

Article

Anomaly Detection Algorithm for Photovoltaic Cells Based on Lightweight Multi-Channel Spatial Attention Mechanism

Aidong Chen ^{1,2,3}, Xiang Li ^{1,2}, Hongyuan Jing ^{2,3}, Chen Hong ^{2,3} and Minghai Li ^{2,3,*}¹ Beijing Key Laboratory of Information Service Engineering, Beijing 100101, China² College of Robotics, Beijing Union University, Beijing 100101, China³ Research Centre for Multi-Intelligent Systems, Beijing 100101, China

* Correspondence: jdtminghai@buu.edu.cn

Abstract: With the proposed goal of “Carbon Neutrality”, photovoltaic energy is gradually gaining the leading role in energy transformation. At present, crystalline silicon cells are still the mainstream technology in the photovoltaic industry, but due to the similarity of defect characteristics and the small scale of the defects, automatic defect detection of photovoltaic cells (PV) by electroluminescence (EL) imaging is a challenging task. In order to better meet the growing demand for high-quality photovoltaic cell products in intelligent manufacturing and use, and ensure the safe and efficient operation of photovoltaic power stations, this paper proposes an improved abnormal detection method based on Faster R-CNN for the surface defect EL imaging of photovoltaic cells, which integrates a lightweight channel and spatial convolution attention module. It can analyze the crack defects in complex scenes more efficiently. The clustering algorithm was used to obtain a more targeted anchor frame for photovoltaic cells, which made the model converge faster and enhanced the detection ability. The normalized distance between the prediction box and the target box is minimized by considering the DIoU loss function for the overlapping area of the boundary box and the distance between the center points. The experiment shows that the average accuracy of surface defect detection for EL images of photovoltaic cells is improved by 14.87% compared with the original algorithm, which significantly improves the accuracy of defect detection. The model can better detect small target defects, meet the requirements of surface defect detection of photovoltaic cells, and proves that it has good application prospects in the field of photovoltaic cell defect detection.

Keywords: photovoltaic cell; electroluminescence; defect detection; image recognition



Citation: Chen, A.; Li, X.; Jing, H.; Hong, C.; Li, M. Anomaly Detection Algorithm for Photovoltaic Cells Based on Lightweight Multi-Channel Spatial Attention Mechanism.

Energies **2023**, *16*, 1619. <https://doi.org/10.3390/en16041619>

Academic Editors: Mauro D’Arco and Francesco Bonaventura

Received: 4 January 2023

Revised: 26 January 2023

Accepted: 3 February 2023

Published: 6 February 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

With the increasing contradiction between economic development and natural resources, green development has become an important trend in global development and is becoming more and more deeply integrated with all areas of human society, economy and politics. With the global trend towards decarbonization, the photovoltaic industry is developing rapidly and is becoming a key force in driving the global energy transition. The quality of the photovoltaic cells has a direct impact on the power generation efficiency and operation of the plant. Crystalline silicon is the main photovoltaic material due to the nature of the crystalline silicon structure; photovoltaic cells are prone to defects such as cracks, scratches and fingers during the production process. Moreover, as the temperature rises rapidly, the photovoltaic panels will also be subjected to extreme high temperatures. Moreover, continuous exposure to the sun and work will bring irreversible damage to the PV panels, leading to a reduction in the output of the PV modules, seriously affecting the power output of the PV plant and may even lead to a fire. PV operation and maintenance are important to guarantee improvements in the efficiency of the plant, reduce the cost of electricity and operate safely.

PV cell defect detection aims to predict the class and location of multiscale defects in EL near-infrared images. As shown in Figure 1, the three most frequently occurring

types of PV cell damage are cracks, fingers and black cores with complex background interference. The areas of the images that show high brightness are areas of crystalline silicon with high conversion efficiency, while the areas that show darker areas are defective areas with inactive and non-luminous light. In addition to the defects, the images are difficult to distinguish because of the color, and the dislocations and the four busbars also appear as dark areas, which may sometimes overlap with the defects, making automatic defect detection in the EL images of PV cells more difficult.

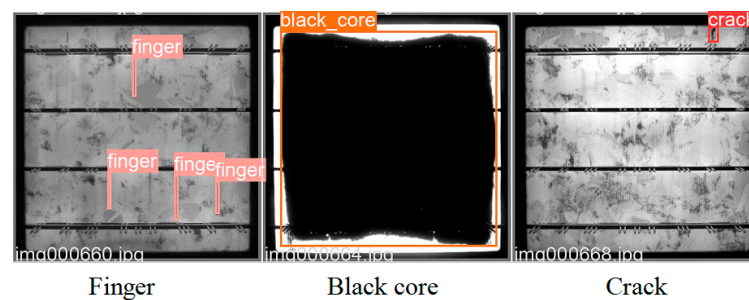


Figure 1. The three most common types of PV cell damage.

Traditional surface defect detection methods have played a huge role for some time. Surface defect features are divided into three main categories: the first, texture feature-based methods, which reflect the organizational structure and alignment characteristics of the image surface through the grey-scale distribution of pixels and their nearby spatial neighbors; the second, color feature-based methods, where color features are less computationally intensive, less dependent on factors such as the size, orientation and viewing angle of the image itself, and highly robust. Specific methods include color histograms [1], color coherence vectors [2], etc.; the third category, methods based on shape features, effectively use the targets of interest in the image for retrieval. However, the traditional methods have poor generalization and robustness. Wen et al. [3] proposed the use of electronic speckle interference (ESPI); that is, laser speckle is used as the carrier of the field change information of the measured object, and the phase change between the front and back of the double beam wave is detected by using the correlation fringes of the speckle field generated by the measured object after laser irradiation to identify cracks and defects in photovoltaic cells. By analyzing the continuity of the speckle pattern, the existence of crack defects is determined. Dhimish et al. [4] used discrete Fourier transform to conduct two-dimensional spectral analysis of binary images of EL images of solar cells, and adjusted the required Fourier transform component in the intermediate frequency domain by analyzing the geometric characteristics of the binary images, so as to improve the detection ability of cracks in solar cells.

In recent years, convolutional neural networks (CNN), a very powerful kind of deep neural network, have been a great success in image recognition and classification, natural language processing (NLP), etc., and are also widely used in industry [5–8]. The advantages are more obvious when the input to the network is an image, making it possible to use the image directly as the input to the network, avoiding the complex process of feature extraction and data reconstruction in traditional recognition algorithms, and having great advantages in the processing of 2D images, such as the network being able to extract the features of the image including color, texture, shape and the topology of the image by itself, in the problem of processing 2D images, especially recognition. The network has good robustness and computational efficiency in processing two-dimensional images, especially in the application of displacement, scaling and other forms of distortion invariance. Akram et al. [9] proposed a method of EL image defect recognition based on a lightweight convolutional neural network structure, which greatly improved the recognition accuracy on solar cell EL image data sets. HUSSAIN et al. [10] observed the similarity between the original EL images and the filter output images obtained via gradient-guided filter tuning,

introduced a mechanism for generating PV cell images based on EL modelling, termed filter fused data scaling. As artificial intelligence and image recognition technologies continue to evolve, deep learning-based object detection techniques are being used in a wide range of industries. Object detection is the identification of objects of interest in an image, determining their class and location, and is one of the core problems in the field of computer vision, with applications in video surveillance, virtual reality, human–computer interaction and other fields. The target detection algorithm R-CNN [11] is a popular detection method in the field of deep learning in recent years, which generates candidate regions of images that may contain targets and then classifies the candidate regions using CNN with high accuracy, but performs poorly in terms of speed. Fast R-CNN [12] solves the problem of a large number of repeated calculations when R-CNN extracts features from all regions, and introduces ROI Pooling. However, Fast R-CNN uses selective search to select candidate regions, which involves a large number of calculations. The Faster R-CNN [13] algorithm has currently the highest detection accuracy and fastest target detection in the R-CNN family algorithm. A region proposal network (RPN) is proposed, in which an image is input and a set of rectangular object suggestion boxes are output. Each box has a typical full convolutional network with objectiveness score. Compared with the selective search method in R-CNN, the generation time of RPN is reduced by 200 times.

The main contributions of this paper include three aspects:

1. Integrating the convolutional block attention module (CBAM) into Faster R-CNN to modify the feature extraction part to assign greater weights to the features of photovoltaic cell defects, so that the network can better distinguish the target and background of crack defects in the image;
2. The K-means clustering algorithm was used to train targeted anchors to cluster the width and height dimensions of the anchors for the three labeled defect boxes to be detected in the photovoltaic cell surface defect dataset, which made it easier for the detection network to learn accurate defect detection anchors, to improve detection accuracy;
3. The traditional $smooth_{L1}$ loss function was replaced by the calculation method of the DIoU loss function, and the normalized distance between the candidate frame and the target frame was directly minimized to achieve a faster convergence speed, so that the regression could overlap with the target frame for even more accuracy and speed when included.

The remainder of this paper is organized as follows. Section 2 reviews and introduces the Faster R-CNN, the convolutional block attention mechanism, the clustering algorithm K-means, and the DIoU loss function. Section 3 describes in detail the design of adding the above three methods to the original Faster R-CNN model. Section 4 conducts experiments to validate the methods in this paper. Finally, Section 5 concludes the paper.

2. Related Work

2.1. Introduction to Faster R-CNN

The task of object detection is to find objects of interest in an image or video and simultaneously determine their location and size. Target detection not only solves the classification problem, but also solves the positioning problem, which is a multitask problem. The target detection algorithm based on deep learning has been developed in two technical ways: two stage and one stage. The two-stage target detector based on the candidate region first focuses on finding the location of the target object, and obtains the proposed frame to ensure sufficient accuracy and recall. It then focuses on classifying the proposed frames to find more precise locations. Common two-stage target detection algorithms include R-CNN, Fast R-CNN, Faster R-CNN and Mask R-CNN [14], which are characterized by high accuracy but lower speed. One stage directly predicts the coordinates of the prediction box, the confidence level of the object contained in the prediction box, and the probability of the object category from the whole image. Common one-stage target detection algorithms include SSD [15] and the YOLO series [16–19], which are characterized by high speed but

lower accuracy. Since battery anomaly detection pays more attention to detection accuracy, this paper improves it based on Faster R-CNN.

Faster R-CNN consists of three parts: feature extraction network, candidate region generation network, classification and location adjustment. The basic structure of Faster R-CNN is shown in Figure 2. First, the feature extraction network. Faster C-NN, as a CNN network target detection method, first uses a convolutional neural network (such as ResNet [20], etc.) to extract image features. This feature map is shared for subsequent RPN layers and full connection layers. Second, the regional suggestion network. The RPN network is used to generate candidate regions. This layer determines whether the anchor belongs to the foreground or background, and then uses the bounding box regression to correct the anchor to obtain the accurate prediction box. Third, region of interest pooling (ROI pooling). ROI pooling is used to convert features of the region of interest of different sizes into the same feature map output to ensure the same size of feature map after flattening. Fourth, classification. The feature map is used to calculate the categories of the prediction box, and the final accurate position of the prediction box is obtained by using the bounding box regression again.

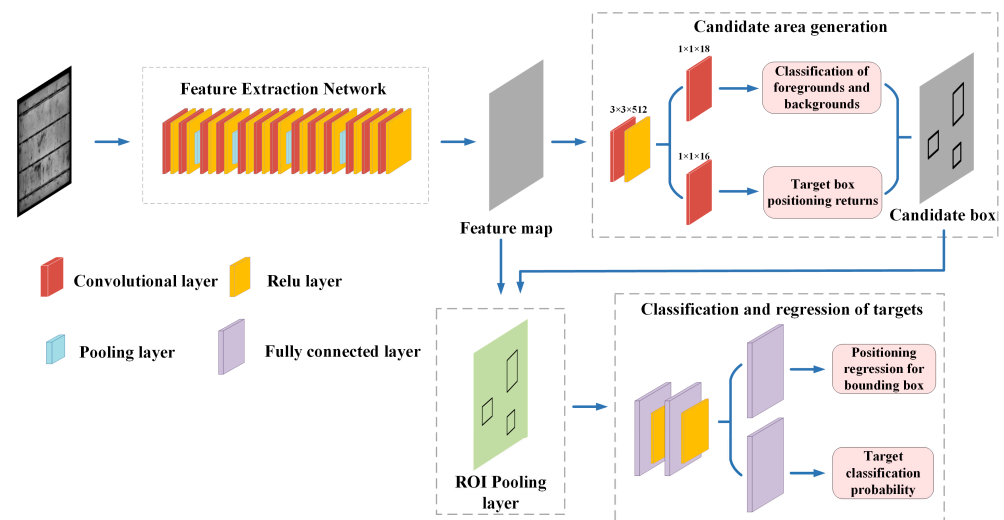


Figure 2. Region-based convolutional neural network (Faster R-CNN) architecture.

2.2. Convolutional Block Attention Mechanism

In addition to these factors, researchers have also looked at another aspect of network design—attention. When using a convolutional neural network to process images, researchers would prefer the convolutional neural network to pay attention to what should be paid attention to, rather than to all information. In reality, it is impossible to manually adjust the places that need attention. At this time, how to make the convolutional neural network adaptively pay attention to important objects becomes extremely important. The attention mechanism is a way to realize the adaptive attention of the network. The convolutional block attention module (CBAM) is a lightweight attention module proposed by Woo et al. [21] which combines channel and spatial attention mechanism modules, as shown in Figure 3. CBAM includes two submodules, the channel attention module (CAM) and the spatial attention module (SAM), which perform channel and spatial attention, respectively. The CBAM module includes four parts: input, CAM, SAM and output. The first is the input feature $F \in R^{C \times H \times W}$, then the CAM one-dimensional convolution $M_C \in R^{C \times 1 \times 1}$ is performed, the convolution result is multiplied by the original image, and the CAM output result is used as the input of the SAM, and then the two-dimensional convolution of SAM, $M_S \in R^{1 \times H \times W}$, is performed, and the output is multiplied by the original map, and then the output result is multiplied by the original image in a process such as Equations (1) and (2). Li et al. [22] optimized the model on the basis of YOLOv5, which

included the use of CBAM to add attention mechanisms to the network layer. The model has a general feasibility value for target recognition of global storage tanks.

$$F' = M_C(F) \otimes F \quad (1)$$

$$F'' = M_S(F') \otimes F' \quad (2)$$

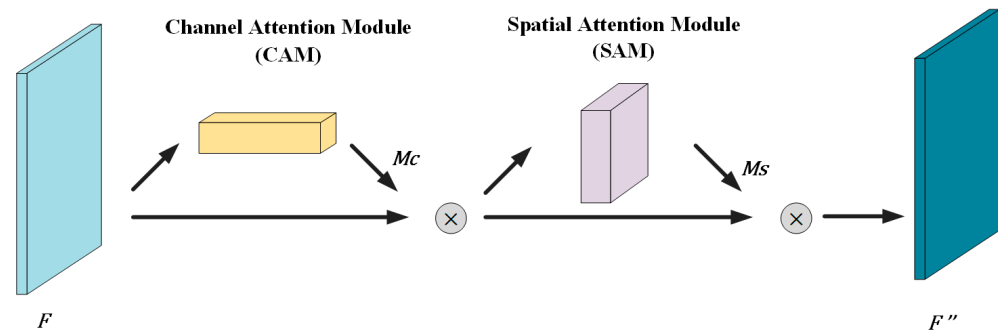


Figure 3. CBAM block architecture.

In these, F represents the input of the feature map ($C \times H \times W$), M_C is the one-dimensional ($C \times 1 \times 1$) channel attention map, M_S is the two-dimensional $1 \times H \times W$ channel attention map, \otimes represents the multiplication operation, F' represents the intermediate output ($C \times H \times W$), and F'' represents the final output ($C \times H \times W$).

2.3. Clustering Algorithm K-Means

The purpose of the K-means algorithm is to divide the set into several categories, so that objects belonging to different classes are as different as possible, while objects belonging to the same class are as similar as possible. The principle of this algorithm is simple and easy to implement; the clustering effect is better, the convergence speed is fast and there is no need for too much human intervention, and the interpretation is strong. The Euclidean distance is usually used as an index to measure the similarity between samples.

k in K-means means that the clusters are clustered into k clusters, and *means* signifies that the mean of the data in each cluster is taken as the center of the cluster, or called the centroid; that is, the centroid of each class is used to describe the cluster. The clustering process of the K-means clustering algorithm can be regarded as the process of continuously finding the centroids of the clusters. This process starts from randomly setting k centroids until k real centroids are found. Li et al. [23] used k-means on the basis of the Faster R-CNN model to improve the cluster analysis of the eel head detection frames annotated in the training set to achieve accurate counting of circulating water farmed eels.

2.4. Loss Function

In existing object detection methods, researchers usually use the $L1$ – norm loss function for bounding box regression. However, the intersection over union (IoU) method is adopted when evaluating the detection performance, so it is more reasonable and effective to use IoU as the loss function of regression. The IoU loss function [24] and the Generalized-IoU (GIoU) loss function [25] have problems of slow convergence and inaccurate regression. The IoU loss is the difference of intersection ratio between prediction box and the target box, expressed as Equation (3). The GIoU loss function adds a penalty term on the basis of IoU loss and is expressed as Equation (4). Distance-IoU (DIoU) loss function is found [26] by combining the normalized distance between the prediction box and the target box; the DIoU loss can provide the moving direction for the prediction box without overlapping with the target box, and the IoU loss in the case of non-overlapping target boxes, the DIoU loss has a large error in both horizontal and vertical cases, DIoU loss has very small

regression error in all cases, and DIOU loss converges faster in training than IoU and GIoU which lose much more.

$$L_{IoU} = 1 - \frac{|B \cap B^{groudtruth}|}{|B \cup B^{groudtruth}|} \tag{3}$$

$$L_{GIoU} = 1 - IoU + \frac{|C - B \cup B^{groudtruth}|}{|C|} \tag{4}$$

For the photovoltaic cell defect dataset in this paper, a DIOU loss function was introduced to regularize the distance between the prediction frame and the target frame, as shown in Equation (11). Liu et al. [27] proposed an improved EfficientDet-based military gesture detection algorithm for the military gesture-detection algorithm, and used DIOU-NMS to remove redundant prediction frames to obtain the final prediction result. Experiments showed that the proposed algorithm had higher prediction accuracy. Li et al. [28] proposed an improved real-time target detection algorithm suitable for insulator dropout fault detection. Based on the YOLOv5s detection network, the DIOU loss function was used to optimize the loss function, and the improved algorithm improved the average accuracy.

3. Research Method

In this section, the feature extraction network incorporating CBAM is first introduced, making the detection network more capable of detecting. Then, the anchor frame parameters obtained by K-means clustering are introduced, making the generated frames more in line with the PV cell defect class proportions. Finally, the calculation method of the DIOU loss function is introduced to make the regression more accurate and faster when it overlaps with the target frame or even contains it.

3.1. Introduction of Feature Extraction Network with CBAM Structure

In order to overcome the problem of unreasonable weight distribution when the original Faster-R-CNN extracts the crack defect features of photovoltaic cells, this paper proposes to introduce the CBAM attention mechanism to optimize the feature extraction network to suppress the features of complex background and grain pseudo-defects. The network structure is shown in Figure 4. In Figure 4, the feature extraction network uses Resnet50 as the basic network, and the photovoltaic cell crack defect image $I \in R^{H \times W}$ is reconstructed and then input into the feature extraction network for feature extraction. After a 3×3 volume, the feature map $F \in R^{H \times W \times C}$ is obtained after product and pooling operation with stride 2. The CAM and SAM are added to the last structural block of the feature extraction network, which are combined in parallel and performed sequentially. After the feature map with the final attention weight is obtained, it is sent to the RPN network to generate anchor boxes.

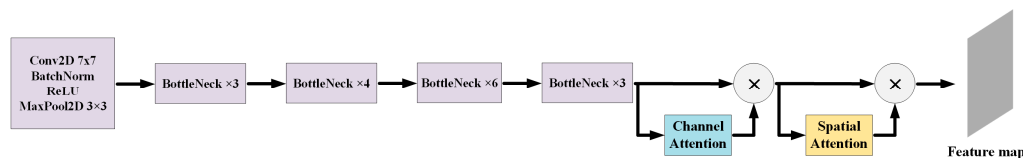


Figure 4. Improved feature extraction network.

The channel attention module is channel-dimension invariant and compresses the spatial dimension, as in Figure 5. This module focuses on the meaningful information in the input image. CAM takes the input feature map through two parallel MaxPool and AvgPool layers, changing the feature map from $C \times H \times W$ to a size of $C \times 1 \times 1$, and then goes through the share MLP module, in which the number of channels is first compressed to $1/r$ times the original number of channels, then expanded to the original number of channels, after the ReLU activation function to obtain two postactivation results. The two output

results are summed element-by-element, and then the output result of channel attention is obtained by a sigmoid activation function. This output result is then multiplied with the original graph to change back to the size of $C \times H \times W$, as in Equation (5), where F denotes the input of the feature graph ($C \times H \times W$), M_C is a one-dimensional ($C \times 1 \times 1$) channel attention graph, MLP is the multilayer perceptron, AvgPool() is the average pooling operation, MaxPool() is the maximum pooling operation, and σ is the sigmoid activation function.

$$M_C(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) = \sigma(W_1(W_0(F_{avg}^C)) + W_1(W_0(F_{max}^C))) \tag{5}$$

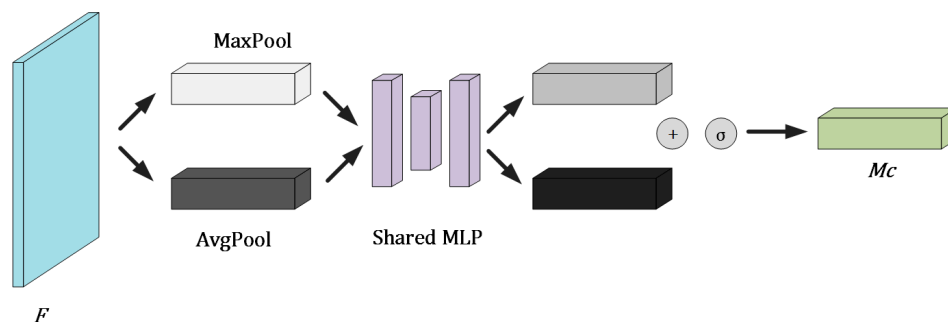


Figure 5. Channel attention module.

The spatial attention module is spatial-dimension invariant and compresses the channel dimension, as shown in Figure 6. This module focuses on the location information of the target. SAM takes the output of channel attention and obtains two $1 \times H \times W$ feature maps by maximum pooling and average pooling, then splices the two feature maps by concat operation, turns them into one-channel feature maps by 7×7 convolution, and then obtains the spatial attention’s feature map, and finally the output result is multiplied with the original map to change back to the size of $C \times H \times W$ as in Equation (6), where F denotes the input of the feature map ($C \times H \times W$), M_S is the two-dimensional $1 \times H \times W$ channel attention map, MLP is the multilayer perceptron, AvgPool() is the average pooling operation, MaxPool() is the maximum pooling operation, σ is the sigmoid activation function, $f^{7 \times 7}$ is the convolution operation with a convolution kernel size of 7×7 , and $[:]$ is the splicing operation of channel dimensions.

$$M_S(F) = \sigma(f^{7 \times 7}([AvgPool(F)]; MaxPool(F))) = \sigma(f^{7 \times 7}(F_{avg}^s; F_{max}^s)) \tag{6}$$

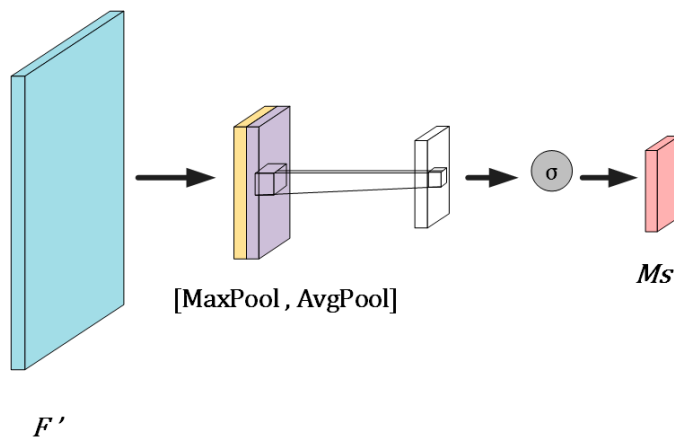


Figure 6. Spatial attention module.

3.2. Anchor Box Scheme Generation Based on K-Means Clustering Algorithm

Clustering the width and height of the labeled boxes in the photovoltaic cell defect data set, and setting the obtained cluster center as the initial anchor box scheme, so that the generated parameter scheme is more representative for the photovoltaic electromagnetic surface defects to be detected, thereby improving the detection accuracy. The pseudocode of the clustering algorithm is shown as Algorithm 1.

Algorithm 1: Anchor frame clustering algorithm

```

1:   Input: Set of target box sizes  $D = (x_1, y_1), \dots, (x_m, y_m)$ , number of clusters  $k$ 
2:   Output: mean vector  $c_1, c_2, \dots, c_k$  as anchor size
3:   Randomly select  $k$  target box sizes as the initial mean vector  $c_1, c_2, \dots, c_k$ 
4:   repeat
5:     let  $D_i = \emptyset$  ( $i = 0, 1, \dots, k - 1$ )
6:     for  $i = 0, 1, \dots, m - 1$  do
7:       for  $j = 0, 1, \dots, k - 1$  do
8:         loss  $\leftarrow 1 - IOU(c_j, (x_i, y_i))$ 
9:         if (loss < min_loss) then
10:            record  $\leftarrow j$ 
11:            min_loss  $\leftarrow$  loss
12:          put  $(x_i, y_i)$  into  $D_{record}$ 
13:        for  $i = 0, 1, \dots, m - 1$  do
14:          calculate a new mean vector  $c_i \leftarrow \frac{1}{length(D_m)} \sum_{i=0}^{length(D_m)-1} d_i$  ( $d_i \in D_m$ )
15:        until none of the current mean vectors are updated

```

The number of clusters in the original anchor frame parameter scheme is five scales and three ratios, so the number of clusters $k = 15$, the clustering algorithm assigns each sample to the nearest class center according to the aspect ratio, and according to the convergence The result adjusts the class centers, repeating until the number of iterations is reached. According to K-means clustering, the coordinates of the 15 anchors and the aspect ratio of photovoltaic cell sheet defects are shown in Figure 7. In this paper, the scale of the anchor box generated by FPN is re-customized from the original $\{32^2, 64^2, 128^2, 256^2, 512^2\}$ to $\{10^2, 30^2, 60^2, 100^2, 800^2\}$. For too large anchor aspect ratios, such as 1:2, etc., add 1:3 and 3:1 anchor aspect ratios, and finally the three aspect ratios of the anchors are $\{1:3, 1:1, 3:1\}$, combined with 5 scales, a total of 15 anchors were customized. The anchors customized by the K-means clustering algorithm are more reasonable for the photovoltaic cell surface defect data set, which can make the defect detection network converge faster and enable the model to obtain better detection performance.

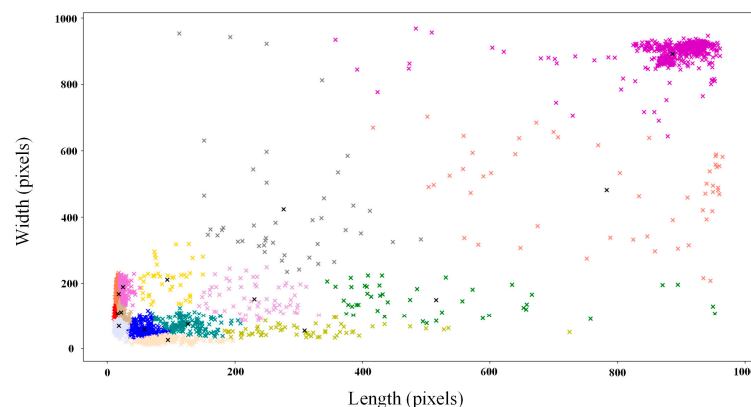


Figure 7. Clustering results for labeled box width and height.

3.3. Loss Function Optimization

The original border regression loss function is replaced by the $smooth_{L1}$ loss function with the DIoU loss function. The DIoU loss function is determined by three parts, namely the IoU, the distance of the center point and the diagonal length of the closure area.

Both the prediction frame and the target frame are represented by four coordinates (x_1, y_1, x_2, y_2) , which are the horizontal and vertical coordinates of the upper left point and the horizontal and vertical coordinates of the lower right point. The algorithm is implemented as follows:

1. Obtain the maximum value of x, y at the upper left point and the minimum value of x, y at the lower right point of the prediction frame and the target frame, find the difference and obtain the two sides of the intersection area, respectively, multiply them together and obtain the intersection value of the prediction frame and the target frame, as shown in Figure 8a;
2. The area of the prediction frame and the target frame are summed and subtracted from the intersection value to the merged value of the prediction frame and the target frame;
3. The IoU can be obtained from the intersection and merge values;
4. The square of the Euclidean distance between the two centroids is obtained by finding the centroid coordinates of the prediction frame and the target frame from their respective coordinates;
5. The minimum value of x, y at the upper left point and the maximum value of x, y at the lower right point of the prediction frame and the target frame are obtained, the difference is found to obtain the two sides of the closed region, and the square of the diagonal distance of the closed region is obtained, as shown in Figure 8b;
6. The DIoU loss value is obtained by Equations (7) and (8).

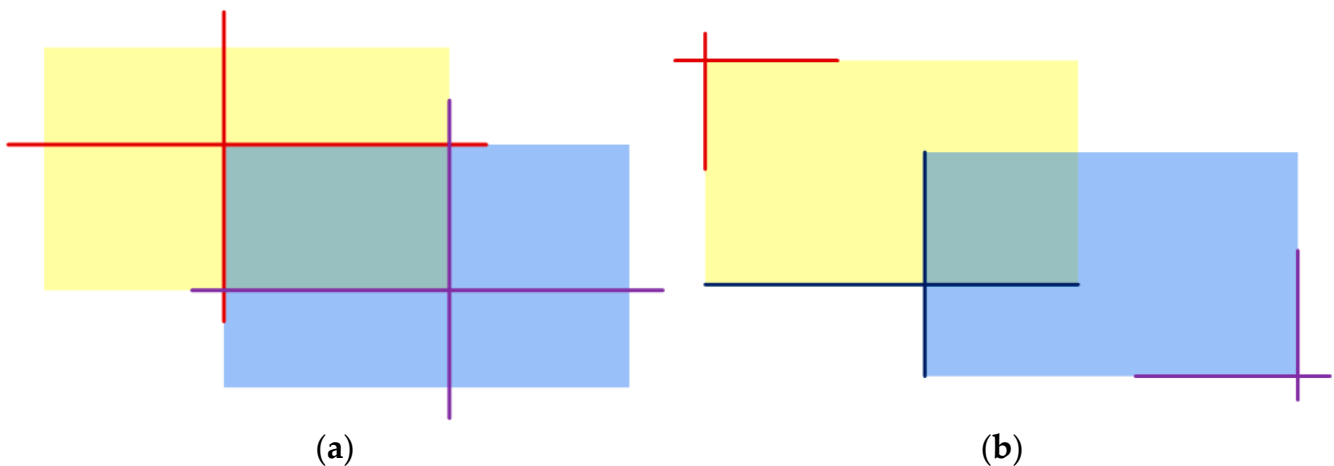


Figure 8. Diagram of the prediction and target boxes. (a) The width and height of the intersection area is obtained from the difference of the specified points. (b) The width and height of the closure area is obtained from the difference of the specified points.

$$DIoU = IoU - \frac{\rho^2(b, b_{gt})}{c^2} \quad (7)$$

$$Loss_{DIoU} = 1 - DIoU \quad (8)$$

where ρ is the Euclidean distance between the two centroids and c is the diagonal length of the closed region.

The pseudocode for DIoU loss function is shown in Algorithm 2.

Algorithm 2: DIoU loss function forward

```

1:   Input: predicted box  $B_1$  coordinate  $(a_1, b_1, c_1, d_1)$ , ground truth  $B_2$  coordinate
       $(a_2, b_2, c_2, d_2)$ 
2:   Output: DIoU loss
3:   area_predicted  $\leftarrow (c_1 - a_1) \times (d_1 - b_1)$ 
4:   area_gt  $\leftarrow (c_2 - a_2) \times (d_2 - b_2)$ 
5:   center_predicted_x  $\leftarrow \frac{(c_1 - a_1)}{2}$ 
6:   center_predicted_y  $\leftarrow \frac{(d_1 - b_1)}{2}$ 
7:   center_gt_x  $\leftarrow \frac{(c_2 - a_2)}{2}$ 
8:   center_gt_y  $\leftarrow \frac{(d_2 - b_2)}{2}$ 
9:   p  $\leftarrow (center\_gt\_x - center\_predicted\_x)^2 + (center\_gt\_y - center\_predicted\_y)^2$ 
10:  width_c  $\leftarrow \max(c_1, c_2) - \min(a_1, a_2)$ 
11:  height_c  $\leftarrow \max(d_1, d_2) - \min(b_2, b_2)$ 
12:  c  $\leftarrow width\_c^2 + height\_c^2$ 
13:  DIoU  $\leftarrow IoU(B_1, B_2) - \frac{p}{c}$ 
14:  return  $1 - DIoU$ 

```

4. Experiments

4.1. Experimental Data and Experimental Setup

This paper uses the PVEL-AD dataset, also called the EL2021 dataset, jointly published by Hebei University of Technology and Beijing University of Aeronautics and Astronautics [29]. The PVEL-AD dataset has near-infrared images of various internal defects and heterogeneous backgrounds, including 1 type of defect-free image and 12 different types of abnormal defects, such as crack (line and star), finger interruption, black core, misalignment, thick line, scratch, fragment, corner, printing error, horizontal dislocation, vertical dislocation and short circuit defects, as shown in Figure 9. Of these, crack (line and star), finger interruption and black core are the most common defects, with other types of defects rarely occurring, which would result in an unbalanced dataset distribution if all classifications were used. In order to obtain a well-trained deep learning model using a balanced dataset distribution, EL defect images and defect-free images with a resolution of 1024×1024 containing the above three common defects were selected in the PVEL-AD dataset to evaluate the three proposed applications and detection results. The annotation tool (LabelImg) was used to mark the different EL defect image data sets, and the defect was closely surrounded by a rectangular box that would reflect the specific location and category of the defect. Table 1 shows the dataset configuration.

Table 1. Dataset configuration.

Category	Training Set	Test Set
Crack	884	376
Finger	2105	853
Black Core	688	340

The experiments were conducted on a workstation with an Intel Xeon (Skylake) Platinum 8163 CPU and 2*NVIDIA Tesla V100S 32G. The model used COCO pre-trained model to initialize ResNet50 to speed up the convergence of the network. The initial learning rate was set to 0.01, the network batch size was set to 4, and the maximum number of iterations was 30, which ensured a complete cycle of the photovoltaic cell EL training data. The RPN network training batch size was set to 256. Other detailed parameters are shown in Table 2.

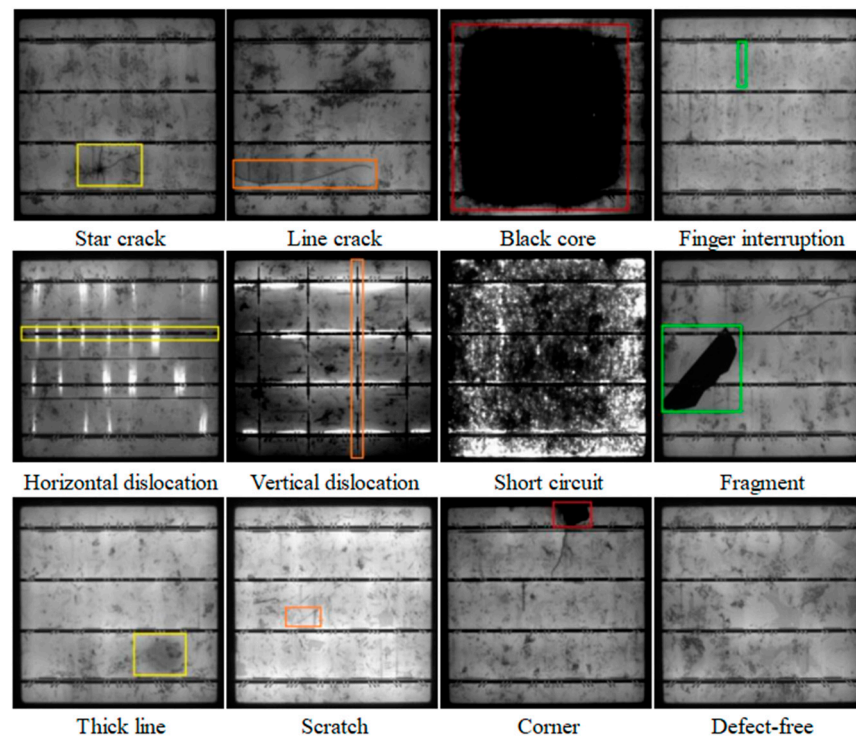


Figure 9. PVEL-AD dataset with 12 different categories of abnormal defects and defect-free images.

Table 2. Parameter details.

Parameter	Choice
Image size	1024 × 1024
Learning rate	0.01
Network batch size	4
Momentum	0.9
RPN batch size	256
Max iteration	30
ROI foreground threshold	(0.5, 1)
ROI background threshold	(0, 0.5)
Image size	1024 × 1024

4.2. Ablation Experiments

To clarify the performance impact of each module in the PV cell surface defect detection model and to verify the effectiveness of each module structure, ablation experiments were designed and trained using a mixed data set, and the test results are shown in Figure 10. The final improved Faster R-CNN network did not improve significantly in the recognition of black cores compared to the original network, because the defect features of black cores were so obvious and simple that even the original network could achieve better detection results. On the other hand, the average accuracy in the two small defect categories of cracks and broken grids improved significantly, which shows that the improved algorithm is more accurate than the original network in locating small and medium-sized defects, and has better results in identifying defects in all three categories.

In order to demonstrate the performance of the algorithms proposed in this paper and to investigate the effectiveness of each of the improvements, the models were trained and tested on the PV cell surface defect dataset using the above improvements, using the same hyperparameters and training techniques for each set of experiments. The experimental results are shown in Table 3, where \checkmark indicates that the module was introduced.

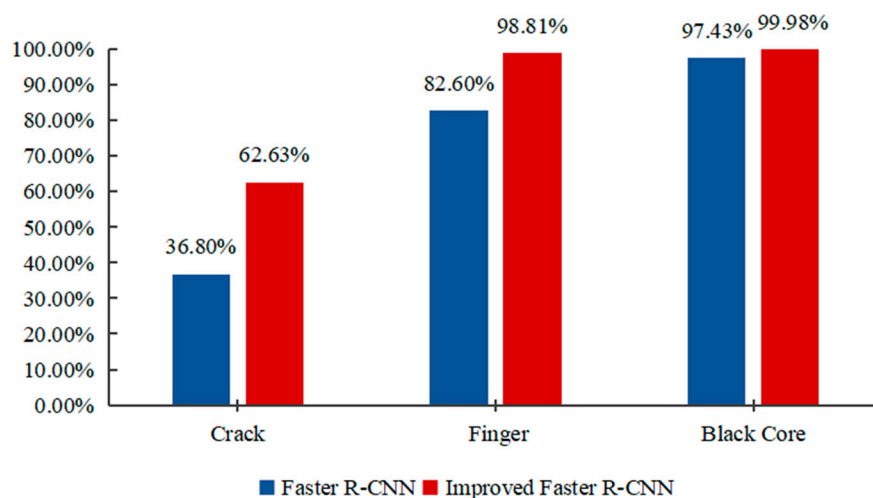


Figure 10. Comparison of the average accuracy of the original Faster R-CNN and this paper.

Table 3. Results of experiments with different improvement methods.

Group	Faster R-CNN	Pre-Training Weights	CBAM	Anchor Clustering	Loss Function	<i>mAP</i> (%)	Crack	Finger	Black Core
1	✓					72.27%	36.80%	82.60%	97.43%
2	✓	✓				78.14%	45.67%	89.95%	98.80%
3	✓	✓	✓			83.10%	55.66%	94.54%	99.10%
4	✓	✓	✓	✓		86.53%	61.93%	98.11%	99.54%
5	✓	✓	✓	✓	✓	87.14%	62.63%	98.81%	99.98%

As can be seen in Table 3, the *mAP* of the original Faster R-CNN was 72.27%, and the *mAP* of the model after using pre-training weights was 78.14%, which is 5.87 percentage points higher than the original model, indicating that the use of pre-training weights can speed up the convergence of the network. On the basis of using pre-training weights, the feature extraction network integrated into CBAM made the *mAP* of the model 83.10%, which is 10.83% higher than the original model and 4.96% higher than that of the second set of experiments, indicating that the strategy combined with CBAM is effective. On the basis of using pre-training weights and incorporating CBAM, the *mAP* of the model after using the improved anchor box generation parameters was 86.53%, which is 14.26% higher than the original model and 3.43% higher than the third group of experiments. It shows that the anchor frame generation parameters proposed in this paper are more suitable for the scale of the defect target than the original anchor point generation parameters, and the positioning is more accurate, which can effectively reduce the situation of missed detection and false detection, and improve the accuracy of the model. After introducing the DIoU loss function, the *mAP* of the model was 87.14%, which is 14.87% higher than the original model and 0.61% higher than that of the third set of experiments, indicating that the replacement of the loss function improves the detection accuracy slightly. In general, each improvement method proposed in this paper effectively improved the model accuracy. After accumulating each improvement method, the model accuracy was gradually improved, and the model obtained better results, indicating that the improved network is less effective for photovoltaic cells. The defect detection capability was significantly improved, and the defect detection effect of photovoltaic cells was better.

As shown in Figure 11, the *AP* values of various types of defects on the test set under different improvement methods can be seen, each improvement method improved the detection accuracy of each defect category, especially the detection accuracy of small defect target categories, significantly. For example, the original Faster R-CNN model had a poor detection effect on the small defect category such as cracks, which was only

36.80%. After adding the improved method in this paper, the final detection accuracy of the algorithm in this paper for this category was 62.63%, which is higher than the original model, respectively, increased by 25.83%. In contrast, for the black core with a large defect area and obvious color contrast, the original model already had high detection accuracy, and the detection accuracy of the improved model can only be slightly improved. The experimental results showed that the improved method in this paper had a certain effect on the detection of small target defects, and could effectively reduce the missed detection rate and false detection rate of defects.

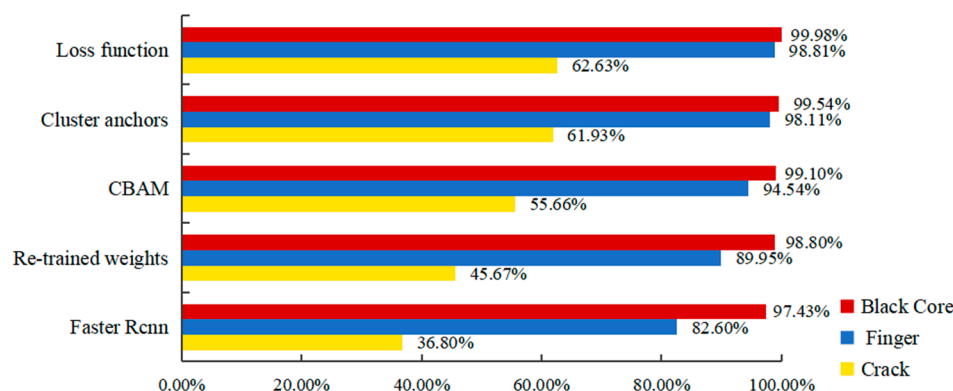


Figure 11. Comparison of AP values of defects with different improvement methods.

4.3. Comparison of Different Target Detection Algorithms

In order to verify the performance of the improved method in this paper on the detection of photovoltaic cell surface defects, the algorithm in this paper was compared with the other five models. The experimental results are shown in Table 4. *AP* and *mAP* were used as evaluation indexes. In the training process, the IoU greater than or equal to 0.5 is a positive sample, and the IOU less than 0.5 is a negative sample. As can be seen from Table 4, the *AP* value of this algorithm for cracks, fingers and black core was better than that of YOLOv5 model and the original Faster R-CNN, especially for cracks and fingers. This is because the algorithm in this paper fully considers the characteristics of targets in the defect data set of photovoltaic cells, and makes targeted improvements for targets with small defects in the data set. At the same time, the *mAP* value of the algorithm in this paper was 87.14%, which was better than the other five models and meets the requirements of higher detection accuracy for the surface defects of photovoltaic cells.

Table 4. Comparison of different algorithms.

Group	<i>mAP</i> (%)	Crack	Finger	Black Core
SSD [15]	78.00%	/	/	/
YOLO v5	68.92%	33.41%	76.41%	96.96%
Faster R-CNN [13]	72.27%	36.80%	82.60%	97.43%
RCA-Faster R-CNN [30]	83.29%	/	/	/
RetinaNet [31]	84.53%	/	/	/
Our method	87.14%	62.63%	98.81%	99.98%

5. Conclusions

Aiming at the characteristics of small defects in electroluminescence images of photovoltaic cells, the low detection accuracy and poor generalization ability of the general models in practical production in various fields, this paper proposes an improved photovoltaic cell based on Faster R-CNN network. The abnormal defect detection method optimizes the feature extraction network by introducing CBAM to suppress the features of complex background and grain pseudo-defects; at the same time, the K-means clustering algorithm is used to train targeted anchors, so that the generated anchor frame can be better

fitted to the defect target detected in this paper, thereby reducing the probability of missed detection and false detection of small defects, and improving the detection accuracy of the model. Finally, the DIoU loss function was used to directly minimize the normalized distance between the anchor boxes and the target box, which speeded up the convergence and made the target box more accurate. The experimental results show that the defect detection algorithm proposed in this paper has a remarkable adaptability to defects of various shapes and small cracks, reduces the probability of missed detection of small target cracks, and improves the detection performance of the entire model. Specifically, the model improved the overall accuracy of PV cell-defect detection by 14.87%. There was also a significant improvement in the detection of small defects, for example, 25.83% for crack defects and 16.21% for finger defects. A small improvement over good detection results was also achieved for obvious defects, for example, a 2.55% improvement for black core defects. It has a good application prospects in the field of photovoltaic cell defect detection.

Author Contributions: Conceptualization, A.C. and X.L.; methodology, A.C.; software, A.C.; validation, H.J., C.H. and M.L.; formal analysis, A.C.; investigation, A.C.; resources, X.L.; writing—original draft preparation, A.C.; writing—review and editing, X.L.; visualization, A.C.; supervision, X.L.; project administration, H.J., C.H. and M.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National key research and development plan “ Multi-dimensional visual information edge intelligent processor chip” (2022YFB2804402).

Data Availability Statement: Not applicable.

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

References

1. Ren, H.W.; Tian, K.G.; Hong, S.X.; Dong, B.Q.; Xing, F.; Qin, L. Visualized Investigation of Defect in Cementitious Materials with Electrical Resistance Tomography. *Constr. Build. Mater.* **2019**, *196*, 428–436. [[CrossRef](#)]
2. Li, Y.B.; Liu, M.J. Aerial Image Classification Using Color Coherence Vectors and Rotation & Uniform Invariant LBP Descriptors. In Proceedings of the 2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Chongqing, China, 12–14 October 2018; pp. 653–656.
3. Wen, T.K.; Yin, C.C. Crack Detection in Photovoltaic Cells by Interferometric Analysis of Electronic Speckle Patterns. *Sol. Energy Mater. Sol. Cells* **2011**, *98*, 216–223. [[CrossRef](#)]
4. Dhimish, M.; Holmes, V. Solar Cells Micro Crack Detection Technique Using State-of-the-Art Electroluminescence Imaging. *J. Sci. Adv. Mater. Devices* **2019**, *4*, 499–508. [[CrossRef](#)]
5. Hu, C.F.; Wang, Y.X. An Efficient Convolutional Neural Network Model Based on Object-Level Attention Mechanism for Casting Defect Detection on Radiography Images. *IEEE Trans. Ind. Electron.* **2020**, *67*, 10922–10930. [[CrossRef](#)]
6. Sassi, P.; Tripicchio, P.; Avizzano, C.A. A smart monitoring system for automatic welding defect detection. *IEEE Trans. Ind. Electron.* **2019**, *66*, 9641–9650. [[CrossRef](#)]
7. Chen, H.Y.; Pang, Y.; Hu, Q.D.; Liu, K. Solar Cell Surface Defect Inspection Based on Multispectral Convolutional Neural Network. *J. Intell. Manuf.* **2020**, *31*, 453–468. [[CrossRef](#)]
8. Han, H.; Gao, C.Q.; Zhao, Y.; Liao, S.S.; Tang, L.; Li, X.D. Polycrystalline Silicon Wafer Defect Segmentation Based on Deep Convolutional Neural Networks. *Pattern Recognit. Lett.* **2020**, *130*, 234–241. [[CrossRef](#)]
9. Akram, M.W.; Li, G.Q.; Jin, Y.; Chen, X.; Zhu, C.A.; Zhao, X.D.; Khaliq, A.; Faheem, M.; Ahmad, A. CNN Based Automatic Detection of Photovoltaic Cell Defects in Electroluminescence Images. *Energy* **2019**, *189*, 116319. [[CrossRef](#)]
10. Hussain, M.; Chen, T.H.; Titrenko, S.; Su, P.; Mahmud, M. A Gradient Guided Architecture Coupled with Filter Fused Representations for Micro-Crack Detection in Photovoltaic Cell Surfaces. *IEEE Access* **2022**, *10*, 58950–58964. [[CrossRef](#)]
11. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
12. Girshick, R. Fast r-cnn. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Santiago, Chile, 7–13 December 2015; pp. 580–587.
13. Ren, S.Q.; He, K.M.; Girshick, R.; Sun, J. Faster r-cnn: Towards Real-Time Object Detection with Region Proposal Networks. In Proceedings of the IEEE Transactions on Pattern Analysis and Machine Intelligence, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 1137–1149.
14. He, K.M.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. In Proceedings of the IEEE International Conference on Computer Vision, Honolulu, HI, USA, 21–26 July 2017; pp. 2961–2969.

15. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. SSD: Single Shot MultiBox Detector. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 10–16 October 2016; pp. 21–37.
16. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 779–788.
17. Redmon, J.; Farhadi, A. YOLO9000: Better, Faster, Stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 7263–7271.
18. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. *arXiv* **2018**, arXiv:1804.02767.
19. Bochkovskiy, A.; Wang, C.-Y.; Liao, H.-Y.M. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934.
20. He, K.M.; Zhang, X.Y.; Ren, S.Q.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778.
21. Woo, S.H.; Park, J.C.; Lee, J.Y.; Kweon, I.S. CBAM: Convolutional Block Attention Module. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; pp. 3–19.
22. Li, X.; Te, R.G.; Yi, F.; Xu, G.C. TCS-YOLO model for global oil storage tank inspection. *Opt. Precis. Eng.* **2023**, *31*, 246–262. [[CrossRef](#)]
23. Li, K.; Jiang, X.L.; Chen, E.K.; Chen, P.; Xu, Z.Y.; Lin, Q. Auto-Counting the Eel Anguilla in Recirculating Aquaculture System Via Deep Learning. *Oceanologia Limnologia Sinica* **2022**, *53*, 664–674.
24. Yu, J.H.; Jiang, Y.N.; Wang, Z.Y.; Cao, Z.M.; Huang, T. UnitBox: An Advanced Object Detection Network. In Proceedings of the 24th ACM international conference on Multimedia, Amsterdam, The Netherlands, 15–19 October 2016; pp. 516–520.
25. Rezatofighi, H.; Tsoi, N.; Gwak, J.Y.; Sadeghian, A.; Reid, I.; Savarese, S. Generalized Intersection Over Union: A Metric and a Loss for Bounding Box Regression. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 658–666.
26. Zheng, Z.H.; Wang, P.; Liu, W.; Li, J.Z.; Ye, R.G.; Ren, D.W. Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; pp. 12993–13000.
27. Zhang, H.T.; Tian, M.; Shao, G.P.; Cheng, J.; Liu, J.J. Research on Military Gesture Detection Algorithm Based on Improved EfficientDet. *Fire Control Command. Control* **2022**, *47*, 97–106.
28. Li, D.P.; Ren, X.M.; Yan, N.N. Real-Time Detection of Insulator Drop String Based on UAV Aerial Photography. *J. Shanghai Jiaotong Univ.* **2022**, *56*, 994.
29. Su, B.Y.; Zhou, Z.; Chen, H.Y. PVEL-AD: A Large-Scale Open-World Dataset for Photovoltaic Cell Anomaly Detection. *IEEE Trans. Ind. Inform.* **2022**, *19*, 404–413. [[CrossRef](#)]
30. Yi, X.T.; Shan, Y.F. Internal Defect Detection of Photovoltaic Cells Based on Improved Faster R-CNN. *J. Electron. Meas. Instrum.* **2021**, *35*, 40–47.
31. Lin, H.; Li, B.; Wang, X.G.; Shu, Y.Y.; Niu, S.L. Automated Defect Inspection of LED Chip Using Deep Convolutional Neural Network. *J. Intell. Manuf.* **2019**, *30*, 2525–2534. [[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.