reg} + L^{conf} + L^{\theta-cls} \tag{8}$$

The optimized loss function adds the angle classification loss on the basis of the original loss in the text, and also uses the sigmoid combined with the binary cross entropy to calculate the angle loss.

4 Experimental Results and Analysis

4.1 Data Set Introduction and Data Enhancement

The experimental data are obtained from aerial image maps, of which 1500 are in the training set and 300 are in the test set. The insulators in the image maps are labeled using the Labelme tool, which is divided into two modes: rectangular box labeling and rotating rectangular box labeling, and the results are shown in Fig. 5. Where a is the rectangular box labeling result, which contains more background, and b is the rotating box labeling result.

(a) (b)

Fig. 5. Annotation mode. (a) Rectangular box labeling; (b) Rotated box annotation.

In view of the difficulty of image data acquisition, tedious and time-consuming labeling, and considering that the neural network model requires more data for fitting in the process of learning target features, the training set is subjected to data enhancement processing. Insulators have directional feature invariance in spatial location, scale variability in shape, and obscured phenomenon in shooting angle. Therefore, this paper uses the data enhancement methods of mirror image enhancement, multi-scale scaling, and random erasure to expand the training data to 3,000 sheets.

4.2 Experimental Environment

The code used for the experiments in this paper runs on a deep learning server configured with an RTX3090 graphics card, 24G of current memory, and 32G of running memory. A neural network model is built using the Tensorflow 2.0.0 deep learning framework.

4.3 Learning Rate Setting

The number of iterations for each model is set to 100, and the Adam optimizer is selected. In the process of searching for the global optimal solution of the model, using a larger learning rate at the early stage of training can avoid the model from falling into the local optimal solution, and at the same time, it should ensure that the learning rate decays at a smaller rate. And in the middle of training, the parameters of each model are more stable, so the learning rate decay should be accelerated. In the late training period, the learning rate should be slowed down in order to prevent the model parameters from oscillating and becoming unstable due to too large a change in the learning rate. Therefore, this paper uses the "cosine annealing" learning rate decay strategy as shown in Fig. 6, and the learning rate decrease helps the model converge.

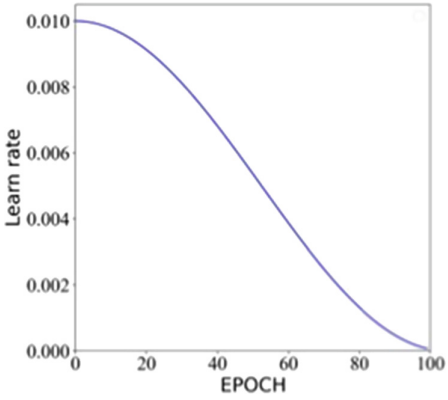

Fig. 6. Cosine annealing.

4.4 Experimental Results

In order to better evaluate the model performance, Precision (P), Recall (R), and Average Precision (AP) evaluation metrics are selected for analysis in this paper. First, TP is defined to denote the number of detected frames with IoU greater than a specified threshold, FP denotes the number of detected frames with IoU less than a specified threshold, and FN denotes the number of undetected true frames. The meaning of each indicator representation is shown below.

Precision (P). The precision represents the probability of the true being an insulator in the sample predicted to be an insulator. The expression is calculated as follows.

$$P = \frac{TP}{TP + FP} \tag{9}$$

Recall (R). The recall represents the probability of predicting a positive sample among the samples that are truly insulators. The expression is as follows.

$$R = \frac{TP}{TP + FN} \tag{10}$$

Average Precision (AP). The average precision is defined as plotting a curve on two-dimensional coordinates using a combination of different P-point coordinates and R-point coordinates, and the area of the curve is the AP of a single category.

4.5 Qualitative and Quantitative Analysis

Comparison of Each Index. The results of model evaluation index comparison are shown in Table 2. In this paper, the improved algorithm is compared with the current mainstream detection algorithm, and the two-stage algorithm Faster RCNN has higher accuracy than YOLOV3-YOLOV5, but the YOLOX algorithm performs better, in which the AP of the improved model reaches 98.7%, which is 1.2% higher than the YOLOX model, and the performance is optimal. Since the evaluation index of rotating frame is different from that of horizontal rectangular frame, its AP reaches 90.8%, which proves that rotating frame detection has better application in transmission line insulator detection.

Table 2. Evaluation indicators

Model	AP%
Faster R-CNN	96.8
YOLOV3	92.4
YOLOV4	93.1
YOLOV5	93.9
YOLOX	97.5
Ours	98.7
R-YOLOX	90.8

Horizontal Frame Detection Results. The effect of the improved model in the horizontal frame detection is shown in Fig. 7, which shows that the insulator as a whole can be well detected. The improved SPP module enables the model to have better characterization capability in terms of multi-scale features, and the smaller insulators in the figure can be well identified. In addition, this paper uses the on-the-fly data erasure enhancement method to effectively alleviate the difficult identification problem caused by the obscured insulators. However, it can be seen from the results that the rectangular box where the insulators are detected contains a large background, which can easily bring interference to the insulator fault detection.

Detection Results of Rotating Frame. In order to reduce the interference of complex background in the recognition results, this paper uses the idea of rotating frame target detection to get the position information of the rotating frame by regressing the position coordinates as well as the angle classification, and its detection results are shown in Fig. 8.

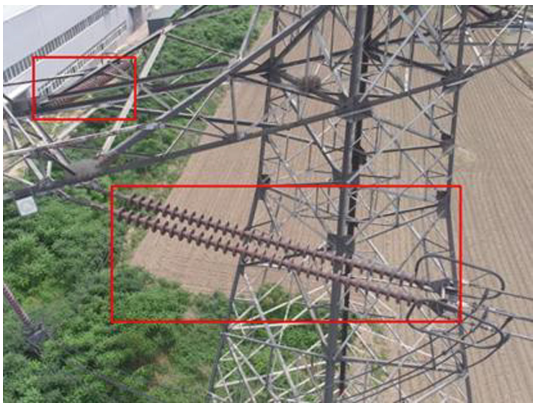

Fig. 7. Horizontal box detection results.

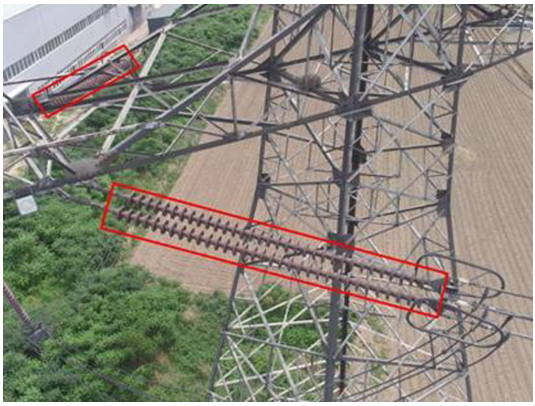

Fig. 8. Detection results of rotating frame

The rotating frame is the main body of the insulator, and there is less background in the frame, which largely reduces the false detection caused by background interference.

4.6 Ablation Experiments

In order to test the insulator recognition effect of the improved network, ablation experiments are set up in this paper, as shown in Table 3, to verify the detection effect of the proposed model by adding improvement steps step by step. The original YOLOX algorithm can achieve 97.5% AP, which is improved by 0.8% points by using ConvNext to improve the backbone feature extraction network and improve the model feature learning capability. Enhancing the output feature fusion by adding channel disorder operation in the SPP module, which achieves 98.7% AP. Finally, using the rotated rectangular frame for target detection, the obtained R-YOLOX model performance can also reach 90.8% with excellent performance.

Table 3. Evaluation indicators

Model	AP%
Original YOLOX	97.5
Improved YOLOX (ConvNext)	98.3
Improved YOLOX (ConvNext + Channel Shuffle)	98.7
R-YOLOX	90.8

5 Conclusion

In this paper, based on the YOLOX detection algorithm, the DarkNet53 is replaced by a backbone network with a larger feeling field, which can make the model learn larger local features, while the use of depth-separable convolution in the backbone network ensures the high-speed of the model. The features output from the spatial pyramid pooling module are independent in the channel dimension, and this paper uses the channel disorder operation to increase the information interactivity between its features and enhance the feature fusion capability. And the idea of rotating frame detection is added to reduce the interference of background in the recognition results and provide good data support for insulator fault recognition, and the final model is optimal in terms of accuracy.

References

1. Chen, F.X., Yang, L., Xie, C., et al.: Unmanned aerial vehicle transmission line patrol technology and its application. Technol. Innov. Appl. **11**(25), 174–176 (2021)
2. Wang, L.F.: Research and Application of UAV in Transmission Line Patrol. Doctor, North China Electric Power University (2018)
3. Liu, Z., Zhang, L.M., Geng, M.X., et al.: Target detection method for high voltage cable based on improved faster R-CNN. J. Intell. Syst. **14**(04), 627–634 (2019)
4. Yao, X.T., Liu, L., Li, Z.Y.: Catenary insulator recognition method based on canny edge feature points. Electr. Porcelain Arrester **01**, 142–148 (2020)
5. Wu, W.H., Sun, L., Wang, G.Z., et al.: Recognition and classification based on synonyms for the allocation of railway catenary insulator. Electr. Porcelain Arrester **01**, 156–160 (2020)
6. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., Berg, A.C.: SSD: single shot multibox detector. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9905, pp. 21–37. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46448-0_2
7. Redmon, J., Farhadi, A.: Yolov3: an incremental improvement, arXiv preprint https://doi.org/10.48550/arXiv.1804.02767
8. Lin T, Y., Goyal, P., Girshick, R., et al.: Focal loss for dense object detection. In: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2980–2988. IEEE (2017)
9. Girshick, R., Donahue, J., Darrell, T., et al.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 580–587.IEEE (2014)
10. Girshick, R.: Fast R-CNN. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1440–1448, IEEE (2015)

11. Ren, S., He, K., Girshick, R., et al.: Faster R-CNN: towards real-time object detection with region proposal networks. In: Advances in Neural Information Processing Systems, vol. 28 (2015)
12. Cheng, H.Y., Zhai, Y.J., Chen, R.: Based on Faster R-CNN insulator in aerial image recognition. J. Mod. Electron. Technol. **42**(02), 98–102 (2019)
13. Zhao, W.Q., Cheng, X.F., et al.: Attention mechanism and Faster RCNN for insulator recognition. J. Intell. Syst. **15**(01), 92–98 (2020)
14. Lai, Q.P., Yang, J., Tan, B.T., et al.: Insulator automatic identification and defect diagnosis model based on YOLOv2 network. China Power **52**(07), 31–39 (2019)
15. Wang, T.: Research on Recognition and Location of Insulators in Aerial Images Based on SSD. Master, School of Control and Computer Engineering (2021)
16. Zhao, W.Q., Zhang, H.M., Xu, M.F.: Insulator recognition based on an improved scale-transferrable network. J. Image Graph. **26**(11), 2561–2570 (2021)
17. Hong, G.: Research on identification and positioning of insulators based on lightweight algorithm UAV images. Inner Mongolia Sci. Technol. Econ. **21**, 83–85 (2019)
18. Tang, L., Wang, S.Q., Jin, H.B., et al.: Intelligent Electric Power,50(02), 69–74 (2022)
19. Liu, Z., Mao, H., Wu, C.Y., et al.: A ConvNet for the 2020s, arXiv preprint https://arxiv.org/abs/2201.03545
20. He, K., Zhang, X., Ren, S., et al.: Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Trans. Pattern Anal. Mach. Intell. **37**(9), 1904–1916 (2015)
21. Yang, X., Yan, J.: Arbitrary-oriented object detection with circular smooth label. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12353, pp. 677–694. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58598-3_40
22. Li, X., Wang, W., Wu, L., et al.: Generalized focal loss: learning qualified and distributed bounding boxes for dense object detection. In: Advances in Neural Information Processing Systems, vol. 33, pp. 21002–21012 (2020)