Multi-angle Identification of Small Target Faults in Transmission Lines Based on Improved YOLOX Algorithm

Shurong Peng, Jieni He(✉), Huixia Chen, Bin Li, Jiayi Peng, and Lijuan Guo

Changsha University of Science and Technology, Changsha 410114, China
1535133864@qq.com

Abstract. In the grid patrol work, there are some fault types with small targets in the line that needs to be detected. For the problem of partial feature loss when the target is small in UAV image recognition, CutMix is used to perform multi-angle image fusion on the line images captured by UAV, which is used to improve the accuracy of target detection. The improved YOLOX-pruning algorithm model is used for deep learning to prune and sparse the network structure, thus removing the redundant nodes of the network to improve the speed of target detection. In this experiment, manually labeled line images are fed into the model to train the features of the faulty components in the images. With a 50% reduction in channel parameter size and multi-angle feature fusion, the algorithm target detection speed is improved by 2.569 frames per second and the mAP value of the faulty data set is improved by 3.378%, reducing the amount of operation while improving the target detection accuracy.

Keywords: Transmission Line Inspection · Deep Learning · Target Detection · Multi Angle · Channel Prunning

1 Introduction

In the power system, the transmission line corridor environment is very complex, the circuit operation state will change with the line element changes, and these changes are partly caused by weather changes, partly caused by foreign body invasion. Transmission lines change more with the terrain, and the power grid personnel need to inspect before and after the peak of electricity consumption and some bad weather before and after, in this case for the personal safety of personnel has a certain adverse impact.

In the process of manual inspection of transmission lines, an infrared thermometer is usually used for local temperature measurement [1] and power robot inspection [2], etc. With the continuous application of UAV inspection in power systems, in addition to general high-voltage transmission lines, UAV inspection has also been continuously applied to the field of new energy generation, such as wind turbine hub center detection and tracking [3] and photovoltaic power generation module temperature detection using infrared thermal imaging UAV [4].

H. Yang et al. (Eds.): SmartGIFT 2022, LNICST 483, pp. 56–72, 2023.
https://doi.org/10.1007/978-3-031-31733-0_6

Traditional target detection methods include HOG gradient histogram, feature pyramid [5], and sliding window technique [6]. The path aggregation network PANet [7] improves the feature pyramid by facilitating the flow of information in an instance-based segmentation framework through bottom-up path enhancement and precise localization of low-level signals. In the context of power system inspection, for the problems of poor image quality, complex background, and poor contrast captured in complex scenes, Ref. [8] proposed a regional convolutional neural network infrared image target detection method incorporating image direction gradient histogram to solve the problem of large scale super-resolution image reconstruction task with large loss of image information, but and to solve the problem of small target fault features under conventional scale image The problem of loss, this study starts from several small target faults in the power grid, combines the fault characteristics, uses multi-angle fusion images, and optimizes the learning of small target fault features by the algorithm.

The target detection feature extraction process is affected by the location and scale of the target in the image, and the detection results will have different degrees of deviation. To address these problems and the features of large computation and high dimensionality, the spatial selection method and multilevel method are proposed to reduce the computation by eliminating regions with less information [9]; and the sliding window in the sliding scan, because it has to be repeatedly scanned under different scales of the same image, will regions overlap and cause redundancy, using a new method that can directly create very small order Markov sets can greatly improve the computational efficiency compared to existing methods [10]. For the current target detection in computer vision, the main function implemented is to identify the target object in the picture and to classify and localize it. Using neural network deep learning for target detection can be achieved based on a large amount of raw data, using a better generalization to fit the nonlinear function [11] and solve the problem of missing features in traditional target detection. In the process of image feature processing, the principal component analysis method PCA (principal component analysis) can reduce the dimensionality of the feature vector, and for the problem of its time-consuming computation of the eigenvector space, Ref. [12] proposed a 2DPCA algorithm, based on the principal component analysis of 2D image matrices, which does not need to transform the image matrix into a vector before feature extraction. In this study, we introduce the channel pruning method to optimize the modal feature extraction channel and remove the redundant nodes and parameters of the network while retaining the main feature network to reduce the computational effort to improve the algorithm running speed.

For various problems arising in transmission lines, many authors have proposed detection methods for specific features of a single fault. In Ref. [13], for aerial transmission line images, a color model conversion method based on grayscale variance normalization is proposed to highlight conductor areas and achieve accurate identification of conductor surface damage areas. For the problem of missing cotter pins in a railroad power installation, Ref. [14] proposed a fault detection method based on a deep convolutional neural network and integrated learning, using an integrated classifier composed of multiple linear SVMs to achieve cotter pin missing fault detection. For the sample class imbalance problem in transmission line insulator defect detection, Ref. [5] used ResNeXt-101 as a feature extraction network to fully extract features and applied an

online hard example mining (OHEM) training strategy to solve the positive and negative sample imbalance problem. For the more difficult to identify transmission line broken strands detection problem, this study further refines the detection targets into conductor loose strands and conductor broken strands and uses multi-scale calibration to change the category imbalance problem and improve the detection effect.

In this study, CSPDarknet [15] is the backbone network algorithm to train the target detection network for several types of faults that are difficult to be detected by image recognition in overhead lines of high-voltage networks, and the model is improved and optimized by channel pruning. Faults containing broken wires, dropped cotter pins, broken insulators, and loose anti-vibration hammers are pre-processed with images, and the target detection is performed after multi-angle image fusion to enhance the input image quality using CutMix for small faults. By comparison, the improved model ensures the accuracy of target detection under the same size data set while the volume is smaller.

2 Transmission Line Fault Characteristics

Transmission lines are exposed to the natural environment for long periods and the components are susceptible to corrosion by rain, snow, high temperatures, and disturbance by strong winds. In transmission lines, cotter pin dislodgement sometimes occurs due to thermal expansion, contraction, or wind [16]. And the cotter pin installation environment is not fixed, the target is small and the background is cluttered, etc., which often leads to missed detection during the identification process.

Anti-vibration hammers are used to prevent fatigue damage to the wire and to absorb wire vibrations during wind [17]. Due to vibration over due to wind often accompanied by rain and erosion resulting in rusting of metal parts, making the anti-vibration hammer loose, fracture, and slip [18], the loss of the original damping effect; long-term operation of the line leads to a certain arc sag and the aging phenomenon of the wire, as well as the line in the air by wind disturbance when the different amplitude of vibration occurs, easily caused by the occurrence of loose strands, half broken strands, and broken strands, resulting in local Current and wire temperature abnormalities; part of the obvious wire breakage due to broken strands produce wire branching, resulting in a straight wire reduction [19], while a small part of the wire produced loose strands and not completely broken strands in the image for the increase in the radius of the wire, which can determine whether there is a wire breakage fault in the image and the extent of the broken strands.

The detection of faults such as missing open pins is a small target detection, which is one of the most time-consuming and labor-intensive parts of manual judgment and is prone to miss detection in the case of large image areas, object occlusion, and complex backgrounds. CutMix was used to achieve multi-angle image fusion, improve the problem of small target fault feature loss when the algorithm samples the image, and conduct deep learning with UAV patrol images so that the neural network is trained to learn the iconographic features of faults such as cotter pin dropped, loose anti-vibration hammer, broken insulator and broken wire in the image.

3 Deep Learning Based on the YOLOX Model

The YOLOX [20] algorithm makes many improvements based on the YOLO series, one of the more important improvements is the use of Anchor-Free detection, and the target detection framework is shown in Fig. 1.

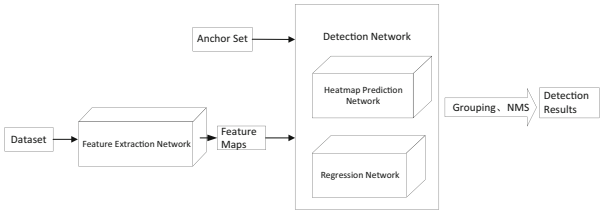


Fig. 1. Anchor-Free target detection framework

Anchor-Free detection is used to solve the problem of the high missed detection rate of objects with serious occlusion and small scale.

Anchor-Free detection is the same as Anchor-Based [21] detection in that both are based on the feature points on the training graph to construct the learned target and the construction process back-calculates the sensory field by the feature points. The difference is that Anchor-Free is an endogenous perspective, i.e., it starts from the properties of the feature points themselves to generate the prediction frame without hyperparameters, so the code complexity is relatively low, and a single feature point responds to a target object, which can directly predict the boundary of the object.

3.1 Model Analysis

The overall structure of the YOLOX model consists of four parts: the input, the backbone network, the neck network, and the prediction. In the input part, methods such as Mosaic [22] data augmentation are used. Deep learning uses the CSPDarknet feature extraction network, which contains the Residual Network, which can be used to improve the accuracy of the algorithm by increasing the number of network layers. The residual convolution is divided into two parts, the backbone part contains two convolutional layers of 1×1 and 3×3, the residual edge directly combines the input and output of the backbone, while the internal residual block uses jump links to alleviate the problem of gradient disappearance caused by increasing depth in deep neural networks with. The SPP structure is used to improve the perceptual field of the network by maximizing the pooling of the same pooling kernel size for feature extraction.

The backbone of the algorithm uses the Focus network structure as shown in Fig. 2.

Compared with the normal convolutional approach of neural networks, Focus sampling does not cause information loss and replaces three normal down-sampling convolutional layers at a time, reducing the number of parameters, Cuda memory, and increasing the speed of forward and backward propagation. In the training process, Focus sampling first slices the image, i.e., every other pixel in a picture is taken, similar to proximity down-sampling, to obtain four independent feature layers, and then the four independent

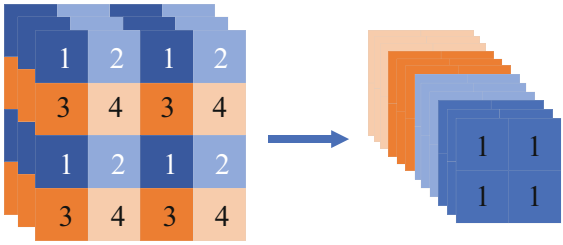


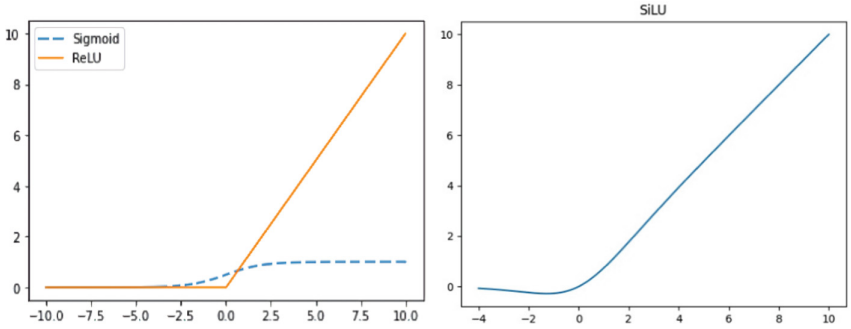
Fig. 2. Principle of focus slicing

feature layers are stacked, making the width and height information concentrated into channel information, thus expanding the input channel four times, and the feature layer becomes twelve channels from the original three channels.

The improvement of the main function of the model, the ReLU activation function, is shown in Fig. 3, by the transformation of Eq. (1).

$$f(x) = x \cdot sigmoid(x) \tag{1}$$

where is the ReLU function. The Sigmoid activation function is combined with the ReLU function to produce the SiLU function, as shown in Fig. 3(b), with no upper bound with a lower bound, smooth, and non-monotonic, i.e., the ReLU function is smoothed.



(a) Origin function (b) Improved function

Fig. 3. Function image of SiLU

The YOLOX algorithm extracts three feature layers for target detection in the feature utilization part, which are located in the middle, lower middle, and bottom layers of CSPDarknet. After obtaining the three effective feature layers, a new enhanced feature layer is obtained by stacking after convolution, up-sampling, and down-sampling.

Since the output channel classification task and regression task are put together, there is a conflict between the two tasks, YOLOX adopts decoupled head to replace the coupled head, which has a better expression effect and the model accuracy will be improved, while the convergence speed of the network is accelerated. The decoupled head network decomposition is shown in Fig. 4.

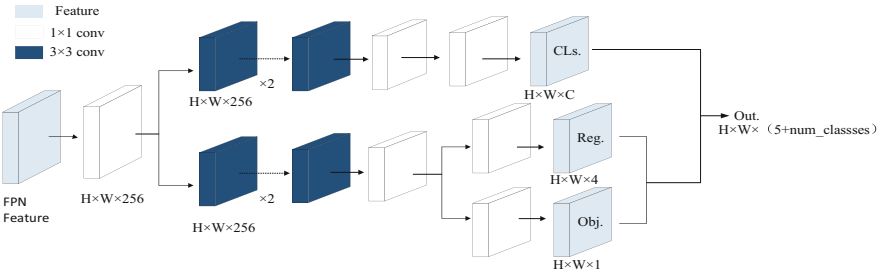


Fig. 4. Decoupled head network decomposition.

FPN Feature denotes the FPN feature layer, and H and W in the figure represent the height (height) and width (width) of the feature map; Reg(H, W, 4) is used to judge the regression parameters of each feature point, and the regression parameters are adjusted to obtain the prediction frame; Obj(H, W, 1) is used to judge whether each feature point contains objects; Cls(H, W, 4 + 1 + num_classes) the first four parameters of the third item are used to judge the regression parameters of each feature point, and the prediction box is obtained after adjustment, the fifth parameter is used to judge whether each feature point contains objects, and finally num_classes is used to judge the kinds of objects contained in the feature points.

The two parallel branches of the decoupling head have two 3×3 convolutional layers for classification and regression tasks respectively. The FPN feature layer extracts better features by fusing feature layers of different morphologies, after which these features are passed into the Yolo Head to obtain prediction results.

3.2 Data Enhancement with Model-Based Channel Cropping Improvements

To improve the accuracy of small target fault recognition and enhance the speed and efficiency of model recognition, we can enhance the learning of target features in images and reduce the redundancy of the network by enhancing the data and optimizing the model channels to obtain a lightweight model while enhancing the target weight mapping.

Image Data Enhancement

In terms of data enhancement, to improve the robustness of the network and reduce the impact of additional factors on recognition, the images are enhanced at the input stage by distorting the color gamut, flipping the images, adding gray bars, etc., and when enhancing the data, both the enhanced images and the positions of the distorted frames are taken into account.

Using stitching the pictures to achieve data enhancement, the Mosaic image enhancement (Fig. 5), operation process will be a non-equivalent stitching combination of four photos, to a certain extent to enrich the background of the detected objects, but in the process of use, Mosaic data enhancement part of the operation will bring the inaccurate annotation box. Therefore, to reduce the redundant boxes generated by the data enhancement operation, this experiment performs mosaic data enhancement on the first 90% of epochs for image data.

Fig. 5. Mosaic picture enhancement effect

Channel Pruning

To obtain a lightweight model suitable for fast detection, this study uses channel pruning to structurally simplify the convolution module of the algorithm. The structural simplification mainly involves tensor decomposition, sparse connectivity, and channel pruning. The parameters of the YOLOX convolutional module are shown in Table 1.

Table 1. Structural parameters of the YOLOX multilayer convolution module.

Convolution	Number	Output feature map	Parameters
Convolution Module	Number of layers	Output feature map size number of channels × length × width	Parameters (convolutional layer size)
Conv2d	1	32, 320, 320	3456
	2	64, 160, 160	22528
	4	32, 160, 160	14336
	8	128, 80, 80	729088
	12	64, 80, 80	212992
	4	256, 40, 40	491520
	23	128, 40, 40	1884160
	4	512, 20, 20	2228224
	11	256, 20, 20	2686976
BatchNorm2d	1	32, 320, 320	64
	2	64, 160, 160	256
	4	32, 160, 160	256
	8	128, 80, 80	2048
	12	128, 40, 40	1536
	5	128, 20, 20	1280
	4	64, 80, 80	2048
	23	256, 40, 40	5888
	4	512, 20, 20	4096
	11	256, 20, 20	5632

For pruning the feature map with c channels, a convolutional filter W of $n \times c \times k_h \times k_u$ is considered to be applied to the mapping X with an input volume of $N \times c \times k_h \times k_w$, resulting in an output matrix Y of $N \times n$. Where N is the number of samples, n is the number of output channels, and k_h, k_w is the size of the convolutional kernel. When performing the pruning of the input channel c to the desired $c_0 (0 \leqslant c_0 \leqslant c)$ value, the bias term is discarded to make the representation simpler while minimizing the reconstruction error, and the representation is shown in Eq. (2).

$$l_0 = \underset{\beta,\ W}{arg\ \min} \frac{1}{2N} \left\| \sum_{i=1}^{c} \beta_i X_i W_i^T - Y \right\|_F^2 \tag{2}$$

where $\|\beta\|_0 \leqslant c'$; $\|\cdot\|_F$ is the Frobenius norm, X_i is the $N \times k_h k_w$ matrix of the channel slice of the ith input mapping X, $i = 1, \cdots, c$; W_i is the $N \times k_h k_w$ filter weight cut from the ith channel of W; β is a coefficient vector of length c for channel selection, and β_i (the ith element of β) is the scalar mask of the ith channel (i.e., whether to discard the entire channel), when $\beta_i = 0$, X_i and W_i can be safely clipped from the feature map. c' is the number of reserved channels, which can be calculated from the desired acceleration ratio and set manually. For overall model acceleration, the acceleration ratio is first assigned to each layer and then calculated for each c'.

To solve the NP puzzle in the optimization equation, the relaxation from to is regularized as shown in Eq. (3).

$$l_1 = \underset{\beta, W}{arg\,\min} \frac{1}{2N} \left\| \sum_{i=1}^{c} \beta_i X_i W_i^{\top} - Y \right\|_F^2 + \lambda \|\beta\|_1, \|\beta\|_0 \leqslant c', \quad \forall_i \|W_i\|_F = 1 \tag{3}$$

where λ is the penalty factor, by increasing the value of λ, there will be more zero terms in β and a higher acceleration ratio can be obtained. Adding the constraint $\forall_i \|W_i\|_F = 1$ to Eq. (3) reduces the computational redundancy caused by the F-parameter of the over-range W_i.

Now perform the channel selection by fixing W and solving for it; secondly, fix the value of and solve for W to reconstruct the error.

Step 1: Fix the W solution, and solve the channel selection problem using LASSO regression as shown in Eq. (4).

$$\hat{\beta}^{LASSO}(\lambda) = \underset{\beta}{arg\,\min} \frac{1}{2N} \left\| \sum_{i=1}^{c} \beta_i Z_i - Y \right\|_F^2 + \lambda \|\beta\|_1, \|\beta\|_0 \leqslant c' \tag{4}$$

where $Z_i = X_i W_i^{\top}$, the ith channel that makes $\beta_i = 0$ is ignored.

Step 2: Fix the solution W and use the selected minimum channel solution W to reconstruct the error as shown in Eq. (5).

$$\underset{W'}{arg\,\min} \left\| Y - X'(W')^{\mathrm{T}} \right\|_F^2 \tag{5}$$

where $X' = [\beta_1 X_1 \beta_2 X_2 \cdots \beta_i X_i \cdots \beta_c X_c]_{N \times ck_h k_w}$, W' is the W that has been reconstructed by $n \times ck_h k_w$, $W' = [W_1 W_2 \cdots W_i \cdots W_c]$. After deriving W', reformulate it to W and then specify $\beta_i \leftarrow \beta_i \|W_i\|_F$, $W_i \leftarrow W_i / \|W_i\|_F$ so that it satisfies the constraint $\forall_i \|W_i\|_F = 1$.

In practice, the repetition of the two steps is time-consuming, and multiple iterations of step one are used until step two is used once after $\|\beta\|_0 \leqslant c'$ is satisfied to obtain the final result.

For channel pruning of the overall model, the above steps are used sequentially, layer by layer. For each layer, the input volume is obtained from the current input feature map and the output volume is obtained from the output feature map of the unpruned model as shown in Eq. (6).

$$\underset{\beta,\ W}{arg\,\min} \frac{1}{2N} \left\| \sum_{i=1}^{c} \beta_i X_i, W_i^{\top} - Y' \right\|_F^2, \ \|\beta\|_0 \leq c' \tag{6}$$

Considering the cumulative error during sequential pruning, Y' of the feature map from the source model is used instead of Y in Eq. (6).

3.3 Multi-angle Image Fusion

To address the problem of feature loss in the recognition of small target faults in images, this experiment uses CutMix for image fusion of multi-angle images with different kinds of faults to enhance the learning of fault features by neural networks. CutMix uses the complete target detection object as a classification label and selectively combines blocks between training images to maintain the regularization effect of region loss by effectively using training pixels. The merging operation of CutMix is Eq. (7) and Eq. (8).

$$\overline{x} = M \odot x_A + (1 - M) \odot x_B \tag{7}$$

$$\overline{y} = \lambda y_A + (1 - \lambda) y_B \tag{8}$$

where $M \in \{0, 1\}W \times H$, denotes the binary mask representing the locations of deletion and fill in the two images, sampled from the bounding box coordinates of the image cropping region; and $\lambda \in (0, 1)$, sampled uniformly from the range of values. A new training sample $(\overline{x}, \overline{y})$ is generated by combining two training samples (x_A, y_A), (x_B, y_B) for training the model of the original loss function.

After the open pin image is fused by CutMix, the areas of different parts are proportionally blended as shown in Fig. 6, and the open pin image is fused with the side image to form a new training image.

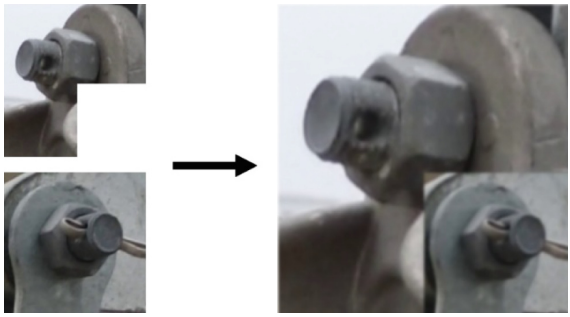


Fig. 6. CutMix Multi-Angle Image Fusion

4 Training Results and Prediction Analysis

The hardware platform used for training this model is Intel Xeon Gold 5217 CPU and NVIDIA Quadro P2200 GPU. The simulation environment is a deep learning virtual environment based on Anaconda with python version 3.6.

4.1 Data Pre-processing and Training Parameters Setting

Before the training, the UAV inspection video of transmission lines in an area of Changsha was selected for the extraction of training images, with a total of 2550 photos, and

the data set contained "cotter pin", "cotter pin off", "insulator broken" To reduce the large loss value caused by the unbalanced category, some of the higher quality images in each fault image were filtered out to maintain a balanced number of samples. The labelImg is used to calibrate the images, and the calibrated images will generate a file in XML format, which contains the fault categories as well as the coordinates of the calibration box. The training set and validation set are generated by the code. To improve the training effect and network learning efficiency, the ratio of the training set and validation set is set to 9:1, and the input images are uniformly processed to 640 × 640 pixels to complete the placement of the dataset.

The data set was increased and supplemented during the second training, with the addition of broken wires and anti-vibration hammers, and the data set contained 3190 UAV patrol images, with some of the new images containing 2 or more types of faults within one image. When training the new dataset, the initial training weight file was used to train on the original architecture to optimize the network weight parameters for the fault features.

Training on the dataset is divided into two phases, the freezing phase, and the thawing phase. Freezing the training can speed up the training due to the common features of the backbone feature extraction network, and also prevent the weights from being destroyed in the early stage of training. In the freezing phase, the backbone of the model is frozen so that the feature extraction network does not change, so that the memory occupied during training is small and only the network is fine-tuned, and the learning rate is set to $1e^{-3}$. For the freezing phase, the backbone of the model is not frozen, so the feature extraction network changes and the memory occupied is large, the backbone parameters of the network also change, and the learning rate is set to $1e^{-4}$.

4.2 Training Results and Model Performance Analysis

The main metrics used to evaluate the target detection algorithm are the speed of recognition as well as the accuracy, and other evaluation metrics developed on this basis. To evaluate the performance of the model after its improvement, the following target detection metrics are used to evaluate the model: Precision (accuracy), which represents the proportion of true positive samples among all samples judged positive; Recall, which represents the proportion of all actual positive samples correctly judged by the algorithm; AP (area under the P-R curve for each category of targets), mAP (Mean Average Precision) is the average value of AP for each category. The formulas for precision and recall are shown in Eq. (9) and Eq. (10).

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

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

where TP (True Positives) are positive samples correctly classified; FP (False Positives) are negative samples incorrectly classified; FN (False Negatives) are positive samples incorrectly assigned, and the sample distribution is as shown in Fig. 7.

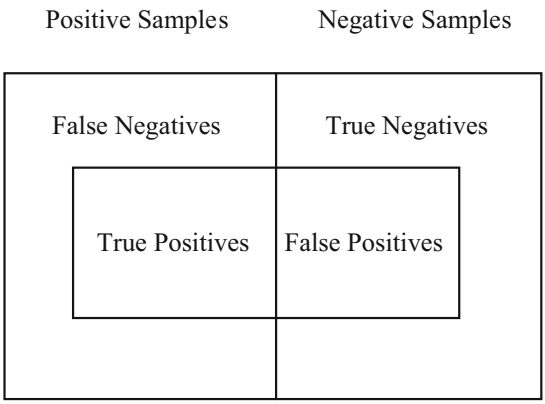


Fig. 7. Sample distribution diagram

Image Data Enhancement

To test the effectiveness of CutMix on the VOC dataset, the YOLOX algorithm in this experiment and the SSD and Faster RCNN target detection algorithms using the backbone network changed to ResNet-50 were considered as controls. The UAV patrol dataset in Pascal VOC format was tested and used mAP values as model evaluation metrics as shown in Table 2. From the table, it can be seen that Mixup and Cutout differ greatly from each model fusion, and both have insignificant and decreasing mAP improvement when combined with YOLOX model, and CutMix performs better in the algorithm adaptation. The combination of CutMix image fusion with the YOLOX algorithm for UAV patrol dataset images can bring greater improvement in target detection performance, with a 1.3% increase in mAP value.

Table 2. Comparison of the effect of CutMix applied to different models

Models	Base (mAP)	Cutout (mAP)	Mixup (mAP)	CutMix (mAP)
SSD	76.7	76.8 (+0.1)	76.6 (−0.1)	77.6 (+0.9)
FasterRCNN	75.6	75.0 (−0.6)	73.9 (−1.7)	76.7 (+1.1)
YOLOX	77.4	77.1 (−0.3)	77.6 (+0.2)	78.7 (+1.3)

Comparison of Channel Pruning Results

In this experiment, the model is channel pruned to achieve structural simplification, and the number of channels is pruned using 1×1 convolution. The parameters of the convolution module after feature channel pruning are shown in Table 3, and the channel reduction rate reaches 50%. Through channel pruning, the redundant nodes in the channels are removed to sparse the weights of the model, and the size of the neural network parameters of the algorithm decreases significantly, saving the memory

when the algorithm is running, thus further reducing the computation during the model operation and reducing the running time.

Table 3. Comparison of channel pruning results

Models	Total Params (trainable)	Forward/backward pass size (MB)	Params size (MB)
YOLOX	25326495	1907.40	96.61
YOLOX+pruning	8968255	1020.29	34.21

Improved YOLOX Evaluation Comparison

The different types of accuracy values output by the evaluation code after training are shown in Fig. 8. The figure shows the accuracies obtained for the prediction frames with a suppression probability of 0.5 or less when the confidence value is set to 0.5. The target detection accuracies are 94.44% for anti-vibration hammer, 95.24% for the cotter pin, 100% for insulator breakage cotter pin drop, 82.05% for wire breakage, and 72% for wire loose strand.

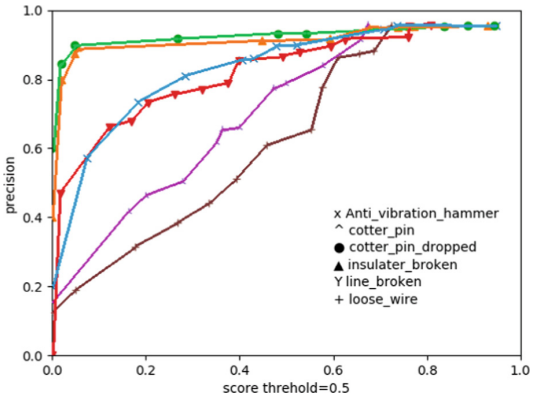


Fig. 8. Different target characteristics recognition accuracy

In the experiments, the loss value images trained with the modified YOLOX model are shown in Fig. 9. Loss represents the loss value of the training set, which is used in the network to update the network parameters; val_loss represents the loss value of the validation set, which is only used for validation, and the output prediction frames of different feature layers are mapped back to the original image for loss calculation.

A generic weight file was added as a pre-training model at the beginning of training, and the backbone network was frozen to prevent the phenomenon that the feature extraction was not effective due to too few random weights during training. In the training to about the 50th time, due to the unfrozen backbone network can be found that the loss

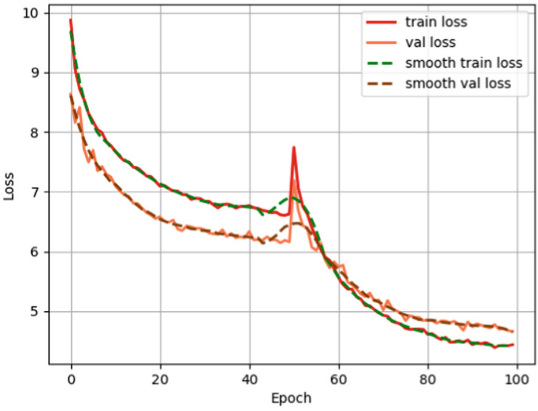


Fig. 9. Training loss value iterative image

function image after the oscillation, the loss value further reduced until convergence, the improved network loss value and val_loss.

This experiment uses CutMix for multi-angle image fusion for data enhancement for grid UAV patrol images, and YOLOX+pruning uses channel pruning to improve the feature channels based on the original model with the same network backbone part.

Table 4 lists the AP enhancement values in, letters in the table correspond to the following target types and models: A: Line broken, B: Insulater broken, C: Cotter pin dropped, D: Cotter pin, E: Anti vibration hammer, F: Loose wire, X: YOLOX, Y: YOLOX+Channel pruning, Z: YOLOX+Channel pruning+Cutmix. It can be seen that the overall decrease in detection accuracy of the algorithm after adding channel pruning is 0.86%, which is due to the temporary decrease in generalization performance caused by channel pruning, and after adding CutMix, the AP of different fault types detection results of the enhanced model are improved to different degrees.

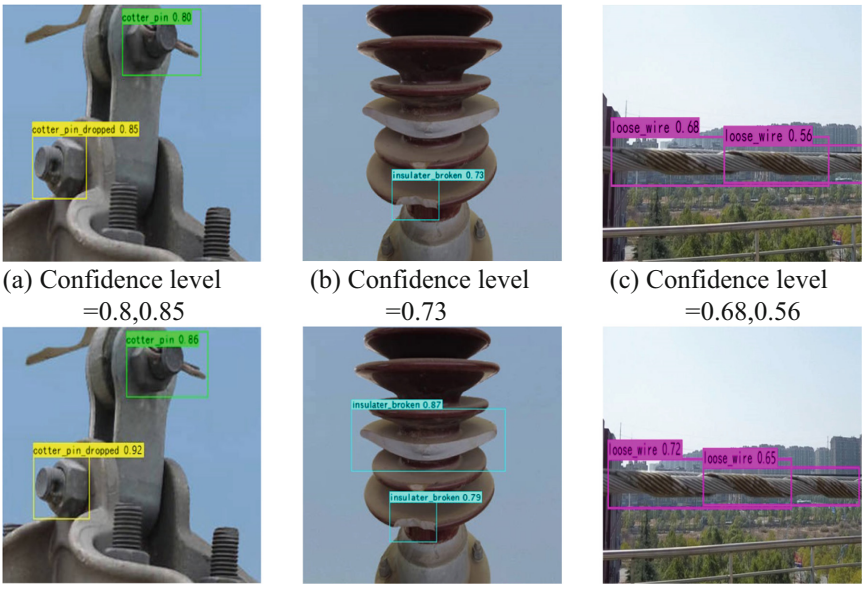
Table 4. Comparison of AP after algorithm improvement

Models	AP/%					
	A	B	C	D	E	F
X	59	71.5	94.82	96.2	89.52	97.83
Y	58.8	70.35 (−1.15)	93.64 (−1.28)	94.87	88.51 (−1.01)	97.73 (−0.1)
Z	62.3	81.47 (+9.97)	96.4 (+1.58)	100	90.89 (+1.47)	98.11 (+0.28)

4.3 Detection Results

The comparison of the results of the improved YOLOX algorithm and the improved target detection model for UAV patrol image detection is shown in Fig. 10, and the

labels contain the categories and confidence levels of the detected targets. Figure 10(a), Fig. 10(b), Fig. 10(c) shows the detection results before the improved model; Fig. 10(d), Fig. 10(e), and Fig. 10(f) show the detection results after the improvement.



(a) Confidence level =0.8,0.85 (b) Confidence level =0.73 (c) Confidence level =0.68,0.56

(d) Confidence level =0.86,0.92 (e) Confidence level =0.87,0.79 (f) Confidence level =0.72,0.65

Fig. 10. Comparison of model target detection results.

Table 5 lists the comparison of the target detection results, it can be seen that compared with the YOLOX algorithm before the improvement, the improved input and channel pruning improved the target detection accuracy in complex backgrounds with a 3.31% increase in accuracy, a 1.106% decrease in the missed detection rate, and a 2.569 increase in the FPS value of the number of images detected per second by the network.

Table 5. Comparison of algorithm

Models	Precision/%	Recall/%	Miss Rate/%	mAP/%	FPS
YOLOX	87.52	87.4	5.327	84.817	39.978
YOLOX+Channel pruning	88.746	85.7	5.365	85.772	42.323
YOLOX+Channel pruning+CutMix	90.83	89.3	4.221	88.195	42.547

5 Conclusions

1. In this study, an improved YOLOX target detection network is used to solve the problems of low recognition rate, slow speed, and large leakage rate of some fault types in power grid inspection images. Based on the feature loss characteristics of small and medium target faults in transmission lines, multi-scale sampling and feature fusion are used to improve; for the fault type data imbalance problem, the wire breakage is further refined into wire scattering and wire breakage faults, and the labels are increased proportionally to solve the data set category imbalance problem.
2. The improved form of channel pruning is used for the algorithm feature acquisition channel, and the experimental results show that the improved YOLOX algorithm reduces the feature channel parameter scale by 50% and increases the FPS by 2.569. The improved algorithm uses CutMix for data improvement and enhancement, which fuses the images of different angles of small target faults while having some improvement in accuracy, and the mAP value increases by 3.378%.
3. The experimental results show that the improved model can improve the target detection speed in small target fault detection in power system, and can be better applied in target detection and power robot overhaul in power system.

References

1. Chen, Z., Bo, W., Le, D., et al.: Application of infrared temperature measurement technology in power system. Electr. Eng. (2017)
2. Tang, L., Fang, L., Wang, H.: Development of an inspection robot control system for 500KV extra-high voltage power transmission lines. In: SICE Conference. IEEE (2005)
3. Dobakhshari, A.S., Ranjbar, A.M.: A circuit approach to fault diagnosis in power systems by wide area measurement system. Int. Trans. Electr. Energ. Syst. **23**, 1272–1288 (2013)
4. Xiang, D.: Application of infrared thermal imaging UAV on board in new energy generation equipment. Heilongjiang Sci. **13**(04), 68–69 (2022)
5. Li, X., Su, H., Liu, G.: Insulator defect recognition based on global detection and local segmentation. IEEE Access **PP**(99), 1 (2020)
6. Zhou, J.Y., Wu, X.P., Zhang, C., et al.: A moving object detection method based on sliding window Gaussian mixture model. J. Electron. Inf. Technol. **35**(7), 1650–1656 (2013)
7. Liu, S., Qi, L., Qin, H., et al.: Path aggregation network for in-stance segmentation. IEEE (2018)
8. Wei, H., Zhang, K., Zheng, L.: Infrared image object detection of power inspection based on HOG-RCNN. Infrared Laser Eng. **49**(S2), 242–247(2020)
9. Dang, L., Bui, B., Vo, P.D., et al.: Improved HOG descriptors. In: Third International Conference on Knowledge & Systems Engineering. IEEE Computer Society (2011)
10. Zhao, K., Zhu, M., Yang, X., et al.: A new method of creating minimal-order Markov set and transition states of M/N sliding window. IEEE Access **PP**(99) 1 (2019)
11. Pan, R., Sun, W.: Deep learning target detection based on pre-segmentation and regression. Opt. Precis. Eng. **25**, 221–227 (2017)
12. Jian, Y., David, Z., Frangi, A.F., et al.: Two-dimensional PCA: a new approach to appearance-based face representation and recognition. IEEE Trans. Pattern Anal. Mach. Intell. **26**(1(1)), 131–137 (2004)

13. Liu, X.: Research on image recognition algorithm for broken strands and damage faults of transmission conductor. Xi'an Polytechnic University (2018)
14. Kang, G., Gao, S., Yu, L., et al.: Fault detection of missing split pins in swivel with clevis in high-speed railway catenary based on deep learning. J. China Rail. Soc. **42**(10), 45–51 (2020)
15. Bochkovskiy, A., Wang, C.Y., Liao, H.: YOLOv4: optimal speed and accuracy of object detection (2020)
16. Wang, H., Shao, Y., Zou, S., Ma, Z., Zhao, S.: Detection of cotter pins missing of connection fittings on transmission lines of power system. In: 2021 40th Chinese Control Conference (CCC), pp. 6873–6879 (2021). https://doi.org/10.23919/CCC52363.2021.9550162
17. Heng, Y., Tao, G., Ping, S., et al.: Anti-vibration hammer detection in UAV image, pp. 204–207 (2017)
18. Zhang, B., Xue, D., Liu, H.: Analysis of the effect of wind loads on the dynamic response of the suspended crossing-tower auxiliary system. IOP Conf. Ser. Earth Environ. Sci. **621**, 012020 (2021)
19. Ai, Z.: Research on overhead transmission lines abnormal detection algorithm based on UAV aerial image. Northeast Electric Power University (2021). https://doi.org/10.27008/d.cnki.gdbdc.2021.000046
20. Ge, Z., Liu, S., Wang, F., et al.: YOLOX: exceeding YOLO series in 2021 (2021)
21. Ma, W., Li, K., Wang, G.: Location-aware box reasoning for anchor-based single-shot object detection. IEEE Access **PP**(99),1 (2020)
22. Lv, H., Zhang, H., Zhao, C., et al.: An improved SURF in image mosaic based on deep learning. In: 2019 IEEE 4th International Conference on Image, Vision, and Computing (ICIVC). IEEE (2019)