

Article

Lightweight Transmission Line Outbreak Target Obstacle Detection Incorporating ACmix

Junbo Hao ^{1,*}, Guangying Yan ², Lidong Wang ², Honglan Pei ², Xu Xiao ³ and Baifu Zhang ⁴¹ State Grid Shanxi Integrated Energy Service Co., Ltd., Taiyuan 030031, China² State Grid Yuncheng Electric Power Supply Company, Yuncheng 044099, China;

yanguangying@sx.sgcc.com.cn (G.Y.); wanglidong@sx.sgcc.com.cn (L.W.); peihonglan@sx.sgcc.com.cn (H.P.)

³ State Grid Gaoping Electric Power Supply Company, Gaoping 048499, China; xiaoxu@sx.sgcc.com.cn⁴ School of Electrical and Power Engineering, Taiyuan University of Technology, Taiyuan 030024, China; zhangbaifu00126@126.com

* Correspondence: hjb00126@126.com

Abstract: To address challenges such as the frequent misdetection of targets, missed detections of multiple targets, high computational demands, and poor real-time detection performance in the video surveillance of external breakage obstacles on transmission lines, we propose a lightweight target detection algorithm incorporating the ACmix mechanism. First, the ShuffleNetv2 backbone network is used to reduce the model parameters and improve the detection speed. Next, the ACmix attention mechanism is integrated into the Neck layer to suppress irrelevant information, mitigate the impact of complex backgrounds on feature extraction, and enhance the network's ability to detect small external breakage targets. Additionally, we introduce the PC-ELAN module to replace the ELAN-W module, reducing redundant feature extraction in the Neck network, lowering the model parameters, and boosting the detection efficiency. Finally, we adopt the SIoU loss function for bounding box regression, which enhances the model stability and convergence speed due to its smoothing characteristics. The experimental results show that the proposed algorithm achieves an mAP of 92.7%, which is 3% higher than the baseline network. The number of model parameters and the computational complexity are reduced by 32.3% and 44.9%, respectively, while the detection speed is improved by 3.5%. These results demonstrate that the proposed method significantly enhances the detection performance.

Keywords: external breakage obstacles; ACmix attention; ShuffleNetv2 network; lightweight



Academic Editor: Vijaya Raghavan

Received: 9 December 2024

Revised: 26 December 2024

Accepted: 16 January 2025

Published: 18 January 2025

Citation: Hao, J.; Yan, G.; Wang, L.; Pei, H.; Xiao, X.; Zhang, B. Lightweight Transmission Line Outbreak Target Obstacle Detection Incorporating ACmix. *Processes* **2025**, *13*, 271. <https://doi.org/10.3390/pr13010271>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The monitoring and maintenance of transmission lines are critical for ensuring the stable operation of power systems and public safety [1]. However, most transmission line wires are exposed in the field, and the environmental conditions around tower installation sites are often complex. These lines are vulnerable to the disturbances caused by factors such as adverse weather, natural disasters, and external damage. With the rapid pace of urbanization, the frequency of faults caused by external damage has been increasing year by year. When transmission lines are damaged due to external factors (e.g., fallen trees, vehicle impacts, or construction mishaps), the success rate of circuit breaker reclosing is low, which significantly undermines the safe and stable operation of the power grid [2]. Therefore, quickly and accurately identifying and detecting the potential hazards to transmission lines caused by external damage has become an urgent issue.

Currently, common external breakage obstacles to transmission lines include trees, mechanical equipment, wild animals, and flying objects. The inspection methods can be

broadly classified into the following two categories: line patrol and online monitoring. The line patrol method includes manual and drone patrols [3,4]. Manual patrols require personnel to walk or use transport to inspect the line, which involves significant resource and manpower investment. This method is particularly inefficient in complex areas where transport cannot be used. Drone patrols, on the other hand, utilize drones equipped with high-definition cameras, thermal imagers, and other sensors to conduct aerial inspections. While the cost is lower, it still relies on the manual inspection of images to detect potential hazards, which reduces the efficiency. The online monitoring method involves installing video surveillance equipment at key nodes for real-time monitoring, coupled with computer vision technology to extract and analyze the data [5]. This approach is an important research direction for preventing external damage to transmission lines [6].

Video surveillance and computer vision technologies for detecting external breakage obstacles to transmission lines face numerous challenges. On one hand, these obstacles are often located in complex, diverse environments, and the types of obstacles are highly variable, and frequently small, which leads to issues such as misdetection or missed detection. On the other hand, existing models tend to have large computational loads and poor real-time detection performance, making them unsuitable for real-time monitoring across a wide range of scenarios. To address these issues, this paper proposes a lightweight external damage obstacle detection algorithm. The main contributions are as follows:

- (1) The lightweight ShuffleNetv2 network is incorporated into the backbone of the YOLOv7 model, reducing the number of model parameters and enhancing the detection speed.
- (2) The ACmix attention mechanism module is embedded into the Neck layer of the network, strengthening the model's feature extraction and integration capabilities, thereby improving the recognition accuracy of small external breakage targets.
- (3) A PC-ELAN module is designed by replacing the standard convolution in the ELAN-W module of the original Neck network with partial convolution (PConv). This modification reduces the influence of irrelevant information on feature learning, decreases the computational costs, and improves the detection efficiency.
- (4) The SIOU loss function is introduced to reduce unstable gradient variations, provide a more stable training process, and accelerate the model's convergence.

The following sections present a detailed description of the methodology used in this paper, including the network architecture design and experimental strategy. Section 2 provides a literature review of the relevant research and introduces the original YOLOv7 model structure. Section 3 details the improved network model and the individual modules proposed in this study. Section 4 outlines the experimental setup, including the environment configuration, dataset, and evaluation metrics. It also presents results from various experiments, including attention experiments, loss function experiments, ablation studies, and comparative experiments. Finally, Section 5 summarizes the key findings, discusses the limitations, and suggests directions for future research.

2. Related Work

2.1. Detection of External Breakage Obstacles Outside Transmission Lines

Target detection algorithms are generally classified into the following two categories based on their design and working principles: two-stage detection networks and single-stage detection networks. Two-stage detection networks first generate candidate regions, followed by region classification and bounding box regression. Examples include algorithms such as Faster R-CNN [7], Mask R-CNN [8], and Cascade R-CNN [9]. In contrast, single-stage detection networks perform target detection and bounding box regression simultaneously within a unified framework, such as the YOLO series [10–15] and SSD [16].

In the complex environment of transmission lines, two-stage detection networks are characterized by a high computational complexity and slow inference speed, making them unsuitable for real-time applications. Additionally, these networks are prone to errors in candidate region detection, especially in complex backgrounds, leading to false or missed detections. On the other hand, while single-stage detection networks are faster, they often struggle with detecting small target obstacles caused by external damage. These networks are also more susceptible to interference from complex backgrounds and may have limited performance when handling multiple targets simultaneously. Thus, improving the detection accuracy for small and multiple targets in the challenging environment of transmission lines, while balancing computational complexity and real-time performance, remains a critical area of research.

Deep learning algorithms, particularly those based on convolutional neural networks (CNNs), have demonstrated a strong potential for detecting external breakage obstacles in transmission lines [17,18]. Zhang Ji et al. [19] applied the Faster R-CNN algorithm to target the identification of transmission line external breach hazards, showing that image recognition technology can effectively detect breach risks. Wei Xianzhe et al. [20] utilized an improved Mask-RCNN network for transmission line external breakage detection, migrating detection branch features to the mask branch, which offered a novel approach for accurately identifying and segmenting external breakage hazard targets. Tian Ersheng et al. [21] employed an enhanced K-means algorithm for target size clustering analysis and applied the YOLOv4 algorithm to detect external breakage hidden targets, improving the identification accuracy. Zheng Hanbo et al. [22] proposed a YOLO-2MCS-based hidden target detection method for transmission line corridors, incorporating a hybrid data augmentation strategy to improve the model's ability to detect multi-scale targets. Sun Yang et al. [23] introduced a transmission line foreign object detection algorithm based on channel pruning, effectively reducing the model size and improving the detection efficiency. Long Leyun et al. [24] enhanced YOLOv5 by incorporating a self-attention module for improved feature extraction and used a multi-scale domain-adaptive network for adversarial learning to boost the model's generalization ability. Wang Yanhai et al. [25] proposed an improved YOLOv7-based target detection algorithm for mechanical breach hazards, integrating the Swin Transformer attention mechanism to enhance the multi-scale feature extraction and using depthwise separable convolution to reduce the model computational costs, resulting in a superior detection accuracy and model efficiency compared to other mainstream algorithms.

Although the above algorithms have shown success in detecting breakage obstacles on transmission lines, they still face challenges in dealing with the diverse and complex environments of transmission line corridors. Due to significant background noise interference, targets are often indistinguishable from the natural surroundings, making it difficult for the model to accurately distinguish between relevant targets and irrelevant background. Additionally, small targets are more likely to be overlooked or misdetected, especially at long distances or from restricted viewing angles, leading to a reduced detection accuracy. Furthermore, when multiple targets appear in the same frame, occlusion and deformation between targets exacerbate the problem of missed detections. In conclusion, existing research has not fully addressed the issues of multi-target misdetection and small target detection in complex environments. Therefore, further exploration is needed to improve the feature extraction, network architecture design, and multi-target differentiation to enhance the accuracy and robustness of transmission line hazard detection.

2.2. YOLOv7 Network Structure

YOLOv7 is an end-to-end target detection model that introduces several optimizations compared to YOLOv5, particularly in the network structure, data augmentation, and activation functions. The specific network structure is shown in Figure 1. YOLOv7 enhances the model's feature learning capabilities through the introduction of a new network architecture called E-ELAN. This architecture maintains the integrity of gradient paths while optimizing feature extraction and fusion, thereby accelerating model convergence and improving its robustness in handling complex scenes and diverse targets. Additionally, YOLOv7 incorporates RepConv, a re-partitioned convolutional layer [26], in the Head layer of the prediction Head. This adjustment modifies the number of output feature channels, optimizing the computational efficiency and enhancing the model's expressive power.

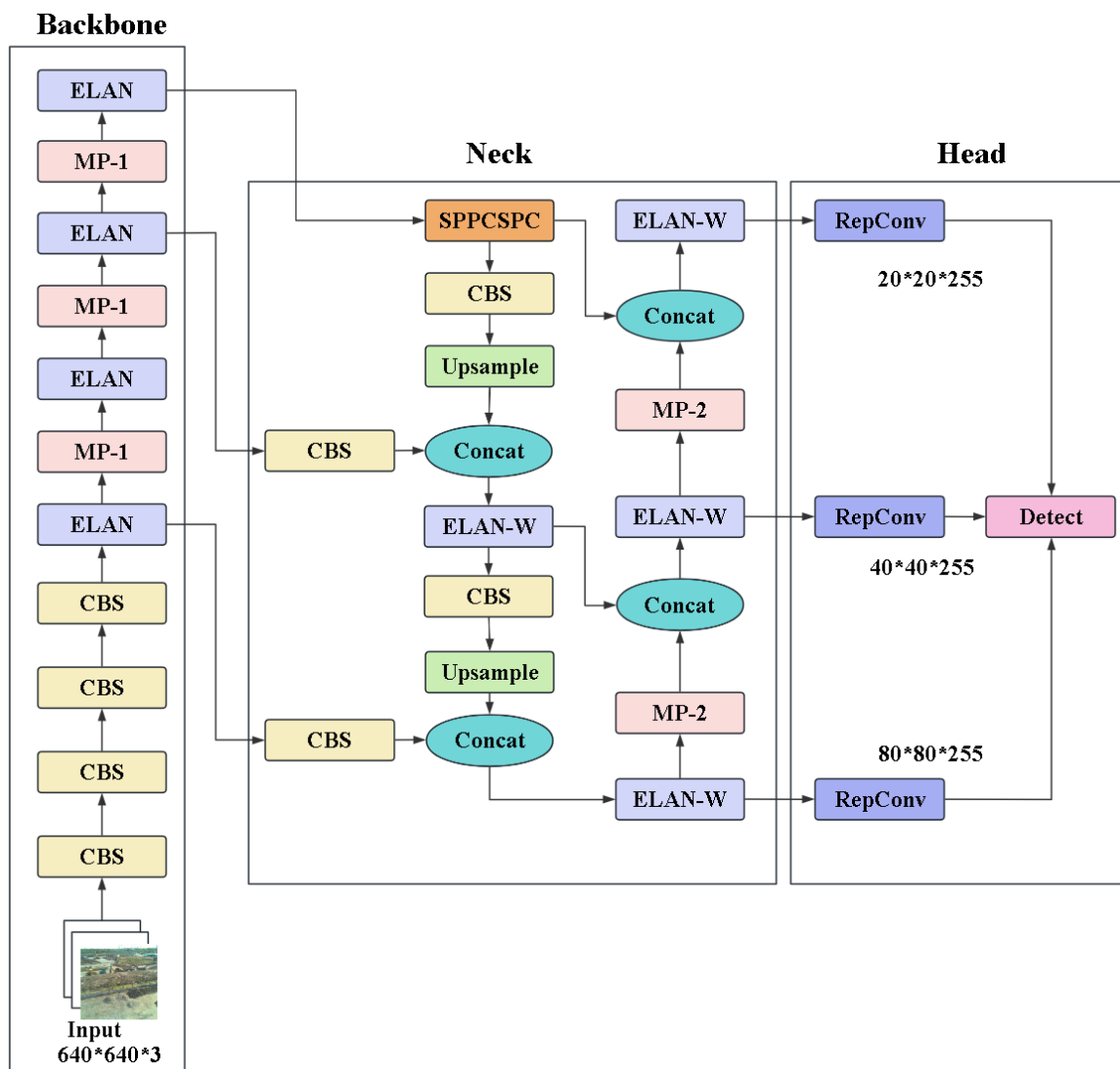


Figure 1. The YOLOv7 network structure. It consists of the backbone, Neck, and Head, where the CBS module consists of the Conv, BN, and SiLU activation functions to extract basic features. The ELAN and ELAN-W modules are used to enhance the gradient flow and multi-scale feature learning, while the SPPCSPC modules integrate the Spatial Pyramid Pooling (SPP) and Cross-Stage Partial (CSP) architectures to improve the feature representation. MP-1 and MP-2 are subsampling modules, which can effectively transmit multi-scale information. The RepConv module is used to classify and locate objects.

3. Proposed Methodology

The network structure of YOLOv7 is relatively complex, requiring significant computational resources, which can lead to the overlooking or misdetection of small targets in complex backgrounds. To address these issues, this paper introduces ShuffleNetv2 as the backbone network for the original YOLOv7 architecture and embeds the ACmix attention mechanism module into the Neck layer. Additionally, we designed the PC-ELAN module to replace the ELAN-W module in the Neck network, optimizing the feature extraction and fusion. Finally, the SIOU loss function is adopted for edge regression, enabling the lightweight and accurate detection of external breakage obstacles. The results of the improved network are shown in Figure 2.

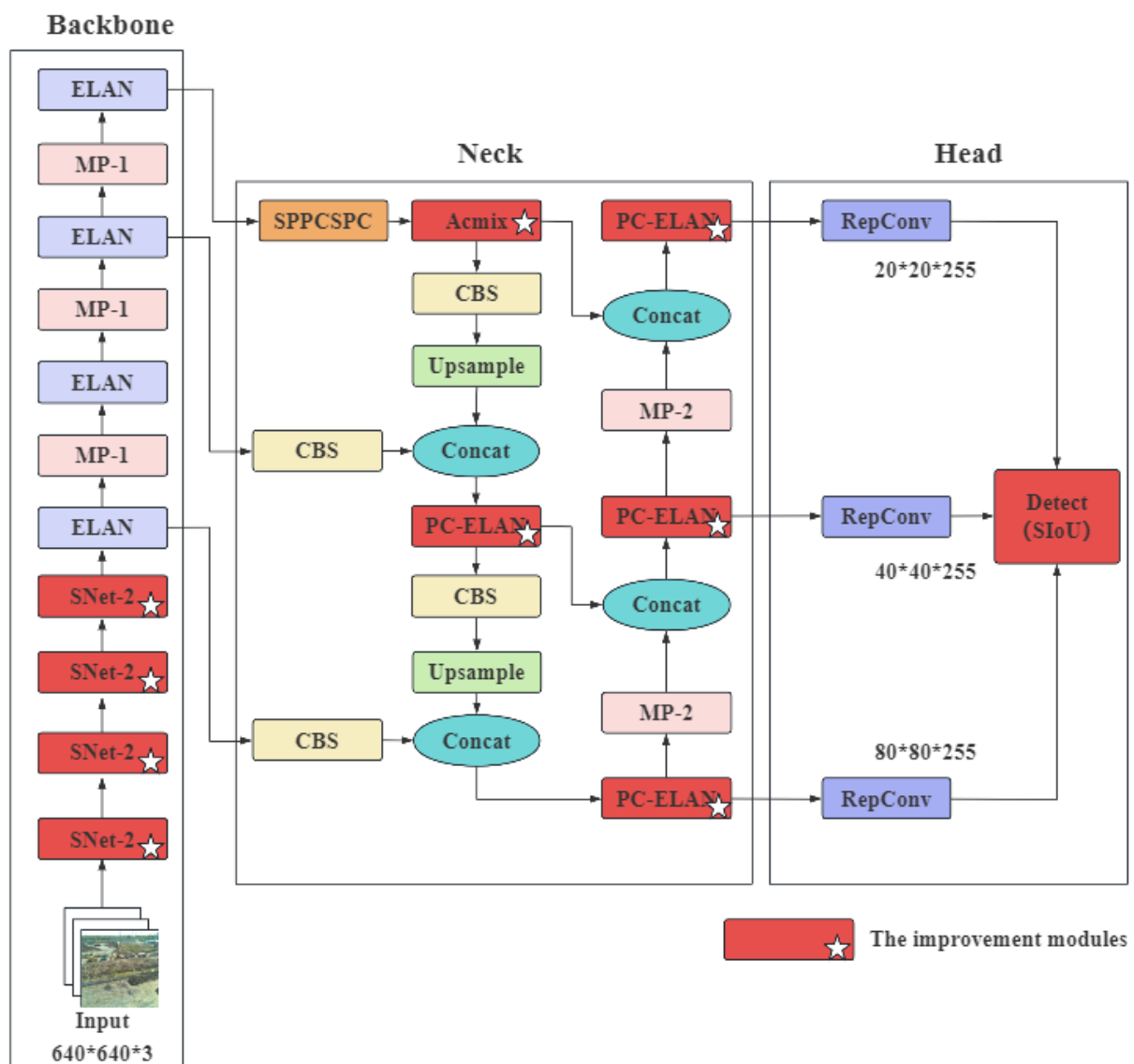


Figure 2. Improved network structure diagram. SNet-2 is a ShuffleNetv2 base unit, which realizes efficient and lightweight feature extraction through a channel grouping and mixing mechanism. ACmix is a newly added attention mechanism that highlights key information areas by dynamically weighting them. PC-ELAN is an ELAN module improved by PConv, which effectively reduces redundant calculations in the feature extraction process. The SIOU loss function is used in Detect to improve the location accuracy of the target.

3.1. Backbone Network Lightweighting Based on ShuffleNetv2

YOLOv7 uses CSP-Darknet53 as its backbone network, which has a large number of model parameters and a slower detection speed. To optimize this, this paper introduces ShuffleNetv2 [27] as a lightweight network to replace the YOLOv7 backbone, reducing the number of parameters and improving the detection efficiency.

ShuffleNetv2 is a lightweight convolutional neural network, and Figure 3 shows its basic unit. This network reduces the computation complexity and the number of parameters while maintaining a high accuracy. Compared to ShuffleNetv1 [28], ShuffleNetv2 optimizes the network structure by introducing channel segmentation. In this operation, the input feature map is split into several groups, with each group undergoing convolution independently, and the resulting feature maps are then reassembled. The process begins with the channel segmentation of the input feature map, which is divided into two branches. The left branch performs no operation, while the right branch contains two standard convolutions and one depth-separable convolution. Finally, the two branches are fused using a Concat operation. This channel mixing operation allows for the exchange of information between feature maps from different groups, reducing the computational complexity and making more efficient use of hardware resources, thereby improving the computational efficiency.

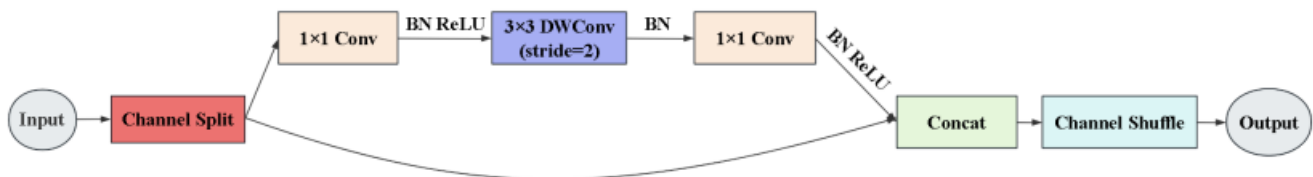


Figure 3. ShuffleNetv2 basic unit.

3.2. Embedded ACmix Attention Mechanism to Capture Global Information

The ACmix attention mechanism [29] (shown in Figure 4) combines traditional convolution with a self-attention mechanism to generate the final feature representation by aggregating the output features from both the convolution path and the self-attention path. This mechanism merges the local feature extraction capability of convolution with the global contextual awareness of self-attention, improving the model's ability to focus on small targets in complex environments. In traditional convolution, the convolution operation effectively captures local features and can control the fine-grainedness of these features by adjusting parameters such as the kernel size and stride. In the self-attention module, multi-Head self-attention is computed through intermediate features, allowing the model to focus on different regions of the input and to capture global information more effectively.

$$\tilde{g}_{ij}^{(p,q)} = K_{p,q} f_{i,j} \quad (1)$$

$$g_{ij}^{(p,q)} = \text{Shift} \left(\tilde{g}_{ij}^{(p,q)}, p - \left\lfloor \frac{k}{2} \right\rfloor, q - \left\lfloor \frac{k}{2} \right\rfloor \right) \quad (2)$$

$$g_{ij} = \sum_{p,q} g_{ij}^{(p,q)} \quad (3)$$

where K represents the kernel size; $K_{p,q}$ represents the center position in terms of (p, q) and K is the nuclear inch scale; f and g represent the input and output feature maps, respectively; $f_{i,j}$, $g_{i,j}$ represent the feature tensor of the corresponding pixel (i, j) for f and g , respectively; (p, q) represents the kernel position.

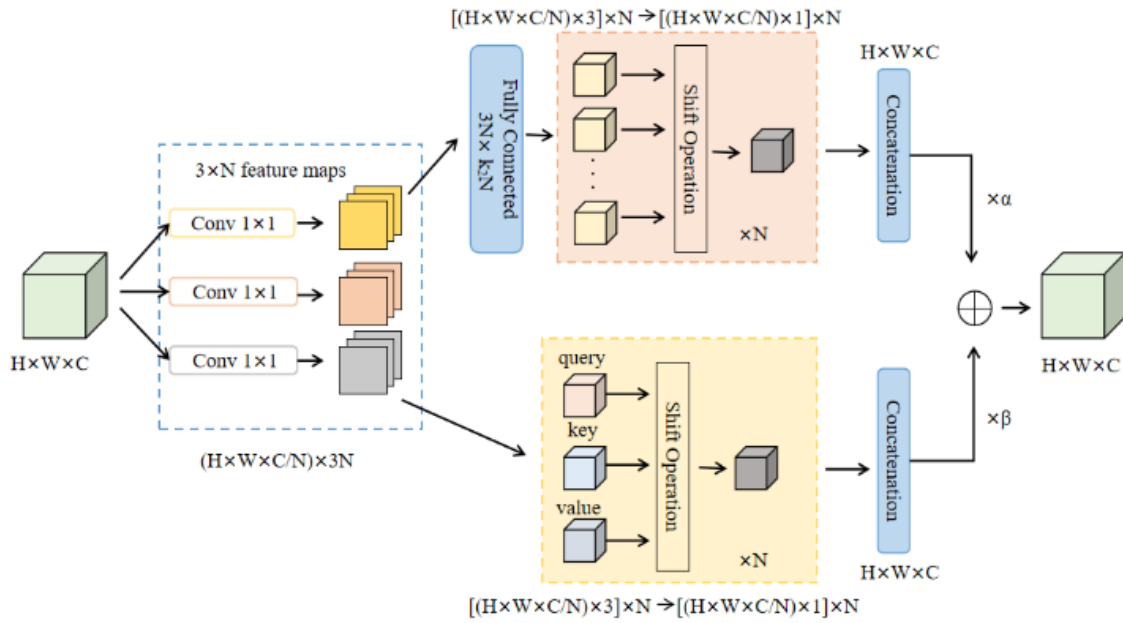


Figure 4. ACmix network structure.

The self-attention mechanism determines the attention weights by dynamically calculating the similarity between relevant pixels using a weighted average operation on the input feature context. This allows the attention module to adaptively focus on different regions, expand its receptive field, and capture more contextual information. As a result, it can more effectively distinguish between the background and the target, capturing more useful features. The computation process can be divided into the following two phases. In the first phase, the input is projected into query and value matrices via 1×1 convolutions, with the features transformed into queries, keys, and values. In the second phase, attention weights are computed, and the value matrices are aggregated. For a standard self-attentive module with N heads, the input tensor is $F \in \mathbb{R}^{C_{in} \times H \times W}$, its input tensor is a , and its output tensor is $G \in \mathbb{R}^{C_{out} \times H \times W}$, where H and W denote the height and width, and $f_{ij} \in \mathbb{R}^{C_{in}}$ and $g_{ij} \in \mathbb{R}^{C_{out}}$ are pixels (i, j) corresponding to F and G , and its output computation can be expressed as

$$g_{ij} = \prod_{l=1}^N \left(\sum_{a,b \in N_k(i,j)} A(W_q^{(l)} f_{ij}, W_k^{(l)} f_{ab}) W_v^{(l)} f_{ab} \right) \quad (4)$$

where $W_q^{(l)}, W_k^{(l)}, W_v^{(l)}$ represent the projection matrices for the queries, keys, and values, respectively; $N_k(i, j)$ denotes the local region centered at (i, j) with a pixel space range of k ; and $\prod_{l=1}^N$ is the concatenation of the outputs from N attention heads. The self-attention weights are calculated as

$$A(W_q^{(l)} f_{ij}, W_k^{(l)} f_{ab}) = \text{softmax} \left(\frac{(W_q^{(l)} f_{ij})^T (W_k^{(l)} f_{ab})}{\sqrt{d}} \right) \quad (5)$$

In summary, the above can be decomposed into two stages and reformulated as the following:

$$q_{ij}^{(l)} = W_q^{(l)} f_{ij}, k_{ij}^{(l)} = W_k^{(l)} f_{ij}, v_{ij}^{(l)} = W_v^{(l)} f_{ij} \quad (6)$$

$$g_{ij} = \prod_{l=1}^N \left(\sum_{a,b \in N_k(i,j)} A(q_{ij}^{(l)}, k_{ab}^{(l)}) v_{ab}^{(l)} \right) \quad (7)$$

Ultimately, the output of the ACmix module is the sum of the results from the two paths, as shown in Equation (8), where α and β represent the learning parameters for convolution and self-attention, respectively, with default values of 1. This fusion process preserves the convolution's sensitivity to local features while incorporating the self-attention mechanism's ability to capture global features. This enhances the model's feature extraction and expression capabilities.

$$F_{\text{out}} = \alpha F_{\text{conv}} + \beta F_{\text{att}} \quad (8)$$

3.3. Designing PC-ELAN Modules to Reduce Memory Consumption

The PC-ELAN module replaces the standard convolution in the ELAN-W of the original YOLOv7 Neck network with a lightweight partial convolution (PConv). The structure of the partial convolution (PConv) is shown in Figure 5, where * represents the convolution operation. In this approach, when performing spatial feature extraction with the convolution kernel, consecutive channels in the front or back segments are selected to represent the entire feature map, while the remaining channels are preserved. This reduces the network size while maintaining efficient spatial feature extraction.

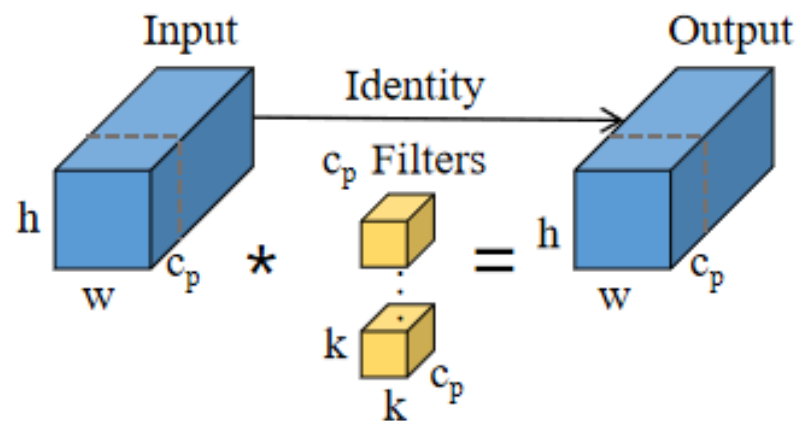


Figure 5. PConv structure.

Compared to standard convolution, PConv applies convolution to only a portion of the feature map, effectively reducing the computational load and memory usage while maximizing the device's computational power. The FLOP formula for PConv [30] is provided in Equation (9), as follows:

$$\text{FLOPs} = h \times w \times k^2 \times c_p^2 \quad (9)$$

where h and w represent the height and width of the feature map, respectively, while k denotes the size of the convolution kernel, corresponding to the number of channels processed by the regular convolution.

3.4. Improving Convergence Speed Using SIoU Loss Function

The coordinate loss function in YOLOv7 is computed using the Complete Intersection over Union (CIoU), as follows:

$$L_{\text{CIoU}} = 1 - \text{IoU} + \frac{\rho^2(\mathbf{b}, \mathbf{b}^{\text{gt}})}{c^2} + \alpha v \quad (10)$$

$$\alpha = \frac{v}{(1 - \text{IoU}) + v} \quad (11)$$

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right)^2 \quad (12)$$

where c denotes the diagonal length of the smallest enclosing rectangle that contains both the predicted and actual bounding boxes. This is commonly used as a measure of the spatial relationship and enclosing properties between the two frames; ρ denotes the Euclidean distance between the centroids of the predicted and actual boxes, quantifying the spatial distance between them; α is a positive trade-off coefficient; v is used as a measure of the consistency in aspect ratios.

The CIoU builds upon the DIoU [31] by adding an additional penalty term to update the loss related to the scale aspect of the bounding boxes. This improves the stability of the target box regression and mitigates the divergence problem that may arise during the training of the IoU [32] and GIoU [33]. However, while the CIoU takes into account the centroid distance, overlap area, aspect ratio, and width-to-height ratio, it does not fully account for true aspect ratio differences or the directional mismatch between the predicted and actual boxes. This limitation restricts the model's ability to optimize the similarity, thus reducing the detection efficiency.

To address these limitations, the loss function is improved by replacing the CIoU with the SIoU [34] to accelerate the convergence. The SIoU loss function incorporates scale sensitivity and angular considerations. It adjusts the predicted box to be more accurately aligned with the target position, reduces the degrees of freedom, and reflects the effect of the width and height on the confidence more realistically. Furthermore, during bounding box regression, the SIoU considers the rotational alignment between the predicted and true boxes, correcting the directional mismatches and improving the model's performance compared to the CIoU. Specifically, the SIoU incorporates the vector angles between the real and predicted boxes and redefines the loss function with the following four components: the angle loss, distance loss, shape loss, and IoU loss.

(1) Angle loss

The angular loss function part, which optimizes the loss calculation by considering the angle of the vectors between the real and predicted frames, is defined as shown in Figure 6 and Equation (13) below:

$$\Lambda = 1 - 2 * \sin^2 \left(\arcsin \left(\frac{C_h}{\sigma} \right) - \frac{\pi}{4} \right) \quad (13)$$

where C_h is the height difference between the center point of the real frame and the predicted frame, σ is the distance between the center point of the real frame and the predicted frame, and $\arcsin \left(\frac{C_h}{\sigma} \right)$ is the angle α .

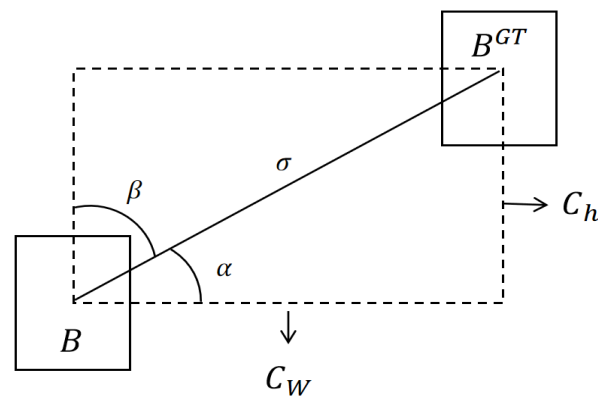


Figure 6. Angle loss. The scheme for the calculation of the angle cost contribution into the loss function.

(2) Distance loss

Distance loss is used to measure the distance between the true box and the predicted box. The purpose of the distance loss is to accelerate the convergence and improve the performance by minimizing the normalized distance between the centroids of the two bounding boxes, as defined in Figure 7 and Equation (14) below:

$$\Delta = \sum_{t=x,y} (1 - e^{-\gamma \rho_t}) \quad (14)$$

where $\rho_x = \left(\frac{b_{cx}^{gt} - b_{cx}}{c_w} \right)^2$, $\rho_y = \left(\frac{b_{cy}^{gt} - b_{cy}}{c_h} \right)^2$, $\gamma = 2 - \Lambda$, and c_w, c_h are the width and height of the smallest outer rectangle of the real and predicted boxes.

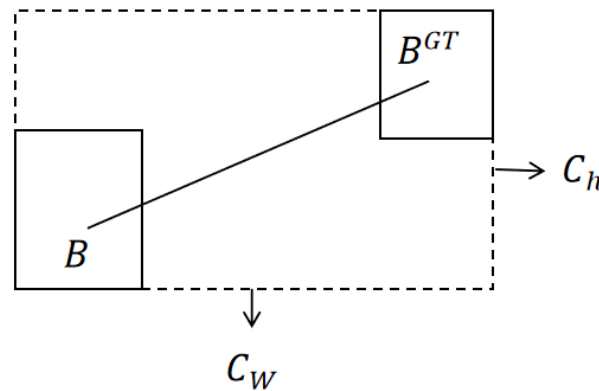


Figure 7. Distance loss. Scheme for the calculation of the distance between the ground truth bounding box and the prediction of it.

(3) Shape loss

The shape loss helps the model to better match the geometric properties of the real and predicted frames by considering the height information of the bounding box, which is defined by Equation (15), as follows:

$$\Omega = \sum_{t=w,h} (1 - e^{-\omega_t})^\theta \quad (15)$$

where $\omega_w = \frac{|w - w^{gt}|}{\max(w, w^{gt})}$, $\omega_h = \frac{|h - h^{gt}|}{\max(h, h^{gt})}$, and θ controls the amount of attention paid to the shape loss.

(4) IoU loss

The IoU is the ratio of the area of the intersection of the prediction frame and the real frame (ground truth) to the area of concatenation, defined in Figure 8 and Equation (16), as follows:

$$IoU = \frac{|B \cap B^{GT}|}{|B \cup B^{GT}|} \quad (16)$$

In summary, the final SIoU loss function is defined as follows:

$$L_{\text{box}} = 1 - IoU + \frac{\Delta + \Omega}{2} \quad (17)$$

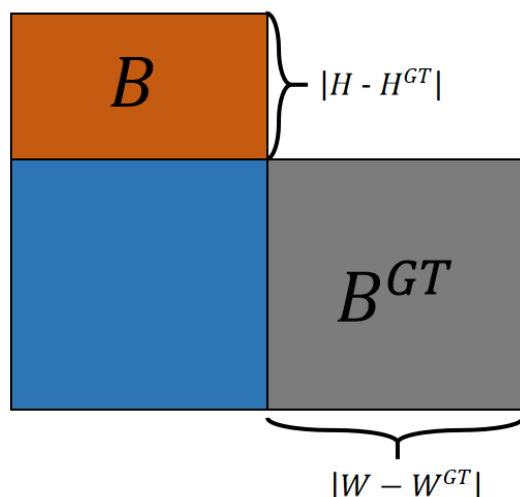


Figure 8. IoU loss. Schematic of the relation of the IoU component contribution.

4. Experiments

4.1. Experimental Environment and Parameter Configuration

The configuration details of the experimental environment are provided in Table 1. The input image size is $640 \times 640 \times 3$. The weight decay is set to 0.0005, and the training batch size is 16, with a total of 200 batches. The initial learning rate is 0.01, and the momentum is set to 0.95.

Table 1. Experimental environment configuration.

Configuration Name	Version/Parameter
Operating system	Ubuntu 20.04LTS
GPU	RTX4090ti \times 2
RAM	48 GB
Memory	2TB SATA
PyTorch	2.0.1
CUDA	11.3
Python	3.9.0

4.2. Experimental Dataset

The experiments in this paper are validated on two video surveillance externally damaged obstacle datasets, which include the general background external broken target dataset and complex background external broken target dataset; the information of the datasets is shown in Figure 9, and Labelme software v4.5.6 (MIT Computer Science and Artificial Intelligence Laboratory, Cambridge, MA, USA) was used to label the data, and the sample dataset is shown in Figure 10, which is briefly introduced as follows:

- (1) General background external broken target dataset: This dataset contains the following five typical hidden target categories: trucks, crane towers, excavators, cranes, and trees. It includes a total of 1307 images with a resolution of 800×600 , yielding 1612 labeled samples. The breakdown is as follows: 356 trucks, 541 crane towers, 298 excavators, 174 cranes, and 243 trees.
- (2) Complex background external broken target dataset: This dataset contains 500 high-definition images of externally broken obstacles under complex backdrop conditions, with a resolution of 1200×900 , totaling 1694 labeled samples. The distribution is as follows: 478 trucks, 641 crane towers, 214 excavators, 123 cranes, and 238 trees.

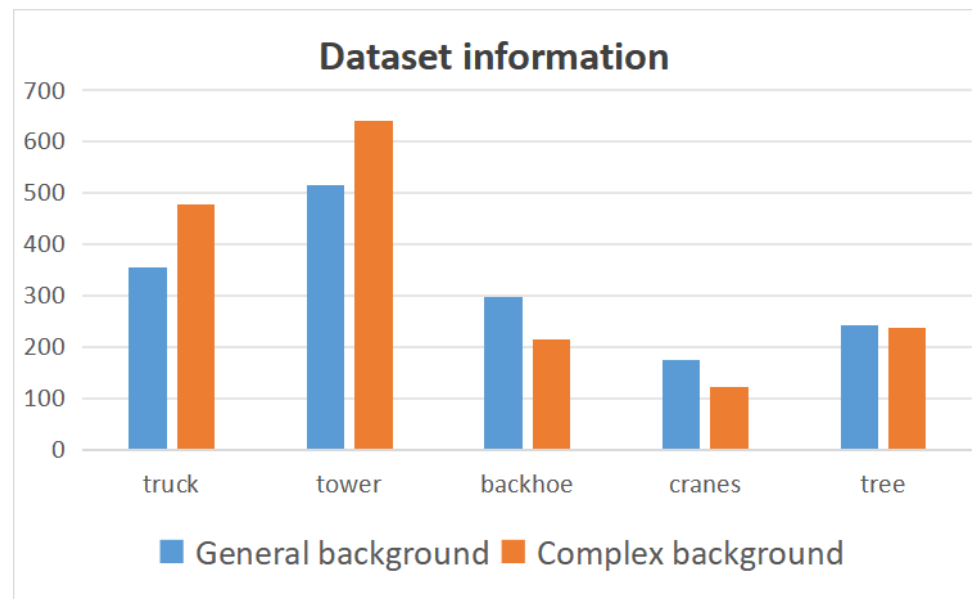


Figure 9. Dataset information.



Figure 10. Sample data.

4.3. Evaluation Index

To validate the detection performance of the model, the experiments used several evaluation metrics, including Precision (P), Recall (R), Average Precision (AP), Mean Average Precision (mAP), Frames Per Second (FPS), Giga Floating Point Operations Per Second ($GFLOPs$), and Number of Parameters ($Params$).

$$P = \frac{TP}{TP + FP} \quad (18)$$

$$R = \frac{TP}{TP + FN} \quad (19)$$

$$AP = \int_0^1 P(r) dr \quad (20)$$

$$mAP = \frac{1}{N} \sum_{n=1}^N AP(n) \quad (21)$$

P represents the proportion of predicted positive samples that are actually positive, ranging from $[0, 1]$, which should be maximized for optimal results. R indicates the proportion of actual positive samples that are correctly predicted as positive, also ranging from $[0, 1]$, with larger values preferred. AP combines P and R to measure the detection precision for a single category, ranging from $[0, 1]$, with higher values indicating better performance. mAP averages the AP across all categories, providing an overall measure of the detection accuracy for multiple categories, with values closer to one being better. FPS indicates the number of frames per second the model can process, which should be maximized for optimal results. $GFLOPs$ refers to the computation required by the model to process an image, with lower values being better. $Params$ indicates the total parameters in the model, with smaller values being preferred, as they indicate a more lightweight model.

4.4. Experimental Results and Analysis

4.4.1. Attention Mechanism Selection Experiments

This experiment was conducted using the general background external broken target dataset. Various attention mechanisms, including ACmix, Squeeze-and-Excitation (SE) [35], Coordinate Attention (CA) [36], and the Convolutional Block Attention Module (CBAM) [37], were integrated into the YOLOv7 model so to analyze their impact on the detection performance. The experimental results are presented in Table 2. Among the tested mechanisms, the ACmix attention mechanism demonstrates the best overall performance, improving the detection accuracy by 1% compared to the CBAM model. Additionally, the number of parameters and computational load are reduced by 0.2M and 0.4 GFLOPs, respectively, resulting in an enhanced detection efficiency.

Table 2. Comparison results of the different attentional mechanisms for the general background external broken target dataset.

Index	Attention Mechanism	Params M	FLOPs G	FPS	mAP %
A	SE	37.8	103.8	64.8	86.2
B	CA	38.1	103.7	63.9	88.4
C	CBAM	37.8	103.9	64.7	89.9
D	ACmix	37.6	103.5	65.2	90.9

4.4.2. Loss Function Selection Experiment

In this paper, experiments were conducted using the complex background external broken target dataset to compare the different loss functions and analyze their impact on model performance. The experimental results are presented in Table 3, while Figure 11a,b illustrate the loss reduction curves during training and validation for each loss function. From the graphs, it can be observed that the SIOU loss reaches the minimum, exhibits smaller fluctuations, converges faster, leading to the best detection performance.

Table 3. Comparison results of the different loss functions for the complex background external broken target dataset.

Index	Loss	P %	R %	mAP0.5 %	mAP0.5–0.95 %
A	CIoU	91.1	85.1	89.7	56.3
B	DIoU	89.7	85	89.6	56.1
C	EIoU	91.2	85.8	89.9	56.3
D	SIOU	91.3	85.1	90.4	56.1

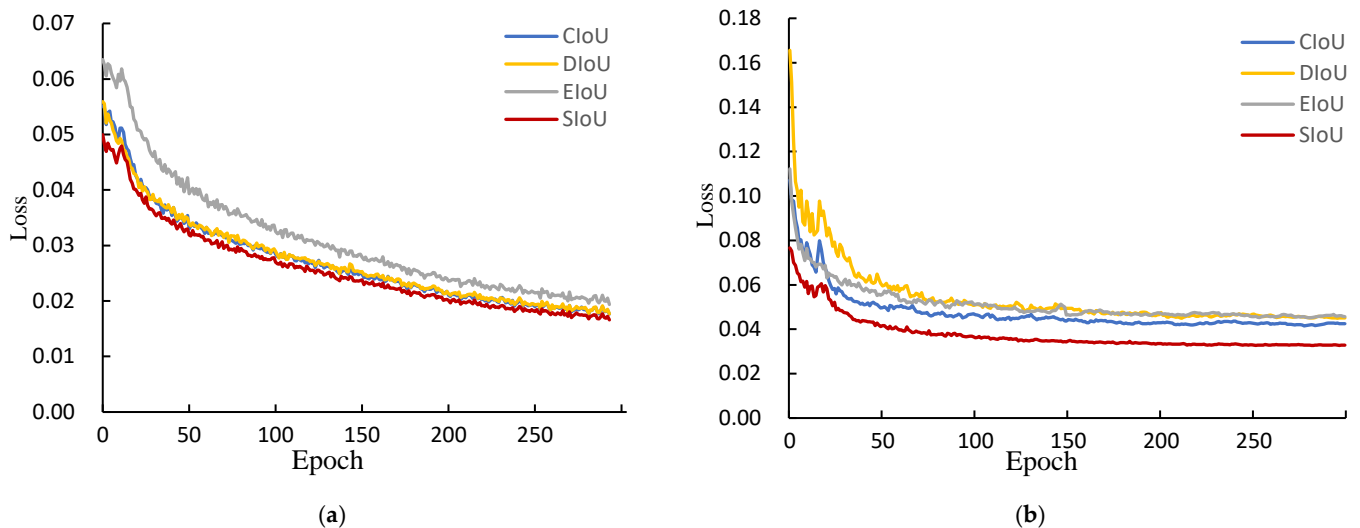


Figure 11. Different IoU loss function curves for the complex background external broken target dataset. (a) Training set; (b) Validation set.

4.4.3. Ablation Experiment

To evaluate the performance of the improved modules, ablation experiments were conducted based on the proposed network structure. The experimental data used were from the general background external broken target dataset, and the results are shown in Table 4, where a ‘√’ indicates the corresponding method was applied.

Group A represents the results of the original YOLOv7 algorithm. Group B introduces ShuffleNetv2 as the backbone network. This improves the model’s detection speed by 1.7 frames/s and reduces the number of parameters and FLOPs by 12M and 18.6G, respectively. However, detection accuracy decreases by 1.3%. This is because ShuffleNetv2 employs techniques such as grouped convolution and channel blending to reduce the computational complexity, sacrificing some model capacity and limiting its ability to capture rich feature information compared to the original network. Group C embeds the ACmix attention mechanism in the SPPCSPC layer of the Neck network, improving the detection accuracy by 1.2%. This is due to ACmix’s ability to combine channel and spatial attention, thereby better capturing important features in the image. Group D replaces the ELAN-W modules in both the backbone and Neck networks with the PC-ELAN modules. This reduces the number of parameters and FLOPs by 9.8M and 30G, respectively. The improvement is attributed to PConv in PC-ELAN, which effectively reduces the computational cost of feature maps. Group E adopts the SIoU loss function, increasing the detection speed by 2.9 frames/s. This improvement results from the SIoU’s ability to reduce unstable gradient changes, leading to a more stable training process. Groups F and G incorporate the ACmix attention mechanism and optimize the loss function within the ShuffleNetv2 backbone network, respectively. Both groups achieve reductions in parameters and FLOPs. Group H is the proposed improved algorithm, achieving an FPS of 69.3 frames/s, with a significant reduction in both parameters and computational load, while achieving the highest detection accuracy. In summary, the reduction in the detection accuracy caused by introducing the ShuffleNetv2 backbone is fully compensated—and further improved—by embedding the ACmix attention mechanism, employing the PC-ELAN module, and adopting the SIoU loss function. These enhancements strike a balance between detection accuracy and model lightweighting, effectively improving the speed, parameter efficiency, and computational cost.

Table 4. Results of the ablation experiments on the general background external broken target dataset.

Index	SNetv2	ACmix	PC-ELAN	SIoU	Params M	FLOPs G	FPS	mAP %
A					36.5	103.2	65.8	89.7
B	✓				24.5	84.6	67.5	88.4
C		✓			37.6	103.5	65.2	90.9
D			✓		26.7	72.3	66.4	89.3
E				✓	36.5	103.2	68.7	90.4
F	✓	✓			29.2	67.8	65.3	89.8
G	✓			✓	23.7	66.4	68.3	91.2
H	✓	✓	✓	✓	24.7	56.8	69.3	92.7

4.4.4. Comparative Experiment

To further validate the performance of the proposed algorithm, comparison experiments were conducted against mainstream models, including Faster R-CNN, SSD, YOLOv3, YOLOv5m, YOLOv5s, YOLOX, YOLOv7-tiny, YOLOv7, and YOLOv8s, using the general background external broken target dataset.

As shown in Table 5, the proposed method achieves an average accuracy of 92.7%, which is 3% higher than the original YOLOv7. Additionally, the number of parameters and computational cost are reduced by 32.3% and 44.9%, respectively, compared to YOLOv7, resulting in optimal overall performance. Although the YOLOv8s model has a slightly lower number of parameters and computational cost than the proposed algorithm, its accuracy is significantly lower. In summary, the improved model presented in this paper demonstrates a faster detection speed and higher accuracy, making it more suitable for deployment on resource-constrained edge devices. This makes it particularly advantageous for industrial applications, supporting the intelligent development of modern transmission line inspection systems.

Table 5. Comparison experiment results on the general background external broken target dataset.

Index	Model	P %	R %	Params M	FLOPs G	FPS	mAP %
A	Faster R-CNN	61.1	82.1	52.7	95.7	22.6	82.6
B	SSD	57.6	74.5	31.9	67.83	38.9	65.4
C	YOLOv3	87.5	61.5	78.5	134.6	37.6	85.2
D	YOLOv5m	81.7	74.2	30.8	68.3	59.2	82.2
E	YOLOv5s	83.1	78.8	17.2	35.8	72.4	84.6
F	YOLOX	86.3	79.2	18.3	41.26	58.7	86.2
G	YOLOv7-tiny	85.5	84.1	13.7	26.8	58.8	87.7
H	YOLOv7	87.9	86.9	36.5	103.2	65.8	89.7
I	YOLOv8s	91.1	75.6	11.1	28.4	75.2	84.3
J	Ours	90.4	87.7	24.7	56.8	69.3	92.7

To provide a more intuitive comparison of the detection performance, representative images from both the general and complex background target datasets were selected for visual analysis. Figure 12 shows the detection results for each comparative model under the general background dataset. The Faster R-CNN algorithm exhibits severe issues with false positives and missed detections, particularly in scenarios involving dense and overlapping small targets. In such cases, the network tends to either merge multiple targets into one or incorrectly split similar regions, leading to detection errors. Although the YOLOv5m model shows improvements in the detection accuracy, its performance on small objects remains limited. This is due to the sparse pixel representation of small objects in the images, which diminishes the models' ability to recognize them, resulting in continued false positives and missed detections.

The algorithm proposed in this paper enhances the detection accuracy and reduces the occurrence of false positives and missed detections by improving the model's long-range dependency and feature fusion capabilities. As shown in Figure 9, the proposed model achieves superior detection performance compared to the other tested models.

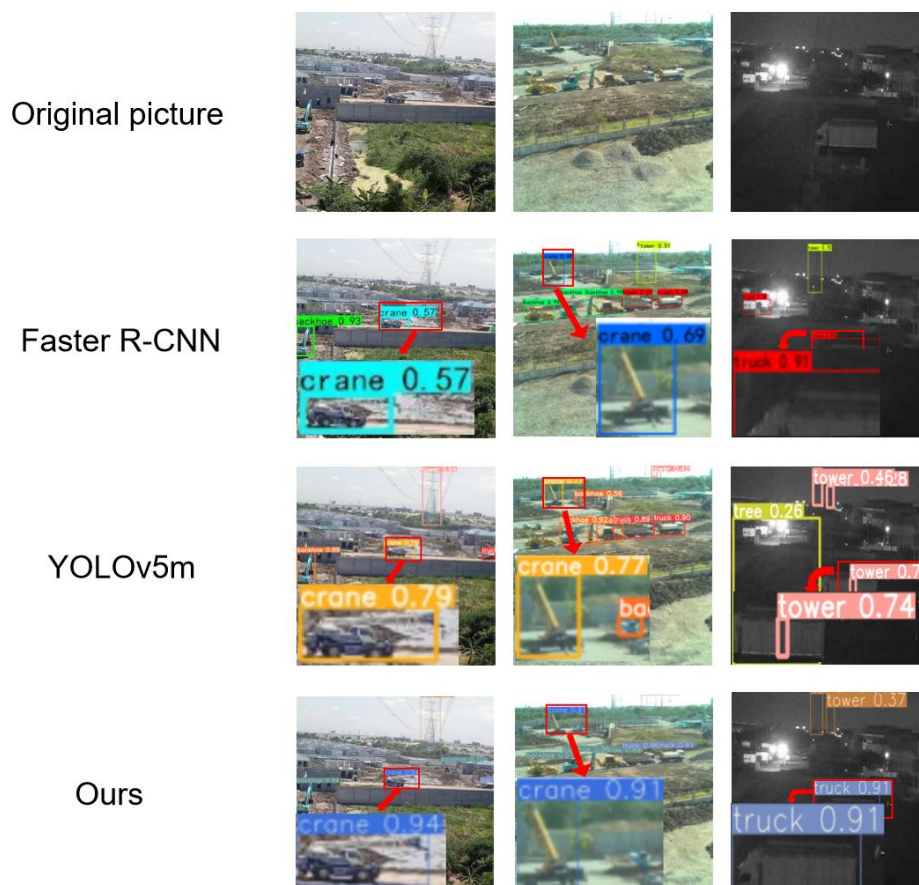


Figure 12. Detection results on the general background external broken target dataset.

4.4.5. Generalization Experiment

To evaluate the generalization ability of the proposed method in complex backgrounds, validation experiments were conducted using the complex background external broken target dataset. This dataset includes a variety of challenging conditions and interference factors, such as different terrain types (e.g., mountainous areas, forests, and fields) and adverse weather conditions, to simulate the diversity and complexity of real-world scenarios. The detection results are presented in Table 6. As can be seen from the table, the model proposed in this paper achieves the highest detection accuracy of the target, while significantly reducing the number of parameters and computational cost, and improves its ability to deal with complex backgrounds by utilizing global contextual information.

The results of visual analysis in complex backgrounds are shown in Figure 13. In scenes with intricate backgrounds and numerous interfering targets, the original algorithm exhibits severe false positives and missed detections, particularly in the areas marked by red circles. The detection accuracy for external breakage obstacles deteriorates in such cases. In contrast, the improved algorithm effectively suppresses background noise interference and substantially mitigates the issues of false positives and missed detections for small targets and multiple overlapping targets. In summary, the proposed algorithm better adapts to real-world transmission line scenarios by integrating multi-scale feature information of

externally breakage obstacles. As a result, it achieves superior detection performance for such targets under complex background conditions.

Table 6. Experiments on the generalizability of the complex background external broken target dataset.

Index	Model	P %	R %	Params M	FLOPs G	FPS	mAP %
A	Faster R-CNN	59.5	80.9	52.7	95.7	21.4	80.1
B	SSD	54.7	72.1	31.9	67.83	36.2	63.7
C	YOLOv3	84.5	60.2	78.5	134.6	39.3	81.6
D	YOLOv5m	80.2	72.5	30.8	68.3	57.1	79.2
E	YOLOv5s	81.5	76.3	17.2	35.8	74.2	83.6
F	YOLOX	84.1	77.5	18.3	41.26	57.3	81.4
G	YOLOv7-tiny	84.3	83.4	13.7	26.8	68.5	84.7
H	YOLOv7	87.2	85.3	36.5	103.2	64.5	88.6
I	YOLOv8s	88.7	73.1	11.1	28.4	72.1	84.1
J	Ours	90.4	86.5	24.7	56.8	69.3	91.4



(a) YOLOv7



(b) Ours

Figure 13. Comparison of the effect before and after the improvement of the complex background external broken target dataset.

5. Discussion and Conclusions

In this paper, we proposed an improved lightweight obstacle detection algorithm for small targets outside of broken transmission lines, which enhances the detection performance by incorporating ShuffleNetv2, the ACmix attention mechanism, the PC-ELAN module, and the SIOU loss function. The experimental results show that the algorithm achieves a 92.7% and 91.4% detection accuracy on two datasets with general and complex backgrounds, respectively, which is 3.0% and 2.8% higher than the original YOLOv7 model. Additionally, the model's parameters are reduced by 32.3%, and the computational cost is lowered by 44.8%. These improvements alleviate issues such as misdetection, omissions, and inefficiency in detecting small obstacle targets in transmission lines, providing reliable technical support for enhancing the safety and stability of power transmission systems, with strong practical application prospects.

Although the model proposed in this paper alleviates small target misdetection and omission, it still has the following limitations:

- (1) **FLOPs Optimization:** While the model demonstrates lower FLOPs compared to most other models, as shown in Tables 5 and 6, its computation still exceeds that of YOLOv5s, YOLOv7-tiny, and YOLOv8s. Future work will focus on further reducing the computation and model parameters while maintaining the detection accuracy, enabling the model to be deployed efficiently in industrial environments.
- (2) **Handling Complex Scenes:** The model performs well in general background scenarios, as shown in Figure 12, but there is room for improvement in the accuracy when dealing with complex backgrounds. Future work will aim to enhance the model's robustness in extremely complex environments, ensuring good detection performance in more challenging scenes.

Author Contributions: Conceptualization, J.H.; methodology, G.Y.; validation, L.W.; formal analysis, H.P.; investigation, X.X.; resources, B.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Science and Technology Project of State Grid Shanxi 440 Electric Power Company, Grant number: 5205M0230006.

Data Availability Statement: The data presented in this study is available on request from the corresponding authors, and the dataset was jointly completed by the team, so the data is not publicly available.

Acknowledgments: The authors wish to thank the editor and reviewers for their suggestions.

Conflicts of Interest: Author Junbo Hao was employed by State Grid Shanxi Integrated Energy Service Co., Ltd. Authors Guangying Yan, Lidong Wang and Honglan Pei were employed by State Grid Yuncheng Electric Power Supply Company. Author Xiao Xu was employed by State Grid Gaoping Electric Power Supply Company. The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

PConv	Partial Convolution
CNNs	Convolutional Neural Networks
SPP	Spatial Pyramid Pooling
CSP	Cross-Stage Partial
IoU	Intersection over Union
CIoU	Complete Intersection over Union
GIoU	Generalized Intersection over Union
DIoU	Distance Intersection over Union
EIoU	Enhanced Intersection over Union
SIoU	Scalable Intersection over Union Loss
mAP	Mean Average Precision
FPS	Frames Per Second
GFLOPs	Giga Floating Point Operations Per Second
Params	Number of Parameters
SE	Squeeze-and-Excitation
CA	Coordinate Attention
CBAM	Convolutional Block Attention Module

References

1. Zhang, J.; Zeng, Q.; Duan, L. Analysis of Factors Influencing the Safety of Transmission Line Operation and Preventive Measures. *Autom. Today* **2023**, *8*, 85–87.
2. Zhao, X.; Jiao, Y.; Liu, Y. Research on Key Technologies of Laser Anti-External Damage System for Transmission Lines and Its Engineering Application. *High Volt. Eng.* **2023**, *49*, 72–77.
3. Zhou, Z.; Yuan, Y.; Zhang, C. A lightweight power inspection method based on object detection model. *Inf. Technol.* **2023**, 87–93. [[CrossRef](#)]
4. Du, W.; Wang, J.; Yang, G. Pylon Object Detection of Helicopter Cruise Based on YOLOv7. *J. Shanghai Univ. Electr. Power* **2023**, *39*, 383–386.
5. Li, J.; Shuang, F.; Huang, J. Safe distance monitoring of live equipment based upon instance segmentation and pseudo-LiDAR. *IEEE Trans. Power Deliv.* **2023**, *38*, 2953–2964. [[CrossRef](#)]
6. Song, L.; Liu, S.; Wang, K. Identification Method of Power Grid Components and Defects Based on Improved EfficientDet. *Trans. China Electrotech. Soc.* **2022**, *37*, 2241–2251.
7. Shaoqing, R.; Kaiming, H.; Ross, G. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *TPAMI* **2017**, *39*, 1137–1149.
8. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. *arXiv* **2017**, arXiv:1703.06870.
9. Cai, Z.; Vasconcelos, N. Cascade r-cnn: Delving into high quality object detection. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 6154–6162.
10. Redmon, J.; Farhadi, A. *Yolo9000: Better, Faster, Stronger*; CVPR: Honolulu, HI, USA, 2017; pp. 6517–6525.
11. Redmon, J.; Farhadi, A. Yolo3: An incremental Improvement. Available online: <https://arxiv.org/abs/1804.02767> (accessed on 8 April 2018).
12. Bochkovskiy, A.; Wang, C.; Liao, H. Yolo4: Optimal Speed and Accuracy of Object Detection. Available online: <https://arxiv.org/abs/2004.10934> (accessed on 23 April 2020).
13. Ge, Z.; Liu, S.; Wang, F.; Li, Z.; Sun, J. YoloX: Exceeding YOLO Series in 2021. Available online: <https://arxiv.org/abs/2107.08430> (accessed on 6 August 2021).
14. Li, C.; Li, L.; Jiang, H.; Weng, K.; Geng, Y.; Li, L.; Ke, Z.; Li, Q.; Cheng, M.; Nie, W.; et al. Yolo6: A Single-Stage Object Detection Framework for Industrial Applications. Available online: <https://arxiv.org/abs/2209.02976> (accessed on 7 September 2022).
15. Wang, C.; Bochkovskiy, A.; Liao, H. Yolo7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 7464–7475.
16. Wei, L.; Anguelov, D.; Erhan, D. *Ssd: Single Shot Multibox Detecto*; Computer Vision-ECCV: Amsterdam, The Netherlands, 2016; pp. 21–37.
17. Wu, X.; Sahoo, D.; Hoi, S.C.H. Recent advances in deep learning for object detection. *Neurocomputing* **2020**, *396*, 39–64. [[CrossRef](#)]
18. Zheng, H.; Cui, Y.; Yang, W. An infrared image detection method of substation equipment combining iresgroup structure and CenterNet. *IEEE Trans. Power Deliv.* **2022**, *37*, 4757–4765. [[CrossRef](#)]
19. Zhang, J.; Yu, J.; Wang, J. Image Recognition Technology for Transmission Line External Damage Based on Depth Learning. *Comput. Syst. Appl.* **2018**, *27*, 176–179.
20. Wei, X.; Lu, W.; Zhao, W. Target detection method for external damage of a transmission line based on an improved Mask R-CNN algorithm. *Power Syst. Prot. Control.* **2021**, *49*, 155–162.
21. Tian, E.; Li, C.; Zhu, G. Identification Algorithm of Transmission Line External Hidden Danger Based on YOLOv4. *Comput. Syst. Appl.* **2021**, *30*, 190–196.
22. Zheng, H.; Hu, S.; Liang, Y. Hidden danger detection method of transmission line corridor based on YOLO-2MCS. *J. China Electrotech. Soc.* **2024**, 1–12. [[CrossRef](#)]
23. Sun, Y.; Li, J. Foreign body Detection Algorithm of YOLOv7-tiny Transmission Lines based on channel pruning. *J. Comput. Eng. Appl.* **2024**, *60*, 319–328.
24. Long, L.; Zhou, L.; Liu, S. Identification of hidden damage targets by external forces based on domain adaptation and attention mechanism. *J. Electron. Meas. Instrum.* **2022**, *36*, 245–253.
25. Wang, Y.; Guo, C.; Wu, D. Hidden target detection method for mechanical external damage of transmission line based on improved YOLOv7. *Electr. Meas. Instrum.* **2024**, 1–10. Available online: <http://kns.cnki.net/kcms/detail/23.1202.th.20240428.1950.002.html> (accessed on 7 September 2022).
26. Ding, X.; Zhang, X.; Ma, N. Repvgg: Making vgg-style convnets great again. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 13733–13742.
27. Ma, N.; Zhang, X.; Zheng, H. ShuffleNetv2: Practical guidelines for efficient CNN architecture design. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 116–131.

28. Zhang, X.; Zhou, X.; Lin, M. Shufflenet: An extremely efficient convolutional neural network for mobile devices. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 6848–6856.
29. Pan, X.; Ge, C.; Lu, R. On the Integration of Self-Attention and Convolution. *arXiv* **2021**, arXiv:2111.14556.
30. Chen, J.; Kao, S.; Hao, H. Run, Don't Walk: Chasing Higher FLOPS for Faster Neural Networks. In Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 17–24 June 2023.
31. Zheng, Z.; Wang, P.; Liu, W. *Distance-IoU loss: Faster and better learning for bounding box regression*. *Proceedings of the 2020 AAAI Conference on Artificial Intelligence*; AAAI Press: Palo Alto, CA, USA, 2020; pp. 12993–13000.
32. Yu, J.; Jiang, Y.; Wang, Z. *UnitBox: An Advanced Object Detection Network*; Association for Computing Machinery (ACM): New York, NY, USA, 2016.
33. Rezatofighi, H.; Tsoi, N.; Gwak, J. Generalized Intersection Over Union: A Metric and a Loss for Bounding Box Regression. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019.
34. Gevorgyan, Z. SloU Loss: More powerful learning for bounding box regression. *arXiv* **2022**, arXiv:2205.12740.
35. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation net-works. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
36. Hou, Q.; Zhou, D.; Feng, J. Coordinate attention for efficient mobile network design. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 13713–13722.
37. Woo, S.; Park, J.; Lee, J. CBAM: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.