



## OPEN ACCESS

## EDITED BY

Yikui Liu,  
Stevens Institute of Technology, United States

## REVIEWED BY

Ge Cao,  
Xi'an University of Technology, China  
Wei Gao,  
University of Denver, United States  
Siyuan Chen,  
Wuhan University, China

## \*CORRESPONDENCE

Ziyong Ye,  
✉ zjlszy97@163.com

RECEIVED 17 September 2023

ACCEPTED 08 January 2024

PUBLISHED 19 January 2024

## CITATION

Zou H, Ye Z, Sun J, Chen J, Yang Q and Chai Y (2024), Research on detection of transmission line corridor external force object containing random feature targets. *Front. Energy Res.* 12:1295830. doi: 10.3389/fenrg.2024.1295830

## COPYRIGHT

© 2024 Zou, Ye, Sun, Chen, Yang and Chai. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Research on detection of transmission line corridor external force object containing random feature targets

Hongbo Zou<sup>1,2</sup>, Ziyong Ye<sup>2\*</sup>, Jialun Sun<sup>2</sup>, Junting Chen<sup>2</sup>,  
Qinhe Yang<sup>2</sup> and Yanhui Chai<sup>2</sup>

<sup>1</sup>Hubei Provincial Key Laboratory for Operation and Control of Cascaded Hydropower Station and New Energy, China Three Gorges University, Yichang, China, <sup>2</sup>College of Electrical Engineering and New Energy, China Three Gorges University, Yichang, China

With the objective of achieving “double carbon,” the power grid is placing greater importance on the security of transmission lines. The transmission line corridor has complex situations with external force targets and irregularly featured objects including smoke. For this reason, in this paper, the high-performance YOLOX-S model is selected for transmission line corridor external force object detection and improved to enhance model multi-object detection capability and irregular feature extraction capability. Firstly, to enhance the perception capability of external force objects in complex environment, we improve the feature output capability by adding the global context block after the output of the backbone. Then, we integrate convolutional block attention module into the feature fusion operation to enhance the recognition of objects with random features, among the external force targets by incorporating attention mechanism. Finally, we utilize EIoU to enhance the accuracy of object detection boxes, enabling the successful detection of external force targets in transmission line corridors. Through training and validating the model with the established external force dataset, the improved model demonstrates the capability to successfully detect external force objects and achieves favorable results in multi-class target detection. While there is improvement in the detection capability of external force objects with random features, the results indicate the need to enhance smoke recognition, particularly in further distinguishing targets between smoke and fog.

## KEYWORDS

transmission line corridor, external force, object detection, random feature targets, attention mechanism

## 1 Introduction

China is undergoing rapid modernization, and one of the fundamental aspects of this process is the expansion of the power transmission network. The electric power industry, as a crucial component of the national infrastructure, plays a significant role in the development of the country's economy. China has established explicit goals of achieving “carbon peaking” by 2030 and “carbon neutrality” by 2060. Consequently, the country will prioritize promoting adjustments in its industrial and energy structure. Additionally, the power grid will increasingly integrate a significant share of clean energy sources, such as wind and photovoltaic power (Hu, et al., 2022; Tian et al., 2022; Xiao and Zheng, 2022). As a result, there will be an increased need for secure operation of the transmission line

corridor. These transmission line corridors connect power production sources to the load, necessitating their secure operation. Given the vastness of China, variations in climate, and other environmental conditions, the establishment of transmission line corridor poses unique challenges. Consequently, the security of the transmission line corridor is frequently compromised. The destruction of transmission line is caused not only by fires but also by various other factors (Gu, et al., 2020; Sheng et al., 2021). Additionally, the presence of large tower cranes and engineering machinery also poses a significant risk. Statistical data indicates that the primary causes of transmission line tripping, in order, are lightning strikes, external force damage, wind deviation, and ice damage, among others. Among these causes, external force damage accounts for 21.4% of transmission line trips, second only to those caused by lightning strikes. Furthermore, an analysis of faults in transmission lines of 220 kV and above in the provincial power grid revealed that external force damage accounted for 12.36% of all transmission line tripping, resulting in a low success rate of tripping and reclosing at only 44.7% (Liang, 2014; Lu, et al., 2016). Additionally, the outage durations typically range between 2 and 3 h. It can lead to damage to the grid infrastructure, which can disrupt production, leading to avoidable economic losses. Additionally, it poses a risk to the safety of civilians, staff, and individuals in close proximity to the electrical equipment.

In current method for ensuring safety, manual inspection is predominantly utilized as the primary method for preventing external force damage. Nonetheless, this method is time-consuming, and the constraint of limited number of workers makes it challenging to continuous inspection (Wang, et al., 2019; Wang, et al., 2021; Ma, et al., 2022). The researcher conducted monitoring of external force in transmission lines using helicopter and drone video surveillance (Golightly and Jones, 2003; Larrauri et al., 2013; Lin, et al., 2019; Wei, et al., 2022). This method has the potential to enhance monitoring efficiency and reduce labor requirements. However, it is important to note that the patrol monitoring method is unable to provide real-time monitoring of moving targets within a specified area. Hence, it is not applicable for detecting unauthorized construction machinery operating within transmission line corridors. The reference (Zhang and Deng, 2020) obtains the external force vibration signal of transmission pole and the vibration signals of transmission towers under different wind excitation conditions. The vibration signals are preprocessed by adopting delay inlay technology to turn the original signal into a two-dimensional form and sent that into convolutional neural network to feature extraction, and achieve vibration pattern recognition by employ the relevance vector machine (Cui, et al., 2023). However, this method has limitations when it comes to recognizing different types of external force.

Detecting external force in transmission line corridor through image analysis enables the reduction of manual labor, effectively alerting against potential threats and preventing external force incidents. The continuous development and application of deep learning in the field of image recognition has proven invaluable for identifying external force damages in transmission lines (Krizhevsky, et al., 2012; Redmon, et al., 2016; Zhang, et al., 2018; Liu, et al., 2019; Ma, et al., 2021; Long, et al., 2022; Wu, et al., 2022; Dong, et al., 2023). The reference (Wei, et al., 2021) uses bounding box annotation instead of partial mask annotation in the process of dataset annotation and improves the average accuracy of

recognizing common categories of external force. The reference (Tian, et al., 2021) employs an enhanced K-means algorithm to determine suitable anchor box from the image. Subsequently, the CSP Darknet-53 residual network is used to extract the deep-seated network feature data of the images, and the feature map is processed by the SPP algorithm. The algorithm, when applied to real-time monitoring pictures of transmission line, demonstrates its ability to detect external force damage accurately and promptly.

Nowadays, the monitoring devices can collect and transfer image. The devices equipped with target recognition model enable target detection in real-time images. Nevertheless, most image monitoring systems solely recognize targets exhibiting distinct features, such as prominent engineering machinery and construction scenes. They often fail to identify targets with irregular characteristics like smoke. In this study, we enhance the performance of the YOLOX-S model by incorporating the global context block (GC block) to enhance the feature output capacity of the backbone network. Additionally, we introduce the convolutional block attention module (CBAM) in the feature fusion process to improve the recognition of randomly featured targets, among externally damaged objects. Finally, we employ the EIoU to enhance the accuracy of target detection box and achieve the detection of complex external force targets in transmission lines corridor.

## 2 Improved YOLOX-S

### 2.1 YOLOX-S

The YOLOX-S is a high-performance one-stage object detection network (Ge, et al., 2021). The network incorporates significant advancements in object detection, including decoupled heads, data augmentation, and anchor free, into the YOLO architecture. The model is composed of three main components. The backbone feature extraction network utilizes the CSP darknet architecture. The Neck enhances the feature extraction network through the use of the path-aggregation network (PANet). The prediction part employs three decoupled heads.

The backbone conducts low-level feature extraction on the input image, resulting in three feature layers. The Neck subsequently performs high-level feature extraction on these layers. Finally, three decoupled heads are employed for object detection, and the detection results are obtained accordingly.

The mosaic data augmentation algorithm is applied in the input layer. Its primary purpose is to combine four images, each accompanied by its respective box. Once the four images are spliced, a new image is generated, along with its corresponding box. Subsequently, the newly generated image is fed into the network for learning.

In addition to YOLOX-S, YOLOX has other types of networks, namely, YOLOX-Nano, YOLOX-Tiny, YOLOX-M, YOLOX-L, YOLOX-X. The performance of these models on the coco dataset is shown in Table 1. As shown in the table, YOLOX-Nano and YOLOX-Tiny are lightweight networks, so there are fewer network parameters and poorer performance at AP (%). In the rest models, the model accuracy improves as the model parameters increase, but the latency also increases. The application scenario in this study is

TABLE 1 Comparison of YOLOX in terms of AP (%) on COCO. YOLOX-Nano and YOLOX-Tiny are tested at 416 × 416 resolution, else are tested at 640 × 640 resolution.

Models	AP (%)	Parameters (M)	Latency (tested on tesla V100)
YOLOX-Nano	25.3	0.91	/
YOLOX-Tiny	32.8	5.06	/
YOLOX-S	39.6	9.0	9.8 ms
YOLOX-M	46.4	25.3	12.3 ms
YOLOX-L	50.0	54.2	14.5 ms
YOLOX-X	51.2	51.2	17.3 ms

characterized by two main features: 1) it is applied in outdoor environments, close to the edge end, and 2) the detection requires high real-time performance. For these two reasons, it is necessary to choose a base model that is moderate in model size, recognition accuracy and latency. The two lightweight models have fewer parameters and there is no dataset enhancement step in the training phase, which makes it difficult to meet the requirements for recognition accuracy. By comparing the rest of the models, YOLOX-S with fewer parameters is easier to deploy on devices used outdoors and meets the real-time requirements for recognition as well as the needs in detection accuracy.

## 2.2 Global context block

In transmission line corridors characterized by complex environments and a multitude of targets to be identified. The convolutional operation of the backbone network models the context within a limited range, and creating a confined receptive field. The global context block incorporates both the non-local network (NL Net) and squeeze excitation networks (SE Net). Additionally, in the GC block, the NL block is simplified to decrease the computational load. The GC block serves a dual purpose: extracting global information from the backbone convolutional network, thereby facilitating feature fusion with the linked neck part, and reducing computation cost. Therefore, incorporating the GC block into the network enables the extraction of global contextual information pertaining to external force targets in complex background (Cao, et al., 2019).

### 2.2.1 Simplified NL block

The NL block conducts inter-pixel correlation analysis by utilizing the current pixel in the given position along with feature-similar pixels of equal size to establish connections between features and global information. The output of the NL block  $z_i$  and the inter-pixel correlation  $w_{ij}$  are calculated using Eq. 1 and Eq. 2:

$$z_i = x_i + W_z \sum_{j=1}^{N_p} \frac{f(x_i, x_j)}{C(x)} (W_v \cdot x_j) \quad (1)$$

$$w_{ij} = \frac{f(x_i, x_j)}{C(x)} = \frac{e^{(W_q x_i)^T (W_k x_j)}}{\sum_{\forall m} e^{(W_q x_i)^T (W_k x_m)}} \quad (2)$$

Where,  $W_v, W_k, W_q$  and  $W_z$  represent the convolution operations.  $x_i$  represents the pixel at the current location, while  $N_p$  represents the set of all pixels. Additionally,  $x_j$  comprises pixels with similar features and an equal size as the location of  $x_i$ .  $f(x_i, x_j)$  denotes the relationship between position  $i$  and  $j$ , and has a normalization factor  $C(x)$ .

The NL block computational cost in the global feature extraction process necessitates structural simplification before fusion with the GC block. As can be seen in Figure 1, the main simplifying operations include: omitting any further operations on  $x_i$  and solely considering the global pixel-feature correlation; removing the  $W_q$  convolution operation; substituting  $W_z$  with  $W_v$ , and not retaining the convolution operation of  $W_v$ . The context modeling module is composed of the simplified NL block. The mathematical expression for the simplified block is presented in Eq. 3.

$$z_i = x_i + W_v \sum_{j=1}^{N_p} \frac{e^{W_k x_j}}{e^{W_k x_m}} x_j \quad (3)$$

### 2.2.2 Transform module

Simplifying the operation of the NL block decreases the computational effort of the module but has an impact on the accuracy of the training results. To address this accuracy loss and effectively utilize the feature information from the convolution operation on the channel, the transform module within the SE block is introduced. Additionally, incorporating layer normalization before applying the ReLU nonlinear activation can enhance the generalization capability of the network. The output of transform module  $z_{i,tran}$  is as Eq. 1:

$$z_{i,tran} = W_{v2} \text{ReLU} \left( \text{LN} \left( W_{v1} \left( \sum_{j=1}^{N_p} \frac{e^{W_k x_j}}{e^{W_k x_m}} x_j \right) \right) \right) \quad (4)$$

Where,  $W_{v1}, W_{v2}$  represent the convolution operations. ReLU represents the nonlinear activation function, and LN denotes layer normalization. The reconstruction of the transform module in the GC block is shown in Figure 2.

The GC block integrates the context modeling module from the simplified NL block, introduces the standardized transform module within the SE block, substitutes the  $W_v$  convolution operation in the simplified NL block, and conducts feature fusion through a sum operation. Figure 3 illustrates the final module, and its output is obtained as Eq. 5:

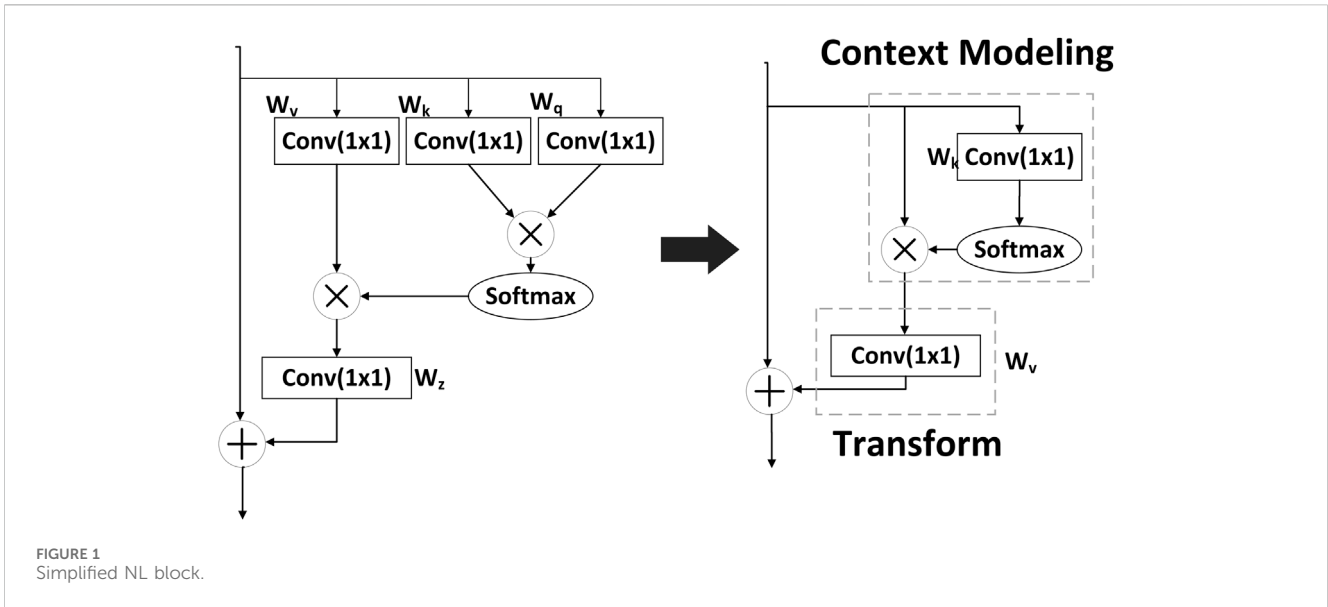


FIGURE 1 Simplified NL block.

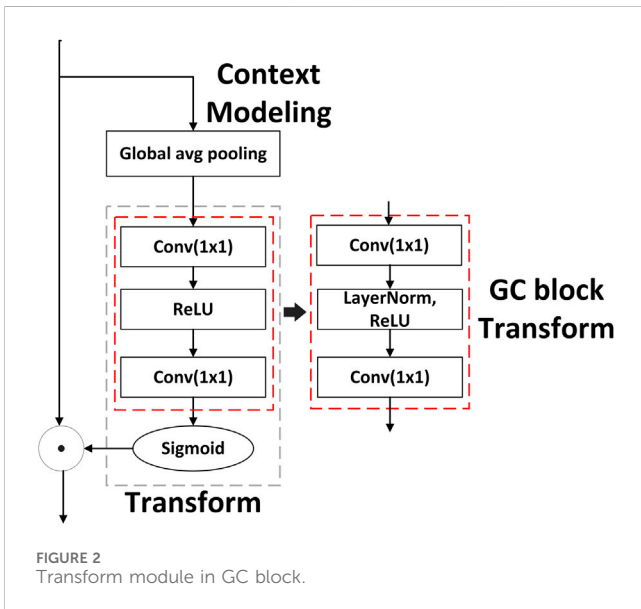


FIGURE 2 Transform module in GC block.

$$z_i = x_i + W_{v2} \text{ReLU} \left( \text{LN} \left( W_{v1} \left( \sum_{j=1}^{N_p} \frac{e^{W_k x_j}}{e^{W_k x_m}} x_j \right) \right) \right) \quad (5)$$

The GC block reduces computational parameters and computation by simplifying the NL block. The simplified block still learn the global context. The SE block adopts rescaling to recalibrate the importance of channels but inadequately models long-range dependency. Finally, the GC block completes the feature fusion by addition to the original feature map.

### 2.3 Convolutional block attention module

In the context of recognizing external force object in transmission lines corridor, the task involves identifying both

obvious features like engineering machinery and random features like smoke. To enhance the recognition ability for smoke, a convolutional attention mechanism is incorporated into the network. This module enables the network to simultaneously attend to multiple types of external force targets across channels and spatial dimensions during feature fusion.

The structure of the convolutional block attention module model incorporates the channel attention module (CAM) and the spatial attention module (SAM). Figure 4 depict the convolutional block attention module (Woo, et al., 2018). Channel attention focuses on the classification of the object in the image through the channel relations of the features. Max pooling can strengthen the unique object feature, so CAM completes the object feature extraction by average pooling and max pooling. Spatial attention focuses on the localization of the target in the image. Pooling operations along the channel axis can effectively highlight information regions, so average pooling and max pooling are applied along the channel axis in SAM. CAM and SAM are placed in a sequential manner, highlighting the location of the target in the image. The sequential arrangement enables the channel and spatial attention modules to achieve complementary attention and accomplish attentional enhancement.

#### 2.3.1 Channel attention module

The channel attention module starts by applying a global average pooling and a global max pooling method to the input's individual feature layers. The outcomes of these pooling methods are then fed into a shared fully connected layer. The results from both pooling methods are summed, and the Sigmoid activation function is applied to obtain a weighted value (from 0 to 1) for each channel in the incoming feature layer. After obtaining the weight, we multiply it by the original input feature layer to obtain the feature map processed by the channel attention module. The formula for the output of channel attention module  $M_c(F)$  is as Eq. 6:

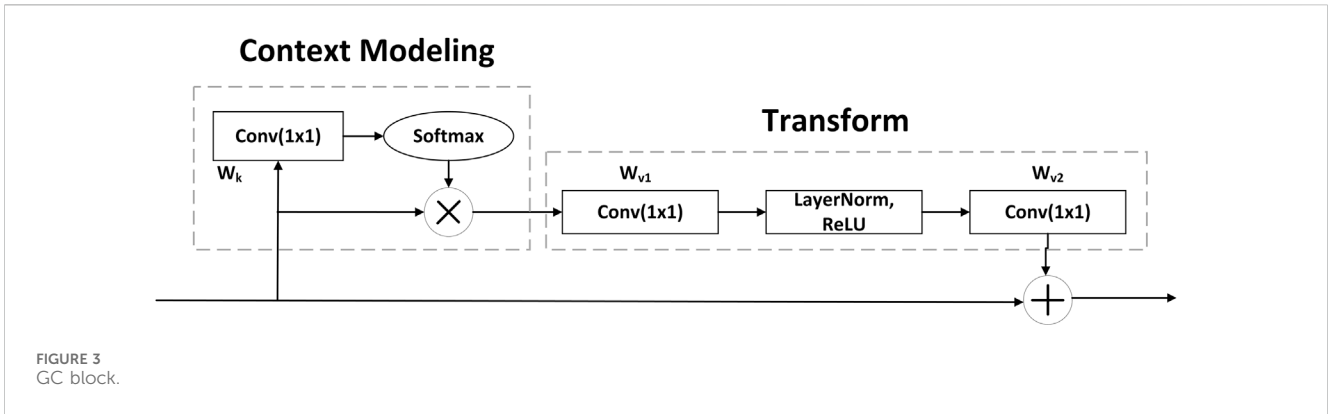


FIGURE 3 GC block.

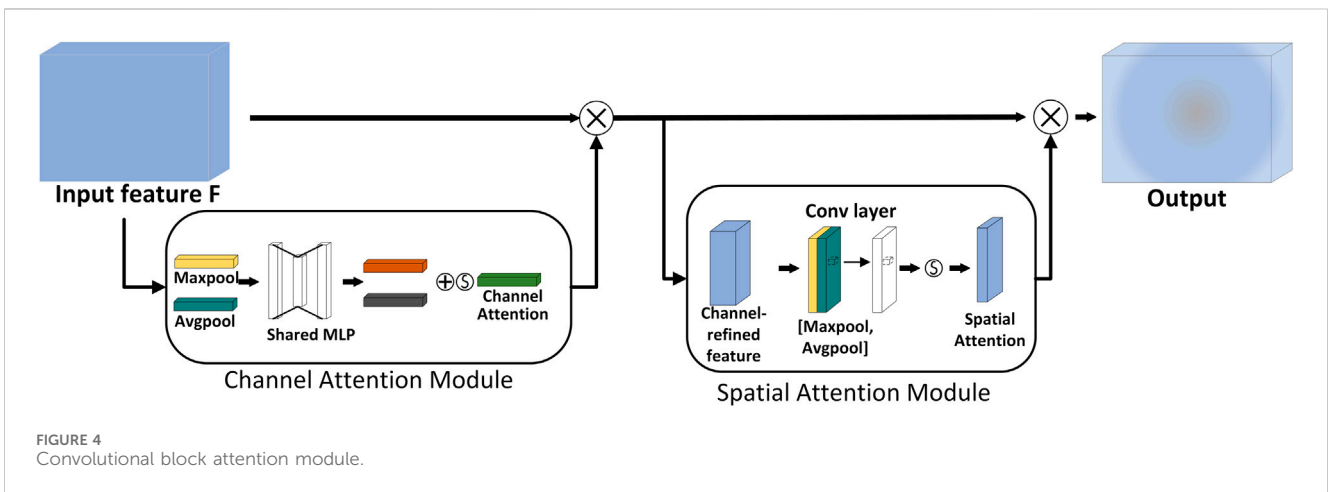


FIGURE 4 Convolutional block attention module.

$$M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) = \sigma(W_1(W_0(F_{avg}^c)) + W_1(W_0(F_{max}^c))) \quad (6)$$

Where,  $F$  represents input feature,  $\sigma$  represents the Sigmoid function,  $F_{avg}^c$  and  $F_{max}^c$  refer to the feature maps obtained through average pooling and max pooling, respectively, multilayer perceptron (MLP) is the shared network in CBAM, and  $W_1$  and  $W_2$  represent the weights of the hidden layer and output layer in the multilayer perceptron, respectively.

### 2.3.2 Spatial attention module

Max pooling and average pooling are performed over the feature layer channels which come from the channel attention module. The pooling results are concatenated, and the channel number is adjusted using a convolution kernel with the size of  $7 \times 7$ . After applying the Sigmoid activation function, the weights (from 0 to 1) for each feature of the input layer are obtained. The formula for the spatial attention module  $M_s(F')$  and the output of CBAM  $F''$  is as is as Eqs 7–9:

$$F' = M_c(F) \otimes F \quad (7)$$

$$M_s(F') = \sigma(f^{7 \times 7}([AvgPool(F'); MaxPool(F')])) = \sigma(f^{7 \times 7}([F_{avg}^s; F_{max}^s])) \quad (8)$$

$$F'' = M_s(F') \otimes F' \quad (9)$$

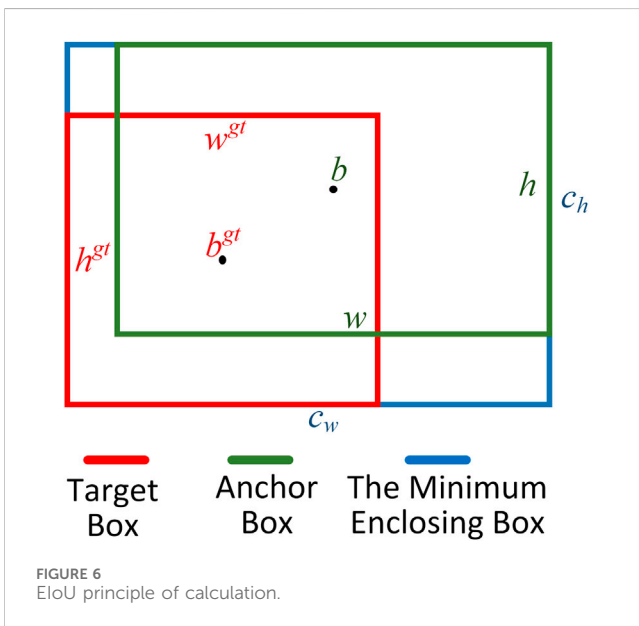
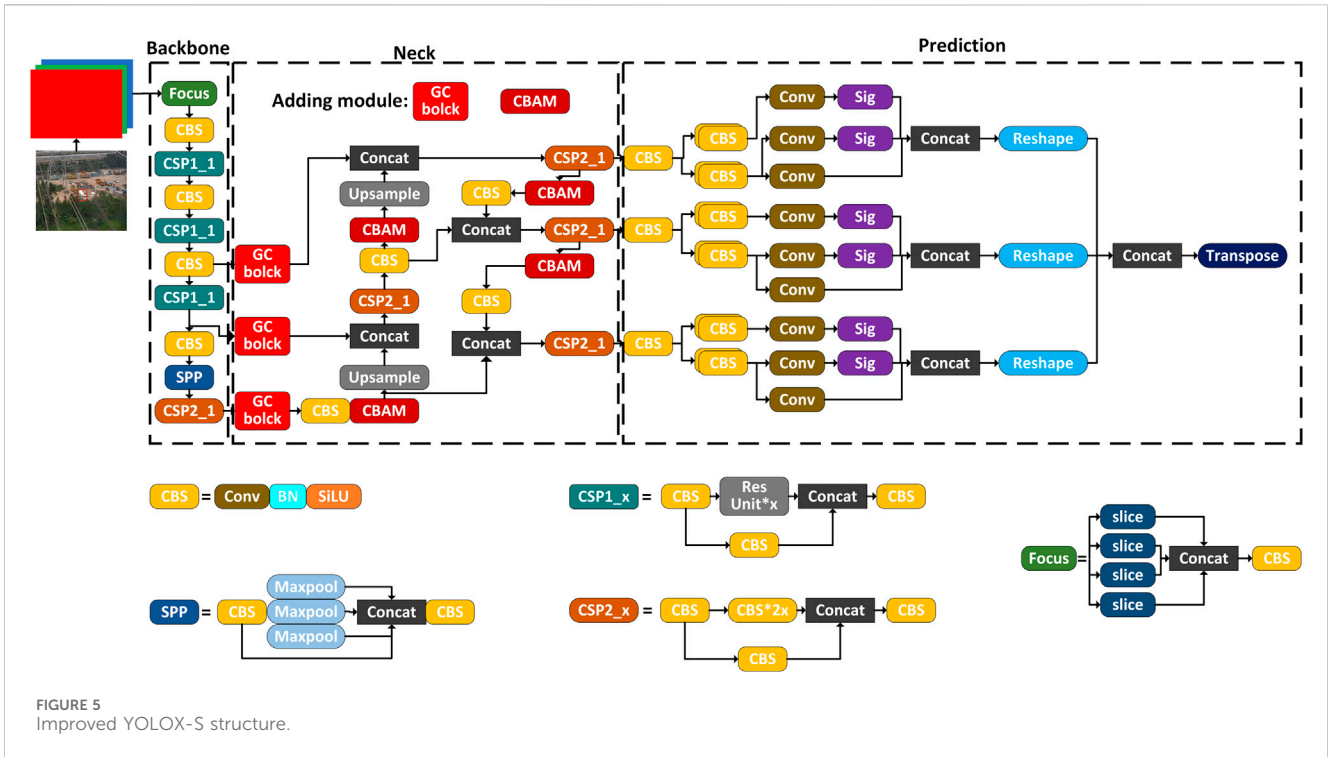
Where,  $F'$  represents channel-refined feature,  $\sigma$  donates the Sigmoid function,  $F_{avg}^s$  and  $F_{max}^s$  refer to the feature maps after two

pooling operations,  $f^{7 \times 7}$  represents a convolution operation with the filter size of  $7 \times 7$ .

In the improved YOLOX-S model, GC blocks are incorporated after the output of the backbone to augment the model's ability to perceive global features. Additionally, the CBAM is introduced to improve attention towards the external force target before the up-sampling in the feature fusion operation and before the feature contact. The modification leads to the final improved network structure of YOLOX-S, as depicted in Figure 5.

### 2.4 Loss function

The loss function in the YOLOX-S model is derived by combining the bounding box loss, object classification loss, and confidence loss. The IoU loss function is the earliest one used in the bounding box loss function. It is computed by taking the intersection over union ratio of target box and anchor box. In order to accurately represent the relative positions of the two boxes, the GIoU loss function was proposed. This function considers the non-box area by setting the minimum bounding rectangle that encloses both target box and anchor box. It comprehensively considers the overlapping areas between target box and anchor box. However, if the anchor box completely contains the target box, the GIoU will be



equivalent to the IoU. So, the paper focuses on optimizing the accuracy of the loss function by introducing the EIoU bounding box loss function, which provides a comprehensive description of the positional relationship between target box and anchor box (Zhang, et al., 2022). Figure 6 illustrates the principle of EIoU calculation. The EIoU loss function comprises three components: overlap loss, center distance loss, and width and height loss between the target box and the anchor box, which are computed by Eq. 10.

$$L_{EIoU} = L_{IoU} + L_{dis} + L_{asp} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{(c_w)^2 + (c_h)^2} + \frac{\rho^2(w, w^{gt})}{c_w^2} + \frac{\rho^2(h, h^{gt})}{c_h^2} \quad (10)$$

Where,  $b$  and  $b^{gt}$  represent the centroids of the anchor box and the target box, respectively;  $\rho(\cdot)$  denotes the Euclidean distance between the two centroids;  $c_w$  and  $c_h$  refer to the width and height of the smallest outer rectangle of the target box and the anchor box, respectively;  $w$  and  $h$  represent the width and height of the anchor box, while  $w^{gt}$  and  $h^{gt}$  represent the width and height of the target box.

### 3 Experiment

#### 3.1 Data set

In order to validate the effectiveness of the proposed model for recognizing the external force objects and smoke objects, transmission towers, excavators, bulldozers, concrete mixer, tower cranes, cranes, and engineering trucks; and smoke targets are selected as the recognition objects. Images captured by monitoring equipment in a specific province serve as the data source for constructing a sample library. This library comprises 15,485 images depicting the external force scenario of the transmission line corridor. We selected clear targets with diverse angles and backgrounds for labeling in the collected images. For smoke, smoke generated by a fire source was selected as a labeled target. A total of 4,003 images containing recognized object were labeled through object setting and screening, resulting in a total of

TABLE 2 A detailed description of image database of external force object.

Object	Label name	Number
Pole and Tower	ganta	2,587
Excavator	wajueji	2,183
Bulldozer	tuituji	849
Concrete mixer	jiaobanche	683
Tower crane	tadio	1,035
Crane	diaoche	840
Trucks	fandouche	1,359
Smoke	yanwu	593

10,129 labeled targets. The label names and corresponding numbers of these labeled targets are displayed in the Table 2.

The training and validation sets were derived from the image sample library, with 80% of the job images randomly assigned to the training set and the remaining 10% allocated to the validation set. Consequently, 3,202 images were used for training, while the validation set comprised 401 images. The rest of the images serve as a test set. In order to speed up training and prevent the weights from being corrupted, the model is trained with pre-trained weights and the 50 epochs are set to freeze training. The model was trained using python 3.7 and PyTorch on Ubuntu 18.04. The training was run on a single GeForce GTX 1080 with CUDA 10.1.

### 3.2 Visualization of attention maps

In order to demonstrate the enhanced effect on smoke which have random features, we utilize grad class activation map (Grad-CAM) for attention visualization (Selvaraju et al., 2017). The Grad-CAM is employed to depict the attention of target positions within

the convolutional layers. By pooling the average of gradients across the entirety of the final convolution layer, weights can be calculated for each channel. These weights are then applied to the feature map to generate a class activation map. The CAM assigns importance to each pixel in relation to the classification result. The Figure 7 depicts a comparison between the pre-improvement and post-improvement states (the darker middle portion of the figure indicates increased attention on the object). It is evident from the figure that the improved model extends attention to a wider range of features within the smoke region. This indicates that the improvement effectively enhances the model's feature extraction capability and enables improved extraction of irregular features.

### 3.3 Training result

The training process is shown in Figure 8. The loss function of the proposed model continuously decreases with iterative training, exhibiting a rapid decrease at the beginning and at the 50th epoch, then stabilizes around the 200th epoch, indicating good convergence performance. During the model training, the overall accuracy curve follows a similar trend as the loss function curve. The overall precision stabilizes around 200 epochs, reaching approximately 81.7%.

As shown in Figure 9, the experimental results demonstrate that the improved YOLOX-S achieves external force target recognition in complex scenarios. To validate the effectiveness of the improved model, a sample of images from the dataset is randomly selected and tested for detecting external force targets. The recognition results obtained using the original network and the improved network are presented below.

As depicted in the Figure 10, both YOLOX-S and the improved YOLOX-S successfully recognize the external force targets in scenario A. However, when compared to the YOLOX-S, the improved YOLOX-S exhibits superior confidence in detecting each target. Nonetheless, the results from the YOLOX-S network contain instances of missed detections. The missed targets

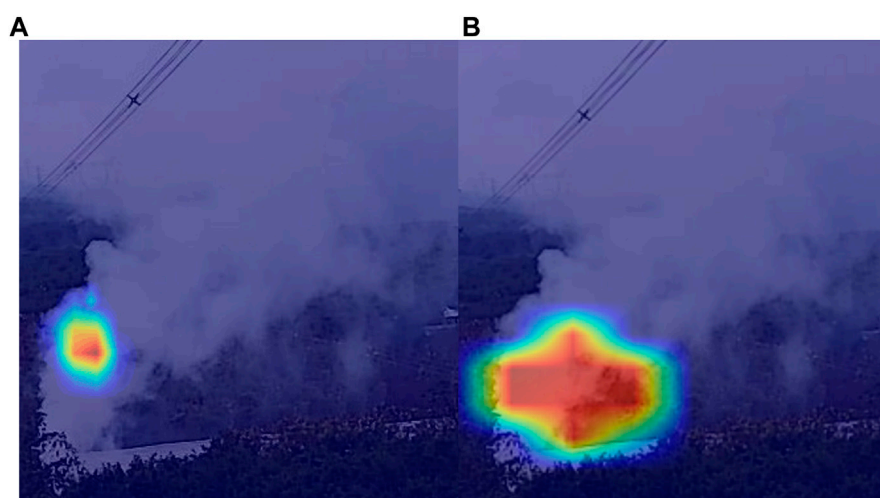


FIGURE 7 Grad-CAM attention map. (A) YOLOX-S, (B) Improved YOLOX-S.

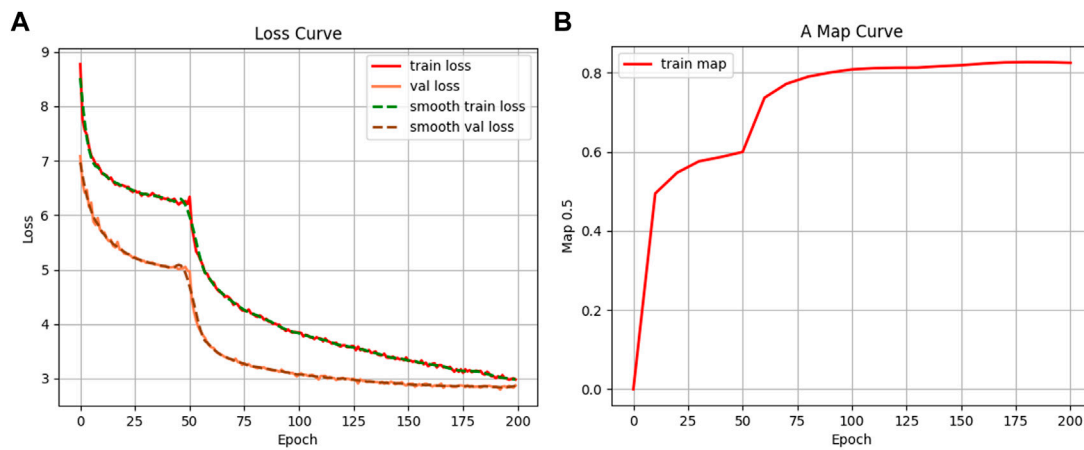


FIGURE 8 Variation curve of loss function (A) and accuracy curve (B) in training process.



FIGURE 9 External force object detection result.

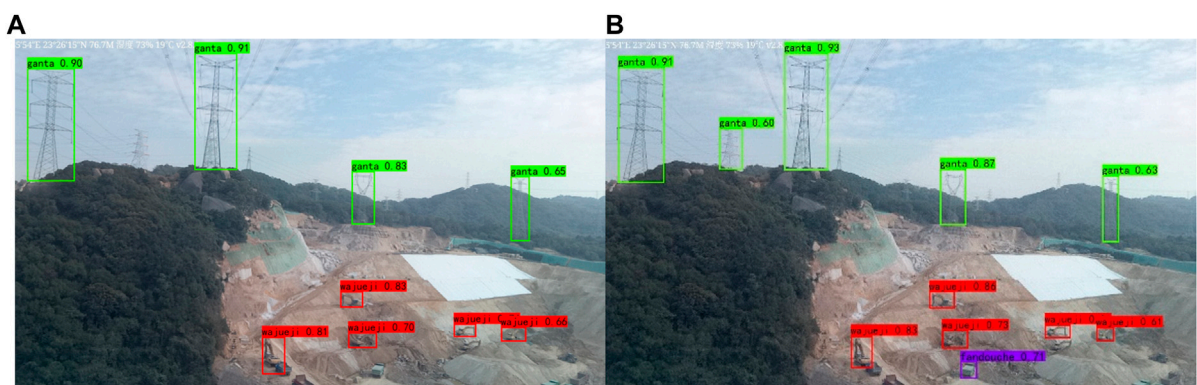


FIGURE 10 External force object detection result in scenario (A) YOLOX-S detection result (B) Improved YOLOX-S detection result.

correspond to smaller external force objects (trucks) or objects with color features resembling the environment (towers), resulting in missed detections. These findings highlight the efficacy of the global context block and attention module in enhancing target recognition.

For comparison, the two-stage network Faster R-CNN is selected (Ren, et al., 2015). In external force scenario B, both Faster R-CNN and the improved YOLOX-S successfully recognize all the targets in the Figure 11. However, the original network exhibits a missed detection for the smaller





FIGURE 11 External force object detection result in scenario B (A) Original image (B) Faster R-CNN detection result (C) YOLOX-S detection result (D) Improved YOLOX-S detection result.

target at a distant location (Excavator). Conversely, the double-stage network misdetects the closer bulldozer target, while both the original network and the improved YOLOX-S accurately detect the target. The improved YOLOX-S demonstrates higher confidence. The Table 3 presents the average recognition precision of various network for identifying targets. The table reveals that the improved YOLOX-S achieves superior average precision in identifying external force object. The improved YOLOX-S surpasses the Faster RCNN in detecting all external force recognition targets. The improved YOLOX-S enhances the recognition precision of each external force target compared to the original network. For engineering apparatus targets such as excavators and bulldozers, target recognition with rich feature information can be improved by 4%–5%. Due to the variety of tower types and differences in their structure, the recognition accuracy improvement is not very good, only 2.09%. Specifically, there is a 3.39% improvement in accuracy in recognizing smoke. This suggests that the inclusion of the GC block and CBAM in the network can enhance recognition accuracy for irregularly featured targets. The improved network demonstrates slightly higher overall recognition accuracy compared to Yolov7 in external force scenarios (Wang et al., 2023). However,

Yolov7 achieves slightly higher accuracy in concrete mixer recognition. The improved YOLOX-S outperforms the other three networks in terms of recognizing smoke.

Smoke exhibits varying concentrations and profiles at different stages. The Figure 12 demonstrates that the improved YOLOX-S effectively captures the smoke and locates its main components. This aids in evaluating the potential risks of smoke within transmission corridors.

To further evaluate the model's recognition precision of targets in the transmission corridors, an ablation study is conducted. The ablation study compares the impact of each improvement component on object recognition precision. The Table 4 illustrates that GC block enhances network precision by 1.38%, while CBAM improves it by 2.24%. GC block enables targets in complex outdoor contexts to be attended to globally, slightly improving recognition accuracy. The attention mechanism improves the model recognition accuracy more, making the model pay more attention to the external broken target in both channel and space. Conversely, EIou has a minor impact on precision, affecting it by merely 0.45%. The EIou is mainly designed to describe the position of the anchor and target boxes well and provide more help to the model training process.

TABLE 3 Comparison of detection results under different target detection models.

Model	Average recognition precision of different external force targets (AP%)								mAP %
	Pole and tower	Excavator	Bulldozer	Concrete mixer	Tower crane	Crane	Trucks	Smoke	
Faster RCNN	80.65	87.37	88.21	77.36	83.45	76.26	76.37	58.13	78.47
YOLOX-S	81.34	85.16	87.11	76.73	84.63	75.46	74.78	57.72	77.87
YOLOv7	82.12	87.63	88.42	79.24	86.24	77.48	78.19	58.21	79.69
Improved YOLOX-S	83.43	90.77	91.96	78.44	88.93	79.35	79.91	61.11	81.74

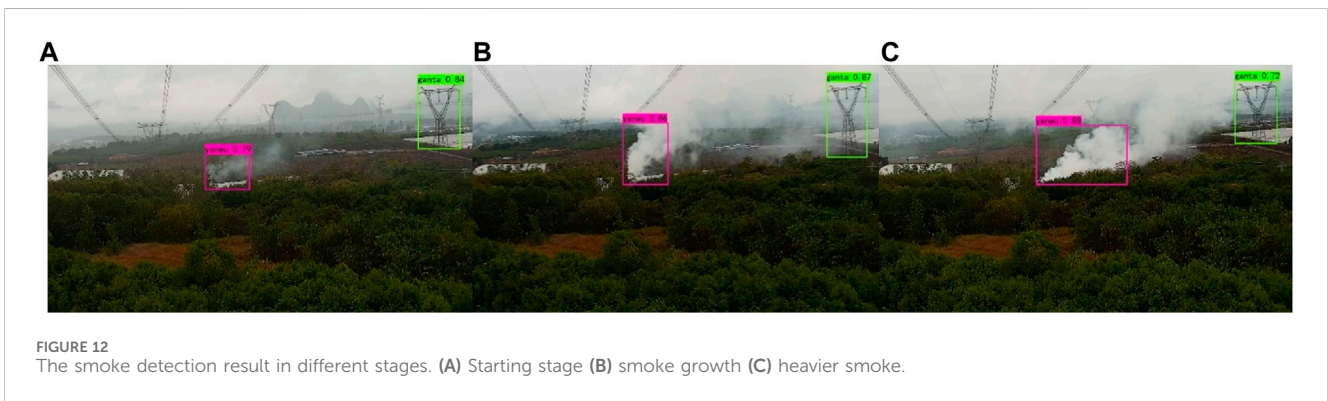


FIGURE 12 The smoke detection result in different stages. (A) Starting stage (B) smoke growth (C) heavier smoke.

TABLE 4 Ablation study.

GC block	CBAM	EIoU	mAP%
\	\	\	77.87
✓	\	\	79.25
\	✓	\	80.11
✓	✓	\	81.29
✓	✓	✓	81.74

## 4 Conclusion

This study focuses on the recognition of external force targets in transmission line corridors, characterized by complex backgrounds, various object types, and irregular features. To achieve this, the model was improved. Conclusions can be drawn based on the model training and recognition results obtained from the dataset:

(1) The improved YOLOX-S effectively identifies external force targets, exhibiting superior performance in complex environments and multi-target scenarios. The enhancements provided by the module improve global perception and target recognition capabilities, particularly for targets with distinct features. The recognition precision of such targets is enhanced by approximately 3% compared to the original model.

(2) The improved network significantly enhances the recognition precision of smoke with random features. The attention heat map generated by Grad CAM demonstrates that the improved module effectively focuses on irregular targets, further refining the network's ability to recognize smoke with non-uniform characteristics. The model demonstrates improved performance in recognizing smoke at various stages and exhibits enhanced tracking capabilities for dynamically changing smoke.

This study incorporates the global context block, attention mechanism, and a new loss function to enhance the YOLOX-S network, resulting in an improved ability to recognize external force object. In light of the aforementioned research, the ensuing research will concentrate on two primary objectives. Firstly, to enhance the recognition precision of external force object and broaden the spectrum of recognizable types. Secondly, to develop an effective assessment of threats to external force transmission lines corridor and conduct an in-depth analysis of their operational environment. Specifically, it was found that the accuracy of smoke recognition needs to be further improved. With frequent changes in ambient wind direction, it is difficult to identify and localize the location of smoke generation. The situation is similar to natural fog, so there is a need to more clearly distinguish between fog and smoke and to more precisely locate smoke. We will analyze the effect on the light reflection effect in terms of the difference between the material composition and composition ratio of smoke and fog, and to use it as a new feature input to distinguish between smoke and fog. Next, we will use the number of various types of external force in the image, the distance and the working range of engineering machinery as the main transmission line external force threat assessment basis.

## Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

## Author contributions

HZ: Writing–review and editing. ZY: Writing–original draft. JS: Writing–review and editing. JC: Writing–review and editing. QY: Writing–review and editing. YC: Writing–review and editing.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research is supported by a project from Yunnan Provincial Department of

## References

- Cao, Y., Xu, J., Lin, S., Wei, F., and Hu, H. (2019). GCNet: non-local networks meet squeeze-excitation networks and beyond, IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), Seoul, Korea (South), 11–17 Oct. 2021. IEEE, 1971–1980. doi:10.1109/ICCVW.2019.00246
- Cui, Y., Fang, C., Wen, Z., Fang, M., You, H., and Guo, J. (2023). Vibration signal identification of external force failure based on time-frequency spectrum and adaptive dynamic weight PSO-CNN algorithm. *Foreign Electron. Meas. Technol.* 42 (01), 144–152. doi:10.19652/j.cnki.femt.2204368
- Dong, Z., Gao, Y., Yuan, B., Yao, X., Zhang, J., and Zeng, J. (2023). An external damage detection method of transmission lines based on improved YOLOv4. *Power Syst. Clean Energy* 39 (06), 17–25. doi:10.3969/j.issn.1674-3814.2023.06.003
- Ge, Z., Liu, S., Wang, F., Li, Z., and Sun, J. (2021). YOLOX: exceeding yolo series in 2021. arXiv preprint, 2021: arXiv:2107.08430. doi:10.48550/arXiv.2107.08430
- Golightly, I., and Jones, D. (2003). Corner detection and matching for visual tracking during power line inspection. *Image & Vis. Comput.* 21 (9), 827–840. doi:10.1016/S0262-8856(03)00097-0
- Gu, K., Xia, Z., Qiao, J., and Lin, W. (2020). Deep dual-channel neural network for image-based smoke detection. *IEEE Trans. Multimedia* 22 (2), 311–323. doi:10.1109/TMM.2019.2929009
- Hu, J., Xu, X., Ma, H., and Yan, Z. (2022). Distributionally robust Co-optimization of transmission network expansion planning and penetration level of renewable generation. *J. Mod. Power Syst. Clean Energy* 10 (3), 577–587. doi:10.35833/MPCE.2021.000156
- Krizhevsky, A., Ilya, S., and Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Commun. ACM* 60, 84–90. doi:10.1145/3065386
- Larrauri, I. J., Sorrosal, G., and González, M. (2013). “Automatic system for overhead power line inspection using an Unmanned Aerial Vehicle-RELIFO project,” in International Conference on Unmanned Aircraft Systems, Atlanta, GA, June 13–14 2013 (IEEE), 244–252. doi:10.1109/ICUAS.2013.6564696
- Liang, Z. (2014). Statistical analysis of transmission line fault tripping in state grid corporation of China in 2011–2013. *East China Power* 42 (11), 2265–2270.
- Lin, G., Wang, B., Peng, H., Wang, X., Chen, S., and Zhang, L. (2019). Multi-target detection and location of transmission line inspection image based on improved Faster-RCNN. *Electr. Power Autom. Equip.* 39 (05), 213–218. doi:10.16081/j.issn.1006-6047.2019.05.032
- Liu, J., Huang, H., Zhang, Y., Luo, J., and He, J. (2019). Deep learning based external-force-damage detection for power transmission line. *2018 3rd Int. Conf. Commun. Image Signal Process. Sanya, China IOPscience* 1169 (1), 012032. doi:10.1088/1742-6596/1169/1/012032
- Long, L., Zhou, L., Liu, S., Huang, B., and Fan, K. (2022). Identification of hidden damage targets by external forces based on domain adaptation and attention mechanism. *J. Electronic Meas. Instrum.* 36 (11), 245–253. doi:10.13382/j.jemi.B2205638
- Lu, J., Zhou, T., Wu, C., Li, B., Tan, Y., and Zhu, Y. (2016). Fault statistic and analysis of 220kV and above power transmission line in province-level power grid. *High. Volt. Eng.* 42 (01), 200–207. doi:10.13336/j.1003-6520.hve.2016.01.026
- Ma, F., Wang, B., Dong, X., Yao, Z., and Wang, H. (2022). Safety image interpretation of power industry: basic concepts and technical framework. *Proc. CSEE* 42 (02), 458–475. doi:10.13334/j.0258-8013.psee.210315
- Ma, F., Wang, X., and Zhou, Y. (2021). Receptive field vision edge intelligent recognition for ice thickness identification of transmission line. *Power Syst. Technol.* 45 (06), 2161–2169. doi:10.13335/j.1000-3673.pst.2020.1015
- Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). “You only look once: unified, real-time object detection,” in Conference on Computer Vision and Pattern Recognition (CVPR), NV, USA, 17 Jun, 2024 (IEEE), 779–788. doi:10.1109/CVPR.2016.91
- Ren, S., He, K., Girshick, R., and Sun, J. (2015). Faster r-cnn: towards real-time object detection with region proposal networks. *Adv. neural Inf. Process. Syst.* 39, 1137–1149. doi:10.1109/TPAMI.2016.2577031
- Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2017). “Grad-CAM: visual explanations from deep networks via gradient-based localization,” in IEEE International Conference on Computer Vision (ICCV), Venice, Italy, October 2–3, 2023 (IEEE), 618–626. doi:10.1109/ICCV.2017.74
- Sheng, D., Deng, J., and Xiang, J. (2021). Automatic smoke detection based on SLIC-DBSCAN enhanced convolutional neural network. *IEEE Access* 9, 63933–63942. doi:10.1109/ACCESS.2021.3075731
- Tian, E., Li, C., Zhu, G., Su, Z., Zhang, X., and Xu, X. (2021). Identification algorithm of transmission line external hidden danger based on YOLOv4. *Comput. Syst. Appl.* 30 (7), 190–196. doi:10.15888/j.cnki.csa.008082
- Tian, K., Sun, W., and Han, D. (2022). Strategic investment in transmission and energy storage in electricity markets. *J. Modern Power Syst. Clean Energy* 10 (1), 179–191. doi:10.35833/MPCE.2020.000927
- Wang, B., Ma, F., Dong, X., Wang, P., Ma, H., and Wang, H. (2019). Electric power depth vision: basic concepts, key technologies and application scenarios. *Guangdong Electr. Power* 32 (09), 3–10. doi:10.3969/j.issn.1007-290X.2019.009.001
- Wang, B., Ma, F., Ge, L., Ma, H., Wang, H., Mohamed, M. A., et al. (2021). Icing-EdgeNet: a pruning lightweight edge intelligent method of discriminative driving channel for ice thickness of transmission lines. *IEEE Trans. Instrum. Measurement* 70, 1–12. Art no.2501412. doi:10.1109/TIM.2020.3018831
- Wang, C. Y., Bochkovskiy, A., and Liao, H. Y. M. (2023). “YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors,” in IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, Fri Jun 21st, 2024 (IEEE), 7464–7475. doi:10.1109/CVPR52729.2023.00721
- Wei, Y., Li, M., Xie, Y., and Dai, B. (2022). Transmission line inspection image detection based on improved Faster-RCNN. *Electr. Power Eng. Technol.* 41 (02), 171–178. doi:10.12158/j.2096-3203.2022.02.023

Science and Technology Major Science and Technology Special Projects (202202AD080004).

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Wei, Z., Lu, W., Zhao, W., and Wang, D. (2021). Target detection method for external damage of a transmission line based on an improved Mask R-CNN algorithm. *Power Syst. Prot. Control* 49 (23), 155–162. doi:10.19783/j.cnki.pspc.210482
- Woo, S., Park, J., Lee, J. Y., and Kweon, I. S. (2018). “Cbam: convolutional block attention module,” in *Proceedings of the European conference on computer vision (ECCV)* (Munich, Germany: Springer), 3–19. doi:10.48550/arXiv.1807.06521
- Wu, M., Guo, L., Chen, R., Du, W., Wang, J., Liu, M., et al. (2022). Improved YOLOX foreign object detection algorithm for transmission lines. *Wirel. Commun. Mob. Comput.* 2022, 1–10. doi:10.1155/2022/5835693
- Xiao, X., and Zheng, Z. (2022). New power systems dominated by renewable energy towards the goal of emission peak & carbon neutrality: contribution, key techniques, and challenges. *Adv. Eng. Sci.* 54 (1), 46–59. doi:10.15961/j.jsuese.202100656
- Zhang, A., and Deng, F. (2020). External vibration identification scheme of transmission line tower based on CNN-rvm. *Comput. Simul.* 36 (04), 76–80.
- Zhang, J., Yu, J., Wang, J., and Tan, S. (2018). Image recognition technology for transmission line external damage based on depth learning. *Comput. Syst. Appl.* 27 (8), 176–179. doi:10.15888/j.cnki.csa.006458
- Zhang, Y., Ren, W., Zhang, Z., Jia, Z., Wang, L., and Tan, T. (2022). Focal and efficient IOU loss for accurate bounding box regression. *Neurocomputing* 506, 146–157. doi:10.1016/j.neucom.2022.07.042