

## Article

# A Visual Fault Detection Algorithm of Substation Equipment Based on Improved YOLOv5

Yuezhong Wu <sup>1</sup>, Falong Xiao <sup>1</sup>, Fumin Liu <sup>1</sup>, Yuxuan Sun <sup>1</sup>, Xiaoheng Deng <sup>2</sup>, Lixin Lin <sup>2,\*</sup> and Congxu Zhu <sup>3</sup>

<sup>1</sup> College of Railway Transportation, Hunan University of Technology, Zhuzhou 412007, China; wuyuezhong@hut.edu.cn (Y.W.); m20085400029@stu.hut.edu.cn (F.X.); m22085400032@stu.hut.edu.cn (F.L.); m22081101018@stu.hut.edu.cn (Y.S.)

<sup>2</sup> School of Electronic Information, Central South University, Changsha 410083, China; dxh@csu.edu.cn

<sup>3</sup> School of Information Science and Engineering, Central South University, Changsha 410083, China; zhucx@csu.edu.cn

\* Correspondence: llx@csu.edu.cn

**Abstract:** The development of artificial intelligence technology provides a new model for substation inspection in the power industry, and effective defect diagnosis can avoid the impact of substation equipment defects on the power grid and improve the reliability and stability of power grid operation. Aiming to combat the problem of poor recognition of small targets due to large differences in equipment morphology in complex substation scenarios, a visual fault detection algorithm of substation equipment based on improved YOLOv5 is proposed. Firstly, a deformable convolution module is introduced into the backbone network to achieve adaptive learning of scale and receptive field size. Secondly, in the neck of the network, a simple and effective BiFPN structure is used instead of PANet. The multi-level feature combination of the network is adjusted by a floating adaptive weighted fusion strategy. Lastly, an additional small object detection layer is added to detect shallower feature maps. Experimental results demonstrate that the improved algorithm effectively enhances the performance of power equipment and defect recognition. The overall recall rate has increased by 7.7%, precision rate has increased by nearly 6.3%, and mAP@0.5 has improved by 4.6%. The improved model exhibits superior performance.

**Keywords:** YOLOv5; fault detection; deformable convolution; BiFPN



**Citation:** Wu, Y.; Xiao, F.; Liu, F.; Sun, Y.; Deng, X.; Lin, L.; Zhu, C. A Visual Fault Detection Algorithm of Substation Equipment Based on Improved YOLOv5. *Appl. Sci.* **2023**, *13*, 11785. <https://doi.org/10.3390/app132111785>

Academic Editors: Rodolfo Haber, Krzysztof Ejmont, Aamer Bilal Asghar and Yong Wang

Received: 3 October 2023

Revised: 20 October 2023

Accepted: 24 October 2023

Published: 27 October 2023



**Copyright:** © 2023 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 power system is one of the most critical infrastructures in modern society and substations are indispensable components of the power system. In the power system, a large number of power terminal equipment may experience wear or damage due to factors such as service life or environmental conditions. If these defects reach a high severity level, they can lead to equipment failure [1–4]. With the continuous increase in electricity consumption, the operational stability and level of intelligence of the power system face higher demands. To ensure the reliable operation of the power system, regular inspections of substations are particularly important. This can promptly identify and address potential safety hazards [5].

In recent years, computer vision technology [6–8] has been widely applied in various fields such as image classification [9,10] and video analysis [11]. It has provided numerous feasible solutions for defect detection problems [12–15], enabling computers to perceive and recognize images or video information like humans. In the field of power equipment maintenance, the Internet of Things (IoT) deployed by the company consists of a large number of power terminal devices. Edge computing-based power monitoring and application systems [16,17] are generally used for analyzing and processing the collected images of power equipment to monitor their operational status. Therefore, object detection

techniques can serve as an edge service to handle and analyze a large volume of device status images.

Traditional object detection algorithms typically start by selecting candidate regions from the input image, followed by manual feature extraction and classification using feature classifiers [18–20]. However, these traditional models suffer from high time complexity and large window redundancy due to the exhaustive sliding window approach, resulting in long detection time and poor detection performance, especially on unfamiliar datasets. With the introduction of convolutional neural networks (CNNs) [21], the paradigm of visual perception has undergone a significant transformation.

Currently, deep learning-based object detection algorithms can be mainly divided into two categories. The first category is the region-based R-CNN-series algorithms [22–24], which achieve high recall rates by extracting image feature information with a minimal number of candidate boxes. Lei X. et al. [25] used a deep convolutional neural network method based on Faster R-CNN to identify and locate broken insulators and bird nests. Jiang A. et al. [26] proposed the Mask R-CNN model for pixel-level recognition of substation equipment, accurately identifying various types of substation devices. Liu Z. et al. [27] proposed an improved Mask R-CNN method that utilizes infrared image semantic segmentation for device feature extraction, making it suitable for infrared image power equipment detection tasks. Yin Z. et al. [28] proposed a detection system that combines edge computing and an enhanced Faster R-CNN for detecting device defects in substation scenarios using collected substation equipment video images. Ou J. et al. [29] presented an improved Faster R-CNN model for automatically detecting 16 types of electrical equipment in substations.

The YOLO series is another type of object detection algorithm [30–32]. These algorithms combine target localization and recognition based on global information, instead of using information from each candidate box, resulting in fast and efficient detection performance [33,34]. Wang X. et al. [35] achieved real-time detection of anomalies in power equipment infrared images by improving the Single Shot MultiBox Detector (SSD) algorithm. Cheng Y. [36] proposed an enhanced algorithm for insulator defect localization and detection in the YOLO (You Only Look Once) model to reduce interference during unmanned aerial vehicle inspections. Hu X. et al. [37] introduced the Path Aggregation Network (PANet) into YOLOv4 to enhance insulator defect detection and used focal loss as the loss function. Zheng H. et al. [38] improved the YOLOv3 model to enhance the detection accuracy of power equipment infrared images with similar ripple patterns. They introduced the Cross-Stage Partial module, which integrates the pathway aggregation network into the feature pyramid structure of the original model. Peng J. et al. [38] proposed the model compression algorithm ED-YOLO, which achieves obstacle avoidance in power line inspections with high accuracy and fast recognition speed. The logistic regression-based object detection algorithm solves detection as a regression problem, simplifying the training process and improving detection speed [39]. It is a practical solution for performing object detection tasks that meet the real-time requirements of power recognition scenarios, especially in edge computing applications.

Currently, deep neural networks face challenges in object detection, particularly in complex scenarios like power systems, industrial equipment, and outdoor environments. The distinction between the target and its surroundings is not significant. The original model struggles to differentiate pixels and detect distant targets that cannot be approached closely. Factors such as outdoor weather make it difficult to effectively capture the target for detection. Issues include poor robustness, weak ability to detect small targets in noisy environments, and ineffective target recognition. There is a need for a method applicable in substation scenarios to identify target image features and enhance model capabilities.

In the early stage, we investigated the latest works of the YOLO series, YOLOv5, YOLOv7, and YOLOv8. Compared to YOLOv5, YOLOv7 added Spatial Pyramid Pooling (SSP) module and Spatial Attention Module (SAM) modules. However, compared to YOLOv5, v7 has slightly slower detection speed and requires more training time but has higher accuracy. YOLOv8 is larger than the YOLOv5 model and has relatively slower

training speed. Although it is similar to the comparison between YOLOv7 and v5 in terms of accuracy, for detection speed and deployment scenarios, such as inspection tasks, autonomous driving, intelligent video surveillance, etc., YOLOv5 is more suitable for fast detection scenes.

YOLOv5 is a deep learning-based object detection model that operates in real time and with high efficiency. It offers advantages such as fast training speed, high detection accuracy, and adaptability to various datasets. Figure 1 illustrates the process of reading image and video data through input. The backbone network employs convolution operations to extract features from the input image. The neck section further extracts features by utilizing multiple layers of convolutional operations and merging feature maps from different levels. Ultimately, three feature maps of varying scales are produced as output. The output component converts these feature maps into object detection results, which include target positions and class information.

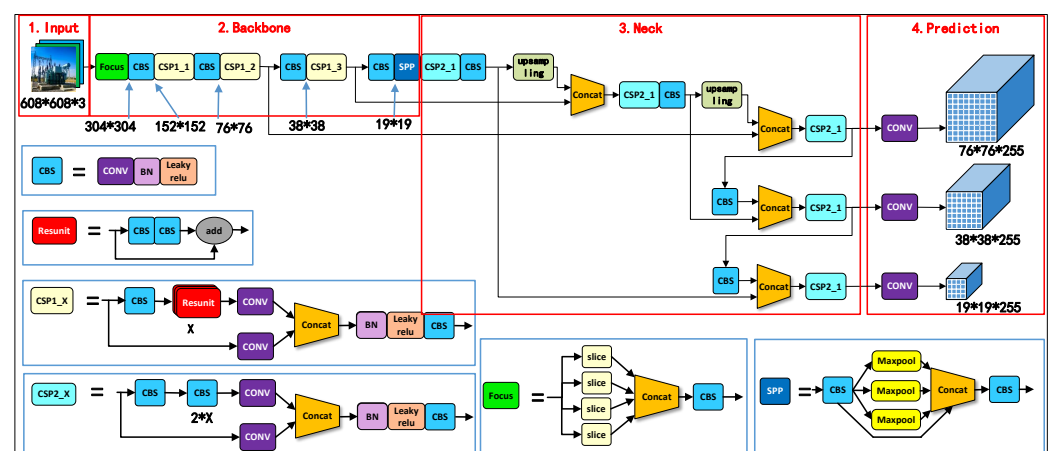


Figure 1. Overall structure diagram of YOLOv5.

Compared to traditional object detection algorithms, the YOLO-series algorithms are highly efficient, fast, and scalable. Among the YOLO-series algorithms, YOLOv5 stands out for its high stability, wide range of applications, and utilization of multiple resources. This paper presents improvements to the YOLOv5 algorithm without significantly increasing its parameter size compared to the original version.

In light of the issues faced by current defect detection models, such as susceptibility to external environmental interference, inadequate recognition performance for devices of varying scales in real power scenarios, and incomplete feature extraction, this paper presents the following contributions: (1) Incorporating a variable convolution module into YOLOv5's backbone network to enhance the capture of finer details in the feature map. (2) Introducing sampling between different levels within YOLOv5's feature extraction component to facilitate information transmission and fusion across diverse scales. (3) Experimental validation and algorithm analysis are conducted to demonstrate the feasibility of the proposed method.

The subsequent sections are organized as follows. Section 2 introduces the original YOLOv5 algorithm and explains the improvements made in this paper. Firstly, variable convolution is added to the network structure, and secondly, upsampling is added between different levels to facilitate information transmission and fusion across scales, enhancing the feature network. Lastly, a prediction layer is included. Section 3 discusses the ablation experiments conducted on the proposed model in this paper and analyzes and validates the comparative results. Section 4 summarizes the research conducted in this paper and provides prospects for future work.

## 2. Framework

### 2.1. YOLOv5 Algorithm Improvements

In light of the complex nature, varying defect morphologies, and significant differences in equipment scenes at substations, this paper proposes an improved YOLOv5 object detection algorithm to enhance semantic distinguishability and alleviate global class confusion. The overall network structure is illustrated in Figure 2.

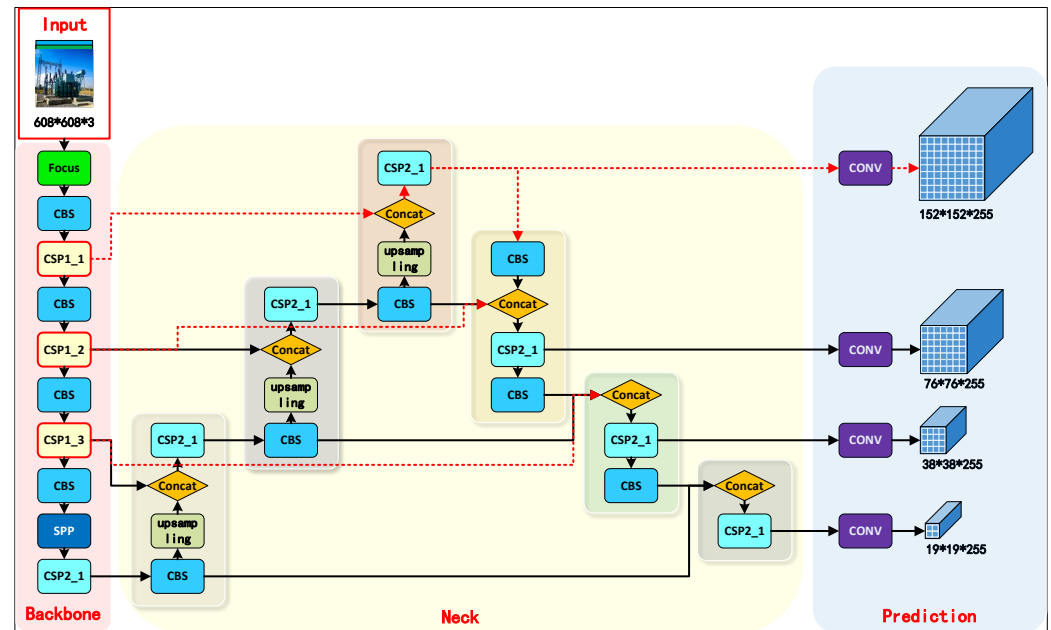


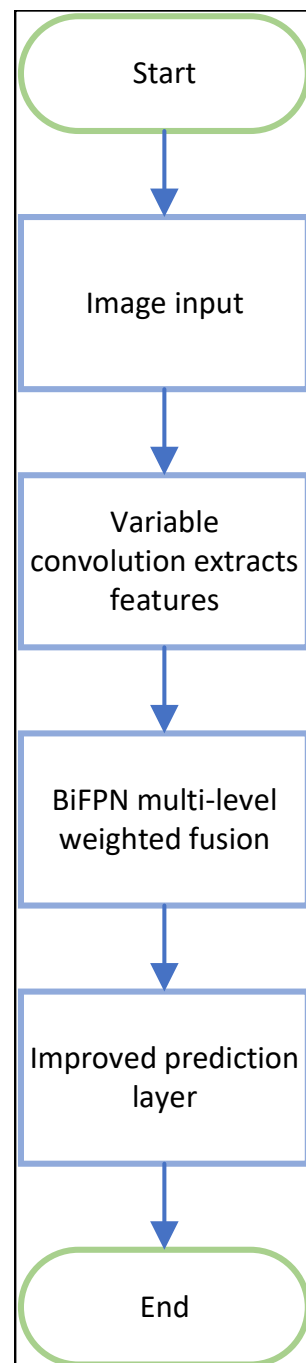
Figure 2. Improved YOLOv5 framework.

The main improvements include the backbone network, feature layers, and prediction layers. Firstly, we introduce deformable convolution in the backbone network section to better adapt to changes in target shape by deforming the input feature map. Secondly, instead of PANet, we use a simple and effective BiFPN structure in the neck of the network to fuse multi-level features from the backbone with weighted fusion. The BiFPN structure fully utilizes multi-scale information while having less computational cost and better performance. Finally, we improve the structure of the model's prediction layer by adding a prediction head to handle large-scale variations in targets. These improvements enhance performance in recognizing small targets and improve the adaptability of object detection algorithms for complex scenes and diverse device configurations in substation equipment images. Figure 3 shows the overall algorithm flowchart.

#### 2.1.1. Fusion-Deformable Convolutional Modules

In order to enhance the semantic distinguishability of the YOLOv5 model for substation equipment defect detection, we propose an improvement measure, which is the introduction of deformable convolutions. Traditional convolution operations can only sample the input feature maps in a fixed manner and cannot handle variations in target shape. However, in substation equipment defect detection, targets with different scales and shapes may coexist at different positions within the same feature layer, which limits the flexibility of object detection. To address this issue, we introduce deformable convolutions. Deformable convolutions can adaptively learn the sampling positions and shapes to better accommodate the shape variations of the targets. It transforms the sampling point positions in traditional convolutions into learnable forms, enabling the network to learn more expressive feature representations. In the improved YOLOv5 detection model, we incorporate deformable convolutions into the backbone network. This allows for better

feature extraction and small object detection in substation equipment defect detection, thus improving the performance of the model.



**Figure 3.** Algorithm flowchart.

Figure 4 shows a schematic diagram of a  $3 \times 3$  deformable convolution. In deformable convolutions, the convolution layer can calculate offsets relative to the input feature map and use these offsets to sample the input feature map. This allows for modeling transformations of the targets. The deformable convolution kernel is distributed spatially in the same layer as the current convolution layer, with offset size matching that of the input feature map. These offsets can be directly applied to each pixel in the input feature map for position shifting and adjustment.

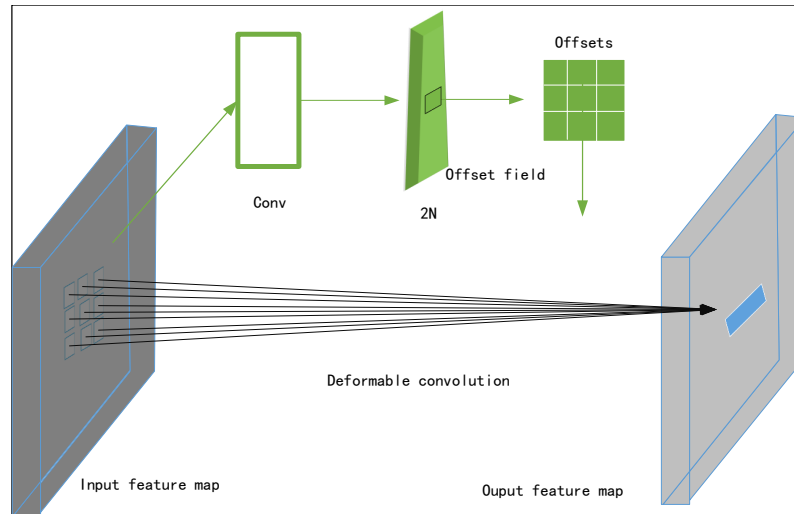


Figure 4. Schematic diagram of 3 × 3 deformable convolution.

In previous work [40], the convolution operation was performed using a fixed-size matrix  $R$  to sample the input feature map. This approach had limitations in terms of receptive field size and dilation. Subsequently, convolutional kernel weighting was applied to the sampled points.

$$R = \begin{bmatrix} (-1, -1) & (-1, 0) & (-1, 1) \\ (0, -1) & (0, 0) & (0, 1) \\ (1, -1) & (1, 0) & (1, 1) \end{bmatrix} \tag{1}$$

For each position  $P_n$  in the output feature map, its expression is as follows:

$$y(p_0) = \sum w(p_n)x(p_0 + p_n) \tag{2}$$

where  $p_n$  represents an unknown enumeration listed in  $R$ .

In the operation of deformable convolution, a regular network is used to convolve and sample the input feature map. The collection of sampled positions is shifted by combining the offset with the predicted weight  $\Delta m_n$  for each sampling point. It can be represented as:

$$y(p_0) = \sum_{p_n \in R} w(p_n)x(p_0 + p_n + \Delta p_n) \Delta m_n \tag{3}$$

where  $\Delta p_n = 1, 2, \dots, N$  represents the offsets,  $p_n + \Delta p_n$  is the sampling point after the offset, and  $\Delta p_n$  is typically a fractional value that cannot accurately obtain the pixel value of that position. Therefore, the pixel value of  $x$  can be calculated using bilinear interpolation as follows:

$$g(a, b) = \max(0, 1 - |a - b|) \tag{4}$$

$$G(q, p) = g(q_x, p_x) \times g(q_y, p_y) \tag{5}$$

$$x(p) = \sum_q G(q, p) \times x(q) \tag{6}$$

where  $p = p_0 + p_n + \Delta p_n$  is the offset position,  $x(q)$  is the pixel value of the four known neighboring pixels in the feature map, and  $G(\cdot, \cdot)$  represents the weights corresponding to the four coordinates.

The traditional Cross-Stage Partial (CSP) module divides the feature map into two branches to extract features and then merges them hierarchically, reducing computational

complexity while maintaining accuracy. In the YOLOv5 network, there are two variants of the CSP structure: CSP1\_X and CSP2\_X, as shown in Figure 5. The backbone network employs the sequential connection of CSP1\_X, which consists of four convolutional layers and X residual modules, effectively reducing computational and memory overhead. On the other hand, CSP2\_X is used in the neck network with a parallel connection. The main difference between CSP2\_X and CSP1\_X in the backbone network is that ordinary convolutional modules replace residual modules in CSP2\_X, enhancing feature fusion capability. This paper focuses on improving the backbone network’s CSP1\_X module by replacing traditional convolution in its lower branch with deformable convolution, as depicted in Figure 6. This modification ensures that the improved model can adaptively accomplish target sampling with only a slight increase in computational complexity.

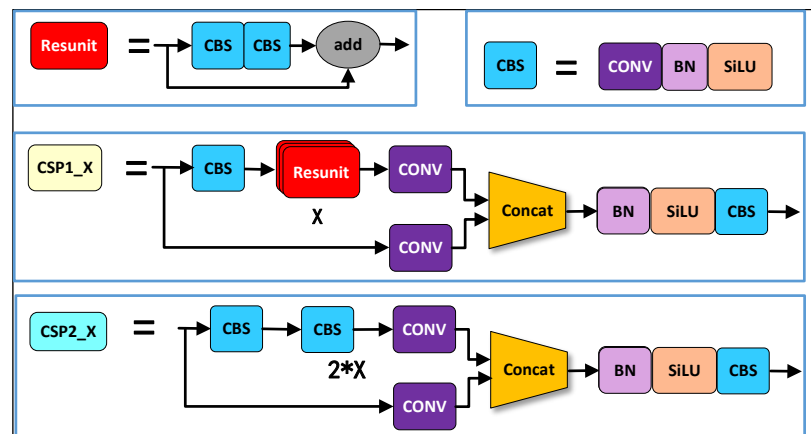


Figure 5. Traditional CSP module.

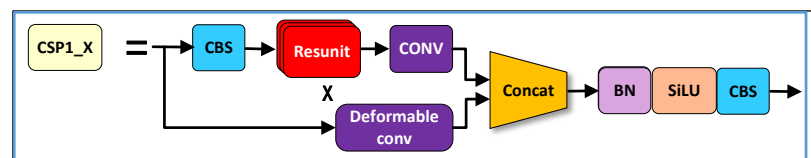


Figure 6. Improved CSP module.

### 2.1.2. Feature Network Improvements

Neck is commonly used in models for better feature fusion and image feature extraction, reducing redundant information by reducing the dimension of feature maps and enhancing feature representation capabilities by fusing features at different levels. As shown in Figure 7a, an efficient PANet method is proposed and adopted for the first time in YOLOv5, which proves the effectiveness of the bidirectional fusion scheme by performing quadratic feature fusion from the bottom up. However, in this paper, BiFPN achieves better results as a complex cross-level and cross-scale bidirectional fusion method, as shown in Figure 7b. Due to the small number of individual type pixels of defect images in the past dataset, the backbone network can effectively identify some shallow features defects, and when it comes to difficult to identify defects, deeper features may produce a large number of feature maps due to the complexity and depth of the network, containing rich information but also containing some redundant or irrelevant information. Through BiFPN, the information transfer and fusion between different levels of the feature pyramid introduces more connections and backpropagation, which can better capture targets at different scales.

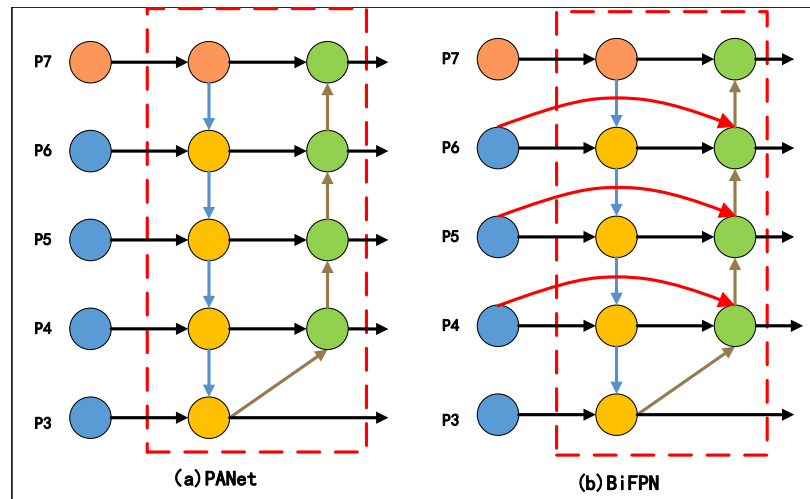


Figure 7. Feature network improvement.

In the feature network design, BiFPN changes the feature extraction and feature fusion methods on the basis of PANet and adds upsampling between different levels to enable the feature map to transmit and fuse information between different scales. This fusion method is not a simple series or superposition, but an adaptive fusion weight calculation method, which can automatically adjust the weight according to the resolution of the feature map to improve the fusion effect. On top of PANet, if the original input feature and the output feature are expressed at the same level, BiFPN does not add extra edges between them to fuse more features and therefore does not add too much extra computation. In this paper, BiFPN adjusts and optimizes the network by using the adaptive weighted fusion strategy of fast normalization fusion.

Fast normalized fusion and softmax-based fusion share similar learning methods and accuracy, as shown in the following equation:

$$O = \sum_i \frac{w_i}{\mathcal{E} + \sum_j w_j} I_i \tag{7}$$

where  $w_i$  is a learnable weight,  $w_i > 0$  is passed after the ReLU, and each normalization weight is between 0 and 1.

As a specific example, the fusion of two features shown in the sixth layer of Figure 7b can be described as follows:

$$p_6^{td} = Conv\left(\frac{w_1 P_6^{in} + w_2 Resize(P_7^{in})}{w_1 + w_2 + \mathcal{E}}\right) \tag{8}$$

$$p_6^{out} = Conv\left(\frac{w'_1 P_6^{in} + w'_2 P_6^{td} + w'_3 Resize(P_5^{out})}{w'_1 + w'_2 + w'_3 + \mathcal{E}}\right) \tag{9}$$

where  $P_6^{td}$  is the intermediate feature of level 6 in the up and down path, and  $P_6^{out}$  is the output feature of level 6 in the bottom-up path,  $\vec{p}^{in} = (p_1^{in}, p_1^{in}, \dots, p_l^{in})$ ,  $\vec{p}^{in}$  represents the input multiscale feature list,  $\vec{p}^{in}$  represents the input feature of horizontal  $l_i$ ,  $\vec{p}^{in} = (p_1^{td}, p_1^{td}, \dots, p_l^{td})$  represents the intermediate feature list on the path, and  $\vec{p}^{in}$  outputs a new  $\vec{p}^{out}$  by aggregating a series of different characteristics.

### 2.1.3. Add Prediction Layer

One reason why YOLOv5 is not suitable for detecting small targets is that there are fewer pixels representing features in small target samples, whereas YOLOv5 has a higher downsampling factor. The dataset of substation equipment contains numerous very small



targets, and as the deeper feature map gradually decreases in resolution, it leads to the loss of detailed information about these small targets. Consequently, accurate detection or classification becomes challenging. To address this issue, this paper introduces a small target detection layer to prevent the loss of details on deep feature maps. Initially, YOLOv5 only performed feature prediction on the last three layers after feature fusion. However, due to continuous downsampling causing small targets to lose their feature information, the performance of small target detection was not satisfactory. Therefore, this experiment adds a feature prediction layer along with the other three prediction layers. This four-headed structure helps mitigate the negative impact caused by variations in object scale (as depicted in Figure 2). Although adding this new prediction layer increases computational and memory overheads by reducing the number of samples for small targets, it enhances both resolution and learning capabilities for these smaller objects within the model.

### 3. Experiments and Analysis

#### 3.1. Datasets

In this paper, we conduct equipment identification and defect detection for various power equipment in substations. We use multiple datasets for experimental verification. Part of the data is collected through self collection, while the other part comes from public datasets such as “China Transmission Line Insulator Dataset (CPLID)”, “Substation Dataset + with Annotation”, “Transformer Equipment Oil Leakage Data Set”, “Safety Supervision Competition Question - Illegal Detection of Anti-high Fall at Typical Work Sites of Power Grid”, and “Manhole Covers, Wire Poles, Electric Boxes, Marking Stones”. In total, there are 15,745 images covering 35 types of equipment images and defect images. These images can be used to train and test models.

In order to simulate the natural interference in the actual operation of power equipment, this paper adopts the data enhancement method of blur, brightness change, and atomization with a certain probability to expand the data of some images. For some categories with fewer defective images, this paper also uses normal images for compositing to construct defective device images. Finally, a total of 13,200 substation equipment defect datasets are constructed and divided into training set, test set, and verification set according to 8:1:1.

#### 3.2. Experimental Configuration and Model Training

This paper uses Pytorch as a deep learning framework to implement improved YOLOv5 in Pytorch 1.8. In order to train the model in this paper, the NVIDIA RTX3060 GPU is used for training and testing, and the corresponding Cuda and Cudnn environments are installed at the same time to support GPU training and improve training efficiency and inference speed.

This paper uses the approach of transfer learning. In the training phase, a partially pre-trained model from YOLOv5 is used. By using these weights, we can save a lot of training time. Then, based on this experiment, analysis is conducted and some parameters are fine-tuned. The total number of iterations is 100 times. Some experimental parameter values are shown in Table 1.

**Table 1.** Experimental parameter settings.

Name	Parameter
learning rate	0.001
batch size	16
decay	0.0005
momentum	0.9

#### 3.3. Experimental Indicators

In this experiment, three indexes were mainly used for accuracy: precision (P), recall (R) and mean average precision (MAP).

Where the precision rate P represents the ratio of the correctly predicted number of positive samples to the actual total number of positive samples, and its calculation formula is as follows:

$$Precision = \frac{TP}{TP + FP} \quad (10)$$

The recall rate R represents the ratio of the correctly predicted number of positive samples to the total actual number of positive samples, which is calculated as follows:

$$Recall = \frac{TP}{TP + FN} \quad (11)$$

MAP is determined by plotting a P–R curve with precision P as the vertical axis and recall R as the horizontal axis, and calculating the area under the curve (AP). The specific calculation method is as follows: for each query of the target category, calculate the mean accuracy AP in order from high to low according to the sorting results of the model, and then average the AP values of all target categories to obtain MAP. The calculation formula is as follows:

$$AP = \int_0^1 PdR \quad (12)$$

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

where TP represents the number of classes that predict positive classes as correct, FP represents the number of classes that predict negative classes as correct, FN represents the number of correct classes predicted as negative classes, and N refers to the total number of detected target classes.

The speed of model detection is evaluated using the FPS value, which refers to the number of frames transmitted per second in a video. The higher the frame rate, the smoother the motion and consequently, the smoother the model detection operation [41,42].

### 3.4. Ablation Experiments

The performance of YOLOv5 is influenced by various factors in actual application processes, including network architecture, training strategy, and data augmentation. To investigate the impact of these factors on YOLOv5's performance, this paper conducts a series of ablation experiments to systematically analyze different components of the algorithm. The ablation experiments gradually modify the network architecture and training strategy using the control variable method to evaluate their influence on detection performance. Experiment 1: Assess the effectiveness of the proposed variable convolutional CSP module on the original detection algorithm. Experiment 2: Evaluate the effectiveness of replacing PANet with the BiFPN structure in the original detection algorithm. Experiment 3: Examine how adding a detection layer affects performance improvement in the original detection algorithm. After completing these experiments, it is necessary to analyze and explain how different components affect inspection performance. Throughout this experiment's training and testing process, consistent configurations are used for accurate comparison as shown in Table 2.

**Table 2.** Comparison of ablation experiments.

	CSP	BiFPN	Head	Precision (%)	Recall (%)	mAP@0.5 (%)
YOLOv5	-	-	-	70.5%	65.2%	68.5%
	√	-	-	74.4%	69.9%	70.1%
	-	√	-	75.3%	70.8%	71.5%
pro-YOLOv5	-	-	√	74.1%	68.2%	70.6%
	√	√	√	76.8%	72.9%	73.1%

In Table 2, row 1 shows the performance of the benchmark YOLOv5s detection algorithm on the dataset in this paper, which shows that the average accuracy is 70.5%, the recall rate is 65.2%, and the mAP@0.5 is 68.5%.

Experiment 1: Verify the impact of introducing variable convolutional CSP modules on benchmark detection in backbone networks. This experiment uses two different networks: one using standard convolutional layers and one using variable convolutional layers. In the second row of Table 2, the experimental results of the introduction of the variable convolution module are shown, and it is found that the introduction of the variable convolution CSP module improves the extraction ability of the network to target features and the performance improvement is significantly compared with the original algorithm, with the mAP@0.5 reaching 70.1%, an increase of 1.6%. Therefore, the variable convolutional CSP module proposed in this paper has better performance than the original detection algorithm, and the performance can be further improved by adding more variable convolutional layers to the backbone network. However, when there are too many variable convolutional layers, performance begins to degrade. Through the analysis, it was also found that the use of variable convolution can increase the depth and width of the network with a small amount of computation. In addition, the performance differences between two networks using variable convolution and standard convolution are compared, and the results show that under the same conditions, networks using variable convolution are superior to networks using standard convolution.

Experiment 2: Verify the influence of replacing PANet with BiFPN structure on improving model performance. The third row in Table 2 shows the experimental results of replacing the original PANet with the BiFPN structure, which improves the model recall to 70.8%, the accuracy to 75.3%, and the mAP@0.5 also improves to a small extent. Experimental results show that the use of BiFPN in YOLOv5 can significantly improve the performance of small target detection. It is found that when using a large number of small target datasets, the network using BiFPN can improve detection performance. In addition, the performance differences between the two networks using PANet and using BiFPN are compared, and the results show that the network using BiFPN is better than the network using PANet under the same conditions.

Experiment 3: Verify the effect on model performance caused by an additional detection head to predict the effect of the position and size of smaller targets. Row 4 in Table 2 shows the performance of adding a detection head before pooling in small target detection, and compared to the original algorithm, the recall rate is improved by 3%, the accuracy is nearly 3.6%, and the mAP is increased to 70.6%. Further analysis found that due to the addition of an additional detection head, the model was able to more accurately predict the position and size of smaller targets, resulting in improved accuracy and recall. Secondly, this paper also improves the recognition ability of the model for small targets by adjusting the training data to make the model pay more attention to the recognition of small targets.

Experimental conclusion: In the test set, the performance of the original YOLOv5 model and the improved model in small target detection was compared, including average accuracy, average recall, and mAP@0.5. The comparison results show that compared with the original algorithm, the improved model performs better in the same dataset and the same environment, with a recall increase of 7.7%, an accuracy improvement of nearly 6.3%, and an average accuracy improvement of 4.6%. The improved model adopts the variable convolutional CSP module and uses a simple and effective BiFPN structure to replace PANet, which has a good improvement on the model performance. At the same time, an additional detection head is added to predict the position and size of smaller targets, which can improve the problem of missing detection of small targets.

#### 4. Results and Analysis

Figure 8 shows an example of the experimental results. The four sets of images in the figure demonstrate the outcomes obtained by utilizing YOLOv5 and enhanced YOLOv5 for detecting substation equipment and its defects in different power equipment environments.

These examples evaluate the effectiveness of the proposed algorithm in enhancing object detection performance specifically for substation equipment scenarios.

Comparing the results in Figure 8, it is evident that in group A images, both algorithms identify the oil pillow. However, due to pixel clarity issues, only the improved algorithm in this paper successfully detects the meter on the oil pillow, while the original YOLOv5 algorithm fails to detect it. In group B images, the algorithm in this paper detects defective insulators with inconspicuous features. In group C images, despite background complexity and confusion between the nest and background, the original YOLOv5 algorithm detects the insulator but not the nest, whereas the improved YOLOv5 algorithm successfully detects both. In group D images, silica gel discoloration defects were missed by the original YOLOv5 algorithm but accurately detected by our improved algorithm. Additionally, for all detected targets, our improved algorithm achieved an average confidence of 0.797, an improvement when compared to 0.748 for the original YOLOv5 algorithm. Therefore, our proposed improved YOLOv5 algorithm outperforms its predecessor in detecting small targets against complex backgrounds by reducing missed detections and false positives while maintaining high detection confidence levels.

The most advanced detectors YOLOv8, YOLOv5, Faster R-CNN, RetinaNet, and YOLOv3 was selected as a control algorithm for experimental analysis. Among them, YOLOv8 was chosen as the latest model in the YOLO series. The improved YOLOv5 model was compared with it to verify the advancement of this model and demonstrate the performance improvement effect compared to the original YOLOv5 model. The evaluation criteria selected include accuracy, recall rate, mAP@0.5, and mAP@0.5:0.95 as mainstream indicators. mAP@0.5 refers to the mAP at IOU set at a threshold of 0.5; mAP@0.5:0.95 is obtained by averaging different threshold mAPs between 0.5 and 0.95 with a step size of 0.05. Table 3 shows the performance of the improved algorithm and mainstream algorithms on our dataset in this paper.

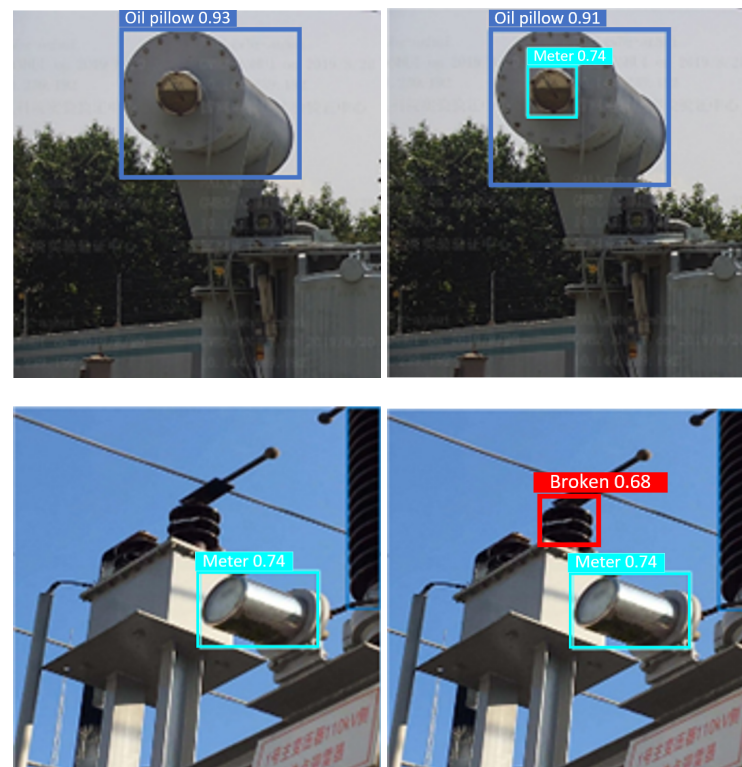
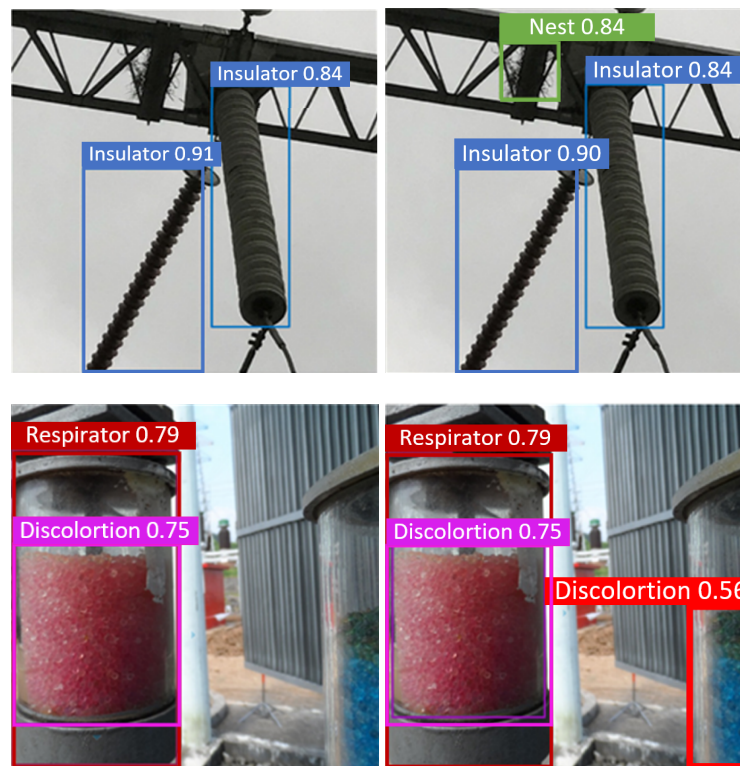


Figure 8. Cont.



**Figure 8.** Comparison of test results.

**Table 3.** Performance of improved algorithms and mainstream algorithms.

Algorithm	Precision (%)	Recall (%)	mAP@0.5 (%)	mAP@0.5:0.95 (%)
Faster R-CNN	72.16	62.25	61.85	39.23
RetinaNet	68.43	60.26	61.85	39.23
YOLOv3	61.63	59.43	61.85	39.23
YOLOv4	66.28	60.45	58.56	37.12
YOLOv5	70.5	65.2	68.5	39.85
YOLOv8	77.5	66.1	72.3	40.21
pro-YOLOv5	76.8	72.9	73.1	42.97

As can be seen from Table 3, the proposed improved algorithm achieves 76.8% accuracy, 72.9% recall, 73.1% mAP@0.5 and 42.97% mAP@0.5:0.95, which is better than other mainstream algorithms. Compared to the original YOLOv5 algorithm, the proposed algorithm contributes improvements of 4.6% for mAP@0.5 and 3.12% for mAP@.5:0.95 and is also better than the comparison algorithm in terms of detection accuracy and recall. This shows that the proposed algorithm can effectively improve the object detection performance on substation equipment defect dataset.

Using the bootstrap method, the average detection accuracy and average FPS of the original YOLOv4, YOLOv5 models, YOLOv8 model, and improved YOLOv5 model were evaluated to obtain Table 4.

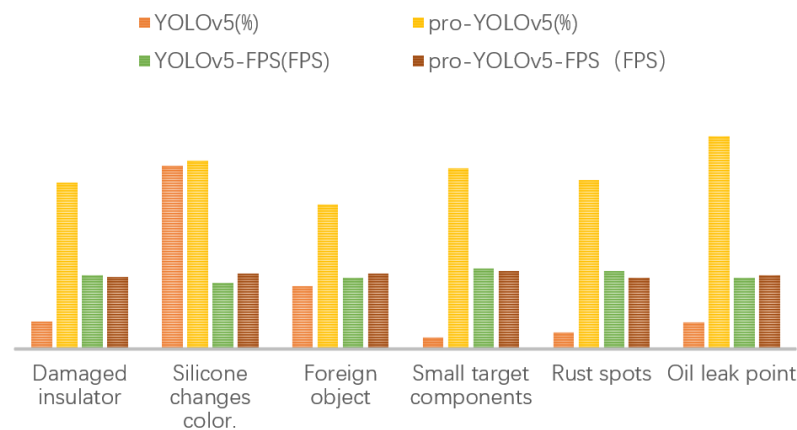
**Table 4.** Bootstrap valuation evaluation model.

Algorithm	Average Precision (%)	Average FPS (%)
YOLOv4	65.4	53
YOLOv5	69.8	49
YOLOv8	76.5	46
pro-YOLOv5	76.9	48

It can be seen that, compared to other contrast algorithm models, the improved YOLOv5 model ensures both detection accuracy and FPS value.

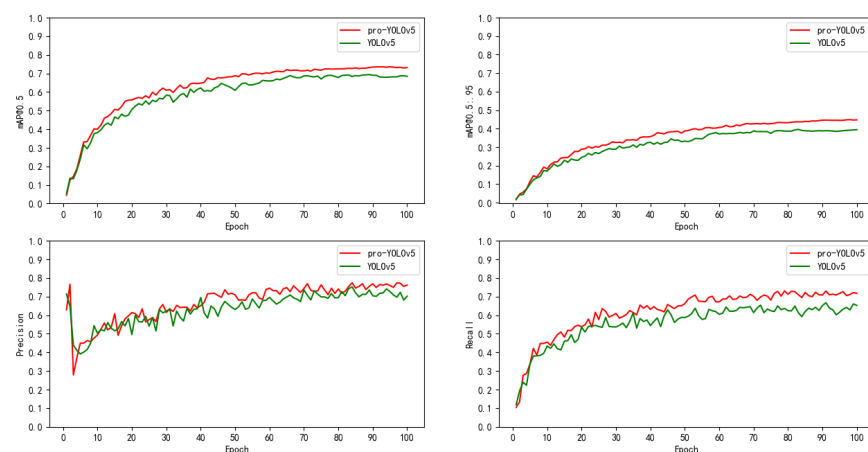
Given that the improved YOLOv5 algorithm performs better than the original YOLOv5 algorithm in detecting small targets but considering the complex and diverse application scenarios, we analyzed the performance of detecting different types of defects in different application scenarios to verify the advantages and disadvantages of the improved YOLOv5 algorithm compared to the original algorithm.

According to the data in Figure 9, it can be concluded that the improved YOLOv5 algorithm performs better than the original YOLOv5 algorithm in detecting small targets such as small components, rust spots, and oil leakage points. Additionally, it maintains a similar FPS value to the original algorithm, ensuring detection efficiency.



**Figure 9.** Performance comparison of improved YOLOv5 and original YOLOv5 in different defect scenarios.

Figure 10 shows a comparison of the mAP@0.5, mAP@0.5:0.95, precision, and recall of the proposed improved YOLOv5 algorithm and YOLOv5. It can be seen from the figure that all four models converge around 100 epochs, and when compared with YOLOv5, the improved algorithm not only has a high detection accuracy but also has a high recall rate, which verifies the fast convergence and high recognition rate of the proposed improved algorithm.



**Figure 10.** Comparison of the original YOLOv5 with the improved algorithm.

## 5. Conclusions

This paper presents an enhanced YOLOv5 algorithm for detecting faults in substation equipment. It specifically addresses the challenge of accurately recognizing small targets in complex substation scenes. The improved algorithm incorporates deformable convolution

modules to dynamically adjust the scale and receptive field size of the target, thereby meeting the requirements of multi-scene object detection. Additionally, it utilizes a simple and effective BiFPN structure in the network's neck section to achieve dynamic weighted multi-level feature fusion, effectively addressing feature fusion challenges at various levels. Furthermore, a small target detection layer is introduced to handle shallow feature maps. Experimental results demonstrate that this improved algorithm significantly enhances detection performance for small targets and complex scenes encountered in substation scenarios. Future research directions may involve expanding datasets, including samples with minor defects, and enhancing the model's ability to recognize such defects. Moreover, further optimization can be pursued to improve detection speed and facilitate its application on mobile devices.

**Author Contributions:** Conceptualization, Y.W.; Methodology, Y.W. and F.X.; Validation, F.X. and Y.S.; Resources, X.D. and C.Z.; Data curation, F.X. and F.L.; Writing—original draft, F.X. and F.L.; Writing—review & editing, Y.W., X.D., L.L. and C.Z.; Visualization, F.X. and Y.S.; Supervision, Y.W.; Funding acquisition, Y.W., F.X. and L.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported in part by the National Key R&D Program of China under Grant no. 2022YFE010300, in part by the Natural Science Foundation of Hunan Province under Grant no. 2021JJ50050, in part by the Scientific Research Fund of Hunan Provincial Education Department under Grant no. 22A0422 and in part by the Hunan Provincial Innovation Foundation For Postgraduate under Grant no. CX20220835.

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

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Cong, S.; Pu, H.; Yao, F. Review on Application of Infrared Detection Technology in State Detection of Electrical Equipment. In Proceedings of the 16th Annual Conference of China Electrotechnical Society, Beijing, China, 24–26 September 2021; Springer: Singapore, 2022; Volume II, pp. 1254–1261.
2. Dongxu, L.; Li, W.; Jinxuan, L.; Shaoyu, X.; Long, W.; Wenwang, X. Design and Research of Intelligent Operation Inspection and Monitoring System of Substation Based on Image Recognition Technology. In Proceedings of the 2022 2nd Asia-Pacific Conference on Communications Technology and Computer Science (ACCTCS), Shenyang, China, 25–27 February 2022; pp. 448–453.
3. Bashkari, M.S.; Sami, A.; Rastegar, M. Outage cause detection in power distribution systems based on data mining. *IEEE Trans. Ind. Inform.* **2020**, *17*, 640–649. [[CrossRef](#)]
4. Jia, Y.; Ying, L.; Wang, D.; Zhang, J. Defect prediction of relay protection systems based on LSSVM-BNNDT. *IEEE Trans. Ind. Inform.* **2020**, *17*, 710–719. [[CrossRef](#)]
5. Bin, C. Research on Intelligent Reconfiguration and Recognition Technology of Mobile Environment for Substation Operation. In *Advancements in Mechatronics and Intelligent Robotics*; Springer: Singapore, 2021; pp. 125–131.
6. Le, N.; Rathour, V.S.; Yamazaki, K.; Luu, K.; Savvides, M. Deep reinforcement learning in computer vision: A comprehensive survey. *Artif. Intell. Rev.* **2022**, *55*, 2733–2819. [[CrossRef](#)]
7. Tian, H.; Wang, T.; Liu, Y.; Qiao, X.; Li, Y. Computer vision technology in agricultural automation—A review. *Inf. Process. Agric.* **2020**, *7*, 1–19. [[CrossRef](#)]
8. Li, Y.; Zhang, Y. Application research of computer vision technology in automation. In Proceedings of the 2020 International Conference on Computer Information and Big Data Applications (CIBDA), Guiyang, China, 17–19 April 2020; pp. 374–377.
9. Chandra, M.A.; Bedi, S. Survey on SVM and their application in image classification. *Int. J. Inf. Technol.* **2021**, *13*, 1–11. [[CrossRef](#)]
10. Rao, Y.; Zhao, W.; Zhu, Z.; Lu, J.; Zhou, J. Global filter networks for image classification. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 980–993.
11. Hernández, G.; Rodríguez, S.; González, A.; Corchado, J.M.; Prieto, J. Video analysis system using deep learning algorithms. In *Ambient Intelligence—Software and Applications, Proceedings of the 11th International Symposium on Ambient Intelligence, Cham, Switzerland, L'Aquila, Italy, 17–19 June 2020*; Springer: Cham, Switzerland, 2021; pp. 186–199.
12. Wang, T.; Zhang, Z.; Tsui, K.L. A deep generative approach for rail foreign object detections via semisupervised learning. *IEEE Trans. Ind. Inform.* **2022**, *19*, 459–468. [[CrossRef](#)]

13. Ye, C.; Zhang, N.; Peng, L.; Tao, Y. Flexible array probe with in-plane differential multichannels for inspection of microdefects on curved surface. *IEEE Trans. Ind. Electron.* **2021**, *69*, 900–910. [[CrossRef](#)]
14. Chu, Y.; Feng, D.; Liu, Z.; Zhao, Z.; Wang, Z.; Xia, X.G.; Quek, T.Q. Hybrid-Learning-Based Operational Visual Quality Inspection for Edge-Computing-Enabled IoT System. *IEEE Internet Things J.* **2021**, *9*, 4958–4972. [[CrossRef](#)]
15. Shang, H.; Wu, J.; Sun, C.; Liu, J.; Chen, X.; Yan, R. Global prior transformer network in intelligent borescope inspection for surface damage detection of aero-engine blade. *IEEE Trans. Ind. Inform.* **2022**, *19*, 8865–8877. [[CrossRef](#)]
16. Minh, Q.N.; Nguyen, V.H.; Quy, V.K.; Ngoc, L.A.; Chehri, A.; Jeon, G. Edge Computing for IoT-Enabled Smart Grid: The Future of Energy. *Energies* **2022**, *15*, 6140. [[CrossRef](#)]
17. Song, C.; Xu, W.; Han, G.; Zeng, P.; Wang, Z.; Yu, S. A cloud edge collaborative intelligence method of insulator string defect detection for power IoT. *IEEE Internet Things J.* **2020**, *8*, 7510–7520. [[CrossRef](#)]
18. Zhou, X.; Wang, Y.; Zhu, Q.; Mao, J.; Xiao, C.; Lu, X.; Zhang, H. A surface defect detection framework for glass bottle bottom using visual attention model and wavelet transform. *IEEE Trans. Ind. Inform.* **2019**, *16*, 2189–2201. [[CrossRef](#)]
19. Ni, X.; Liu, H.; Ma, Z.; Wang, C.; Liu, J. Detection for rail surface defects via partitioned edge feature. *IEEE Trans. Intell. Transp. Syst.* **2021**, *23*, 5806–5822. [[CrossRef](#)]
20. Wen, L.; Wang, Y.; Li, X. A new cycle-consistent adversarial networks with attention mechanism for surface defect classification with small samples. *IEEE Trans. Ind. Inform.* **2022**, *18*, 8988–8998. [[CrossRef](#)]
21. Kalchbrenner, N.; Grefenstette, E.; Blunsom, P. A convolutional neural network for modelling sentences. *arXiv* **2014**, arXiv:1404.2188.
22. Xu, Y.; Fu, M.; Wang, Q.; Wang, Y.; Chen, K.; Xia, G.S.; Bai, X. Gliding vertex on the horizontal bounding box for multi-oriented object detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *43*, 1452–1459. [[CrossRef](#)]
23. Zheng, Z.; Wang, P.; Ren, D.; Liu, W.; Ye, R.; Hu, Q.; Zuo, W. Enhancing geometric factors in model learning and inference for object detection and instance segmentation. *IEEE Trans. Cybern.* **2021**, *52*, 8574–8586. [[CrossRef](#)]
24. Tian, Z.; Shen, C.; Chen, H.; He, T. FCOS: A simple and strong anchor-free object detector. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *44*, 1922–1933. [[CrossRef](#)]
25. Lei, X.; Sui, Z. Intelligent fault detection of high voltage line based on the Faster R-CNN. *Measurement* **2019**, *138*, 379–385. [[CrossRef](#)]
26. Jiang, A.; Yan, N.; Wang, F.; Huang, H.; Zhu, H.; Wei, B. Visible image recognition of power transformer equipment based on mask R-CNN. In Proceedings of the 2019 IEEE Sustainable Power and Energy Conference (iSPEC), Beijing, China, 21–23 November 2019; pp. 657–661.
27. Liu, Z.; Fu, H.; Li, Y.; Zhang, G.; Hu, C.; Zhang, G. Electrical Equipment Detection in Infrared Images Based on Transfer Learning of Mask-RCNN. *J. Data Acquis. Process.* **2021**, *36*, 176–183.
28. Yin, Z.; Meng, R.; Fan, X.; Li, B.; Zhao, Z. Typical visual defect detection system of substation equipment based on edge computing and improved Faster R-CNN. *China Sci.* **2021**, *16*, 343–348.
29. Ou, J.; Wang, J.; Xue, J.; Wang, J.; Zhou, X.; She, L.; Fan, Y. Infrared image target detection of substation electrical equipment using an improved faster R-CNN. *IEEE Trans. Power Deliv.* **2022**, *38*, 387–396. [[CrossRef](#)]
30. Zhang, M.; Yin, L. Solar cell surface defect detection based on improved YOLO v5. *IEEE Access* **2022**, *10*, 80804–80815. [[CrossRef](#)]
31. Guo, Z.; Wang, C.; Yang, G.; Huang, Z.; Li, G. Msft-yolo: Improved yolov5 based on transformer for detecting defects of steel surface. *Sensors* **2022**, *22*, 3467. [[CrossRef](#)]
32. Yang, R.; Li, W.; Shang, X.; Zhu, D.; Man, X. KPE-YOLOv5: An Improved Small Target Detection Algorithm Based on YOLOv5. *Electronics* **2023**, *12*, 817. [[CrossRef](#)]
33. Wang, Y.; Zhang, X.; Li, L.; Wang, L.; Zhou, Z.; Zhang, P. An Improved YOLOv7 Model Based on Visual Attention Fusion: Application to the Recognition of Bouncing Locks in Substation Power Cabinets. *Appl. Sci.* **2023**, *13*, 6817. [[CrossRef](#)]
34. Hsu, W.Y.; Lin, W.Y. Ratio-and-scale-aware YOLO for pedestrian detection. *IEEE Trans. Image Process.* **2020**, *30*, 934–947. [[CrossRef](#)] [[PubMed](#)]
35. Xuhong, W.; Hao, L.; Shaosheng, F.; Zhipeng, J. Infrared image anomaly automatic detection method for power equipment based on improved single shot multi box detection. *Trans. China Electrotech. Soc.* **2020**, *35*, 302–310.
36. Cheng, Y. Detection of power line insulator based on enhanced Yolo Model. In Proceedings of the 2022 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC), Dalian, China, 14–16 April 2022; pp. 626–632.
37. Hu, X.; Zhou, Y. Insulator defect detection in power inspection image using focal loss based on YOLO v4. In Proceedings of the International Conference on Artificial Intelligence, Virtual Reality, and Visualization (AIVRV 2021), Sanya, China, 19–21 November 2021; Volume 12153, pp. 90–95.
38. Zheng, H.; Li, J.; Liu, Y.; Cui, Y.; Ping, Y. Infrared Target Detection Model for Power Equipment Based on Improved YOLOv3. *Trans. China Electrotech. Soc.* **2021**, *36*, 1389–1398.
39. Zhu, X.; Lyu, S.; Wang, X.; Zhao, Q. TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 2778–2788.
40. Dai, J.; Qi, H.; Xiong, Y.; Li, Y.; Zhang, G.; Hu, H.; Wei, Y. Deformable convolutional networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–27 October 2017; pp. 764–773.



41. Heda, L.; Sahare, P. Performance Evaluation of YOLOv3, YOLOv4 and YOLOv5 for Real-Time Human Detection. In Proceedings of the 2023 2nd International Conference on Paradigm Shifts in Communications Embedded Systems, Machine Learning and Signal Processing (PCEMS), Nagpur, India, 5–6 April 2023; pp. 1–6.
42. Saxena, A. Optimized fractional overhead power term polynomial grey model (OFOPGM) for market clearing price prediction. *Electr. Power Syst. Res.* **2023**, *214*, 108800. [[CrossRef](#)]

**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.