


## Article

# Research on Lightweight Method of Insulator Target Detection Based on Improved SSD

Bing Zeng<sup>1,\*</sup> , Yu Zhou<sup>1,\*</sup>, Dilin He<sup>1</sup>, Zhihao Zhou<sup>1</sup>, Shitao Hao<sup>1</sup>, Kexin Yi<sup>1</sup>, Zhilong Li<sup>2</sup>, Wenhua Zhang<sup>1</sup> and Yunmin Xie<sup>1</sup>

<sup>1</sup> Nanchang Institute of Technology, Nanchang 330099, China; nathan997@163.com (D.H.); 13164162010@163.com (Z.Z.); 18331317390@163.com (S.H.); superstephen@sina.com (K.Y.); zhangwenhua\_610@163.com (W.Z.); xie\_yunmin@163.com (Y.X.)

<sup>2</sup> State Grid Shanghai Municipal Electric Power Company Maintenance Company, Shanghai 200063, China; ynotbn@sina.com

\* Correspondence: zengbing\_wuhu@whu.edu.cn (B.Z.); zjj1825819376@163.com (Y.Z.)

**Abstract:** Aiming at the problems of a large volume, slow processing speed, and difficult deployment in the edge terminal, this paper proposes a lightweight insulator detection algorithm based on an improved SSD. Firstly, the original feature extraction network VGG-16 is replaced by a lightweight Ghost Module network to initially achieve the lightweight model. A Feature Pyramid structure and Feature Pyramid Network (FPN+PAN) are integrated into the Neck part and a Simplified Spatial Pyramid Pooling Fast (SimSPPF) module is introduced to realize the integration of local features and global features. Secondly, multiple Spatial and Channel Squeeze-and-Excitation (scSE) attention mechanisms are introduced in the Neck part to make the model pay more attention to the channels containing important feature information. The original six detection heads are reduced to four to improve the inference speed of the network. In order to improve the recognition performance of occluded and overlapping targets, Diou-NMS was used to replace the original non-maximum suppression (NMS). Furthermore, the channel pruning strategy is used to reduce the unimportant weight matrix of the model, and the knowledge distillation strategy is used to fine-adjust the network model after pruning, so as to ensure the detection accuracy. The experimental results show that the parameter number of the proposed model is reduced from 26.15 M to 0.61 M, the computational load is reduced from 118.95 G to 1.49 G, and the mAP is increased from 96.8% to 98%. Compared with other models, the proposed model not only guarantees the detection accuracy of the algorithm, but also greatly reduces the model volume, which provides support for the realization of visible light insulator target detection based on edge intelligence.

**Keywords:** SSD; insulator; lightweight; channel pruning; target detection



**Citation:** Zeng, B.; Zhou, Y.; He, D.; Zhou, Z.; Hao, S.; Yi, K.; Li, Z.; Zhang, W.; Xie, Y. Research on Lightweight Method of Insulator Target Detection Based on Improved SSD. *Sensors* **2024**, *24*, 5910. <https://doi.org/10.3390/s24185910>

Academic Editor: Guangcai Sun

Received: 12 June 2024

Revised: 29 August 2024

Accepted: 6 September 2024

Published: 12 September 2024



**Copyright:** © 2024 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

Insulators are essential components in power transmission lines, primarily functioning to provide electrical insulation and mechanical anchorage. However, due to their prolonged exposure to outdoor environments and the effects of overvoltage, they are susceptible to damage, which can impair the stable operation of the transmission lines and potentially lead to large-scale power outages, resulting in significant economic and social losses [1]. Therefore, to ensure the secure operation of power transmission lines, conducting the safety inspections of these lines is imperative. Conventional inspection methods entail visual examination by human operators or the use of specialized equipment to detect defects in insulators. Nonetheless, such approaches are prone to missing defects, making erroneous judgments, and are inefficient [2]. With the ongoing development of deep learning and drone technologies, they have come to play a crucial role in the safety inspections of power transmission lines [3]. Initially, drones equipped with high-definition cameras are utilized

to conduct aerial photography of insulators on power transmission lines, capturing high-resolution images. Subsequently, deep learning techniques are employed to accurately identify and analyze defects within these images of the insulators.

In the realm of insulator object detection in power transmission lines, numerous scholars have conducted related research. The two-stage target detection algorithms include R-CNN [4], Fast R-CNN [5], and Faster R-CNN [6], which show an excellent performance and high detection accuracy in the field of target detection. Zhao [7] et al. present an improved Faster R-CNN-based method for insulator object detection, which achieves the precise detection of insulators under various aspect ratios, scales, and conditions of mutual occlusion by enhancing the anchor generation method within the Region Proposal Network (RPN) and refining the non-maximum suppression (NMS) strategy. Haijian [8] et al. propose a transmission line insulator detection method based on an improved Faster R-CNN, substituting the original VGG-16 network with a deeper Resnet-50 network and incorporating attention mechanism modules, and the target detection accuracy is improved by 1.63%, albeit at a cost of a slower detection speed. To address the issue of the slow recognition speed inherent in R-CNN series algorithms, Redmond et al. [9] introduced the YOLO (You Only Look Once) series of algorithms, which, as a type of single-stage object detection algorithm, offer a high detection speed but compromise on the detection precision to some extent. Juping [10] et al. proposed an overhead power transmission line object detection method based on an improved YOLOv5, which enhances the detection accuracy of small objects by incorporating larger-scale detection layers and skip connections into the algorithm. In addition, a small object-enhanced Complete Intersection over Union (CIoU) is put forward as the loss function of the bounding box regression. And pruning methods are adopted to lighten the model. The results indicate that this method achieves a 4% increase in the detection accuracy, a 58% reduction in the model size, and a 3.3% improvement in the detection speed. Wang [11] et al. introduced an insulator defect detection method based on ML-YOLOv5, in which the depthwise separable convolutions is employed as the backbone feature extraction network and the feature fusion module is improved by adopting an Enhanced Feature Pyramid Network (MFPN), and utilizing YOLOv5m as a teacher model for knowledge distillation. The experimental outcomes demonstrate that this algorithm boasts high detection accuracy and rapid detection speeds. In addition, Liu [12] et al. proposed a multi-scale detection algorithm, the SSD (Single-Shot MultiBox Detector) algorithm, which overcomes the shortcomings of the R-CNN series and YOLO series and has advantages in speed and accuracy. Xuan [13] et al. presented an improved SSD for the online detection of Insulators and Spacers Based on a UAV System. This approach utilizes the lightweight MnasNet network as the feature extraction network to generate feature maps and employs two multi-scale feature fusion strategies to integrate multiple feature maps. The outcomes illustrate that the algorithm excels in both a high detection accuracy and fast detection speed; however, there remains room for further reductions in the algorithm's size.

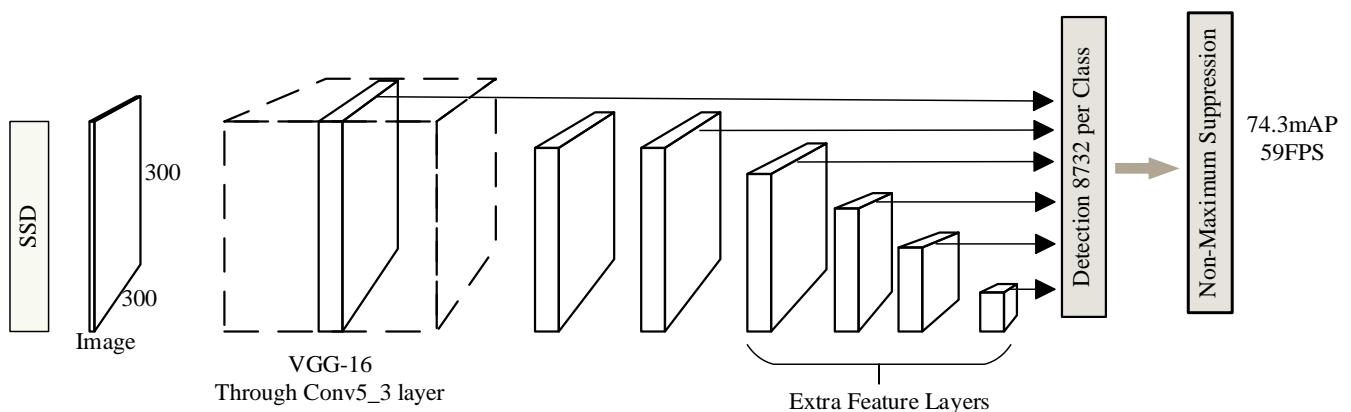
Prior to inspecting defects on insulators, a preprocessing stage is indispensable: the identification and localization of insulators within images through object detection techniques. This preliminary step lays the groundwork for subsequent defect detection, enabling the system to focus its analysis on areas of potential defects in insulators, thereby enhancing the overall efficiency and accuracy of the inspection process. In pursuit of a balance between the detection accuracy, recognition speed, and a smaller algorithmic footprint, numerous scholars have adopted more lightweight convolutional neural network models; however, these still entail substantial computational loads and parameter counts [14]. Against this backdrop, this paper proposes a lightweight visible-light insulator object detection algorithm based on an improved SSD. First, improvements are made to the base model to enhance its detection accuracy, followed by pruning operations to achieve model lightweighting. To mitigate the decline in precision typically associated with pruning, knowledge distillation is employed to fine-tune the lightweight model. Ultimately, the performance of the proposed algorithm is validated using a visible-light insulator

dataset, with a comparative analysis against classic object detection algorithms to confirm the efficacy of the improvement strategies outlined herein. The algorithm proposed in this paper can effectively solve the problem that the detection accuracy of the algorithm is not high in the insulator target detection task and the algorithm is too large to be deployed to a mobile terminal, such as the UAV.

## 2. SSD Network Model

The SSD object detection algorithm is characterized by multi-prediction layers and multi-scale features [15]. Its network architecture can be divided into three parts: First, the base network utilizes the VGG-16 structure to extract multi-scale feature information from the target. Second, auxiliary convolutional layers are connected to the final feature map of VGG-16, constructing deeper output layers for object detection. Third, the prediction convolutional layers obtain feature information from the feature maps and utilize NMS to derive the detection results [16]. The model architecture of the SSD algorithm is depicted in Figure 1. The input image size for SSD is  $300 \times 300$  pixels, with VGG-16 serving as the feature extraction layer. Through six convolutional layers, it constructs multi-scale detection layers to capture feature information at various scales including  $38 \times 38$ ,  $19 \times 19$ ,  $10 \times 10$ ,  $5 \times 5$ ,  $3 \times 3$ , and  $1 \times 1$ , forming a multi-scale feature extraction network [17]. By setting prior boxes on feature maps of different depths and resolutions, and performing category prediction and location refinement for each prior box boundary, objects are precisely matched. The fact that different convolutional layers in a CNN have distinct receptive fields enables the network to effectively recognize targets of different sizes. Ultimately, the network computes the coordinates and class of candidate boxes through regression [18]. In the SSD algorithm, the formula for calculating the anchor box scales corresponding to each feature map is shown in Equation (1). Here,  $S_m$  denotes the scale of candidate boxes for the  $m$ -th feature map;  $S_{max}$  represents the maximum scale of candidate boxes, typically set at 0.9;  $S_{min}$  signifies the minimum scale of candidate boxes, usually set to 0.2; and  $r$  denotes the total number of feature maps. The main symbols and their meanings are shown in Table 1.

$$S_m = S_{min} + \frac{S_{max} - S_{min}}{r - 1}, m \in \{1, 2, \dots, r\} \quad (1)$$



**Figure 1.** SSD algorithm model.

**Table 1.** Main symbols and their meanings.

Symbol	Description
$S_m$	the scale of candidate boxes for the $m$ -th feature map
$S_{max}$	the maximum scale of candidate boxes
$S_{min}$	the minimum scale of candidate boxes
$r$	the total number of feature maps

Table 1. Cont.

Symbol	Description
$h$	the picture height
$w$	the picture width
$c$	the picture length
$y'_i$	channel feature map
$\Phi_{i,j}$	undergoes a linear operation
$d \times d, k \times k$	the size of the linear operation kernel
$\sigma(\cdot)$	the compressed feature maps are then normalized by the sigmoid function
$\delta(\cdot)$	the ReLU function
$M$	the prediction box with a higher prediction score
$B_i$	the other prediction boxes
$\rho$	the Euclidean distance between $M$ and $B_i$
$\varepsilon$	non-zero constant
$X$	the input to the $BN$ layer
$Y$	the output from the $BN$ layer
$\gamma$	represents the normalized scale factor
$\sigma$	the variance computed over a mini-batch for the $BN$ layer
$\mu$	the mean computed over a mini-batch for the $BN$ layer
$\beta$	a bias compensation in the normalization process
$\Gamma$	encompassing all prunable channels
FPS	Frame Per Second
$Tp$	true positive predictions
$Fp$	false positive predictions
$Fn$	indicates false negative predictions

### 3. Improved SSD Network Model

This paper proposes the following improvements to the SSD network model based on the actual characteristics of power transmission line insulator images, with the enhanced SSD network structure illustrated in Figure 2.

- (1) The original feature extraction network, VGG-16, is replaced with a lightweight Ghost Module network to initially achieve model lightweighting.
- (2) The Neck part of the SSD network adopts an FPN+PAN structure to enhance feature extraction capabilities. To facilitate the fusion of local and global features, a SimSPPF structure is introduced at each input end of the Neck.
- (3) Multiple Spatial and Channel Squeeze-and-Excitation (scSE) attention mechanism modules are incorporated into the Neck section, enabling the network to better focus on channels containing critical feature information while preserving positional information of feature layers.
- (4) The original six detection heads are reduced to four to accelerate the network's inference speed. To improve the recognition of occluded and overlapping objects, DIOU-NMS replaces the conventional non-maximum suppression.
- (5) Channel pruning strategies are employed to eliminate unimportant weight matrices, further lightweighting the constructed network model and achieving model compression objectives.
- (6) To mitigate the impact of channel pruning on detection accuracy, knowledge distillation is applied to fine-tune the lightweight network model, ensuring detection precision is maintained.

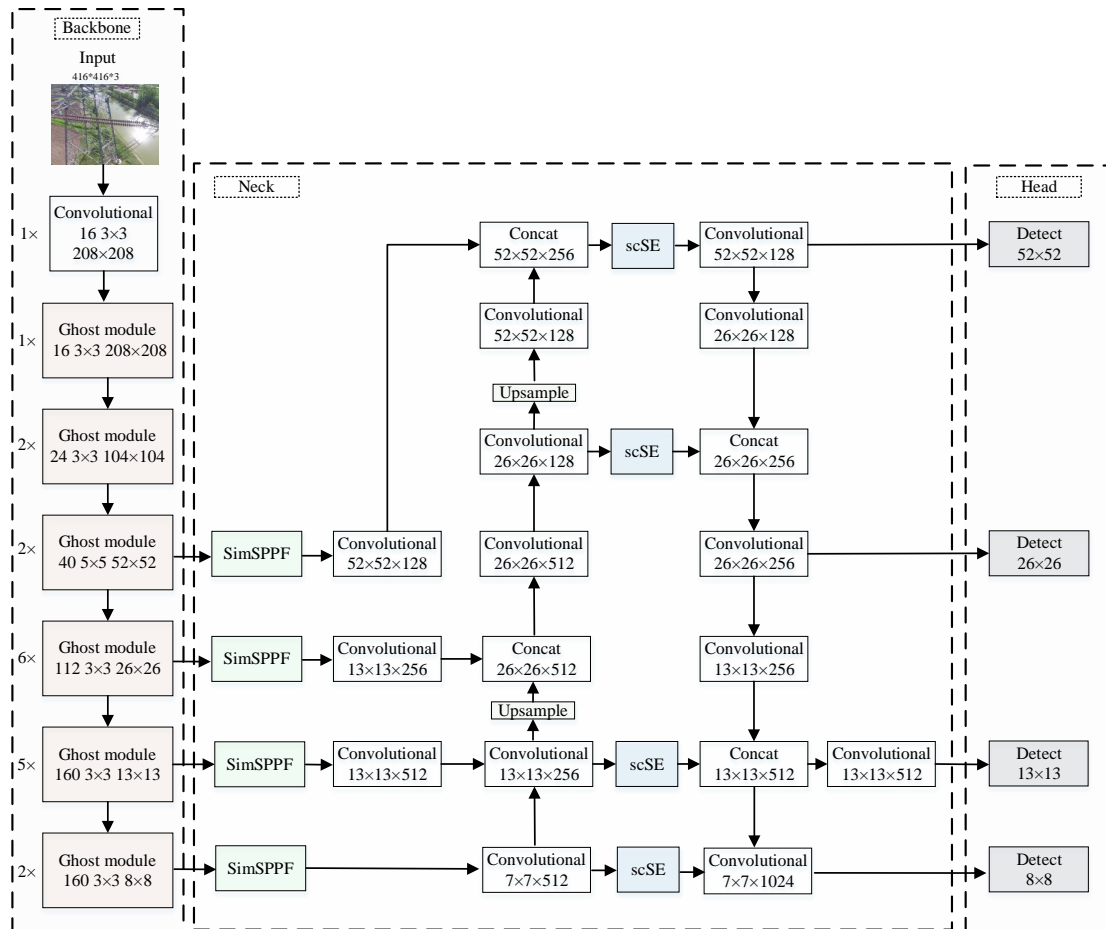


Figure 2. Improved SSD network model.

### 3.1. Feature Extraction Network

The original SSD network model employs VGG-16 as its feature extraction backbone, which comprises a stack of numerous convolutional and pooling layers, leading to a deep network architecture. However, this model is characterized by a substantial number of parameters, necessitating longer training times and posing significant challenges in the tuning process, thereby hindering its deployment on mobile devices. Consequently, in pursuit of maintaining the detection performance while reducing the model size, this study eschews the VGG-16 network in favor of adopting a lightweight Ghost Module to construct the primary feature extraction backbone.

The Ghost Module functionally substitutes conventional convolution [19], capable of generating an equivalent number of feature maps to standard convolutional layers through a two-step process. Initially, a  $1 \times 1$  convolution with fewer output channels is employed to perform dimensionality reduction, thereby creating a condensed feature map from the input feature layer. Subsequently, depthwise separable convolution is applied to this condensed map to yield similar feature maps. Finally, by concatenating the condensed feature map with its corresponding similar feature maps, an output feature map is attained, mirroring the structure of those produced by standard convolutions. The Ghost Module's convolution operation encompasses two primary components. The first part involves obtaining intrinsic feature maps through conventional convolutional operations. If the input image size is  $h \times w \times c$ , the computational cost of this part is  $h \times w \times c \times m \times w' \times h'$ . The other part employs a simple linear operation to generate multiple feature maps, as illustrated by Equation (2), where depthwise separable convolution is applied to the original

features. Each channel feature map,  $y'_i$ , undergoes a linear operation,  $\Phi_{i,j}$ , to produce ghost feature maps.

$$y_{ij} = \Phi_{i,j}(y'_i), \forall i = 1, \dots, m, j = 1, \dots, s \quad (2)$$

The theoretical acceleration ratio of replacing conventional convolutional modules with the Ghost Module is given by Equation (3), where  $d \times d$  denotes the size of the linear operation kernel, which is comparable in magnitude to  $k \times k$  and  $s \ll c$ . Consequently, Equation (3) can be approximated by Equation (4), indicating that the Ghost Module entails significantly fewer parameters and computational costs compared to standard convolutions.

$$r_s = \frac{n \cdot h' \cdot w' \cdot c \cdot k \cdot k}{\frac{n}{s} \cdot h' \cdot w' \cdot c \cdot k \cdot k + (s-1) \cdot \frac{n}{s} \cdot h' \cdot w' \cdot d \cdot d} \approx \frac{s \cdot c}{s + c - 1} \approx s \quad (3)$$

$$r_c = \frac{n \cdot c \cdot k \cdot k}{\frac{n}{s} \cdot c \cdot k \cdot k + (s-1) \cdot \frac{n}{s} \cdot d \cdot d} \approx \frac{s \cdot c}{s + c - 1} \approx s \quad (4)$$

### 3.2. Feature Fusion Network

In order to solve the influence of different target sizes on the model detection accuracy caused by the change in shooting angle in the process of transmission line insulator image acquisition, the FPN+PAN structure is integrated into the Neck part of the SSD network to enhance feature extraction capabilities, particularly catering to objects of diverse scales [20]. Furthermore, to facilitate the fusion of local and global features, both the SimSPPF structure and scSE attention mechanism modules are introduced at the inputs of the Neck [21]. The FPN+PAN module is depicted in Figure 3, wherein the Feature Pyramid Network (FPN) structure performs upsampling from higher to lower dimensions of the backbone network's outputs, thereby capturing strong semantic information. Conversely, the Path Aggregation Network (PAN) structure conducts downsampling from lower to higher dimensions, acquiring robust location information across various scales. Ultimately, these features are concatenated across dimensions, enabling the superior recognition of objects across different scales.

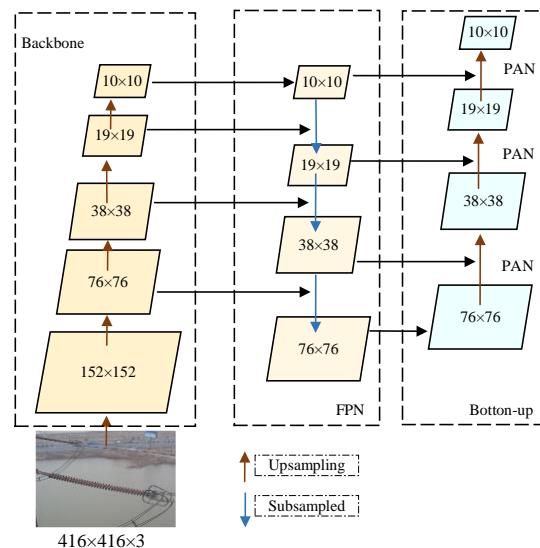
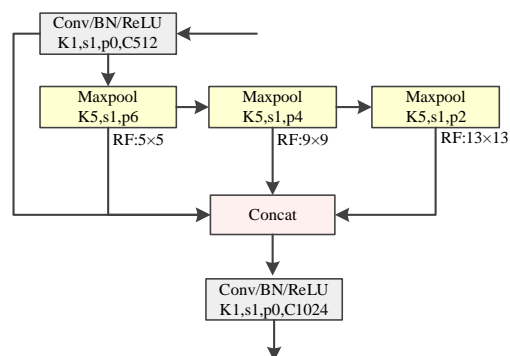


Figure 3. FPN+PAN Module.

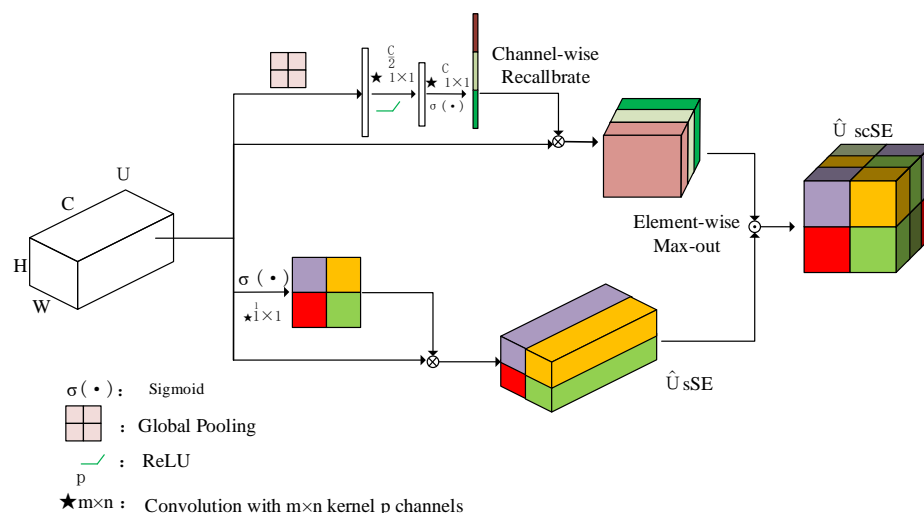
In the process of feature extraction from power transmission line insulator images, issues arise due to inconsistencies in image scales and distortions caused by operations such as resizing, cropping, and grayscale transformations. To circumvent these issues, this study integrates the SimSPPF module at the input end of the Neck section, facilitating the fusion

of multi-scale insulator feature maps and global feature maps. The structure of the SimSPPF module is illustrated in Figure 4. This module processes the input data sequentially through several Maxpool layers with  $5 \times 5$  kernel sizes. Notably, the combined outputs of two sequential  $5 \times 5$  Maxpool layers equate to that of a single  $9 \times 9$  Maxpool layer, and similarly, the combined output of three sequential  $5 \times 5$  Maxpool layers matches that of a single  $13 \times 13$  Maxpool layer. Consequently, the SimSPPF structure requires only three  $5 \times 5$  convolution kernels to achieve the integration of local and global features, thereby enhancing computational efficiency and reducing computational overhead. Moreover, the SimSPPF module employs the ReLU activation function to expedite network inference, further boosting the detection efficiency.



**Figure 4.** SimSPPF network structure.

To enhance the SSD network model's capability in capturing and focusing on critical features in insulator images while preserving positional information in feature layers, this study incorporates multiple scSE attention mechanism modules into the Neck section [22]. As depicted in Figure 5, the scSE attention mechanism module is comprised of a parallel combination of a spatial squeeze–excitation (sSE) module and a channel squeeze–excitation (cSE) module. The sSE module reduces the channel information in feature maps to perform dimensionality reduction, and the compressed feature maps are then normalized by the sigmoid function  $\sigma(\cdot)$  to obtain important spatial information, thereby invigorating key spatial features and increasing focus on crucial channel features. Meanwhile, the cSE module adjusts feature maps based on feature correlations across different channels. It compresses the feature map of size  $H \times W \times C$  through global average pooling, followed by activation through the ReLU function  $\delta(\cdot)$  and sigmoid normalization  $\sigma(\cdot)$  to derive the importance of channel features, thereby enhancing attention to vital features [23].



**Figure 5.** Attention mechanism model.

When UAVs capture images of insulators, variations in the shooting angles often lead to the occlusion of the insulators in the photographs. To enhance the recognition of the occluded objects, this paper adopts DIoU-NMS in place of the conventional NMS technique [24]. The definition of DIoU-NMS is outlined by the following formula:

$$s_i = \begin{cases} s_i & PIoU - RDIoU(M, B_i) < \varepsilon \\ 0 & PIoU - RDIoU(M, B_i) \geq \varepsilon \end{cases} \quad (5)$$

$$RDIoU(M, B_i) = \frac{\rho^2(M, B_i)}{c^2} \quad (6)$$

In the equation,  $M$  represents the prediction box with a higher prediction score,  $B_i$  denotes the other prediction boxes,  $\rho$  is the Euclidean distance between  $M$  and  $B_i$ , and  $c$  is the diagonal distance of the smallest enclosing rectangle covering both  $M$  and  $B_i$ . DIoU-NMS effectively determines whether two overlapping boxes belong to the same object and efficiently suppresses bounding boxes. Compared to ground-level natural perspectives, the overlap rate of insulators is lower when viewed from a UAV perspective; hence, a small threshold  $\varepsilon$  is employed in this study to enhance the accuracy of the SSD algorithm in detecting insulator targets [25].

### 3.3. Model Compression and Fine-Tuning

To mitigate the reliance of the SSD network model on computational power, storage space, and other resources of edge intelligent terminals, this study employs channel pruning strategies to compress the enhanced SSD network model [26]. In order to accelerate the model convergence, *BN* layers are introduced after convolutional layers. The *BN* layers process the input data through shift and scaling parameters, normalizing the outputs of each convolutional layer within a reasonable range, as depicted in Equations (7) and (8).

$$Y = BN(X) = \gamma \frac{X - \mu}{\sqrt{\sigma^2 + \varepsilon}} + \beta \quad (7)$$

$$Y = \lim_{\gamma \rightarrow 0} \gamma \frac{X - \mu}{\sqrt{\sigma^2 + \varepsilon}} + \beta = \beta \quad (8)$$

In this context,  $X$  represents the input to the *BN* layer and  $Y$  denotes the output from the *BN* layer.  $\sigma$  and  $\mu$  are, respectively, the variance and mean computed over a mini-batch for the *BN* layer.  $\beta$  serves as a bias compensation in the normalization process, while  $\gamma$  acts as a scaling factor post-normalization, signifying the importance of channels.  $\varepsilon$  is a small non-zero constant to prevent division by zero. When  $\gamma$  approaches zero, the activation function following the *BN* layer maps the channel inputs to smaller output values [27], suggesting that the corresponding channel contributes minimally to the *BN* layer's output. Consequently, this redundant channel can be pruned, leading to a lightweight network architecture.

During conventional training, the model's loss function does not incorporate  $\gamma$ , resulting in a post-training distribution of  $\gamma$  that tends towards a normal distribution with most values close to 1, making pruning of the model challenging. To identify redundant channels,  $\gamma$  needs to be incorporated into the loss function for sparsification training, with an L1 regularization imposed on  $\gamma$  to drive the model parameters towards structured sparsity, thereby facilitating the identification of crucial channels [28]. The modified loss function is expressed as Equation (9).

$$L = \sum_{(x, y)} l(f(x, W), y) + \lambda \sum_{\gamma \in \Gamma} g(\gamma) \quad (9)$$

In the equation, the first summation represents the loss function of the conventionally trained model, while the second summation denotes the L1 regularization penalty term.  $L$  signifies the loss function for sparse training, with  $\Gamma$  encompassing all prunable channels.



The function  $g(\gamma)$  embodies the L1 regularization, here defined as  $g(\gamma) = |\gamma|$ . Initially, the model undergoes standard training. Following this, the well-trained model is subjected to sparse training via the loss function  $L$ , promoting sparsity. Upon the completion of the sparse training, the relevance  $\gamma$  of redundant channels diminishes towards zero, thus accomplishing model lightweighting.

To address the degradation in performance resulting from model pruning, this study employs knowledge distillation to fine-tune the model post-channel pruning [29]. Leveraging transfer learning, complex teacher networks guide simpler student networks, migrating knowledge to the student model. In this work, YOLOv5 is selected as the teacher network, as depicted in Figure 6, with the student network adopting a hint-based learning strategy to imbibe pertinent features from the teacher network [30]. To counteract the imbalance between insulator targets and the background during object recognition, a weighted cross-entropy loss is employed in the knowledge distillation network. In pursuit of further enhancing the network performance, the regression outputs from the teacher network are used as upper bounds, ensuring that the student network is not penalized even if it outperforms the teacher, thus fostering an environment conducive to learning without constraints [31]. This methodology promotes the preservation and enhancement of critical detection capabilities in the distilled, lightweight model [32,33].

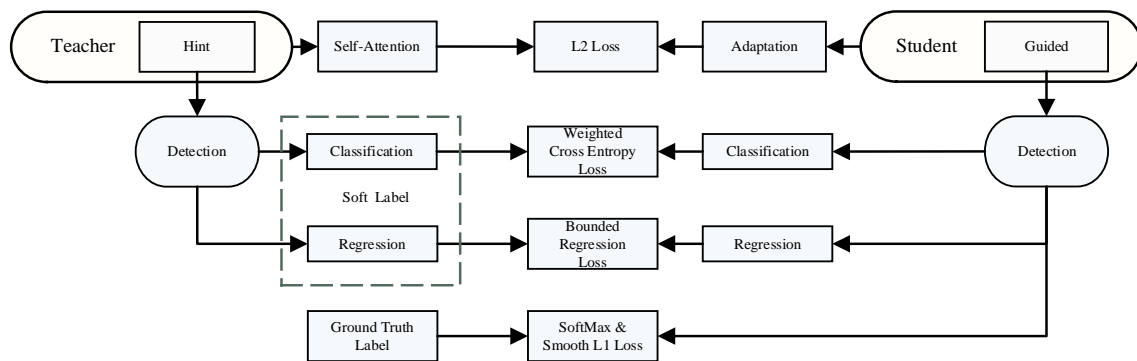


Figure 6. Knowledge distillation diagram.

## 4. Experimental Results and Analysis

### 4.1. Experimental Environment

The experiments reported herein were conducted on a 64-bit Windows 11 operating system, utilizing the deep learning framework PyTorch. The detailed configuration of the experimental environment is presented in Table 2.

Table 2. Experimental environment configuration.

Category	Parameter
CPU	12th Gen Intel(R) Core(TM) i7-12700KF 3.6 GHz
Memory	32 G
GPU	NVIDIA GeForce RTX 3090Ti
GPU memory	24 G
OS	Windows 11
CUDA version	CUDA 11.0
cuDNN	cuDNN 7.6.5
Language	Python 3.6

### 4.2. Datasets and Training

To validate the effectiveness of the algorithm presented in this paper, an open-source visible light insulator dataset was selected. In order to enhance the generalization capability of the model, data augmentation techniques were applied to a portion of the sample images, thereby increasing the diversity of the dataset. These techniques included image cropping,

stitching, color space transformation, resizing, among others. The augmented dataset was then annotated using the LabelImg (version 1.8.6) tool, with insulators in the dataset labeled consistently as 'Insulator'. The annotation format adhered to PascalVOC standards, ensuring uniformity across similar objects. Upon the completion of the labeling process, XML files were generated and stored within a label directory, each corresponding to an annotated image. The annotated files and their respective dataset images were meticulously paired and subsequently split into training and testing sets at an 8:2 ratio. The contents of the segmented annotation files were further converted into a training text and testing text, formatted according to predefined specifications, to facilitate model training.

The training set was fed into an improved SSD network model, with the maximum learning rate initialized at 0.01 and decreased to a minimum of 0.0001 throughout training. A batch size of 16 was employed to balance the computational efficiency and memory utilization. The model underwent 300 iterations of training, during which the optimal weights were saved for future deployment. The input image resolution was standardized to  $416 \times 416$  pixels to accommodate the architecture's requirements and enhance feature extraction.

The convergence curve of the training loss for the enhanced SSD model is illustrated in Figure 7, demonstrating the model's learning progression and stability over the course of the training epochs. The pseudo-code of the proposed model (Algorithm 1) is presented in the following form:

---

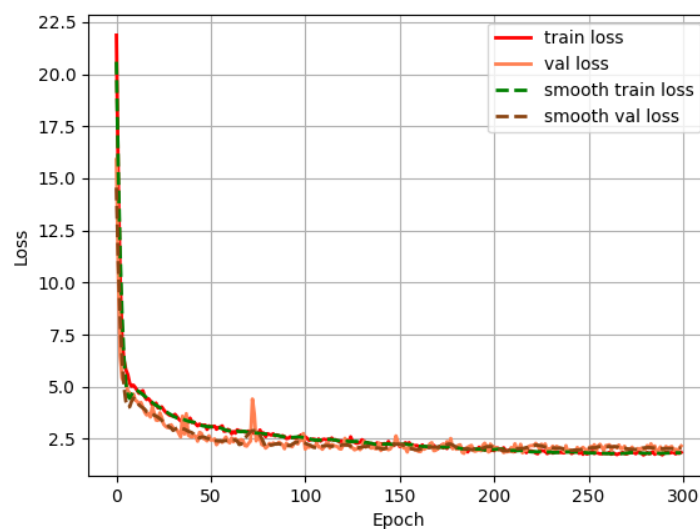
**Algorithm 1:** pseudo-code of the proposed model

---

Input: An image to be detected

Output: An image with detection results

- 1: Resize the input image to  $416 \times 416$  and normalize it.
  - 2: Pass the processed image through the backbone network to extract features.
  - 3: Feed the extracted features into the network model (backbone, neck, and head) to obtain candidate bounding boxes.
  - 4: For each candidate bounding box:
    - Perform classification and bounding box regression;
    - Decode the regression results to determine the final position of the bounding box;
    - Apply DIoU-NMS to filter out overlapping detections;
    - Map the detection result onto the original image.
  - 5: Return the image with the overlaid detection results.
- 



**Figure 7.** Training loss curve of improved SSD model.

### 4.3. Evaluating Indicator

This paper evaluates the enhanced SSD algorithm as using metrics including mean Average Precision ( $mAP$ ), Precision ( $P$ ), Recall ( $R$ ), Frames Per Second (FPS), the number of parameters, and Floating-point Operations Per Second (FLOPs). A higher  $mAP$  indicates greater detection accuracy, while larger numbers of parameters and higher computational loads signify a bulkier algorithm. Particularly, smart edge devices impose stringent constraints on the model size in terms of both the parameter count and computational demands. The term  $mAP@0.5$  signifies the average precision across all classes when the Intersection over Union (IoU) threshold is set to 0.5, reflecting the trend of precision as recall varies.  $R$  measures the proportion of true positive samples correctly identified, thereby gauging the extent of missed detections.  $P$ , on the other hand, assesses the fraction of predicted positive samples that are indeed true positives, indicating the rate of false alarms [34]. FPS quantifies the speed of detection, with a higher FPS translating to faster detection. FLOPs is used to evaluate the computational complexity of the model. The relevant formulas are as follows, where  $Tp$  denotes true-positive predictions,  $Fp$  represents false-positive predictions (negative samples incorrectly labeled as positive), and  $Fn$  indicates false-negative predictions (positive samples mislabeled as negative).

$$P = Tp / (Tp + Fp) \quad (10)$$

$$R = Tp / (Tp + Fn) \quad (11)$$

$$AP = \int_0^1 P(R) dR \quad (12)$$

$$mAP = \frac{\sum_{i=1}^N AP_i}{N} \quad (13)$$

### 4.4. Ablation Experiment

To verify the performance of the improved SSD model in detecting insulator targets, ablation experiments were conducted to compare the original SSD network with the model proposed in this paper. The setup for these ablation experiments is summarized in Table 3, where “√” denotes the inclusion of a module, and “×” indicates its exclusion. The outcomes of these ablation experiments are presented in Table 4.

**Table 3.** Ablation experimental design.

Models	VGG-16	Ghost Module	SimSPPF	scSE	FPN+PAN	Lightweight
SSD	√	×	×	×	×	×
A	×	√	×	×	×	×
B	×	√	√	×	×	×
C	×	√	√	√	×	×
D	×	√	√	√	√	√

**Table 4.** Ablation experimental results.

Models	Parameters	FLOPs	$P$	$R$	FPS f/s	$mAP@0.5/\%$
SSD	26.15 M	118.95 G	0.95	0.81	132	96.8%
A	5.07 M	3.21 G	0.96	0.79	111	93.2%
B	5.20 M	3.38 G	0.97	0.77	101	94.8%
C	7.04 M	3.39 G	0.94	0.81	101	95.6%
D	0.61 M	1.49 G	0.78	1	67	98.0%

According to Table 3, comparing Model A with the original SSD algorithm reveals that after replacing the SSD's backbone feature extraction network, VGG-16, with the Ghost Module, the model size is reduced by 80.6%, albeit at the expense of a decrease in the detection accuracy. This confirms that while the Ghost Module employs fewer parameters, leading to a smaller model size, it also has an adverse effect on the detection precision. Comparing Models A, B, and C illustrates that the introduction of the SimSPPF structure and scSE attention mechanism leads to negligible changes in the model size, but improves the accuracy, validating that these enhancements strengthen the model's comprehension and processing of input data, thereby enhancing the detection accuracy. The comparison between Models C and D shows that Model D, which incorporates the FPN+PAN structure and undergoes channel pruning and knowledge distillation, achieves a 91.3% reduction in model size while improving the detection accuracy by 2.4%. This evidence supports the notion that the FPN+PAN structure, along with knowledge distillation and channel pruning, significantly reduces the model size while effectively boosting the detection accuracy.

#### 4.5. Comparison of Different Algorithm Effects

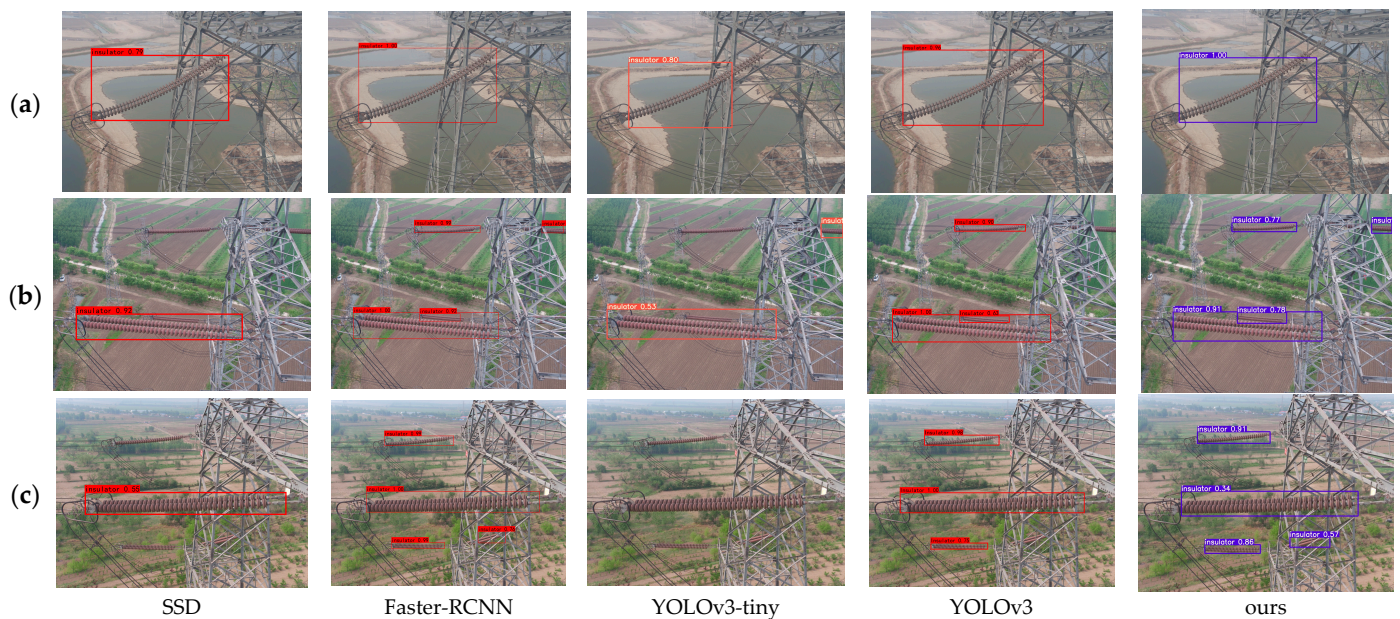
To validate the efficacy of the proposed algorithm, comparative experiments were conducted against classical object detection algorithms using the same dataset. The experimental environment was consistent with that of the ablation experiments, with all models trained for 300 iterations. The results of these comparative experiments are summarized in Table 5.

**Table 5.** Improved SSD algorithm compared with mainstream target detection algorithm.

Models	Lightweight	FPS/f/s	P	R	Parameters	FLOPs	mAP@0.5/%
SSD		132	0.95	0.81	26.15 M	118.95 G	96.8%
YOLOv3		71	0.80	0.98	61.52 M	65.60 G	96.9%
YOLOv5		94	0.94	0.99	47.06 M	115.92 G	98.2%
Faster-RCNN		22	0.70	1	137.10 M	370.21 G	98.6%
Ghost-YOLOv3	✓	55	0.77	0.98	46.45 M	25.32 G	95.5%
YOLOv3-Tiny	✓	142	0.62	0.81	8.67 M	5.49 G	73.8%
Ours	✓	67	0.78	1	0.61 M	1.49 G	98.0%

According to the results in Table 5, it can be seen that the algorithm in this paper has the smallest computational and parameter requirements. Compared with algorithms such as YOLOv3, YOLOv5, and Faster RCNN, the model has significantly reduced computational and parameter requirements, with an average accuracy of 98%, slightly lower than YOLOv5 and Faster RCNN. Compared with lightweight algorithms such as Ghost-YOLOv3 and YOLOv3-Tiny, our algorithm has a lower parameter and computational complexity, and higher detection accuracy. In summary, while maintaining high accuracy, the algorithm proposed in this article has the lowest number of model parameters and computational complexity.

Figure 8 illustrates the detection results of various algorithms. The figure indicates that the proposed algorithm and Faster-RCNN achieve the highest accuracy in detecting insulators in Figure 8a. In the insulator images depicted in Figure 8b,c, the original SSD, YOLOv3-tiny, and YOLOv3 all exhibit varying degrees of missed detections. Faster-RCNN achieves the highest detection accuracy; however, Table 4 reveals that it is larger in size and slower in detection speed. Upon a comprehensive comparison, the proposed algorithm shows a superior overall performance, confirming the effectiveness of the improvement strategies outlined herein.



**Figure 8.** Detection results of insulator detection models on different data sets. (a) shows the results of different algorithms on a single insulator. (b) shows the detection results of different models when the insulators in the photos are incomplete and the scales are different. (c) shows the detection results of different models with different insulator scales.

## 5. Conclusions

Addressing the challenge of balancing the detection accuracy with the model size in power transmission line insulator inspection algorithms, which hampers their deployment on embedded devices, this paper presents a lightweight insulator target detection model based on improved SSD. The main contents of this paper are as follows:

(1) Through the introduction of the lightweight Ghost Module network, the initial lightweight of the model is realized. (2) By introducing a SimSPPF structure and FPN+PAN structure, the fusion of the local and global features of the model is promoted, and the feature extraction capability of the model is enhanced. (3) By introducing multiple spatial and channel squeeze incentive (scSE) attention mechanism modules, the model's ability to focus on key features is improved. (4) By reducing the number of detection headers and introducing the DIOU-NMS mechanism, the detection speed is improved, and the model's recognition ability of occluded and overlapping targets is improved. At the same time, through channel pruning and knowledge distillation, the number of model parameters is further reduced and the accuracy of model detection is improved.

The major conclusions in this paper are listed as follows.

By using the lightweight Ghost Module network to replace the original feature extraction network, VGG-16, the model size is reduced by 80.6%, indicating that the lightweight module can effectively reduce the model size. However, the model detection accuracy is also reduced.

With the introduction of the SimSPPF structure and scSE attention mechanism, although the model size is not significantly reduced, the detection accuracy is effectively improved, which proved that the SimSPPF structure and scSE attention mechanism can effectively improve the model's ability to understand and process input data, thus effectively improving the detection accuracy.

Combined with the FPN+PAN structure, channel pruning and knowledge distillation were performed on the model. The number of parameters decreased by 91.3%, from 7.04 M to 0.61 M, and the detection accuracy increased by 2.4%, from 95.6% to 98.0%, which verified the effectiveness of the improvement measures taken in this paper.

Compared with other models, the proposed model has the lowest number of parameters while maintaining high detection precision, and other parameters have been effectively improved, which shows that the improvement strategy in this paper can significantly improve the model performance. The model in this paper not only ensures the detection accuracy, but also minimizes the consumption of hardware resources, and can meet the requirements of the deployment and application on edge intelligent terminals. However, the resource utilization and energy consumption of the algorithm have not been deeply studied in this paper. In the future, how to reduce the resource utilization rate and energy consumption of the algorithm at the edge intelligent terminal will be deeply studied, and the endurance time of the UAV can be further improved while ensuring the accuracy and efficiency of target detection and defect identification.

**Author Contributions:** Conceptualization, B.Z.; methodology, Y.Z.; software, Y.Z.; validation, D.H., Z.Z., S.H. and K.Y.; formal analysis, W.Z.; investigation, Y.X.; resources, B.Z. and Z.L.; data curation, Y.Z.; writing—original draft preparation, D.H.; writing—review and editing, D.H. and B.Z.; visualization, Z.Z. and K.Y.; All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported in part by the Science and Technology Project of Education Department of the Jiangxi Province (Grant. GJJ211942).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The raw data supporting the conclusions of this article will be made available by the authors on request.

**Acknowledgments:** The authors gratefully acknowledge the support of the Science and Technology Project of the Education Department of the Jiangxi Province (Grant. GJJ211942).

**Conflicts of Interest:** The authors declare no conflicts of interest. Author Zhilong Li was employed by the State Grid Shanghai Municipal Electric Power Company Maintenance Company. The remaining 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.

## References

1. Wei, L.; Jin, J.; Deng, K.; Liu, H. Insulator defect detection in transmission line based on an improved lightweight YOLOv5s algorithm. *Electr. Power Syst. Res.* **2024**, *233*, 110464. [[CrossRef](#)]
2. Qi, Y.; Mu, S.; Wang, J.; Wang, L. Intelligent Recognition of Transmission Line Inspection Image Based on Deep Learning. *J. Phys. Conf. Ser.* **2021**, *1757*, 012056. [[CrossRef](#)]
3. Shahrzad, F.; Azam, K.; Hossein, N. PTSRGAN: Power transmission lines single image super-resolution using a generative adversarial network. *Int. J. Electr. Power Energy Syst.* **2024**, *155*, 109607.
4. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Region-Based Convolutional Networks for Accurate Object Detection and Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *38*, 142–158. [[CrossRef](#)]
5. Arora, N.; Kumar, Y.; Karkra, R.; Kumar, M. Automatic vehicle detection system in different environment conditions using fast R-CNN. *Multimed. Tools Appl.* **2022**, *81*, 18715–18735. [[CrossRef](#)]
6. Zhao, W.; Xu, M.; Cheng, X.; Zhao, Z. An insulator in transmission lines recognition and fault detection model based on improved faster RCNN. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 5016408. [[CrossRef](#)]
7. Zhao, Z.; Zhen, Z.; Zhang, L.; Qi, Y.; Kong, Y.; Zhang, K. Insulator Detection Method in Inspection Image Based on Improved Faster R-CNN. *Energies* **2019**, *12*, 1204. [[CrossRef](#)]
8. Haijian, H.; Yicen, L.; Haina, R. Detection of Insulators on Power Transmission Line Based on an Improved Faster Region-Convolutional Neural Network. *Algorithms* **2022**, *15*, 83. [[CrossRef](#)]
9. Redmon, J.; Divvala, K.S.; Girshick, B.R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. *arXiv* **2015**, arXiv:1506.02640.
10. Gu, J.; Hu, J.; Jiang, L.; Wang, Z.; Zhang, X.; Xu, Y.; Zhu, J.; Fang, L. Research on Object Detection of Overhead Transmission Lines Based on Optimized YOLOv5s. *Energies* **2023**, *16*, 2706. [[CrossRef](#)]
11. Wang, T.; Zhai, Y.; Li, Y.; Wang, W.; Ye, G.; Jin, S. Insulator Defect Detection Based on ML-YOLOv5 Algorithm. *Sensors* **2023**, *24*, 204. [[CrossRef](#)]
12. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. SSD: Single Shot MultiBox Detector. *arXiv* **2015**, arXiv:1512.02325.

13. Liu, X.; Li, Y.; Shuang, F.; Gao, F.; Zhou, X.; Chen, X. ISSD: Improved SSD for Insulator and Spacer Online Detection Based on UAV System. *Sensors* **2020**, *20*, 6961. [[CrossRef](#)] [[PubMed](#)]
14. Liu, J.; Hu, M.; Dong, J.; Lu, X. The application of a lightweight model FA-YOLOv5 with fused attention mechanism in insulator defect detection. *Front. Energy Res.* **2023**, *11*, 1283394. [[CrossRef](#)]
15. Deng, X.; Li, S. An Improved SSD Object Detection Algorithm Based on Attention Mechanism and Feature Fusion. *J. Phys. Conf. Ser.* **2023**, *2450*, 012088. [[CrossRef](#)]
16. Zhai, S.; Shang, D.; Wang, S.; Dong, S. DF-SSD: An Improved SSD Object Detection Algorithm Based on DenseNet and Feature Fusion. *IEEE Access* **2020**, *8*, 24344–24357. [[CrossRef](#)]
17. Qian, H.; Wang, H.; Feng, S.; Yan, S. FESSD: SSD target detection based on feature fusion and feature enhancement. *J. Real-Time Image Process.* **2023**, *20*, 2. [[CrossRef](#)]
18. Hong, F.; Lu, C.H.; Tao, W.; Jiang, W. Improved SSD Model for Pedestrian Detection in Natural Scene. *Wirel. Commun. Mob. Comput.* **2022**, *2022*, 1500428. [[CrossRef](#)]
19. Han, K.; Wang, Y.; Tian, Q.; Guo, J.; Xu, C.; Xu, C. GhostNet: More features from cheap operations. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 1577–1586.
20. Li, L.; Zheng, C.; Mao, C.; Deng, H.; Jin, T. Scale-Insensitive Object Detection via Attention Feature Pyramid Transformer Network. *Neural Process. Lett.* **2021**, *54*, 581–595. [[CrossRef](#)]
21. Hu, H.; Zhu, Z. Sim-YOLOv5s: A method for detecting defects on the end face of lithium battery steel shells. *Adv. Eng. Inform.* **2023**, *55*, 101824. [[CrossRef](#)]
22. Nie, P.; Guo, Y.; Lou, B.; Yang, C.; Cao, L.; Pan, W. Tool wear monitoring based on scSE-ResNet-50-TSCNN model integrating machine vision and force signals. *Meas. Sci. Technol.* **2024**, *35*, 086117. [[CrossRef](#)]
23. Yan, P.; Sun, Q.; Yin, N.; Hua, L.; Shang, S.; Zhang, C. Detection of coal and gangue based on improved YOLOv5.1 which embedded scSE module. *Measurement* **2022**, *188*, 110530.
24. Shepley, A.; Falzon, G.; Kwan, P. Confluence: A Robust Non-IoU Alternative to Non-Maxima Suppression in Object Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *10*, 11561–11574. [[CrossRef](#)]
25. Jiang, L.; Nie, W.; Zhu, J.; Gao, X.; Lei, B. Lightweight object detection network model suitable for indoor mobile robots. *J. Mech. Sci. Technol.* **2022**, *36*, 907–920. [[CrossRef](#)]
26. He, Y.; Dong, X.; Kang, G.; Fu, Y.; Yan, C.; Yang, Y. Asymptotic Soft Filter Pruning for Deep Convolutional Neural Networks. *IEEE Trans. Cybern.* **2020**, *50*, 3594–3604. [[CrossRef](#)]
27. Huang, M.; Liu, Y.; Zhao, L.; Wang, G. A lightweight deep neural network model and its applications based on channel pruning and group vector quantization. *Neural Comput. Appl.* **2023**, *36*, 5333–5346. [[CrossRef](#)]
28. Zhang, Z.; Kong, S.; Peng, K. Improved YOLOv4 Power Insulator Fault Detection. *J. Phys. Conf. Ser.* **2021**, *2010*, 012148. [[CrossRef](#)]
29. Arima, K.; Nagata, F.; Shimizu, T.; Otsuka, A.; Kato, H.; Watanabe, K.; Habib, M.K. Improvements of detection accuracy and its confidence of defective areas by YOLOv2 using a data set augmentation method. *Artif. Life Robot.* **2023**, *28*, 625–631.
30. Zhao, Z.; Lv, X.; Xi, Y.; Miao, S. Defect detection method for key area guided transmission line components based on knowledge distillation. *Front. Energy Res.* **2023**, *11*, 1287024.
31. Zhao, Z.; Lyu, J.; Chu, Y.; Liu, K. Toward generalizable robot vision guidance in real-world operational manufacturing factories: A Semi-Supervised Knowledge Distillation approach. *Robot. Comput. Integr. Manuf.* **2024**, *86*, 102639. [[CrossRef](#)]
32. Du, W.; Geng, L.; Zhao, Z.; Wang, C.; Huo, J. Decoupled knowledge distillation method based on meta-learning. *High-Confid. Comput.* **2024**, *4*, 100164. [[CrossRef](#)]
33. Wei, L.; Tong, Y. Enhanced-YOLOv8: A new small target detection model. *Digit. Signal Process.* **2024**, *153*, 104611. [[CrossRef](#)]
34. Li, Y.; Li, J.; Zhai, Y.; Meng, P. Detection of Self-explosive Insulators in Aerial Images Based on Improved YOLO v4. *J. Phys. Conf. Ser.* **2022**, *2320*, 012025.

**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.