Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 5835693, 10 pages
https://doi.org/10.1155/2022/5835693

Research Article

Improved YOLOX Foreign Object Detection Algorithm for Transmission Lines

Minghu Wu,[1,2] Leming Guo,[1,2] Rui Chen,[3] Wanyin Du,[1] Juan Wang,[1] Min Liu,[1] Xiangbin Kong,[1] and Jing Tang[1]

[1]*Hubei Collaborative Innovation Center for High-efficiency Utilization of Solar Energy, Hubei University of Technology, Wuhan 430068, China*
[2]*Hubei Engineering Research Center for Safety Monitoring of New Energy and Power Grid Equipment, Hubei University of Technology, Wuhan 430068, China*
[3]*Institute of Artificial Intelligence Industry Technology, Nanjing Institute of Technology, Nanjing 211167, China*

Correspondence should be addressed to Rui Chen; chenrui@njit.edu.cn

Received 14 June 2022; Revised 28 September 2022; Accepted 29 September 2022; Published 20 October 2022

Academic Editor: Xin Ning

It is quite simple for foreign objects to attach themselves to transmission line corridors because of the wide variety of laying and the complex, changing environment. If these foreign objects are not found and removed in a timely manner, they can have a significant impact on the transmission lines' ability to operate safely. Due to the problem of poor accuracy of foreign object identification in transmission line image inspection, we provide an improved YOLOX technique for detection of foreign objects in transmission lines. The method improves the YOLOX target detection network by first using Atrous Spatial Pyramid Pooling to increase sensitivity to foreign objects of different scales, then by embedding Convolutional Block Attention Module to increase model recognition accuracy, and finally by using GIoU loss to further optimize. The testing findings show that the enhanced YOLOX network has a mAP improvement of around 4.24% over the baseline YOLOX network. The target detection SSD, Faster R-CNN, YOLOv5, and YOLOV7 networks have improved less than this. The effectiveness and superiority of the algorithm are proven.

1. Introduction

The power grid's transmission line serves as the conduit for electric energy, and maintaining its stability and security for power transmission is essential to the grid's efficient and secure operation [1]. Significant statistics show that the foreign objects that regularly appear on the electricity system are bird nests, kites, balloons, and trash. These components are easily capable of causing short-circuit faults or single-phase faults between transmission lines, which can lead to a number of short-circuit accidents, some of which might result in fire and substantial power outages, leading to significant economic losses [2, 3]. Short circuits will cause a domino effect that endangers the lives and property of those who reside close to power lines [4]. Additionally, it endangers the lives and security of maintenance workers who go to inspect the electricity infrastructure. The transmission lines typically travel across a variety of landscapes, through densely populated regions with heavy traffic. Response time will be constrained once any security problems need to be resolved manually. Processes for manual operation and maintenance are also very costly, time-consuming, and challenging to finish on schedule. As a result, by using UAV aerial photography of transmission lines for inspection, intelligent inspection technology [5–8] was developed, which can save a lot of time and resources while also having a high detection efficiency when compared to artificial techniques. To recognize aerial data, nevertheless, still requires manual judgment. As a result, the detection process's overall effectiveness and precision must be improved.

In this big data era, GPU computing power is increasing, and deep learning is gradually proving to be advantageous in many computer vision applications, notably target recognition jobs. Beginning with two-stage networks like R-CNN, Fast R-CNN [9], Mask-R-CNN [10], and Faster R-CNN [11], as well as some excellent algorithms for improving the CNN network model [12–15], which have the advantages of high detection accuracy and low leakage rate, there has been an explosion of deep learning-based target detection networks since 2014. However, the detection speed is slow and the computation is relatively complicated, making it difficult to be used for the detection and prevention of cybercrime. Later, the single-stage detection method [16] emerged, combining prediction frame localization and candidate region feature extraction for direct target class identification and detection frame localization. Researchers have started to take lighter, faster single-stage target detection networks into consideration. These networks have the benefits of quick detection and low computation to meet the demand for real-time detection, ushering in a new era of single-stage target detection networks. Among them, the single-stage target detection model is mainly YOLO series, including YOLOV4 [17, 18], YOLOV5 [19], YOLOV6 [20], and YOLOV7 [21]. At the same time, YOLO series also includes many variants, such as PP-YOLO [22].

The authors presented a way to recognize and detect foreign objects [23], such as birds, on transmission lines while still having the issue of low network detection accuracy by enhancing the YOLOv3 model, which principally uses an improved network of two-scale detection frames. In order to find broken strands and foreign objects, the authors suggested a method using grayscale and conductor width fluctuation for UAV inspection images [24]. However, this method can only find foreign objects on the line itself, not on transmission line towers. For finding bird nests on transmission lines, a dynamic federated learning strategy is recommended [25]. This method necessitates the use of a central server, which has network needs and information transfer delays that are difficult to ensure, to process statistics before returning the detection findings. The aforementioned techniques still have some limitations, when it is used for transmission line foreign object detection.

With detection rates of up to 140 frames per second, YOLOX [26], which was introduced in 2021, stunned the globe and is a strong contender for real-time and mobile deployment scenarios. Without changing the target feature extraction network, the YOLOX-S version has been slightly improved for a few domains in the literature [27–32]. Feature extraction has also been improved by upgrading the FPN (feature pyramid networks), which has led to some gains in target recognition accuracy. The YOLOX-S regression, however, lacks sufficient precision. The mAP (Mean Average Precision) of YOLOX-M, YOLOX-L, YOLOX-X, and other deeper layers of YOLOX can be higher than YOLOX-S. However, the model will contain more data, increasing the hardware requirements of the method. YOLOX Tiny and YOLOX Nano versions can be used on a wider range of computers and have a faster frame rate. Nevertheless, their accuracy mAP has a gap when compared to more intricate network models. Consequently, it is challenging to achieve the demanding requirements for real-time and target frame regression accuracy scenarios. The following are the main contributions of this research, which present a lightweight and accurate target identification model based on YOLOX-S, in order to better balance speed and accuracy and better apply the YOLO model to the transmission line foreign object detection problem:

(1) The Atrous Spatial Pyramid Pooling (ASPP) was utilized to replace the Spatial Pyramid Pooling (SPP) in order to broaden the receptive field and enhance sensitivity to foreign objects of various sizes [33]

(2) A lightweight CBAM (Convolutional Block Attention Module) is embedded in the network in order to make the model pay more attention to the important position information and channel information in the feature map

(3) GIoU loss (generalized crossmerge ratio loss function) is used [34] to replace the original IoU loss function [35], which can solve the problem that the model cannot be optimized without overlapping objectives, realizing the ability to distinguish two objects in different permutations

The above is the focus of this paper. Although the improved object detection model YOLOX in this paper is employed in the foreign object detection of transmission lines, the improved model also can be utilized in a wider range, such as occlusion target detection, super resolution reconstruction, video content segmentation, image repair, and other fields [36–40].

The rest of this paper is organized as follows. Section 2 describes the related work; Section 3 introduces the proposed method in details; Section 4 reveals the experimental results and analysis; Section 5 deals with the conclusion.

2. Related Work

2.1. YOLOX. The YOLOX algorithm, which includes the advantages of the YOLO series network, was proposed by YOLO series in 2021. The three elements that distinguish SimOTA (Sample Optimal Transport Assignment) from the previous YOLO series are its dynamic positive sample matching technique, decoupled head, and anchor-free design. Additionally, a number of advancements are used to integrate a series of innovations that not only achieve APs beyond YOLOv3 [31], YOLOv4, and YOLOv5 but also achieve competitive inference speed. These advancements include the introduction of the focus network structure of YOLOv5 for channel broadening, the use of mosaic data enhancement techniques, and a number of other improvements.

The three components of the YOLOX model are shown in Figure 1: YoloHead, the enhanced feature extraction network, and the backbone feature extraction network. The backbone feature extraction network makes use of the CSPDarknet network through a series of channel modifications to gather feature information from feature layers with different geometries. The enhanced feature extraction network achieves feature fusion using the feature pyramid

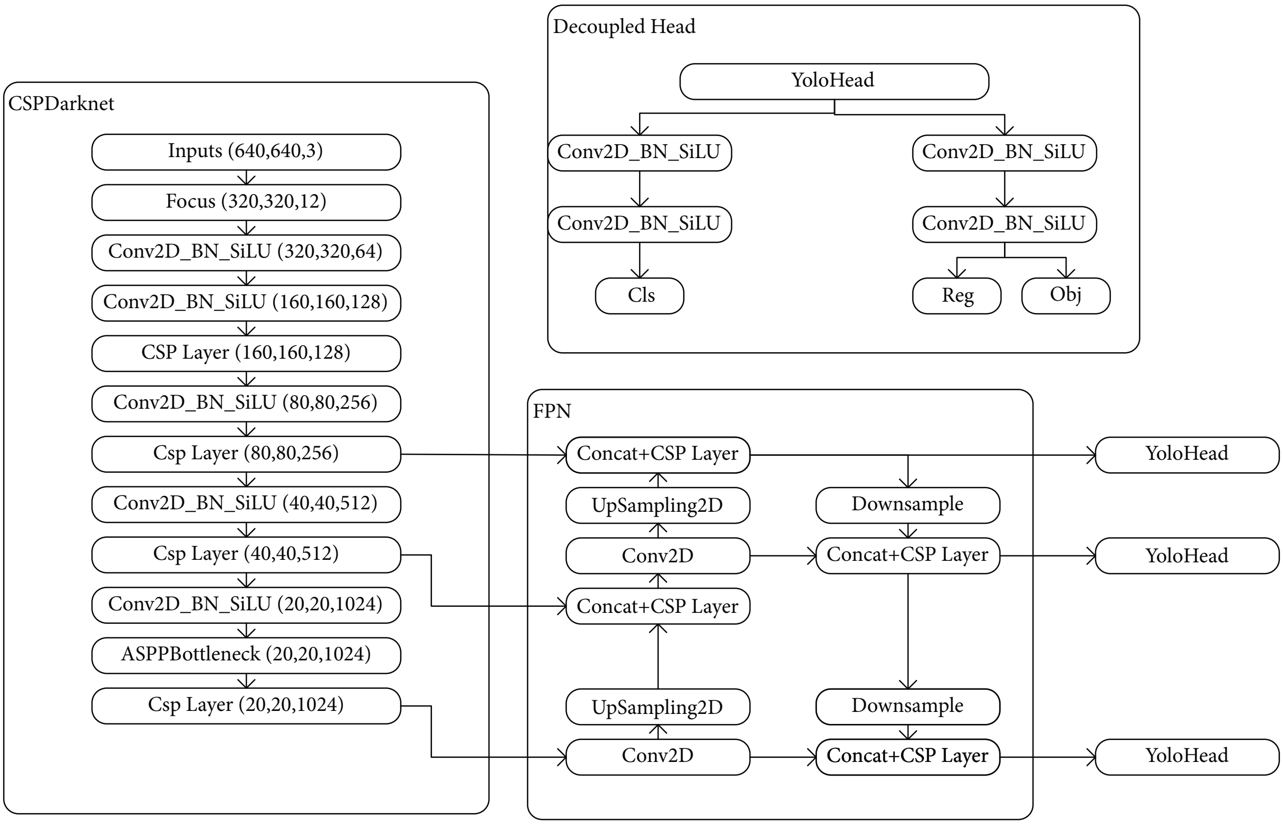

Figure 1: YOLOX structure.

network. Three decoupled heads are used by the detection head, which consists of the classifier and regressor parts, to produce three prediction results: the object type contained in the feature points, the regression parameter of the feature points if the feature points contain objects, and the regression parameter of the feature points otherwise.

The input image will first undergo feature extraction in CSPDarknet, followed by the acquisition of three feature layers in FPN to combine feature information at various scales for feature fusion, and finally, the acquisition of the three enhanced effective feature layers and input to the detection head to determine whether the feature points have objects corresponding to them.

2.2. Atrous Spatial Pyramid Pooling (ASPP). By maximizing the pooling of various pooling kernel sizes for feature extraction and enlarging the network's perceptual field, the SPP (Spatial Pyramid Pooling) technique improves the network's capacity to extract multiscale contexts. However, it is generally accepted that the adjacent pixel places in the image contain redundant information. The spatial resolution will decrease as the perceptual field is widened.

The ASPP module employs a number of parallel atrous convolution layers with various sampling rates. The features that were retrieved for each sampling rate are further processed and combined to create the final output in a different branch. The module generates convolutional kernels with various sensory fields by varying the expansion coefficient (rate), which are then used to gather multiscale object information and broaden the network's sensory field. This improves the network's capacity to acquire multiscale contexts without affecting the shape.

A specific ASPP with four parallel branches, the first of which is a $1 * 1$ regular convolution layer, is shown in Figure 2. A $3 * 3$ expansion convolution is used for the second and third branches, with various expansion coefficients for each branch.

2.3. Convolutional Block Attention Module (CBAM). Finding a set of attention weight coefficients that apply to the model is the major objective of the attention mechanism, which is often described as pulling more significant input from a particular area while ignoring or suppressing unimportant data. Deep learning can use the attention process to separate out the information that is most important, enabling the network's inherent properties to be learned. CBAM's structural layout is shown in Figure 3. As illustrated in Figure 3, CBAM has two distinct submodules that execute attention to channel and space, respectively: CAM (Channel Attention Module) and SAM (Spatial Attention Module). In addition to saving computational resources and parameters, this also makes it simpler to integrate into current network topologies. The addition of CBAM led to features spanning more regions of the recognized item and a higher likelihood of

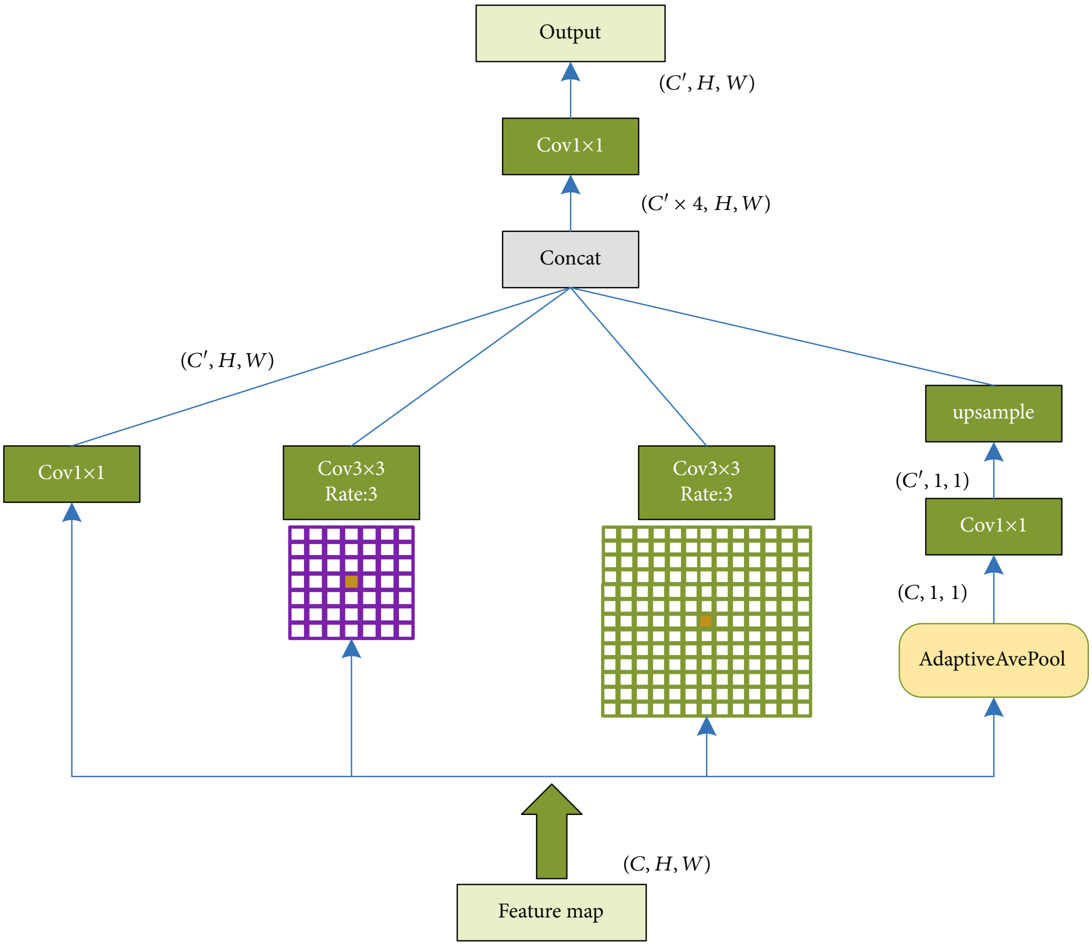

FIGURE 2: Atrous Spatial Pyramid Pooling.

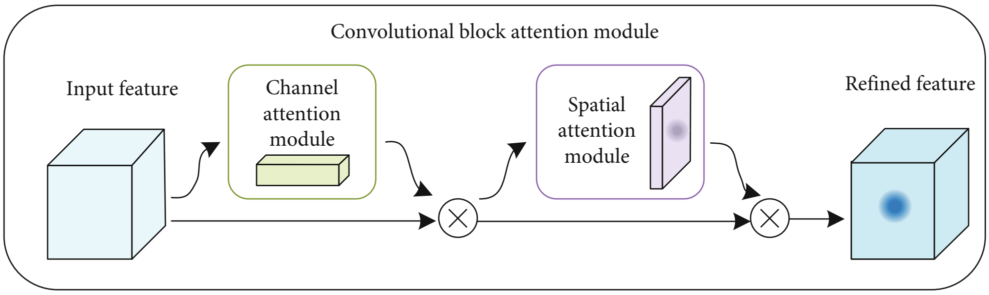

FIGURE 3: Convolutional Block Attention Module.

subsequently discriminating the object, showing that the network's attention mechanism enables it to learn to concentrate on the crucial information.

2.3.1. Channel Attention Module (CAM). CAM is concerned with determining which elements contain critical information. To achieve dimensional compression of the feature map, maximum and average pooling of the input image is used. The information was subsequently fed into a two-layer neural network (MLP) with shared complete connectivity. After a summing process based on the corresponding multiplication of components, the two feature maps are activated by a sigmoid function to obtain the channel attention feature maps with weight. The channel attention $M_C(F)$ is calculated as follows:

$$\begin{aligned} M_C(F) &= \sigma(\mathrm{MLP}(\mathrm{AvgPool}(F)) + \mathrm{MLP}(\mathrm{MaxPool}(F))) \\ &= \sigma\left(W_1\left(W_0\left(F^c_{\mathrm{avg}}\right)\right) + W_1(W_0(F^c_{\max}))\right), \end{aligned} \tag{1}$$

where σ denotes the sigmoid function and W_0 and W_1 are the MLP weights shared for both inputs.

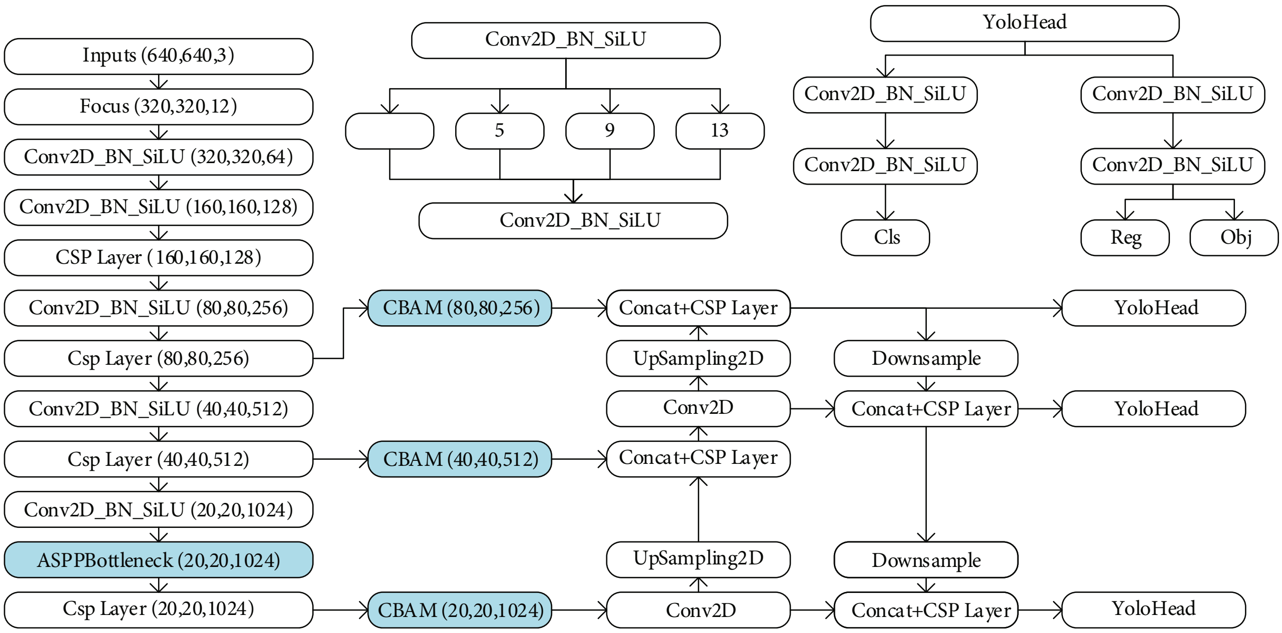

Figure 4: Improved YOLOX-S network structure.

2.3.2. Spatial Attention Module (SAM). SAM is primarily concerned with where the visual information of importance is placed in the picture, and spatial attention can be utilized in conjunction with channel attention to weight distinct spatial elements and aggregate them selectively. The spatial attention feature $M_S(F)$ is calculated by

$$\begin{aligned} M_S(F) &= \sigma\left(f^{7\times7}([\mathrm{AvgPool}(F)\,;\mathrm{MaxPool}(F)])\right) \\ &= \sigma\left(f^{7\times7}\left(\left[F^s_{\mathrm{avg}}\,;F^s_{\mathrm{max}}\right]\right)\right). \end{aligned} \quad (2)$$

After performing global max pooling and global average pooling based on the channel on the feature map produced by the Channel Attention Module, a concat operation is performed on these two H ∗ W ∗ 1 feature maps. The two feature maps are then blended based on the channel. A 7 ∗ 7 convolution method is then used to reduce the dimensionality to 1 channel. After that, the sigmoid function is employed to provide a spatial attention feature ($M_S(F)$). Finally, the input and output images ($M_S(F)$) are multiplied to obtain a CBAM-processed image.

3. The Proposed Method

3.1. Improved YOLOX-S Network Structure. The improved YOLOX-S network structure in this paper is shown in Figure 4. The proposed method switches the YOLOX-S backbone feature extraction network's original SPP structure for an ASPP structure that uses several parallel cavity convolution layers with different sampling rates to produce a better perceptual field than the original structure. Three feature layers reside in the middle, lower middle, and bottom layers of the CSPDarknet of the backbone section, respectively. The three feature layers' shapes are FEAT1 = (80,80,256), FEAT 2 = (40,40,512), and FEAT3 = (640,640,3) when the input is (640,640,3). Prior to each feature layer being output to the CSP net network structure, insert CBAM. This penalizes the attention module for weight sparsity and suppresses less significant weight, allowing attention operations to capture conspicuous characteristics while maintaining performance at the same level.

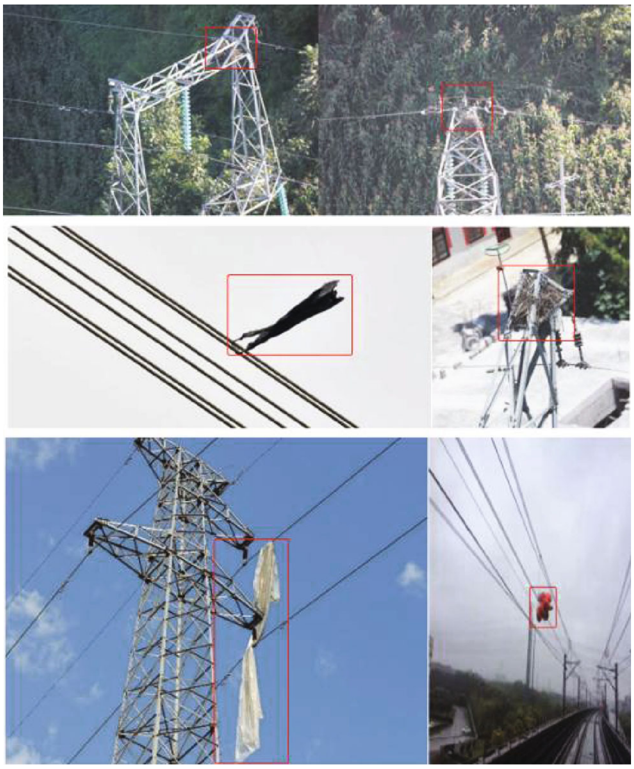

Figure 5: Dataset samples.

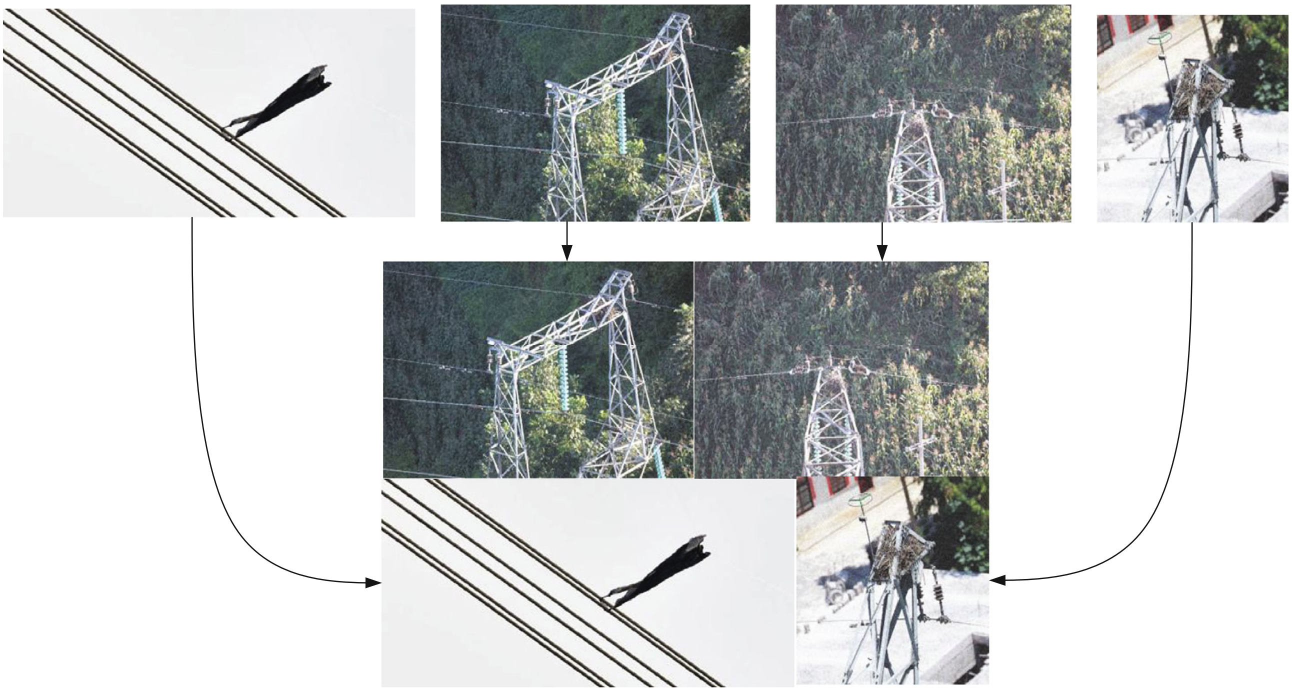

Figure 6: Data enhancement processed by mosaic.

3.2. Improved Loss Function. The feature points' regression parameter (Reg), whether or not they include objects (Obj), and the kind of objects they contain are all predicted using YoloHead in YOLOX (Cls). In Obj and Cls, like in the first YOLOX network, the binary cross loss function (BCE loss) will still be used. For calculating the Reg loss function of the bounding box, the generalized crossmerge ratio loss function (GIoU loss) has been improved. The GIoU loss function offers extra benefits over the IoU loss function.

(1) Like IoU, it has nonnegativity and scale invariance

(2) GIoU is not sensitive to scale

(3) The lower bound of IoU is GIoU, which takes the values [-1, 1]. GIoU adds a penalty term that moves the prediction box toward the target box if the prediction box and the target box do not overlap

(4) In addition to overlapping regions, GIoU focuses on nonoverlapping regions, which can better depict overlap

The IoU is calculated by

$$\mathrm{IoU}(A, B) = \frac{A \cap B}{A \cup B}, \tag{3}$$

where A is the anticipated rectangular box and B is the real rectangular box. The GIoU is calculated by

$$\mathrm{GIoU}(A, B) = \mathrm{IoU}(A, B) - \frac{|\mathrm{C}| - |A \cup B|}{|C|}, \tag{4}$$

Table 1: Experimental environment.

Type	Object	Edition
Hardware	Operating system	Windows 10
	Graphic card	NVIDIA TITAN RTX
	CPU	Intel Xeon Gold
Software	Python	3.7
	Deep learning framework	PyTorch
	CUDA	11.1

where C is the smallest rectangular box that contains A and B. The loss function is

$$\mathrm{Loss}_{\mathrm{giou}} = 1 - \mathrm{GIoU}. \tag{5}$$

When IoU is zero, it means that A and B do not coincide. When A is very far from B, IoU value is infinitely close to zero, and GIoU tends to -1. When IoU equals 1, the two frames overlap, and IoU value equals 1. As a result, GIoU has a value of (-1,1).

To sum up, the loss function formula of the model is given below, which is mainly composed of three parts, namely, location loss, category loss, and confidence loss of positive and negative samples.

In the formula, λ_{coord} represents the positive sample weight coefficient, and λ_{noobj} represents the negative sample weight coefficient; $K \times K$ represents the number of rectangular boxes divided on the feature map; I_{ij}^{obj} represents the existence of positive samples; I_{ij}^{noobj} represents negative samples; w and h represent width and height; and C represents the confidence level of samples. GIoU is used for the location to calculate loss, and crossentropy is used for confidence and category loss.

$$\begin{aligned} L(\text{object}) = {} & \lambda_{\text{coord}} \sum_{i=0}^{K\times K} \sum_{j=o}^{M} I_{ij}^{\text{obj}} (2 - w_i \times h_i)(1 - \text{GIoU}) \\ & - \sum_{i=0}^{K\times K} \sum_{j=o}^{M} I_{ij}^{\text{obj}} \left[\widehat{C}_i \log\left(C_i\right) + \left(1\text{-}\widehat{C}_i\right) \log\left(1 - C_i\right)\right] \\ & - \lambda_{\text{noobj}} \sum_{i=0}^{K\times K} \sum_{j=o}^{M} I_{ij}^{\text{noobj}} \left[\widehat{C}_i \log\left(C_i\right) + \left(1\text{-}\widehat{C}_i\right) \log\left(1 - C_i\right)\right] \\ & - \sum_{i=0}^{K\times K} \sum_{j=o}^{M} I_{ij}^{\text{obj}} \sum_{c\in\text{classes}} \left[\widehat{p}_i(c) \log\left(p_i(c)\right) + (1\text{-}\widehat{p}_i(c)) \log\left(1 - p_i(c)\right)\right]. \end{aligned} \tag{6}$$

4. Experiment

4.1. Data Collection. Because there is not a publicly accessible dataset for transmission line foreign objects, this study employs a manually compiled dataset with four types of bird nests, balloons, garbage, and kites. The transmission line foreign object dataset is created using the LabelImg tool, and it includes the four categories mentioned above. Based on the image sample data gathered, the label information for the images will be recorded in an XML file. After screening out, 4517 images are acquired from the 2D target shots that do not have any noticeable compression marks. The dataset was randomly divided into training, validation, and test sets that were all independent of one another in an 8 : 1 : 1 ratio. There were 3613 photographs in the training set, 452 in the validation set, and 452 in the test set. The most frequent occurrence across the entire dataset is the existence of bird nests atop transmission lines. A small sample of the dataset is shown in Figure 5.

To expand the data amount, the dataset used in this study underwent a number of processing steps, including flipping and rotating the data images. The main objective is to accelerate convergence and enhance generalization properties of the model. Although the initial photographs were of great quality, the network model being trained for this study will reduce them, leading to recognition for 640 ∗ 640 images in the end.

4.2. Mosaic Data Enhancement. In this study, mosaic data enhancement is applied, and the mosaic-processed data enhancement is displayed in Figure 6. The image and frame combination is then carried out by the computer after reading four photos simultaneously, flipping, scaling, and altering each image's color spectrum. This offers a number of benefits.

(1) Enriching the dataset: four images are selected and scaled at random. All of them are distributed randomly for stitching considerably enriching the detection dataset. The random scaling, in particular, adds many small targets, allowing for higher network robustness

(2) Images are saved in GPU RAM by explicitly calculating the data of four photos, eliminating the requirement for a high minibatch size to produce better results

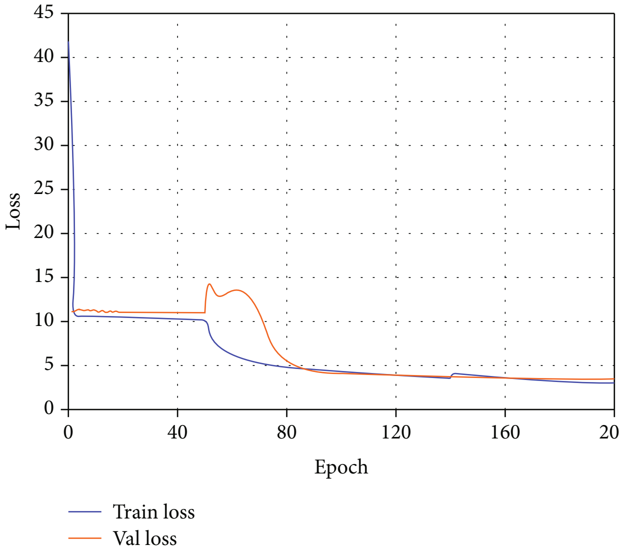

Figure 7: Loss curve.

Table 2: Experiment results with different improvement methods.

Model	ASPP	CBAM	GIoU	mAP	FPS
YOLOX	×	×	×	82.33%	31.43
Improvement 1	✓	×	×	85.67%	30.38
Improvement 2	✓	✓	×	86.04%	30.13
Improvement 3	✓	✓	✓	86.57%	30.08

Table 3: Comparison of performance between main stream target detection models.

Model	mAP	*P*	*R*	FPS
Fast R-CNN	53.58%	50.12%	59.28%	31.33
SSD	73.91%	75.34%	78.60%	26.84
YOLOV4	84.47%	84.01%	86.76%	24.35
YOLOV7-S	83.77%	82.67%	86.01%	34.89
YOLOv5-S	76.42%	75.87%	79.32%	32.13
Proposed	86.57%	85.98%	89.16%	30.08

4.3. Experimental Environment. The operating environment required for the experiment is shown in Table 1. All experiments are conducted in this environment.

The input image tensor and the initialization learning rate are (640, 640, 3). Adam optimizer with cosine annealing learning rate is used for the training. In order to prevent the early model training data from having too much unpredictability, the model's core is frozen for the first 50 iterations and the feature extraction network is left unchanged. After 50 iterations, the frozen component is eliminated, and all network parameters are changed with the training, which is repeated three times for a total of 300 iterations.

4.4. Model Evaluation. For target detection models, the usual evaluation metrics Mean Average Precision (mAP) and frames per second (FPS) are utilized. The area under the precision-recall (PR) curve is hereby designated as mAP,

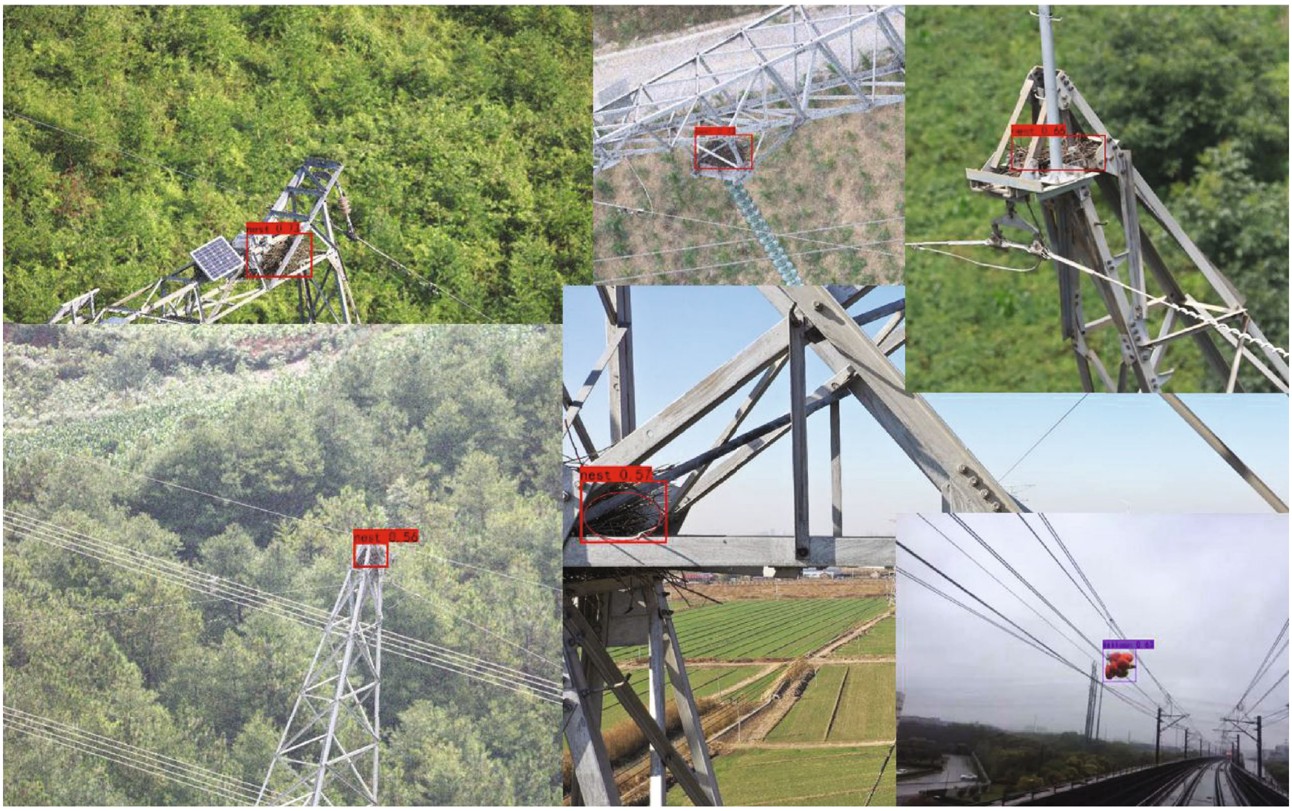

Figure 8: Effect of YOLOX detection.

and the mean value of AP for each category is referred to as mAP. The precision and recall rate are calculated as:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \qquad \text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{7}$$

TP (True Positive) is the positive samples that have been correctly assigned. FP (False Positive) is the positive samples that have been erroneously assigned, and FN (False Negative) is the negative samples that have been incorrectly assigned.

4.5. Experimental Design and Results. The figure shows the training loss of the YOLOX model in terms of the experiment. Although there are some variances in the center, the loss curve finally tends to be smooth as the number of training rounds increases. When epoch exceeds approximately 150, the model begins to progressively converge, and the training procedure does not seem to be overfitting. The loss function is shown in Figure 7.

The "-s" lightweight specification is chosen for this experiment to suit the model deployment requirements. A set of ablation experiments and a set of comparison experiments are constructed in this work. Ablation experiments are utilized to examine the impact of several improvement components of this study on network performance in order to completely investigate the model performance, followed by comparison experiments with mainstream networks (Fast R-CNN, SSD, YOLOv5, YOLOV4, and YOLOV7-S).

4.5.1. Ablation Experiment. To determine how the improved portion of this study affected the model's performance, three sets of trials were developed, each of which used the same training parameters but different model contents. The results of the model performance test are shown in Table 2, where "√" denotes the better model strategy and "×" denotes the improved model strategy that was not applied. Table 2 indicates that improvement 1 expands the perceptual field by using ASPP instead of the original SPP structure, and improvement 2 adds a lightweight attention module (CBAM) to capture salient features through attention operations. Improvement 3 employs a generalized crossmerge ratio loss function (GIoU_loss) to enhance the fit of the prediction frame to the target frame.

4.5.2. Model Comparison. Using the same dataset and hardware setup, the upgraded YOLOX model is compared against the YOLOv5 target detection model from the previous generation as well as the equally effective Fast R-CNN, SSD, and other mainstream target detection techniques. The outcomes of the comparison experiments are shown in Table 3.

The approach in this study detects foreign objects on transmission lines, as shown by the two sets of data in the above table. The detection mAP values have been enhanced to varying degrees when compared to the original YOLOX algorithm. Compared to other popular target identification models (Fast R-CNN, SSD, YOLOv5, YOLOV4, and YOLOV7-S), the proposed method performs better. While providing high precision detection, the model's FPS does not decrease dramatically as accuracy increases, and the detection speed still has certain advantages over the mainstream approach. The visual effects of the proposed YOLOX detection method are shown in Figure 8.

5. Conclusions

In this research, we use the YOLOX method to detect foreign objects on transmission lines using the dataset of foreign objects on transmission lines. Using the ASPP structure based on hole convolution, the suggested strategy increases the network's ability to gather multidimensional input while extending the perceptual field without degrading sampling. The ASPP structure increases the perceptual field without sacrificing sampling and increases target identification accuracy while retaining the current inference speed, which enhances the network's ability to obtain multiscale contexts. Improve the way the CBAM attention mechanism is used to highlight the distinctive qualities of common alien objects. The GIoU loss function is used to improve detection accuracy since it more accurately illustrates the overlap between the prediction frame and the ground truth. The mosaic approach of training the model enhances the model's ability to detect objects in complex backgrounds and gives the model a more lifelike appearance. According to the experimental findings, the improved YOLOX-S model described in this paper performs better in terms of inference speed and detection accuracy and may be applied more successfully in the field of transmission line foreign object identification.

Then, we will focus on minimizing duplicate model components, removing network topologies that are unrelated to the domain used in this article, and increasing model recognition efficiency without sacrificing recognition accuracy. The improvement of the detecting head and the backbone feature extraction network parts has been the main focus of the effort to date. In order to further improve recognition accuracy, the upgraded feature extraction network section will then be changed in accordance with the most current feature extraction network pyramid findings.

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 is 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 (61703201), the Future Network Scientific Research Fund Project (FNSRFP2021YB26), and the Key Research and Development Plan of Hubei Province (2021BGD013), Special Project of Central Government for local Science and Technology Development of Hubei Province (2019ZYYD020), Science and Technology Research Program of Hubei Provincial Department of Education (T201805), Subsidy Project after Demonstration Project of Intellectual Property Application of Hubei Province (2019ZSCQ01), Hubei University of Technology Ph.D. Research Startup Fund Project (BSQD2020015), Open Foundation of Hubei Key Laboratory for High-Efficiency Utilization of Solar Energy and Operation Control of Energy Storage System (HBSEES202214), and Wuhan City and Other Sub-Provincial Cities Projects (2022010801020267).

References

[1] D. E. Yi-min, T. A. Zhi-qian, L. I. Hong-bing, and Y. A. Zhong-ya, "Real-time rendering algorithm optimization for large scale transmission lines based on LOD," *Computer and Modernization*, vol. 1, p. 115, 2017.

[2] O. M. Butt, M. Zulqarnain, and T. M. Butt, "Recent advancement in smart grid technology: future prospects in the electrical power network," *Ain Shams Engineering Journal*, vol. 12, no. 1, pp. 687–695, 2021.

[3] Z. Jing, C. Yu, F. Xi, F. Wu, Z. Tao, and P. Yang, "Reliability analysis of distribution network operation based on short-term future big data technology," *Journal of Physics: Conference Series*, vol. 1584, no. 1, article 012027, 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 4th Mediterranean Conference on Embedded Computing (MECO)*, pp. 159–162, Budva, Montenegro, 2015.

[6] K. Zou, Z. Jiang, and Q. Zhang, "Research progresses and trends of power line extraction based on machine learning," in *2021 2nd International Symposium on Computer Engineering and Intelligent Communications (ISCEIC)*, pp. 211–215, Nanjing, China, 2021.

[7] Y. Wu, Y. Luo, G. Zhao et al., "A novel line position recognition method in transmission line patrolling with UAV using machine learning algorithms," in *2018 IEEE International Symposium on Electromagnetic Compatibility and 2018 IEEE Asia-Pacific Symposium on Electromagnetic Compatibility (EMC/APEMC)*, pp. 491–495, Suntec City, Singapore, 2018.

[8] S. Qi, "Cleaning system based on autonomous patrol of UAV and intelligent detection of foreign matters," in *2020 7th International Forum on Electrical Engineering and Automation (IFEEA)*, pp. 675–678, Hefei, China, 2020.

[9] R. Girshick, "Fast R-CNN," in *Proceedings of the IEEE international conference on computer vision*, pp. 1440–1448, IEEE Computer Society,1730 Massachusetts Ave., NW Washington, DC,United States, 2015.

[10] 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, Venice, Italy, 2017.

[11] 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.

[12] Z. Ghasemi Darehnaei, M. Shokouhifar, H. Yazdanjouei, and S. M. J. R. Fatemi, "SI-EDTL: swarm intelligence ensemble deep transfer learning for multiple vehicle detection in UAV images," *Concurrency and Computation: Practice and Experience*, vol. 34, no. 5, article e6726, 2022.

[13] J. Basha, N. Bacanin, N. Vukobrat et al., "Chaotic Harris hawks optimization with quasi-reflection-based learning: an application to enhance CNN design," *Sensors*, vol. 21, no. 19, p. 6654, 2021.

[14] P. Singh, S. Chaudhury, and B. K. Panigrahi, "Hybrid MPSO-CNN: multi-level particle swarm optimized hyperparameters of convolutional neural network," *Swarm and Evolutionary Computation*, vol. 63, article 100863, 2021.

[15] A. M. Hilal, H. Alsolai, F. N. Al-Wesabi et al., "Fuzzy cognitive maps with bird swarm intelligence optimization-based remote sensing image classification," *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 4063354, 12 pages, 2022.

[16] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 779–788, Las Vegas, NV, USA, 2016.

[17] D. Wu, S. Lv, M. Jiang, and H. Song, "Using channel pruning-based YOLO v4 deep learning algorithm for the real-time and accurate detection of apple flowers in natural environments," *Computers and Electronics in Agriculture*, vol. 178, article 105742, 2020.

[18] J. Yu and W. Zhang, "Face mask wearing detection algorithm based on improved YOLO-v4," *Sensors*, vol. 21, no. 9, p. 3263, 2021.

[19] X. Zhu, S. Lyu, X. Wang, and Q. Zhao, "TPH-YOLOv5: improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 2778–2788, Nashville, TN, USA, 2021.

[20] C. Li, L. Li, H. Jiang et al., "YOLOv6: a single-stage object detection framework for industrial applications," 2022, https://arxiv.org/abs/2209.02976.

[21] C. Y. Wang, A. Bochkovskiy, and H. Y. M. Liao, "YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," 2022, https://arxiv.org/abs/2207.02696.

[22] X. Long, K. Deng, G. Wang et al., "PP-YOLO: an effective and efficient implementation of object detector," 2020, https://arxiv.org/abs/2007.12099.

[23] Y. Chen, L. Sun, Y. Zhang et al., "Research on bird detection technology for electric transmission line based on YOLO v3," *Computer Engineering*, vol. 46, no. 4, pp. 294–300, 2020.

[24] W. Wang, J. Zhang, J. Han, L. Liu, and M. Zhu, "Broken strand and foreign body fault detection method for power transmission line based on unmanned aerial vehicle image," *Journal of Computer Applications*, vol. 35, no. 8, p. 2404, 2015.

[25] S. Haotian, L. Tong, W. Pu, X. Liang, and Z. Hongwei, "Foreign object detection of electric transmission line with dynamic federated learning," *IOP Conference Series: Earth and Environmental Science*, vol. 791, no. 1, article 012159, 2021.

[26] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "YOLOX: exceeding YOLO series in 2021," 2021, https://arxiv.org/abs/2107.08430.

[27] Y. Wang, Y. Li, and K. Tu, "Multi-view convolutional neural networks crowd counting model based on YOLOX," in *2022 7th International Conference on Intelligent Computing and Signal Processing (ICSP)*, pp. 1616–1619, Xi'an, China, 2022.

[28] Z. Tian, C. Shen, H. Chen, and T. He, "FCOS: fully convolutional one-stage object detection," in *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 9627–9636, Seoul, Korea, 2019.

[29] M. Liu and C. Zhu, "Residual YOLOX-based ship object detection method," in *2022 2nd International Conference on Consumer Electronics and Computer Engineering (ICCECE)*, pp. 427–431, Guangzhou, China, 2022.

[30] J. Zhang and S. Ke, "Improved YOLOX fire scenario detection method," *Wireless Communications and Mobile Computing*, vol. 2022, Article ID 9666265, 8 pages, 2022.

[31] G. Wang, H. Zheng, and X. Zhang, "A robust checkerboard corner detection method for camera calibration based on improved YOLOX," *Physics*, vol. 9, p. 828, 2022.

[32] M. Zhang, C. Wang, J. Yang, and K. Zheng, "Research on engineering vehicle target detection in aerial photography environment based on YOLOX," in *2021 14th international symposium on computational intelligence and design (ISCID)*, pp. 254–256, Hangzhou, China, 2021.

[33] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1904–1916, 2015.

[34] H. Rezatofighi, N. Tsoi, J. Y. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized intersection over union: a metric and a loss for bounding box regression," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 658–666, Long Beach, California, 2019.

[35] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, and D. Ren, "Distance-IoU loss: faster and better learning for bounding box regression," *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, no. 7, pp. 12993–13000, 2020.

[36] W. Wang, Q. Lai, H. Fu, J. Shen, H. Ling, and R. Yang, "Salient object detection in the deep learning era: an in-depth survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 6, pp. 3239–3259, 2022.

[37] W. Wang, J. Shen, X. Lu, S. C. H. Hoi, and H. Ling, "Paying attention to video object pattern understanding," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 7, pp. 2413–2428, 2021.

[38] Y. Chen, L. Liu, V. Phonevilay et al., "Image super-resolution reconstruction based on feature map attention mechanism," *Applied Intelligence*, vol. 51, no. 7, pp. 4367–4380, 2021.

[39] R. Xia, Y. Chen, and B. Ren, "Improved anti-occlusion object tracking algorithm using Unscented Rauch-Tung-Striebel smoother and kernel correlation filter," vol. 34, Tech. Rep. 8, Journal of King Saud University-Computer and Information Sciences, 2022.

[40] Y. Chen, H. Zhang, L. Liu et al., "Research on image inpainting algorithm of improved total variation minimization method," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–10, 2021.