

## Research Article

# Object Detection Algorithm of Transmission Lines Based on Improved YOLOv5 Framework

Hao Zhang , Xianjun Zhou , Yike Shi , Xuan Guo , and Hang Liu 

*School of Electrical and Electronic Engineering, Hubei University of Technology, Wuhan 430068, China*

Correspondence should be addressed to Xianjun Zhou; [xjzhou@hbut.edu.cn](mailto:xjzhou@hbut.edu.cn)

Received 13 September 2023; Revised 24 October 2023; Accepted 25 November 2023; Published 9 January 2024

Academic Editor: Jin Liu

Copyright © 2024 Hao Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Foreign objects easily attach to the transmission lines because of the various laying methods and the complex, changing environment. They have a significant impact on the safe operation capability of transmission lines if these foreign objects are not detected and removed in time. An improved YOLOv5 technique is provided to detect foreign objects in transmission lines due to the low-foreign object recognition accuracy image detection. The method first reduces the computation and memory consumption by introducing the RepConv structure, further improves the detection accuracy and speed of the model by embedding the C2F structure. This method finally is further optimized neural network by the Meta-ACON activation function. The results indicate that the average detection accuracy of the improved YOLOv5 network can reach 96.9%, which is 2.2% higher than before. Additionally, corresponding detection speed can reach 258.36 frames/second, which surpasses existing mainstream target detection models, performing better in terms of the balance of inference speed and detection accuracy. Consequently, the effectiveness and superiority of the algorithm have been proved.

## 1. Introduction

Grid transmission lines serve as carriers for the transmission of the electricity we use every day [1]. How to ensure that the transmission lines are safe and stable for power transmission is a necessary condition for the safe and effective operation of the power grid. Given the significant data statistics, the following kinds of foreign objects often appear on the power grid: bird nests, kites, balloons, and garbage. The above objects hanging on the transmission lines or towers can quickly short-circuit or single-phase faults between transmission lines, leading to all kinds of short-circuit accidents, severe fires, and widespread power outages, which brings serious economic losses [2, 3]. The chain reaction caused by short circuits also threatens the lives and property of people living around the transmission line [4]. It dramatically threatens the lives of the maintenance personnel who come to repair the power grid. Transmission lines often span a variety of complex landscapes. Moreover, the places they pass through are generally sparsely populated and difficult to access. On this basis, intelligent inspection technology [5–8] was developed to inspect transmission lines through aerial

photography by UAVs, which can save a lot of human resources and material resources with high-detection efficiency compared with the manual labor. However, the data from aerial photography still need to be distinguished by human judgement, so the detection efficiency and accuracy of the whole process still need to be improved.

As the GPU computing power continues to increase, deep learning is gradually showing its advantages in various fields of computer vision. Since 2014, deep learning-based target detection networks erupted, starting with two-stage networks, such as R-CNN [9], Fast-RCNN [10], Mask-RCNN [11], Faster R-CNN [12], which have the advantages of high-detection accuracy and low-leakage rate. However, the detection speed is slow. The calculation is relatively complicated and challenging for the transmission line foreign object detection and prevention work. Single-stage detection algorithms emerged then to integrate the feature extraction of candidate regions with the positioning of prediction frames and directly perform the judgment of target categories and the positioning of detection frames [13–17].

In 2020, YOLOv5 [18] was introduced, shocking the world with its extremely fast detection speed, making it an

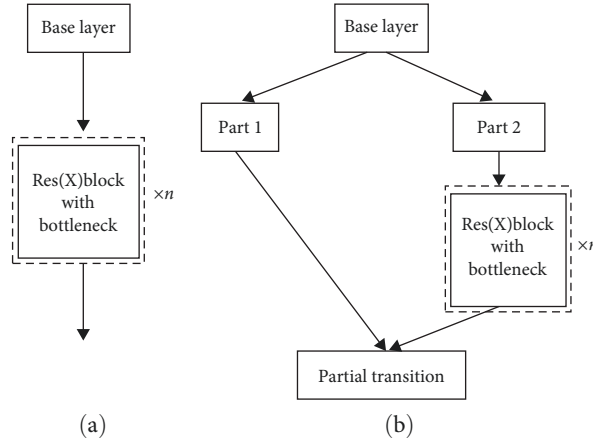


FIGURE 1: CSP module: (a) ResNe(X)t and (b) CSPResNe(X)t.

ideal candidate for real-time conditions and mobile deployment environments. The related studies [19–24] have made lightweight improvements to the YOLOv5s version based on different domains. Without modifying the subject feature extraction network, they enhanced feature extraction by improving the feature pyramid network (FPN) [25–27] in various ways, resulting in some improvement in the accuracy of target recognition. The YOLOv5s regression, however, needs more precision. Although using more profound levels of YOLOv5m [28], YOLOv5l [29], and YOLOv5x [30] have better MAP (mean average precision) [31] compared to YOLOv5s. The amount of data in the model will be more as well as the demand of the algorithm for the hardware will be higher. The use of YOLOv5-Lite version has a faster FPS and can be more easily deployed on more platforms. However, its accuracy mAP has a particular gap compared to more complex network models. Therefore, it is challenging to meet scenarios with high requirements for real-time and target frame regression accuracy. Based on YOLOv5, this paper proposes a target detection model that is lightweight and has a faster detection speed in order to better tradeoff speed and accuracy and to make the YOLO model better applied to the transmission line foreign object detection task, with the following main contributions:

- (1) The RepConv [32–35] structure is introduced, which reduces the computation and memory usage by sharing parameters and adding convolutional layers and improves the target inference speed while guaranteeing the existing target recognition accuracy;
- (2) The C2F [36, 37] structure is used to improve the semantic representation of features, thus enhancing the feature extraction capability of the network and thus facilitating the model to detect small targets more accurately;
- (3) The Meta-ACON [38, 39] function replaces the traditional activation function. The Meta-ACON activation function adaptively adjusts the activation function parameters, thus improving the model's generalization ability and making the model more practical.

## 2. YOLOv5 Network

YOLOv5 is a deep convolutional neural network-based target detection algorithm. Its network structure mainly consists of three parts: backbone network, feature extraction network, and prediction network. Among them, the backbone network uses CSPDarknet53 as the backbone, which can effectively extract image features. The feature extraction network uses feature maps of different scales for effective target detection, while the SPP structure can improve the detection accuracy without increasing the computational effort. The prediction network consists of a series of convolutional and prediction layers to achieve target classification, position, and scale prediction. YOLOv5 balances speed and accuracy by optimizing and improving the network structure, making it an efficient and accurate target detection algorithm.

**2.1. Backbone Network.** The backbone network of YOLOv5 uses a combination of CSP and SPP modules, which can improve detection performance and efficiency. The CSP and SPP modules are described in detail below.

**2.1.1. CSP Module.** The CSP module is the core module of the YOLOv5 backbone network, which can reduce the computational complexity and memory consumption of the model and improve its accuracy of the model. The structure of the CSP module is shown in Figure 1.

The CSP module divides the input feature map into two parts: a branching part and a connecting part. The branching part comes to perform the convolution operation, and the connecting part connects directly to the output. Finally, the feature maps of these two parts are connected. The branching part usually uses multiple convolutional layers to extract features. Moreover, these convolutional layers use different sizes and numbers of convolutional kernels to extract feature information step by step. The connected part usually uses  $1 \times 1$  convolutional layers for feature dimension transformation to ensure that the feature dimensions of the branch, and connected parts are the same. This module can significantly

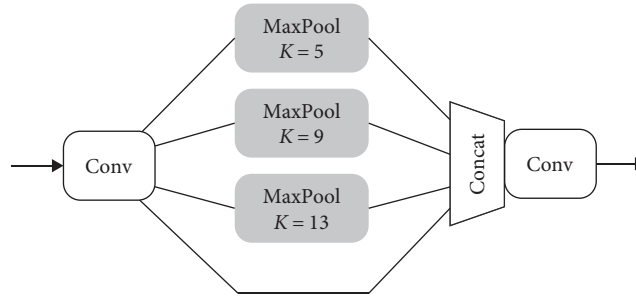


FIGURE 2: SPP structure.

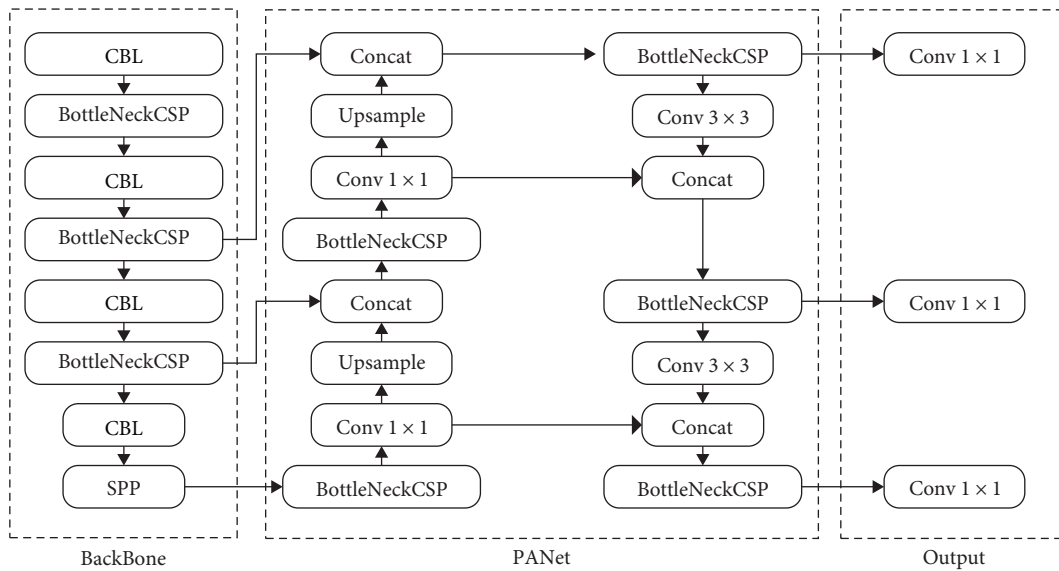


FIGURE 3: YOLOv5 network structure diagram.

reduce computation and memory consumption while maintaining high-detection accuracy.

**2.1.2. SPP Module.** SPP module is a spatial pyramid pooling module that can pool the feature maps at different scales to obtain feature information at different scales. The structure of the SPP module is shown in Figure 2.

The SPP module extracts feature information at multiple scales from the feature map by pooling operations of different sizes. These feature information are then concatenated together to capture target information of various scales as input for the next network layer. The resolution of the feature map will gradually decrease after continuous convolutional pooling, but the number of channels will gradually increase. This method enhances the receptive field of the network, which can capture a broader range of the target information. Eventually, the feature maps are passed to the prediction network to predict the location and class of the target.

The input image is passed through a convolutional layer in the backbone network to get the feature map. A series of CSP modules then are stacked to form the BottleNeckCSP structure and SPP module to gradually extract the feature layer, which is shown in Figure 3.

**2.2. Feature Extraction Network.** In YOLOv5, the feature pyramid structure is used to enhance the feature extraction part to detect targets of different sizes better. The structure comprises multiple feature layers with different spatial resolutions and semantic information.

Precisely, the feature pyramid structure consists of the following components:

**Base feature layer:** features of the image are extracted using a backbone network (e.g., CSPDarknet53) and down-sampled through a series of convolutional and pooling layers.

**Upsampling module:** lower resolution feature maps usually lose detailed information about the target in the feature pyramid structure. Therefore, YOLOv5 uses an upsampling module to recover this detailed information. In the upsampling module, the accuracy of the detected target is improved by interpolating the lower resolution feature maps so that they have the same size as the higher resolution feature maps.

**Feature fusion:** YOLOv5 uses feature fusion to further improve the detection performance, where feature maps of the different resolutions are fused through a series of convolution operations to obtain more accurate detection results.

**Multiscale prediction:** YOLOv5 uses multiple feature layers for target detection. Each feature layer has a different resolutions and semantic information to detect targets of the

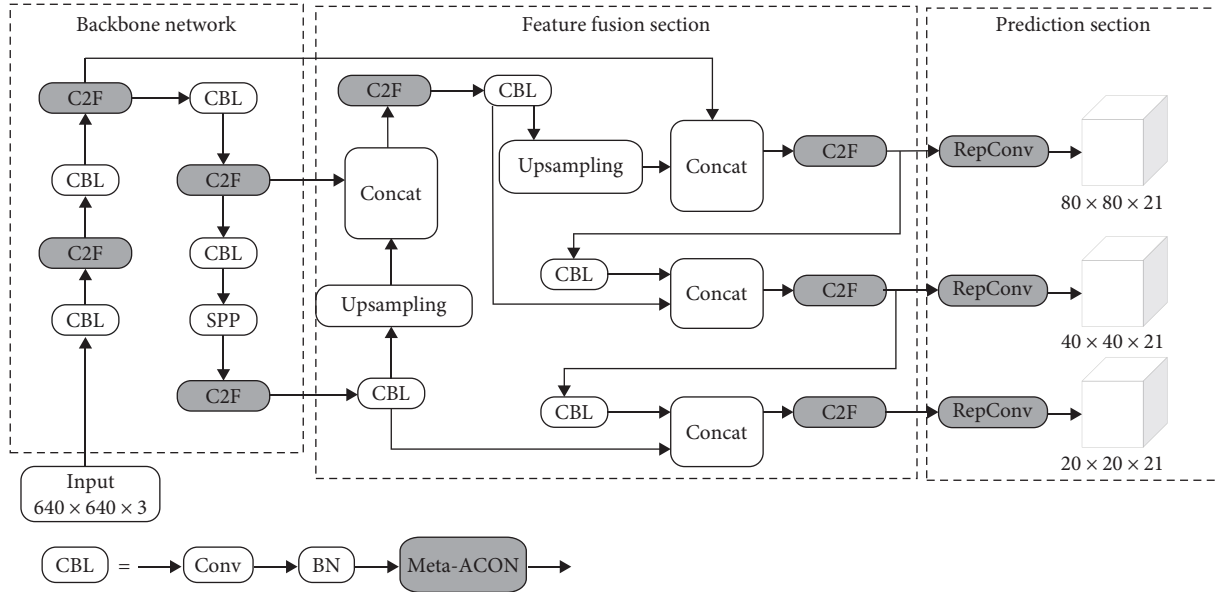


FIGURE 4: The improved YOLOv5 network.

different sizes. Specifically, each feature layer generates a set of prediction frames and then rejects the overlapping prediction frames by a nonmaximum suppression (NMS) algorithm to finally output the final detection results.

In general, the feature pyramid structure in YOLOv5 improves detection performance by using the fusion of multiple feature layers, especially when detecting targets of different sizes. This structure can improve detection accuracy and speed and has become one of the mainstream methods in target detection algorithms.

**2.3. Prediction Network.** In YOLOv5, the YOLO head is a network structure containing multiple convolutional and fully connected layers, which processes the feature extraction results of each feature map and outputs the detection results. The prediction process dividing into three steps: (1) feature mapping: in the YOLO head, the feature extraction results of each feature map need to be compressed for the output detection results. The output section of Figure 3 uses a  $1 \times 1$  convolution layer to compress the number of channels of each feature map to a certain number of values. (2) Prediction: in the YOLO head, the prediction is performed by a set of convolutional layers. These convolutional layers are responsible for predicting each target's class, position, and confidence scores. These scores are used in a subsequent NMS operation to filter out prediction frames with low-confidence scores. (3) Decoding: some decoding functions are used in the YOLO head to convert the predicted results into actual bounding box coordinates and category probabilities. The decoding process usually includes back-calculating the network output values, applying anchor boxes and offsets, and normalizing confidence scores using the sigmoid function.

In YOLOv5, detection results are obtained by using the YOLO head. By processing the feature extraction results of

each feature map, it outputs information such as the position, category, and confidence score of each detected object. It also utilizes decoding to convert this information into actual bounding box coordinates and category probabilities. This method allows YOLOv5 to maintain a high-detection speed while ensuring high-detection accuracy.

### 3. Improved YOLOv5 Network

This paper makes several improvements to the YOLOv5 network structure, including adding the C2F module, the RepConv module, and the Meta-ACON loss function. The C2F module can enhance the semantic correlation between different categories of targets for efficient feature extraction. The RepConv can reduce the computation and memory usage. Furthermore, the Meta-ACON loss function can effectively solve the gradient disappearance and gradient explosion problem, enhance nonlinear expression capability, and improve detection performance and robustness, thus making the target detection improve in both accuracy and efficiency. The improved YOLOv5 network structure is shown in Figure 4.

**3.1. RepConv.** RepConv is a convolutional neural network module based on repeated convolution, which aims to solve the computational complexity and memory consumption problems of convolutional operations in traditional convolutional neural networks. RepConv achieves the effect of reducing the computation and memory consumption by sharing parameters and adding convolutional layers.

In traditional convolutional neural networks, convolutional operations require much computation and memory. For instance, in a convolutional layer, for a  $3 \times 3$  convolutional kernel and 64 channels of input and 128 channels of output,  $64 \times 3 \times 3 \times 128 = 294,912$  multiplication and addition operations are required. The computation mentioned

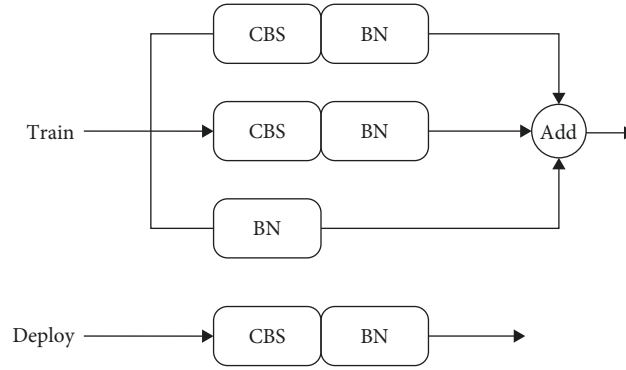


FIGURE 5: RepConv structure.

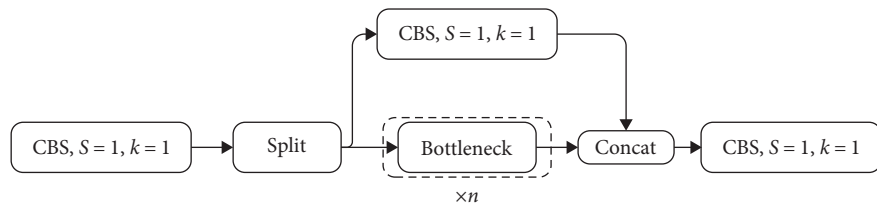


FIGURE 6: C3 structure.

above increases with the convolutional kernel size and number of input and output channels, leading to computational complexity and memory usage problems.

RepConv reduces computation and memory usage by adding convolutional layers and shared parameters. Specifically, RepConv decomposes the original convolutional layer into multiple smaller convolutional layers, each with smaller convolutional kernel size and the number of output channels than the original convolutional layer. Thus, the original convolutional operation that requires a large amount of computation can be turned into multiple small convolutional operations, reducing computation and memory usage.

Meanwhile, RepConv also adopts a shared parameter approach. In other words, multiple small convolutional layers share the same convolutional kernel, which further reduces the number of parameters and memory usage.

RepConv is somewhat different in training and inference. The training has the summed output of three branches. The deployment will reparametrize the parameters of the branches to the main branch, as shown in Figure 5.

**3.2. C2F Module.** The C2F module derives from the C3 module and the ELAN module, and the C3 module is shown in Figure 6. However, the C2F module adopts a more lightweight design scheme, as shown in Figure 7. This module includes two submodules: the Bottleneck module and the CBS module, where two CBS modules stack the Bottleneck module. The CBS module is stacked by Conv2d, Batch-Norm2d, and SiLU activation functions. The above modules can extract features efficiently and reduce computation and model size.

In addition to the convolutional layers, the C2F module also uses residual and jump connections to improve the

model performance and enable the network to learn the features better. The residual connection can avoid the problems of gradient disappearance and gradient explosion. In contrast, the jump connection can improve the semantic representation of the features, thus enhancing the feature extraction ability of the network. The above structures can help the network to learn the features better. Thus, it helps the network to perform better feature extraction and classification, which in turn improves the detection accuracy and speed. Additionally, the C2F module can also improve the network's robustness, enabling the network to perform well when processing various scenes and images.

**3.3. Improved Activation Function.** Meta-ACON is a nonlinear activation function that is based on the idea of adaptive normalization (AN). The formula is shown in Equation (1) as follows:

$$SM(x_1, x_2, \dots, x_n; \beta) = \frac{1}{\beta} \log \left( \sum_{i=1}^n e^{\beta x_i} \right), \quad (1)$$

where  $x_1, x_2, \dots, x_n$  is the input and  $\beta$  is a hyper parameter of "softened parameters." The adaptive switching between linear and nonlinear mappings is learned by defining a switching factor  $\beta$ .

Meta-ACON has the feature of adaptively deciding whether a neuron is activated. Therefore, unlike traditional activation such as ReLU, ACON allows each neuron to be activated or inactivated adaptively. This activation behavior contributes to the generalization and transmission performance of the neural network and avoids the phenomena of "gradient vanishing," "gradient explosion," and "neuron

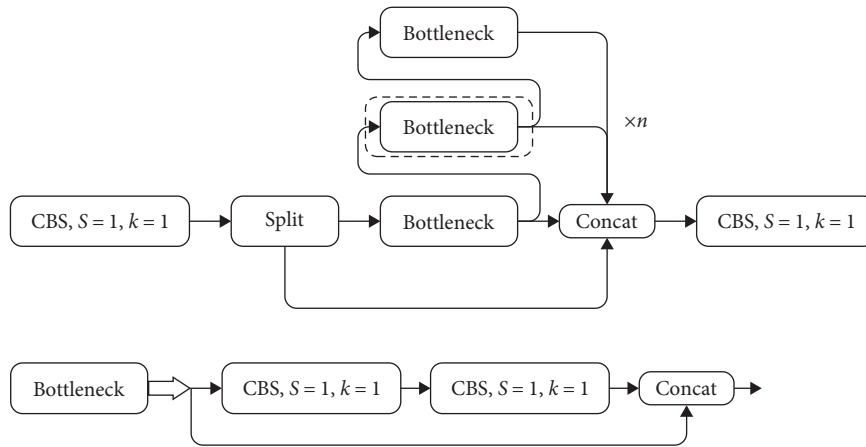


FIGURE 7: C2F structure.

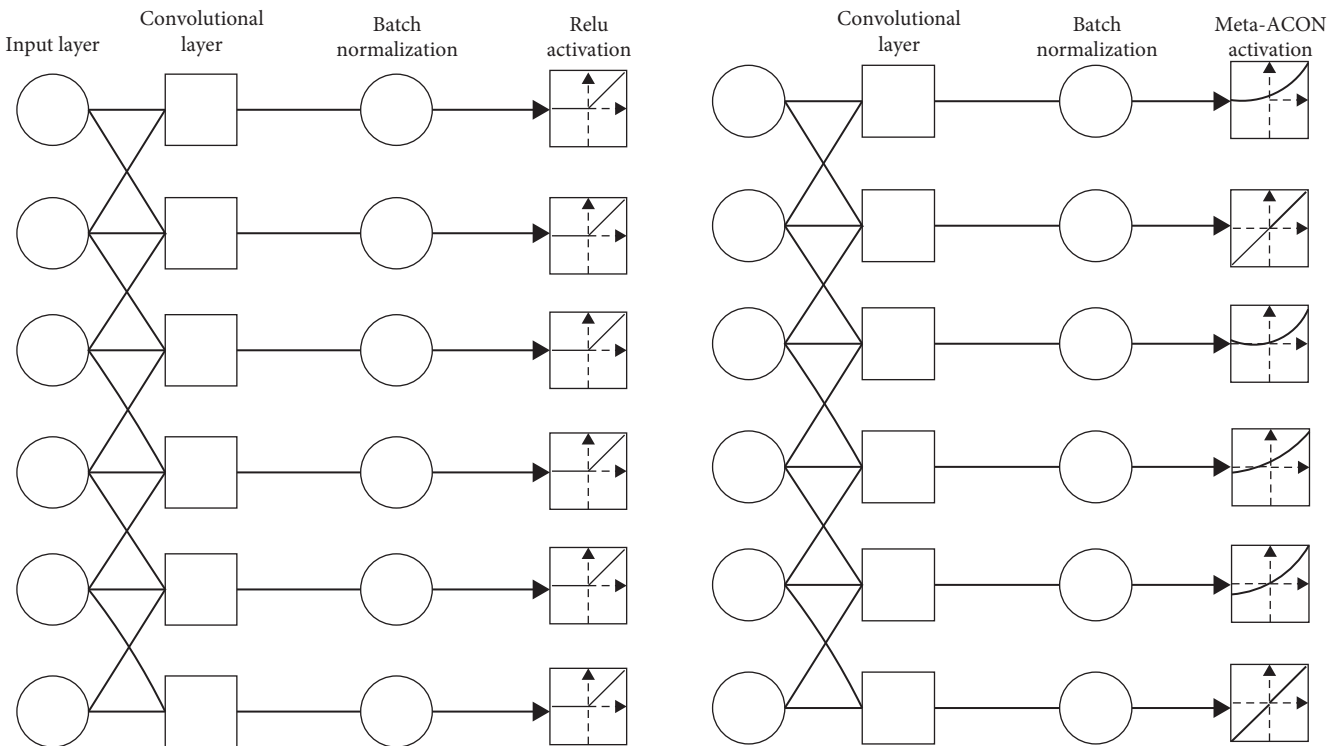


FIGURE 8: Schematic diagram of ReLU activation function and Meta-ACON activation function.

necrosis”. In this paper, we use this activation function to enhance the learning and generalization ability of the model and replace the ReLU activation function in deep convolutional neural networks, which is shown in Figure 8.

In summary, adding the Meta-ACON activation function to YOLOv5 effectively solve the gradient disappearance and gradient explosion problems, enhance the nonlinear expression capability, and improve the detection performance and robustness, thus making the network more powerful and effective.

#### 4. Transmission Line Foreign Object Detection Dataset

Since there is no publicly available data in the field of transmission line foreign object detection, this paper uses the aerial images of transmission lines collected in a project with the Suzhou Power Supply Company of the State Grid Corporation as the basis for constructing the original dataset of this paper, as shown in Figure 9.



FIGURE 9: Original dataset.

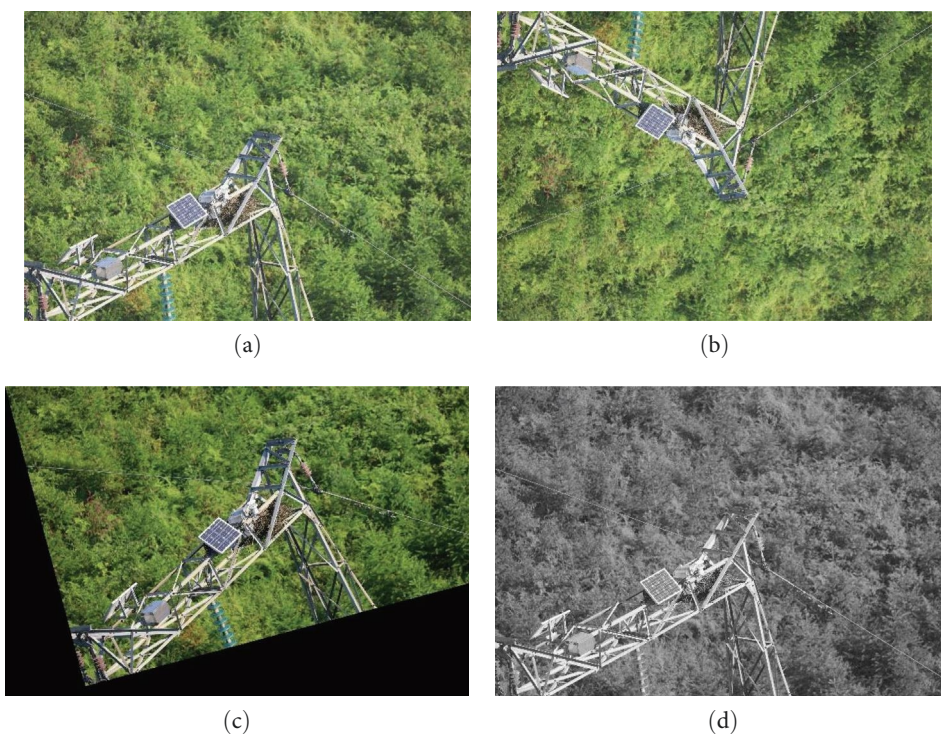


FIGURE 10: Expansion of transmission line foreign object dataset. (a) Original image, (b) rotate 180°, (c) 30° rotation, and (d) grayscale processing.

The tens of thousands of transmission line pictures collected were screened. The blurred images were excluded, and the two-dimensional target pictures, with precise shooting and no evident compression traces, were selected as far as possible. They are screening a total of 1,564 pictures of everyday foreign objects. The deep learning training process requires a large number of training samples. The more data used in the dataset, the higher the accuracy of the model detection after training. The number of selected images is still far from the amount of data in the above common dataset, so the Albumentation image enhancement techniques are used to expand the foreign

object images, including but not limited to rotation, scaling, transposition, contrast adjustment, brightness adjustment, grayscale adjustment, motion blur, and gridding, to ensure the training effect of the model. Figure 10 shows the enhancement process of the dataset.

## 5. Experimental Results and Analysis

*5.1. Experimental Environment.* The experimental environment used in this paper is a computer with an Intel Xeon Gold CPU and an NVIDIA GeForce TITAN RTX GPU with

32 GB of RAM. The operating system used in this paper is Ubuntu 20.04 LTS, Python version 3.8.5. PyTorch 1.7.1 is used as the deep learning framework. Furthermore, other commonly used Python libraries, such as NumPy, Pandas, and Matplotlib.

This paper uses a self-built transmission line foreign object dataset as the training and testing dataset. The common types of foreign objects on transmission lines: bird nests, balloons, kites, and garbage, are included. Moreover, 4,517 photos after expansion are used as the dataset. The data set is divided using a stratified sampling method in the ratio of 8:2. The training set and test set are included, with 3,613 images in the training set and 904 images in the test set. The training set is the data samples fitted to the model. The test set is the samples kept separately in the model training for model hyperparameter tuning and preliminary assessment of the model capability.

**5.2. Evaluation Metrics.** Accuracy and recall are two evaluation metrics commonly used in deep learning to assess the performance and accuracy of classification models. Accuracy and recall are often used together with the confusion matrix.

The confusion matrix is a  $2 \times 2$  matrix in which each element represents the match between the actual and predicted categories. Each of the four elements of the matrix represents:

- (1) TP (true positive): the actual is positive and the predicted is positive.
- (2) FP (false positive): the actual negative and predicted positive cases.
- (3) FN (false negative case): actual positive case predicted negative case.
- (4) TN (true negative): actual negative case and predicted negative case.

Among them, precision refers to the proportion of the actual positive samples among the predicted positive samples, which is calculated as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \quad (2)$$

The recall is the number of positive samples correctly identified by the model as a percentage of all positive samples. The calculation formula is as follows:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (3)$$

where TP denotes true examples and FN denotes false counterexamples. The higher the recall, the better the model can identify the positive cases.

In practical applications, accuracy and recall rates usually need to be considered balanced. For example, the recall rate is more important than the accuracy rate because a higher misdiagnosis rate may lead to missing the disease in oncology diagnosis, while a higher miss rate allows for a second

examination. Conversely, in some scenarios, such as financial fraud detection, the accuracy rate is more important than the recall rate because a misdiagnosis can lead to financial loss.

In conclusion, accuracy rate and recall rate are standard evaluation metrics in deep learning to assess the performance and accuracy of the classification models. In practical applications, suitable evaluation metrics need to be selected according to the specific scenarios and balanced considerations.

**5.3. Experimental Design and Experimental Results.** The same data set and the same parameter settings are used in this experiment in order to examine the performance of the improved model in this paper more intuitively. The training loss graph of the model is drawn based on the log files saved during the training process, it is compared with the training loss graph of the YOLOv5 model before improvement, as shown in Figures 11 and 12. From left to right, each image represents Box\_Loss, Obj\_Loss, Cls\_Loss, Precision, and Recall in turn.

**5.3.1. Ablation Experiments.** RepConv is a repeatable convolutional structure, which can increase the depth and width of the network without increasing the computational effort due to its discretization of the convolutional layers as well as the shared parameter property, and also improve the performance of the model while keeping the model size constant. The C2F module, on the other hand, compared to the C3 module, is designed to pay more attention to the richness of the gradient flow, which can better extract image features and thus improve the performance of the model. In addition, the C2F module adjusts different numbers of channels for different scale models, which further improves the performance of the model. Therefore, the C2F module can obtain richer gradient flow information while ensuring lightweight. In addition, RepConv can reduce the number of parameters in the model, thus reducing the risk of overfitting. Finally, regarding the Meta-ACON function, it can dynamically learn (adaptively) the linearity/nonlinearity of the activation function and control the degree of nonlinearity at each layer of the network, which significantly improves the model's generalization performance. Therefore, theoretically all three modules can improve the performance of the YOLOv5 model.

Next, we further verified this with an ablation test. In the ablation experiments, the C2F module, the RepConv module, and the Meta-ACON module are added to the YOLOv5 model. The performance improvement of these modules is evaluated by comparing them with the original YOLOv5 model. The results reveal that the models with the addition of the C2F module, RepConv module, and Meta-ACON module all achieve performance improvements compared to the original YOLOv5 model. The experimental results are shown in Table 1.

Specifically, the mAP scores improved by 0.6 percentage points in the C2F group, 0.3 percentage points in the RepConv group, and 0.3 percentage points in the Meta-ACON group. This result indicates that adding the C2F, RepConv, and Meta-ACON modules to the YOLOv5 model



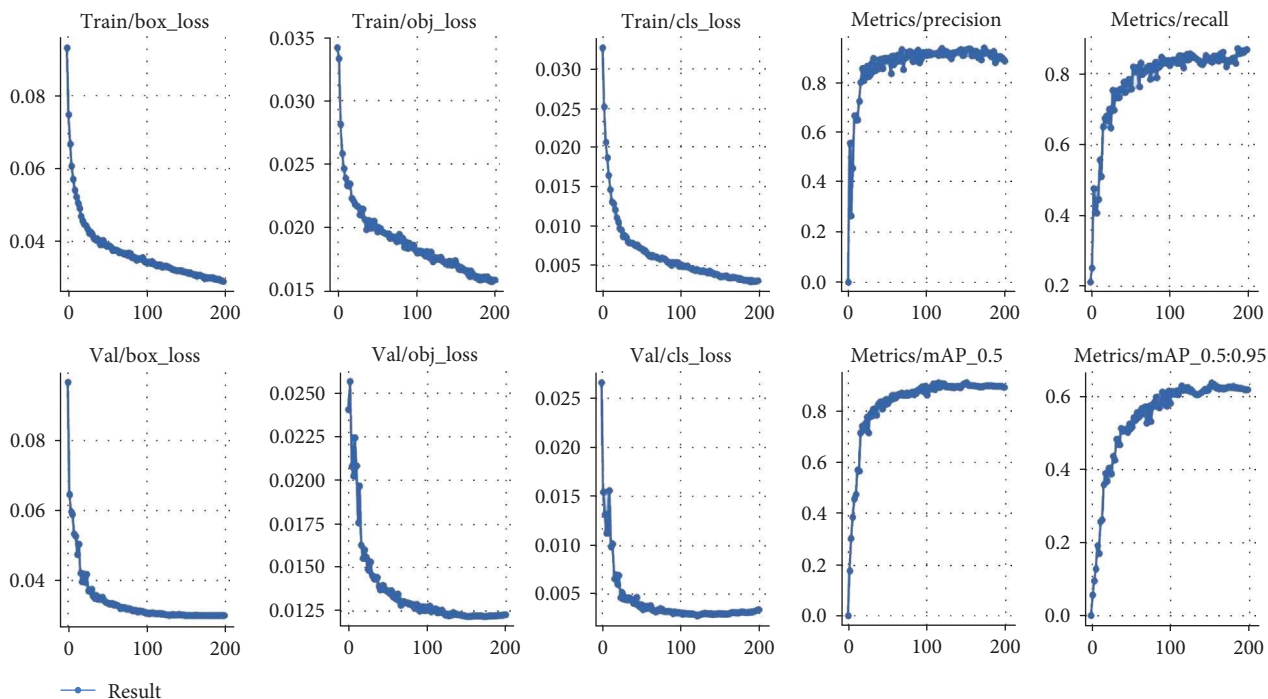


FIGURE 11: Improved previous loss graph.

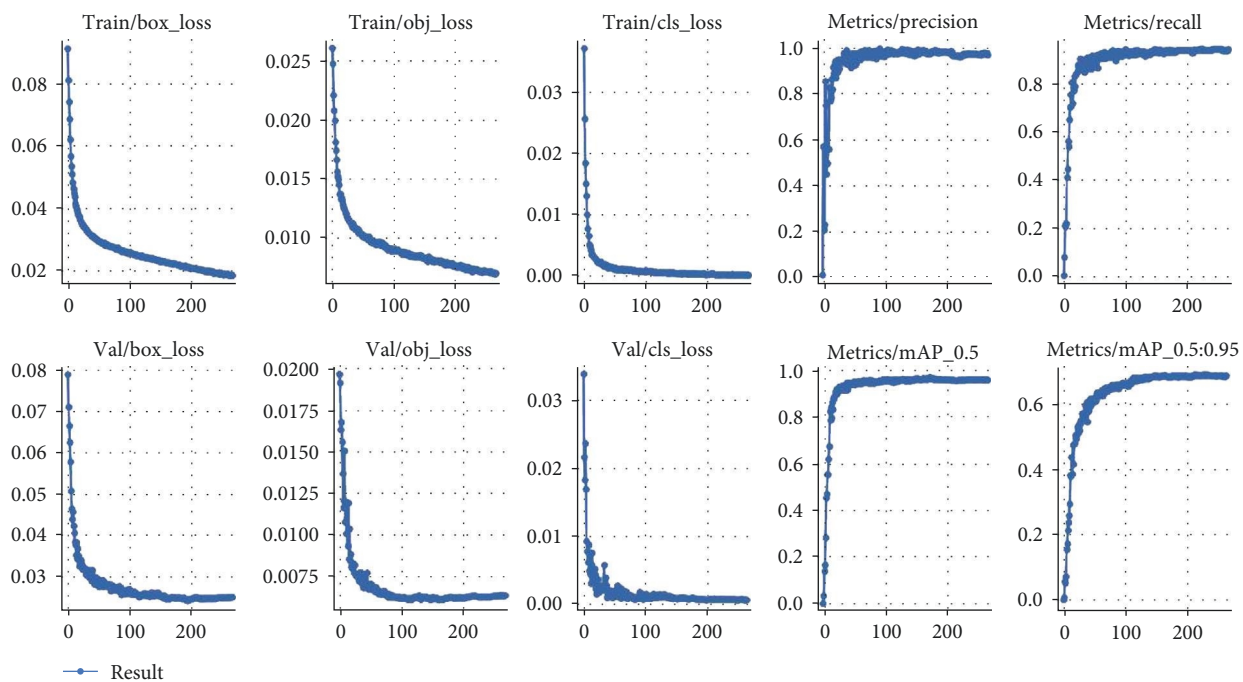


FIGURE 12: Improved loss graph.

TABLE 1: Results of ablation experiments.

Group	C2F	RepConv	Meta-ACON	P	R	mAP@.5	mAP@.5:.95	FPS
G1	—	—	—	0.962	0.925	0.947	0.688	269.72
G2	√	—	—	0.988	0.927	0.963	0.692	265.08
G3	√	√	—	0.978	0.94	0.966	0.691	282.62
G4	√	√	√	0.964	0.949	0.969	0.686	258.36

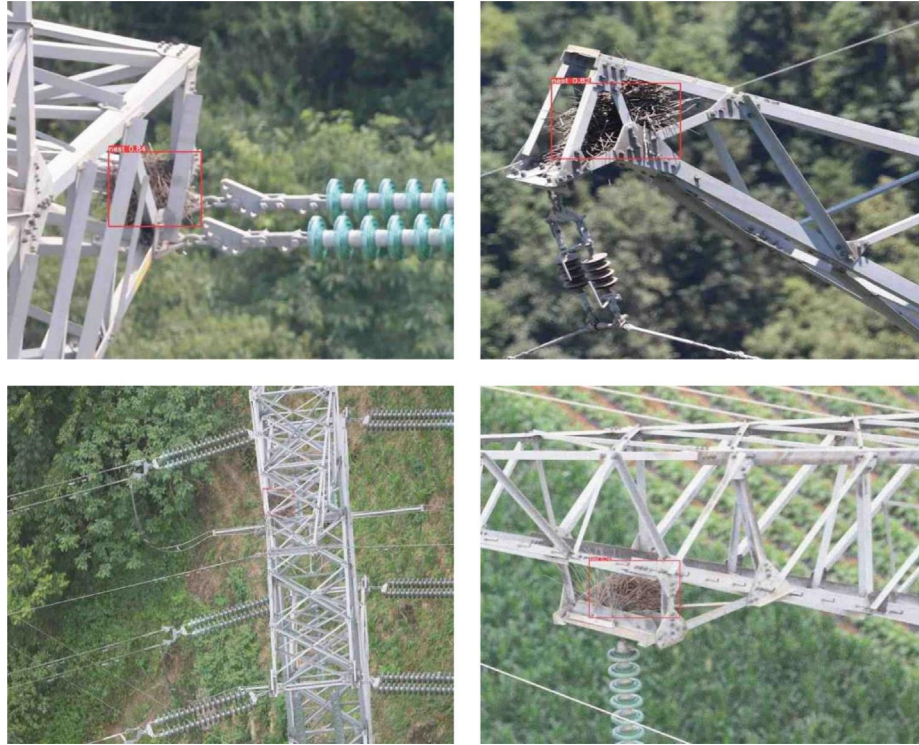


FIGURE 13: Model recognition effect.

can improve the model performance and make it more accurate in detecting target objects. Although it is a minor enhancement, it can still improve the model's performance.

Overall, the results of this ablation experiment show that adding the C2F module, RepConv module, and Meta-ACON module can make the YOLOv5 model perform better in the target detection task, with higher detection accuracy and better performance. The experimental results are shown in Figure 13.

**5.3.2. Comparison Experiments.** This paper compares the accuracy and inference speed of several target detection models, YOLOv5, YOLOv4, YOLOv3, SSD, and Faster R-CNN, in constructing foreign object datasets under the same test environment, which aims at verifying the effectiveness of the improved YOLOv5 algorithm on the scale of model parameters to reduce model complexity and improving model detection speed.

With regard to inference speed, the inference speed of these models was tested using a computer equipped with an NVIDIA GPU. The experimental results show that YOLOv5 has a faster inference speed than faster R-CNN and SSD's inference speed lies between them. The same images for each model is used in the test for inference and recorded the average inference time for each model.

In the test, the inference speed was 3.7-ms per image, corresponding to a frame rate of 269.72 frames/second when using the YOLOv5 model for image inference. In comparison, the inference speed of the SSD model is 13 ms/image. The corresponding frame rate is 75 frames/second under the same hardware environment. In contrast, the

inference speed of the Faster R-CNN model is 91 ms/image, corresponding to a frame rate of 11 fps.

As can be seen from Table 2, the improved YOLOv5 achieves a balance between accuracy and inference speed with high performance. While SSD provides a balanced choice between speed and accuracy, YOLOv5 slightly outperforms SSD on the dataset of this paper. In addition, YOLOv5 employs certain new techniques, such as SPP and PAN, enabling it to improve accuracy and speed. As seen from the Table 2, the improved version of YOLOv5 performs significantly better than SSD on the dataset of this paper. Furthermore, some image enhancement techniques, such as Alburnation image enhancement, are used in this paper, making the improved version of YOLOv5 improve accuracy and robustness. Table 2 suggests that the improved YOLOv5 is much faster than Faster R-CNN in terms of speed. Although the accuracy of the improved YOLOv5 is slightly inferior to that of Faster R-CNN, speed is a more critical factor in some scenarios, in real-time object detection or large-scale target detection tasks, for example. Therefore, an improved version of YOLOv5 is more suitable for these scenarios. Faster R-CNN can be a more accurate model choice if time and resources allow. This paper also compares the performance of two models of the YOLO family such as YOLOv3 and YOLOv4. On the dataset of this paper, the experimental results are shown in Table 2. The improved YOLOv5 also significantly improves computational speed, with an increase in FPS of 223.36 and 232.36 compared to YOLOv3 and YOLOv4, respectively. The improved YOLOv5 is more accurate for feature extraction of different target categories and more sensitive than YOLOv3 and YOLOv4.

TABLE 2: Comparative experimental results.

Algorithm	P	R	mAP@.5	mAP@.5:.95	FPS	Inference/ms
Yolov3	0.98	0.94	0.97	0.70	35	27
Yolov4	0.92	0.88	0.96	0.624	26	38
Yolov5s	0.962	0.925	0.947	0.688	269.72	3.7
Yolov5m	0.968	0.949	0.963	0.694	178	5.6
Yolov5l	0.99	0.939	0.964	0.697	126	7.9
Yolov5x	0.967	0.937	0.957	0.689	79	12.6
SSD	0.76	0.71	0.73	0.53	75	13
Faster R-CNN	0.969	0.90	0.967	0.637	11	91
Ours	0.964	0.949	0.969	0.686	258.36	3.9



FIGURE 14: Improved YOLOv5 heat map.

It indicates that the improved YOLOv5 has fully considered the characteristics of different target classes in the network design and training process to improve the recognition and classification ability of the model. Compared with YOLOv5 models of different sizes, the accuracy of the improved YOLOv5 is similar to YOLOv5l, but the detection speed is much higher than YOLOv5l.

The characteristic heat map is shown in Figure 14. The improved version of YOLOv5 can more accurately determine the target location and size information during the detection of targets in images, thus improving the positioning and detection capability of the model.

Overall, this experiment validates the superior performance of the improved YOLOv5 in the target detection task through a comprehensive comparison with multiple target detection models on the dataset used in this paper. The comparison focuses on accuracy, recall, and computation speed,

highlighting the improved YOLOv5's exceptional performance in these areas. Meanwhile, the analysis of feature heat maps also provides a more intuitive observation of the advantages of the improved YOLOv5 in feature learning and target detection, which provides valuable references for the subsequent research and applications.

The loss plots of training SSD, Faster R-CNN, and YOLO series training multiple epochs are shown in Figure 15.

Some commonalities and differences in the training process of different models can be observed from the loss curves. The YOLOv3 and YOLOv4 models have faster decreasing loss values in the first few epochs after the training starts. The decline of these models slows down after the 120-th epoch. SSD still has a slight decline around the 150-th epoch and finally stabilizes at a smaller value. However, the two-stage Faster R-CNN model has a significant increase in the loss value after the start of thawing training. Although a similar

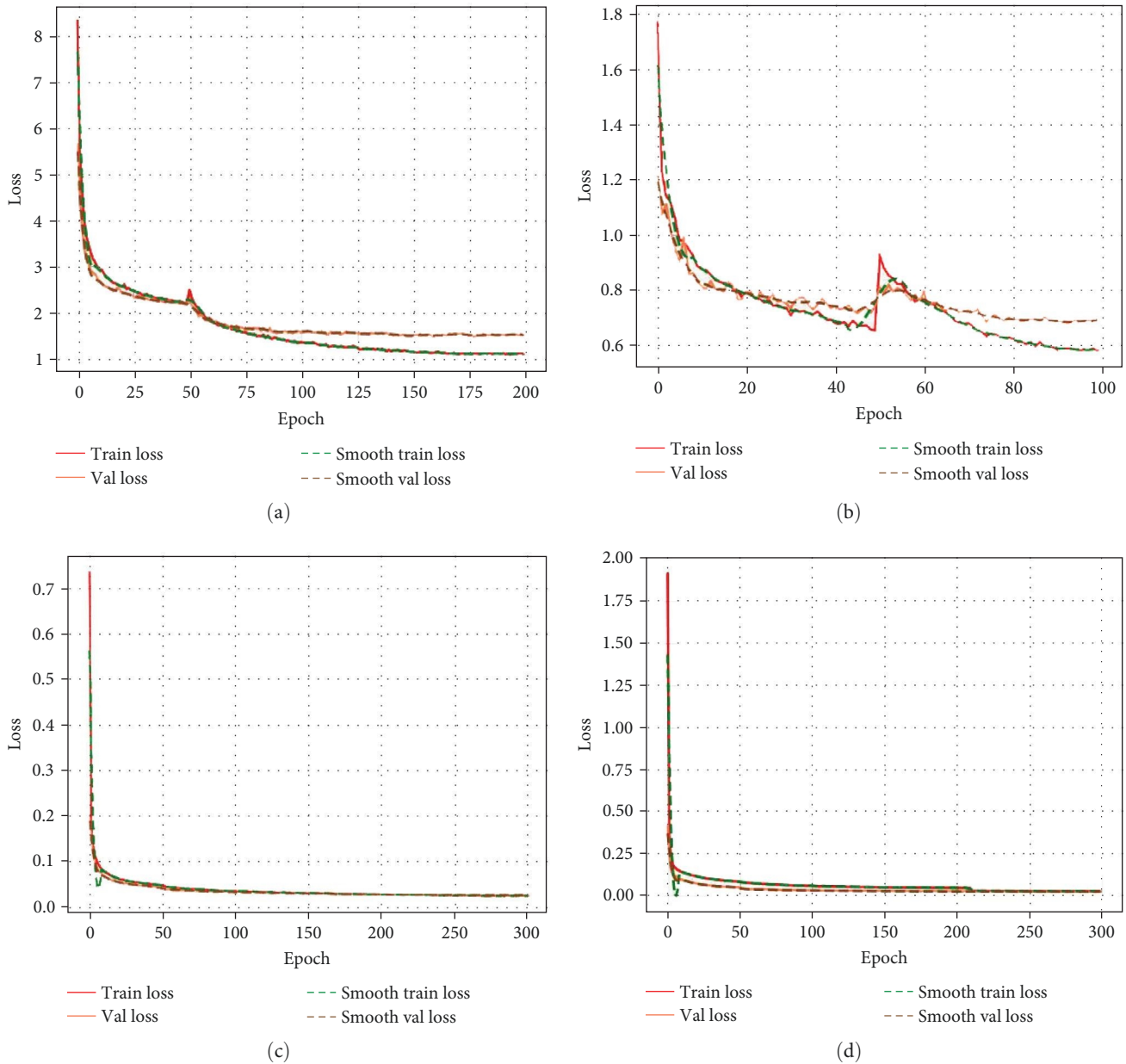


FIGURE 15: Loss curves of different models. (a) Graph of experimental results of SSD, (b) graph of experimental results of Faster R-CNN, (c) graph of experimental results of YOLOv3, and (d) graph of experimental results of YOLOv4.

situation occurred in the SSD model, the Faster R-CNN was significantly more drastic in degree. Eventually, the loss values of all models stabilized.

## 6. Conclusion

This paper addresses the problem that transmission lines are often hung with foreign objects. Furthermore, traditional manual inspection is complicated, inefficient, and inaccurate. It improves YOLOv5's ability to extract features at different scales, introduces the RepConv structure, reduces computation and memory consumption by sharing parameters and increasing convolutional layers, and improves target inference speed while ensuring the accuracy of existing target

recognition. C2F structure is used to improve the semantic expression of features, thus enhancing the feature extraction ability of the network, which is conducive to the model's more accurate detection of small targets. Finally, the Meta-ACON activation function is used to replace the traditional activation function, which can adaptively adjust the parameters of the activation function, thus improving the generalization ability of the model and making it more practical. The experimental results show that the improved YOLOv5 model can effectively improve the performance and robustness of the target detection model and has a better balance between inference speed and detection accuracy, which has specific practical application value and can be better applied to the field of transmission line foreign body detection. Future

research directions can further explore more effective model structures and activation functions to meet the continuous demand for target detection models in the practical applications.

## Data Availability

Due to the nature of this research, participants of this study did not agree for their data to be shared publicly, so supporting data are not available.

## Conflicts of Interest

The authors declare that they have no conflicts of interest to report regarding the present study.

## Acknowledgments

This work was financially supported by the National Natural Science Foundation of China (grant number: 61901165, 61601177) and the Natural Science Foundation of Hubei Province (2019CFB530), Beijing, China, vol. 1750, no. 1, p. 012023, 2021. We would also like to thank Leming Guo for his help in researching this paper and Ding Chen for his kind support in translating this paper.

## References

- [1] Y. Deng, Z. Tang, H. Li, and Z. Yang, "Real-time rendering algorithm optimization for large scale transmission lines based on LOD," *Computer and Modernization*, vol. 1, Article ID 115, 2017.
- [2] M. L. Tuballa and M. L. Abundo, "A review of the development of smart grid technologies," *Renewable and Sustainable Energy Reviews*, vol. 59, pp. 710–725, 2016.
- [3] G. Dileep, "A survey on smart grid technologies and applications," *Renewable Energy*, vol. 146, pp. 2589–2625, 2020.
- [4] J. Katrasnik, F. Pernus, and B. Likar, "A survey of mobile robots for distribution power line inspection," *IEEE Transactions on Power Delivery*, vol. 25, no. 1, pp. 485–493, 2010.
- [5] V. I. Koshelev and D. N. Kozlov, "Wire recognition in image within aerial inspection application," in *2015 IEEE 4th Mediterranean Conference on Embedded Computing (MECO)*, pp. 159–162, IEEE, Budva, Montenegro, 2015.
- [6] T. Zhang and J. Dai, "Electric power intelligent inspection robot: a review," *Journal of Physics: Conference Series*, vol. 1750, Article ID 012023, 2021.
- [7] R. Han, L. Liu, S. Liu, P. Jiang, Y. Han, and Z. Yang, "Research and application of substation intelligent inspection technology based on multi spectral image recognition," in *2020 IEEE International Conference on High Voltage Engineering and Application (ICHVE)*, pp. 1–4, IEEE, Beijing, China, 2020.
- [8] S. Jin, J. Yin, J. Zhou, Y. Cai, and Y. Hu, "Intelligent inspection technology of substation secondary system based on machine vision," *Journal of Physics: Conference Series*, vol. 2428, Article ID 012034, 2023.
- [9] P. Bharati and A. Pramanik, "Deep learning techniques—R-CNN to mask R-CNN: a survey," in *Computational Intelligence in Pattern Recognition*, vol. 999 of *Advances in Intelligent Systems and Computing*, pp. 657–668, Springer, Singapore, 2020.
- [10] R. Girshick, "Fast R-CNN," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1440–1448, IEEE, 1730 Massachusetts Ave., NW Washington, DC, United States, 2015.
- [11] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2961–2969, IEEE, Venice, Italy, 2017.
- [12] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *Advances in Neural Information Processing Systems*, vol. 28, 2015.
- [13] F. Zhou, H. Zhao, and Z. Nie, "Safety helmet detection based on YOLOv5," in *2021 IEEE International Conference on Power Electronics, Computer Applications (ICPECA)*, pp. 6–11, IEEE, Shenyang, China, 2021.
- [14] X. Shen, G. Shi, H. Ren, and W. Zhang, "Biomimetic vision for zoom object detection based on improved vertical grid number YOLO algorithm," *Frontiers in Bioengineering and Biotechnology*, vol. 10, Article ID 905583, 2022.
- [15] D. Li, X. Sun, H. Elkhouchlaa et al., "Fast detection and location of longan fruits using UAV images," *Computers and Electronics in Agriculture*, vol. 190, Article ID 106465, 2021.
- [16] T. Fan, "Research and realization of video target detection system based on deep learning," *International Journal of Wavelets, Multiresolution and Information Processing*, vol. 18, no. 1, Article ID 1941010, 2020.
- [17] S. Chen, D. Liu, Y. Pu, and Y. Zhong, "Advances in deep learning-based image recognition of product packaging," *Image and Vision Computing*, vol. 128, Article ID 104571, 2022.
- [18] Q. Hao, X. Guo, and F. Yang, "Fast recognition method for multiple apple targets in complex occlusion environment based on improved YOLOv5," *Journal of Sensors*, vol. 2023, Article ID 3609541, 13 pages, 2023.
- [19] Z. Xu, X. Huang, Y. Huang, H. Sun, and F. Wan, "A real-time zanthoxylum target detection method for an intelligent picking robot under a complex background, based on an improved YOLOv5s architecture," *Sensors*, vol. 22, no. 2, Article ID 682, 2022.
- [20] G. Wen, S. Li, F. Liu et al., "YOLOv5s-CA: a modified YOLOv5s network with coordinate attention for underwater target detection," *Sensors*, vol. 23, no. 7, Article ID 3367, 2023.
- [21] H. Zang, Y. Wang, L. Ru et al., "Detection method of wheat spike improved YOLOv5s based on the attention mechanism," *Frontiers in Plant Science*, vol. 13, Article ID 993244, 2022.
- [22] G. Chen, H. Zhou, Z. Li et al., "Multi-scale forest fire recognition model based on improved YOLOv5s," *Forests*, vol. 14, no. 2, Article ID 315, 2023.
- [23] W. Bai, J. Zhao, C. Dai, H. Zhang, L. Zhao, Z. Ji, and I. Ganchev, "Two novel models for traffic sign detection based on YOLOv5s," *Axioms*, vol. 12, no. 2, Article ID 160, 2023.
- [24] Z. Tang, L. Zhou, F. Qi, and H. Chen, "An improved lightweight and real-time YOLOv5 network for detection of surface defects on indocalamus leaves," *Journal of Real-Time Image Processing*, vol. 20, Article ID 14, 2023.
- [25] G. Ghiasi, T. Y. Lin, and Q. V. Le, "NAS-FPN: learning scalable feature pyramid architecture for object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7036–7045, IEEE, 2019.
- [26] A. E. Gamal, B. A. Fowler, H. Min, and X. Liu, "Modeling and estimation of FPN components in CMOS image sensors," *Solid State Sensor Arrays: Development and Applications II. SPIE*, vol. 3301, pp. 168–177, 1998.
- [27] H. Kuang, B. Wang, J. An, M. Zhang, and Z. Zhang, "Voxel-FPN: multi-scale voxel feature aggregation for 3D object

- detection from LIDAR point clouds,” *Sensors*, vol. 20, no. 3, Article ID 704, 2020.
- [28] S. Luo and J. Liu, “Research on car license plate recognition based on improved YOLOv5m and LPRNet,” *IEEE Access*, vol. 10, pp. 93692–93700, 2022.
- [29] G. Li, L. Fu, C. Gao et al., “Multi-class detection of kiwifruit flower and its distribution identification in orchard based on YOLOv5l and euclidean distance,” *Computers and Electronics in Agriculture*, vol. 201, Article ID 107342, 2022.
- [30] J.-L. Zhang, W.-H. Su, H.-Y. Zhang, and Y. Peng, “SE-YOLOv5x: an optimized model based on transfer learning and visual attention mechanism for identifying and localizing weeds and vegetables,” *Agronomy*, vol. 12, no. 9, Article ID 2061, 2022.
- [31] M. P. Mathew and T. Y. Mahesh, “Leaf-based disease detection in bell pepper plant using YOLO v5,” *Signal, Image and Video Processing*, vol. 16, no. 3, pp. 841–847, 2022.
- [32] M. Soudy, Y. Afify, and N. Badr, “RepConv: a novel architecture for image scene classification on intel scenes dataset,” *International Journal of Intelligent Computing and Information Sciences*, vol. 22, no. 2, pp. 63–73, 2022.
- [33] D. Wu, S. Jiang, E. Zhao et al., “Detection of *Camellia oleifera* fruit in complex scenes by using YOLOv7 and data augmentation,” *Applied Sciences*, vol. 12, no. 22, Article ID 11318, 2022.
- [34] B. P. Athidhi and V. P. Smitha, “YOLOv7-based model for detecting safety helmet wear on construction sites,” in *Intelligent Sustainable Systems. ICoISS 2023*, vol. 665 of *Lecture Notes in Networks and Systems*, pp. 377–392, Springer, Singapore, 2023.
- [35] J. Zhang, S. Liu, H. Yuan et al., “Deep learning for microfluidic-assisted *Caenorhabditis elegans* multi-parameter identification using YOLOv7,” *Micromachines*, vol. 14, no. 7, Article ID 1339, 2023.
- [36] H. Lou, X. Duan, J. Guo et al., “DC-YOLOv8: small-size object detection algorithm based on camera sensor,” *Electronics*, vol. 12, no. 10, Article ID 2323, 2023.
- [37] R. Bai, F. Shen, M. Wang, J. Lu, and Z. Zhang, “Improving detection capabilities of YOLOv8-n for small objects in remote sensing imagery: towards better precision with simplified model complexity,” *Research Article*, 2023.
- [38] H. Pei, Y. Sun, H. Huang, W. Zhang, J. Sheng, and Z. Zhang, “Weed detection in maize fields by UAV images based on crop row preprocessing and improved YOLOv4,” *Agriculture*, vol. 12, no. 7, Article ID 975, 2022.
- [39] Q. Guo, J. Liu, and M. Kaliuzhnyi, “YOLOX-SAR: high-precision object detection system based on visible and infrared sensors for SAR remote sensing,” *IEEE Sensors Journal*, vol. 22, no. 17, pp. 17243–17253, 2022.