processes

MDPI

Article

Bird Droppings Defects Detection in Photovoltaic Modules Based on CA-YOLOv5

Linjun Liu [1], Qiong Li [1,*], Xu Liao [1] and Wenbao Wu [2]

[1] School of Information Engineering, Nanchang Hangkong University, Nanchang 330063, China; 2204081100018@stu.nchu.edu.cn (L.L.)
[2] China Power Construction Group Jiangxi Power Construction Co., Ltd., Nanchang 330001, China
* Correspondence: liqiong471255611@163.com

Abstract: Due to the characteristics of different forms and small bird droppings-related defects in photovoltaic modules, problems of missing detection, wrong detection, and low detection accuracy often exist in the bird droppings covering the detection of photovoltaic modules. In this paper, a defect identification method based on the improved Coordinate Attention—You Only Look Once (CA-YOLOv5) network is proposed. Firstly, a layer of the Coordinate Attention module is added between the Backbone and Neck network, which takes into account channel and location information to enhance the feature extraction ability of the network model. Secondly, the small target detection layer is added to fuse different feature information from shallow networks and deep networks so as to improve the multi-scale feature detection ability of the network structure and effectively improve the detection effect of small targets. Finally, the experimental results show that, compared with the original algorithm, the proposed improved algorithm based on YOLOv5s has a mean average precision (mAP) value of 92%, an increase of 5.2%, and the model volume is also reduced in different degrees compared with other mainstream algorithms, achieving a good balance between detection accuracy and model volume. The results show that the model has a more accurate detection result for the bird droppings-related defect detection of photovoltaic modules and can provide a reference for the detection of photovoltaic modules in real life.

Keywords: photovoltaic defects; fault detection; YOLO algorithm; deep learning

check for updates

Citation: Liu, L.; Li, Q.; Liao, X.; Wu, W. Bird Droppings Defects Detection in Photovoltaic Modules Based on CA-YOLOv5. *Processes* **2024**, *12*, 1248. https://doi.org/10.3390/pr12061248

Received: 8 May 2024
Revised: 23 May 2024
Accepted: 12 June 2024
Published: 18 June 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

With the increasing energy crisis, solar energy, as a kind of green energy, has received more and more attention [1]. Although the solar photovoltaic power generation method is simple and easy to repair, due to long-term installation in a complex outdoor environment, solar photovoltaic panels are prone to various defects, which has a serious impact on the effectiveness of photovoltaic panels; for example, it can reduce the power generation efficiency of photovoltaic modules, cause hot spot effects, shorten the service life of photovoltaic modules, and increase system maintenance costs, and the bird droppings defects are one of the more easily produced and a larger number of defects. Therefore, it is particularly necessary to check for bird droppings-related defects in photovoltaic panels. However, the traditional manual inspection method has some problems, such as low efficiency, missing detection, and misdetection. With the development of Unmanned Aerial Vehicle (UAV) aerial photography technology, the data foundation is provided for the detection of bird droppings based on image recognition. The identification of bird droppings-related defects by deep learning methods can greatly improve the efficiency and accuracy of inspection.

For this reason, scholars at home and abroad have conducted extensive research on deep learning-based photovoltaic defect detection methods, which can be roughly divided into two categories: single-stage and two-stage object detection algorithms. Single-stage algorithms such as You Only Look Once (YOLO) [2] and Single Shot MultiBox Detector

(SSD) [3] only need to extract features once to achieve target detection, which has a high detection speed. However, the two-stage algorithm needs to generate candidate bounding boxes and then identify the real object from these candidate boxes common R-CNN (Region with CNN feature) [4], Fast R-CNN [5], Faster R-CNN [6], etc. These methods can improve detection accuracy, but relative detection speed is slower.

Kong et al. [7] proposed a global multi-convolution module based on the YOLOv5 algorithm to balance the spatial pyramid structure and improve the attention mechanism and loss function. The improved model improved the fault detection accuracy of photovoltaic modules and made it meet the needs of fault detection in actual industry. Chen et al. [8] proposed a new object detection method, DSM-YOLOv5, which can effectively improve the detection efficiency of UAV aerial images in complex environments and dense scenes by deepening the network depth of the feature fusion part and adding additional small target detection heads. Xie et al. [9] integrated the GhostNetV2 module based on the YOLOv5s network and introduced an improved Dense Convolutional Network (DenseNet) module and attention mechanism, which improved the detection accuracy of photovoltaic modules under complex texture backgrounds and solved the problems of missing detection and false detection. Yu et al. [10] proposed a small target detection algorithm based on YOLOv5, designed Context Feature Module (CFM) and feature specify module (FSM) modules, adopted the Transpose module, and added a detection layer to improve the detection accuracy of small target detection. However, due to the increase in the number of network layers and the increase in the number of parameters in the model, the speed of target detection is reduced. Guo et al. [11] proposed a photovoltaic module defect detection algorithm based on improved YOLOv5. A ghost module was introduced, the squeeze-and-excitation (SE) network was fused, and the Bidirectional Feature Pyramid Network (BiFPN) structure was adopted for multi-scale feature fusion at the same time, which reduced the size of the model and met the demand for real-time detection. However, compared with the original model, the detection accuracy is reduced, and the generalization ability of the model is insufficient. Wang et al. [12] proposed a photovoltaic (PV) module defect detection algorithm based on YOLOv5 LiteX, which enhanced the performance of the model in feature fusion and minimal target detection, basically meeting the requirements of real-time detection in the work site environment, but the detection accuracy of PV module defects was still low. Zhou et al. [13] proposed a photovoltaic module defect identification method based on a multi-scale convolutional neural network (CNN), built three end-to-end convolutional neural network models of different scales, and introduced an attention mechanism, which solved the problem that small defects were not easy to identify and improved the defect identification accuracy but did not solve the defect location problem. Yan et al. [14] proposed a feature fusion detection network based on a multi-source image fusion network to complete robust defect detection tasks for photovoltaic panels and improve the accuracy of defect detection. However, the registration process for image data still requires more manual intervention. Tian et al. [15] proposed a photovoltaic module defect detection model based on multi-scale feature fusion that embedded a Coordinate Attention mechanism, used a bidirectional feature pyramid, and added a micro-target detection layer. This model is more suitable for solving the photovoltaic module defect detection problem under complex background conditions. Kisantal et al. [16] increased the proportion of small targets in the data set through the operation of copy–paste small targets, thereby increasing the contribution of small targets to the network. However, this only increases the number of samples of small target objects in the data set, cannot guarantee the balance of positive and negative sample numbers during training, and does not make use of the feature information of small targets.

In summary, although the current algorithm for photovoltaic detection has improved the detection accuracy to a certain extent, there are still shortcomings, such as increasing the size of the model and the number of parameters in order to improve the accuracy, resulting in slow detection speed [7–10], weak model migration ability and insufficient generalization ability [11], photovoltaic defect location problems [13], and insufficient use of small target information of photovoltaic modules [16]. By analyzing the advantages and disadvantages of the photovoltaic defect detection algorithms used in the above studies and considering that the photovoltaic module bird droppings-related defects collected by the UAV are small and the target is not clear, this paper will use the YOLOv5s model as the basic framework. Firstly, the Coordinate Attention module is introduced to enhance attention to the target area from space and enhance the key feature extraction ability of the network model for the image. Secondly, the detection layer of small targets is added at the same time, and the different characteristics of the shallow and deep networks are merged to enhance the detection ability of small target bird droppings. Finally, the ablation of the improved network and the comparison experiment of different algorithms show that the algorithm can effectively improve the detection ability of bird droppings-related defects of photovoltaic modules and achieve a balance between model size and detection accuracy, which is very suitable for the UAV inspection scene of photovoltaic modules.

2. YOLOv5 Network Model Introduction

YOLO algorithms are known for their high speed and accuracy, and YOLOv5 is a fifth-generation single-stage object detection algorithm that further improves model structure, training strategies, and reasoning speed and is a mainstream single-stage target detection network with excellent detection speed and accuracy [17,18]. The YOLOv5 network consists of four parts: Input, Backbone, Neck, and Output.

At the Input end, YOLOv5 uses Mosaic data enhancement mode to randomly select four images for scaling and stitching to increase data diversity. At the same time, adaptive anchor frame calculation and image scaling technologies are introduced to improve the efficiency of target detection. The Backbone network of YOLOv5 uses the Focus structure, which was removed in version 6.0, to facilitate model deployment. Therefore, the Backbone network of version 6.0 is mainly composed of CBS (Conv + BatchNorm + SiLU), CSP1_X (Cross Stage Partial), and SPPF (Spatial Pyramid Pooling Fast). SPPF structure converts feature graphs of different sizes into feature vectors of fixed sizes, and training images with diverse sizes makes the convergence effect of the network better than training images of single sizes [19]. CSP1_X in Backbone divides the underlying features into two parts, one of which is spliced with the other branch after multiple residual structures and convolution operations to enhance the network's learning ability.

The Neck part is composed of the CSP2_X structure, Feature Pyramid Network (FPN) [20], and Path Aggregation Network (PAN) [21]. In the Neck network, FPN and PAN structures are used to achieve bottom-up and top-down feature fusion, which enhances the feature fusion capability. Three decoupling heads are set at the output end of the network, which is connected to the three outputs of the Neck part, respectively, and the corresponding prediction boundary boxes are generated on the feature graphs with sizes of 20×20, 40×40, and 80×80. This paper adopts a lightweight Yolov5s model for training, and the network structure is shown in Figure 1. In Figure 1, Conv represents a module in the convolutional neural network, BN represents Batch Normalization, SiLU represents the Swish function, CBS represents Conv + BatchNorm + SiLU, and SPPF is an improved version of Spatial Pyramid Pooling, which is used to splice different pooling results.

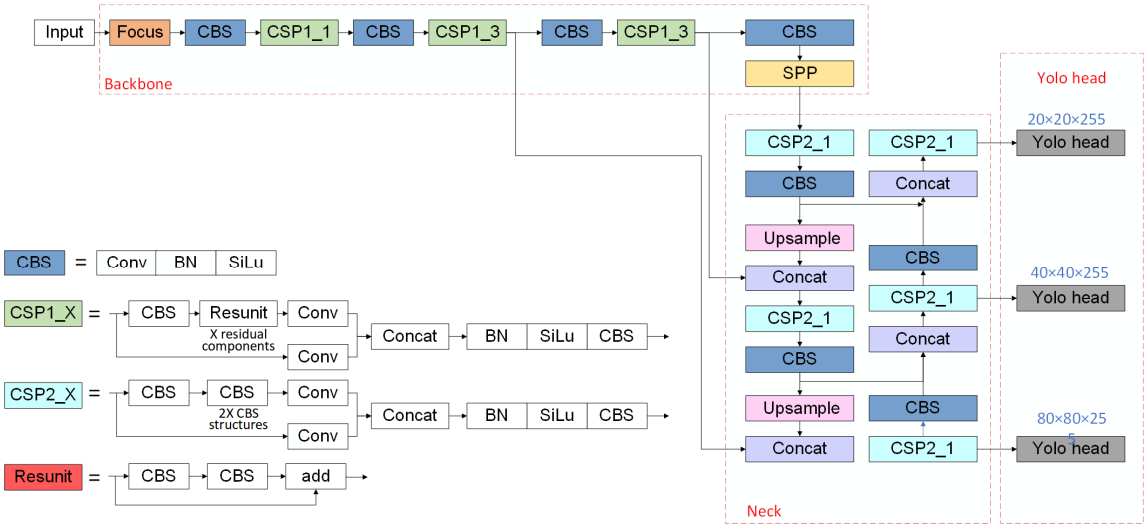


Figure 1. Structure diagram of YOLOv5s.

3. Improved CA-YOLOv5 Algorithm

Due to the large resolution, relatively complex background, and small target defects of the photovoltaic module pictures collected by UAV, the current YOLOv5s network structure model is not effective for multi-scale target detection, and there are often problems of missing detection, wrong detection, and low accuracy for small target detection. Given these problems, in this paper, the following improvements are made to the YOLOv5s model to improve the effect of multi-scale target detection:

1. Coordinate Attention is added to the Backbone network to enable the model to process specific parts of the feature map more accurately to improve the capability of feature extraction and, thus, the accuracy of the network model.
2. Add a new small target detection layer to the original network structure and integrate the location information of the shallow network with the semantic information of the deep network to enhance the multi-scale target detection capability of the network structure and improve the accuracy of small target detection.

3.1. The Coordinate Attention Module

Attention mechanisms include squeeze-and-excitation networks [22], convolutional block attention module (CBAM) [23], ECA-Net [24], Coordinate Attention, etc. Coordinate Attention [25] mainly encodes channel relationships and long-term dependencies through accurate position information. The model can identify and extract the key information in the image more accurately. Compared with other attention mechanisms, Coordinate Attention not only considers the channel information but also the location information related to the direction and is flexible and lightweight enough to be easily inserted into the core module of the lightweight network, which can improve the precision and accuracy of the model without adding additional computational overhead.

In this paper, for photovoltaic module images, bird droppings only account for a small part of the pixels, which is easy to be affected by complex backgrounds. To solve the problem of jumbled information interference, Coordinate Attention is introduced in this paper to enhance the attention of the bird dross region from space and reduce the influence of image brightness imbalance and low contrast on segmentation accuracy. As shown in Figure 2, a layer of Coordinate Attention is added based on the structure diagram of YOLOv5s. The CA with a white background refers to the Coordinate Attention. When incorporating the improvement of the attention mechanism, attention is added to the Backbone network of YOLOv5 feature extraction before the SPPF layer. The attention mechanism is used to encode the remote dependency and location information from the

horizontal and vertical spatial dimensions of the input image, and it embeds the location information into the feature vector so that the network can obtain a wider range of features.

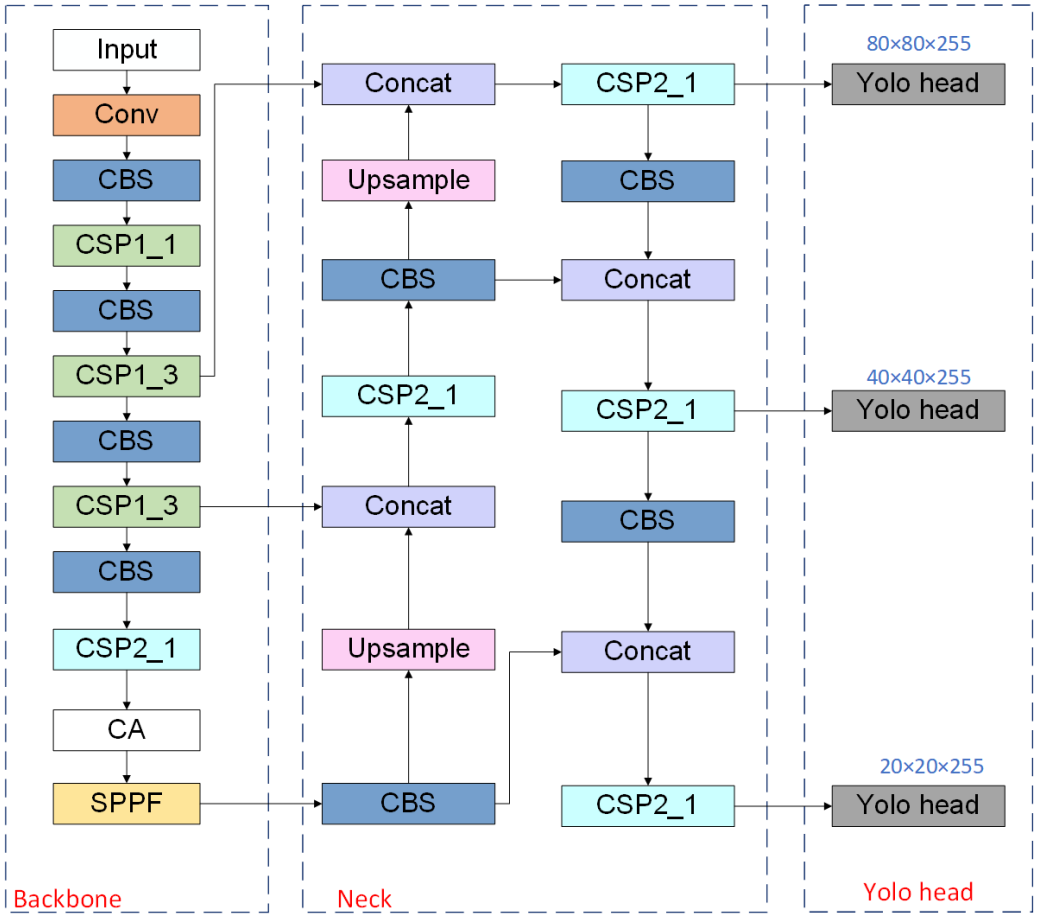


Figure 2. Network structure after Coordinate Attention is added.

The structure of the Coordinate Attention is shown in Figure 3. CA channel attention can be divided into two one-dimensional feature encoding processes and collection features in the direction of two spaces. Long-range correlation can be obtained along one direction, while precise location information can be retained along the other spatial direction. Then, the generated feature map is decoded into a pair of direction-specific and position-cognizant attention maps, which can be used in combination with the original input feature map to enhance the features of the detected object [26]. The CA first extracts the global information of the width and height directions of the input feature graph to obtain the width and height directions of the feature graph. Specifically, the input feature graph with size $C \times H \times W$ is pooled in the X direction and the Y direction, respectively, to generate the feature graph with size $C \times H \times 1$ and $C \times 1 \times W$, respectively. The calculation formula is as follows:

$$z_c^h(h) = \frac{1}{W} \sum_{0 \le i \le W} x_c(h, i) \tag{1}$$

$$z_c^w(w) = \frac{1}{H} \sum_{0 \le j \le H} x_c(j, w) \tag{2}$$

where c is the number of channels, representing different feature channels in the feature graph; h is the height, representing the vertical dimension of the feature map; w is the width, representing the horizontal dimension of the feature map; $z_c^h(h)$ is the output result of the c channel with height h; $z_c^w(w)$ is the output of channel c of width w.

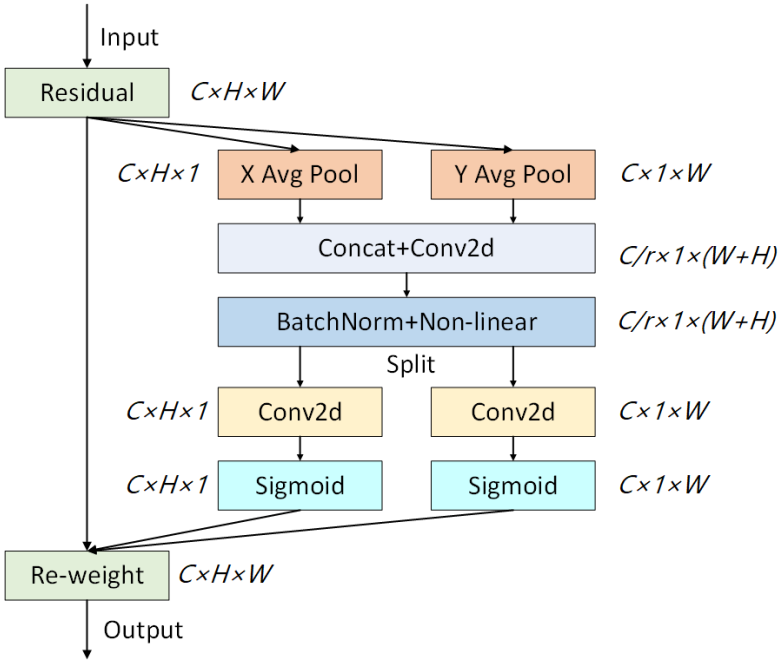


Figure 3. Structure of Coordinate Attention.

Secondly, the output results obtained above are concatenate operations, and then, through convolution transformation, the intermediate feature mapping of spatial information encoded in horizontal and vertical directions is obtained. The formula is as follows:

$$f = \delta\left(F_1\left(\left[z^h, z^w\right]\right)\right) \tag{3}$$

where $\left[z^h, z^w\right]$ is the concatenate operation along the spatial dimension, δ is the nonlinear activation function, F_1 is the 1×1 convolution transformation function, and f is the intermediate feature mapping encoding the spatial information in the horizontal and vertical directions.

f is decomposed into $f^h \epsilon R^{C/r \times H}$ and $f^w \epsilon R^{C/r \times W}$ along the space dimension, and then 1×1 convolution is used to raise the dimension, respectively; both F_h and F_w in Formulas (4) and (5) are 1×1 convolution, and then combined with the sigmoid activation function, the final attention vectors $g^h \epsilon R^{C \times H \times 1}$ and $g^w \epsilon R^{C \times 1 \times W}$ are obtained, the formula is as follows:

$$g^h = \sigma\left(F_h\left(f^h\right)\right) \tag{4}$$

$$g^w = \sigma(F_w(f^w)) \tag{5}$$

Finally, the output g^h and g^w are extended as attention weights respectively, and the feature graph with attention weights in height and width is obtained. The formula is as follows:

$$y_c(i,j) = x_c(i,j) \times g_c^h(i) \times g_c^w(j) \tag{6}$$

Coordinate Attention is a mechanism that integrates location information into channel information, which can be flexibly inserted into classic mobile networks and has a more accurate detection effect, improving the accuracy of network models.

3.2. Multi-Scale Small Target Detection Layer

For the photovoltaic module pictures collected by the UAV, the defect size of the bird droppings to be detected is too small, and the background image is large. For a picture of 4000 × 3000 pixels, the pixel of bird droppings is less than 50 × 150, and the shape is irregular, so the detection is difficult. The original YOLOv5s model has only three detection

layers. The input images are downsampled 8 times, 16 times, and 32 times, respectively, and the generated prediction scales are used to detect small targets, medium targets, and large targets, respectively. However, due to the large downsampling, it is difficult to learn the feature information of small targets from the deep feature map after multiple downsampling steps. The original prediction layer of three scales has been unable to meet the small target detection in the image, and it is easy to miss the detection phenomenon, which leads to the difficulty of bird droppings-related detection of photovoltaic modules.

Therefore, this paper proposes to add a small target detection layer to the network structure, from the original three-layer detection layer to the four-layer detection layer. Based on the original network's three defect prediction scales of 20 × 20, 40 × 40, and 80 × 80, an upsampling layer is added to form a prediction scale of 160 × 160. The newly added prediction layer retains more location information and more detailed features about small targets and has more effective results for detecting small targets. The improved network structure is shown in Figure 4.

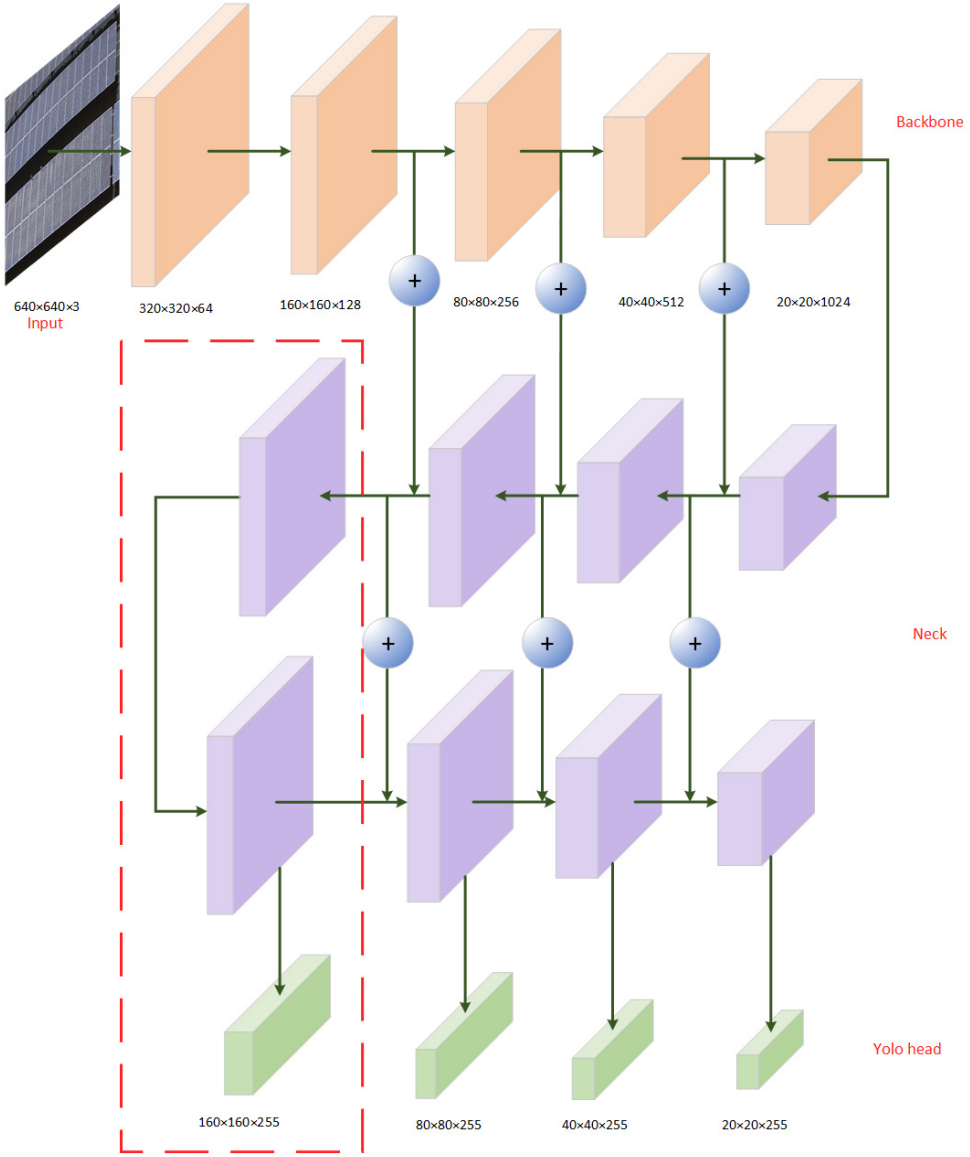


Figure 4. Network structure after adding a small target detection layer.

The red dotted frame in the figure is a newly added small target detection layer. The size of the input image is 640 × 640, and the feature maps of 20 × 20, 40 × 40, and 80 × 80 are obtained after multiple convolutions and 32, 16, and 8 times downsampling.

Downsampling by a factor of 8 means that the width and height of the feature map are both 1/8 of the original input image. Since the size reduction is relatively small, these feature maps still retain a certain level of spatial detail, which helps the model more accurately locate small objects. Downsampling by a factor of 16 indicates that the width and height of the feature map are both 1/16 of the original input image. These features are typically used to detect medium-sized objects, as they possess sufficient spatial information to locate objects while also containing a certain amount of semantic information to aid in classification. Downsampling by a factor of 32 is often used for detecting larger objects. Due to the greater reduction in size, these feature maps possess a larger receptive field, enabling them to capture broader contextual information. The shallow feature map has sufficient location information for small targets, but the original YOLOv5s model structure does not make full use of the shallow feature map in the Neck module.

In this paper, an upsampling layer is added on this basis. By adjusting the grid size in the added upsampling layer, the model can better capture the feature information of small objects. The use of a more refined feature map enhancement algorithm enhances the detection capability for small objects, enabling the identification of even smaller targets at different scales. This significantly improves the algorithm's detection ability for small objects, thus enhancing the accuracy and practicability of the detection algorithm. The size of the feature map obtained by processing the upsampling layer is 160 × 160. The generated feature map is fused with the feature map of the second layer in the Backbone module, and the generated 160 × 160 scale feature map is input into the prediction module to obtain a detection layer for small targets. The feature map processed by shallow networks contains more small target location information, and the feature map processed by deep networks contains more small target feature information. The new detection head integrates the characteristics of the two kinds of networks, deepens the network depth, enhances the multi-scale learning ability of the model, and has a more efficient small target detection effect.

4. Methodology

The photovoltaic module defect identification process based on the improved CA-YOLOv5s algorithm mainly consists of the data processing stage, model training stage, and model testing stage, and the overall process is shown in Figure 5.

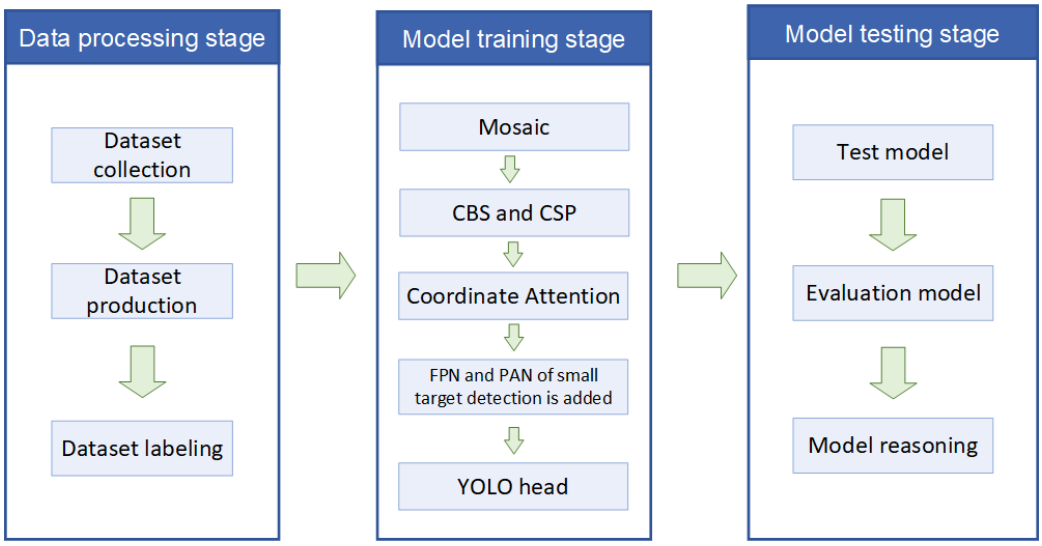


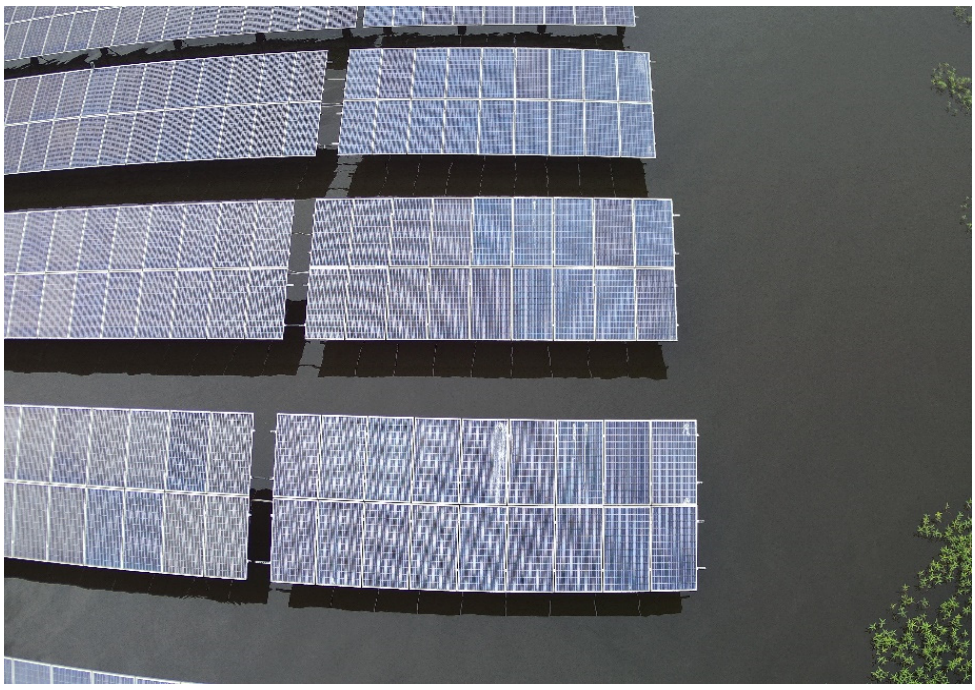
Figure 5. Flowchart of defect identification of photovoltaic modules.

1. Data processing stage

This experiment mainly uses the photovoltaic defect pictures collected by ourselves. The image data have a wide detection range, the environment is complex, and there are

various lighting conditions. This paper mainly cuts and marks the defective photovoltaic data, and the specific operations are as follows:

- Data set collection. The appropriate UAV equipment and camera were selected to collect image data from the photovoltaic site, and a total of 457 photovoltaic picture data from different locations and different illuminations were collected. Post-processing of the captured photos, including flipping, rotating, adjusting color, adjusting contrast, and other operations, was carried out to obtain better image effects. The photovoltaic module pictures collected are shown in Figure 6. The main type of surface defect is bird droppings.
- Data set production. After screening the collected photovoltaic image data, a total of 433 images with photovoltaic module defects were selected. Due to the small number of images collected, it is necessary to preprocess the photovoltaic data and expand the photovoltaic data set. In this paper, batch cropping of these photovoltaic data was carried out, and each image was cut into quarters on average, and a total of 1729 photovoltaic module images were obtained.
- Data set labeling. The LabelImg tool was used to screen and mark 1729 pictures after cutting. The marked pictures of photovoltaic modules are shown in Figure 7. A total of 842 defective photovoltaic module images were screened out, and a category of bird droppings-related defects was created to mark all the bird droppings-related defects in the photovoltaic module images. The program was written to modify the format of all marked data sets and randomly divide them according to the proportion of the training set 80%, test set 10%, and verification set 10%, of which 674 were in the training set, 84 were in the verification set, and 84 were in the test set.



(**a**)

Figure 6. *Cont*.



(**b**)

Figure 6. Pictures and defects in photovoltaic modules. (**a**) Photovoltaic module picture. (**b**) Bird droppings in photovoltaic modules.

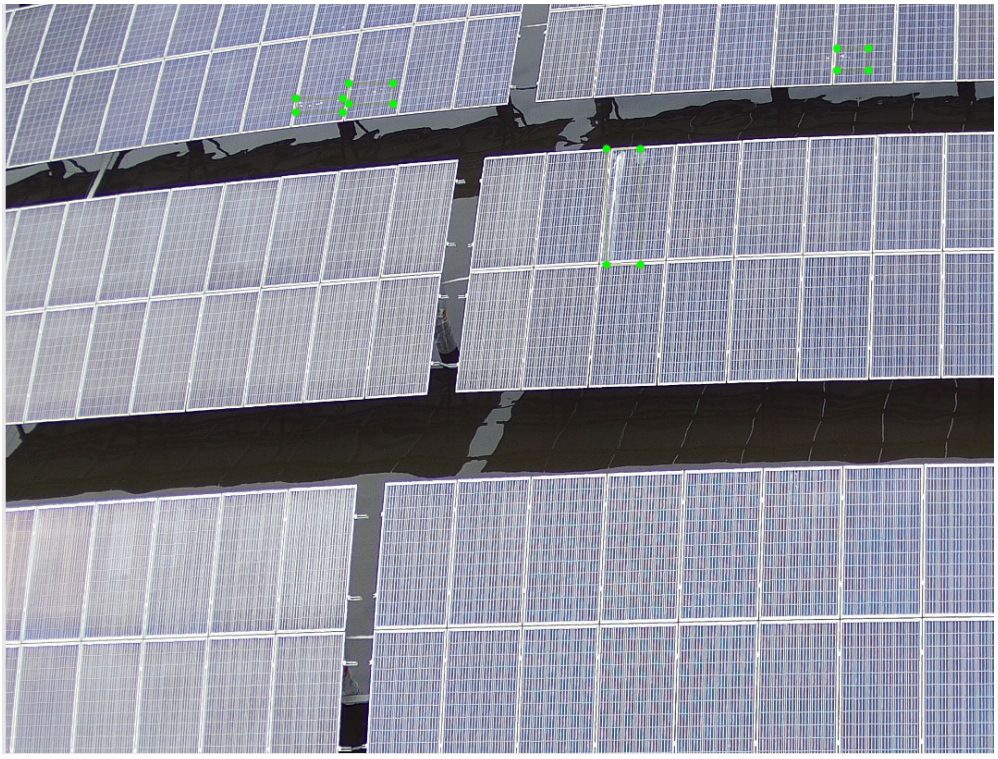


Figure 7. Picture of photovoltaic module marking.

2. Model training stage

The data set created above is used for training. The operations of the data in training are as follows:

- After the Mosaic operation at the input end, images were clipped, scaled, and then randomly distributed for Mosaic to enrich the data set.

- After data are output from the input, a feature map of size 320 × 320 × 64 is output through a layer of convolution.
- After several CBS and CSP structures, the feature mapping of the base layer is divided into two parts first, and then they are combined through the cross-stage hierarchy to output the feature maps of different scales, respectively.
- Through a layer of Coordinate Attention, the location information is embedded into the feature vector so that the network can obtain a wider range of features.
- Using the SPPF structure to serialize inputs through multiple MaxPool layers of 5 × 5 size and concatenate outputs.
- After the FPN and PAN stages, a small target detection layer is added in the modified feature fusion stage, and the feature maps of four scales are output.
- The upper layer output features are finally input to the detection head for specific image detection, and the target box is filtered by NMS to obtain the prediction box.

Finally, data such as Loss, results, and mAP are calculated to save the images detected during training and the training model. The specific structure of the improved CA-YOLOv5s algorithm is shown in Figure 8.

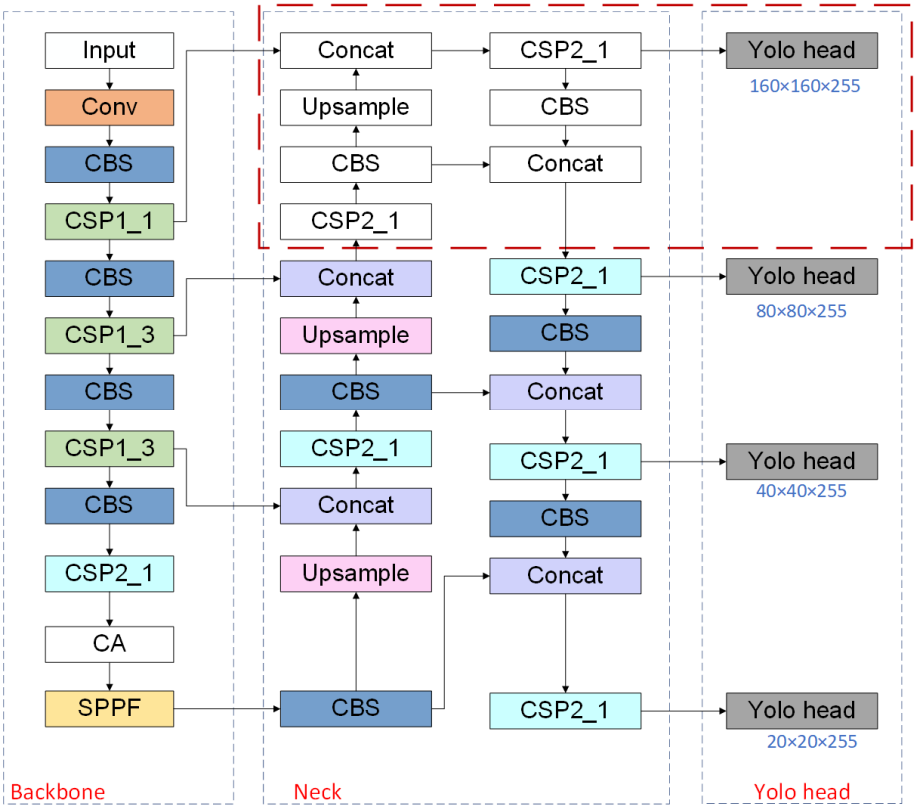


Figure 8. Network structure diagram of the improved CA-YOLOv5s algorithm.

3. Model testing stage

To evaluate the quality of the training model, we use the marked validation set pictures and the trained training model for testing. The test results output the validation pictures with the bird droppings-related defects prediction box, the Precision rate, the Recall rate, and other information, and use the evaluation indicators such as mAP, Recall, and Precision to evaluate the quality of the model. The obtained evaluation results are excellent models for model inference. The data set is used as the test set for pictures that are not labeled, and the photovoltaic module bird droppings-related defect detection is carried out on these pictures. The input model is used for inference, defect categories, and prediction boxes are drawn on the test set pictures, and redundant prediction boxes are eliminated by

non-maximum suppression. Finally, the test picture with the bird droppings-related defect prediction box is output.

5. Results

5.1. Evaluation Index

In the data analysis stage, mAP (mean average precision), Recall, Precision, and other parameters were used to evaluate the model effect, where Precision represents the probability of the sample being positive among all the predicted positive samples, and its formula is shown in (7); Recall refers to the probability of being predicted as a positive sample in an actual positive sample, and its formula is shown in (8); AP is the average accuracy; that is, the Precision value on the PR curve is calculated on average, and the calculation formula is shown in (9). mAP means to calculate the mean value of AP, and its calculation formula is shown in (10). The larger the mAP value, the more accurate the fitting between the detection box and the real box, and the more reliable the result.

$$P = \frac{TP}{TP + FP} \tag{7}$$

$$R = \frac{TP}{TP + FN} \tag{8}$$

$$AP = \int_0^1 P(R)dR \tag{9}$$

$$mAP = \frac{\sum_{i=1}^{N} AP_i}{N} \tag{10}$$

where P represents the precious rate, R stands for recall rate, TP represents the number of samples with correct positive examples, FP represents a positive example of error, FN represents the number of samples of counterexamples of errors, AP stands for average accuracy, and N indicates the total number of target categories to be detected.

5.2. Ablation Study

To verify the effectiveness of the improved algorithm by adding Coordinate Attention and a small target detection layer to the YOLOv5s model structure, ablation experiments were conducted to evaluate the influence of different modules on the performance of the target detection algorithm under the same experimental conditions. YOLOv5s was selected as the basic algorithm in the ablation experiment, and the experiment was carried out in the processed PV module image data set. The input image resolution was set to 640 × 640, and the results of training 300 epochs are shown in Table 1.

Table 1. Ablation results.

Model	Coordinate Attention	The Small Target Detection Layer	P	R	mAP@.5	Model Size
YOLOv5s	×	×	0.836	0.859	0.868	14.4 MB
YOLOv5s+CA	√	×	0.924	0.864	0.917	15.4 MB
YOLOv5s+Detect	×	√	0.874	0.87	0.914	16.3 MB
CA-YOLOv5s	√	√	0.893	0.864	0.92	16.1 MB

As can be seen from the results in the table, the Precious value of the original YOLOv5s algorithm on the photovoltaic data set is 83.6%, the Recall value is 85.9%, and the mAP value is 86.8%. After the Coordinate Attention is added, the Precious value is 92.4%, the Recall value is 86.4%, the mAP value is 91.7%, the Precious value is increased by 8.8%, the Recall value is increased by 0.5%, the mAP value is increased by 4.9%, and the algorithm detection effect is improved. After the small target detection layer is added, the Precious value is

87.4%, the Recall value is 87%, the mAP value is 91.4%, the Precious value is increased by 3.8%, the Recall value is increased by 1.1%, the mAP value is increased by 4.6%, and the detection effect of small targets is improved. Finally, the two modules are applied to the network model at the same time, and the Precious value of the improved algorithm is 89.3%, the Recall value is 86.4%, and the mAP value is 92%. Compared with the original algorithm, the Precious value is increased by 5.7%, the Recall value is increased by 0.5%, and the mAP value is increased by 5.2%, which proves that the algorithm improvement is effective. The mAP values of each experimental result are shown in Figure 9.

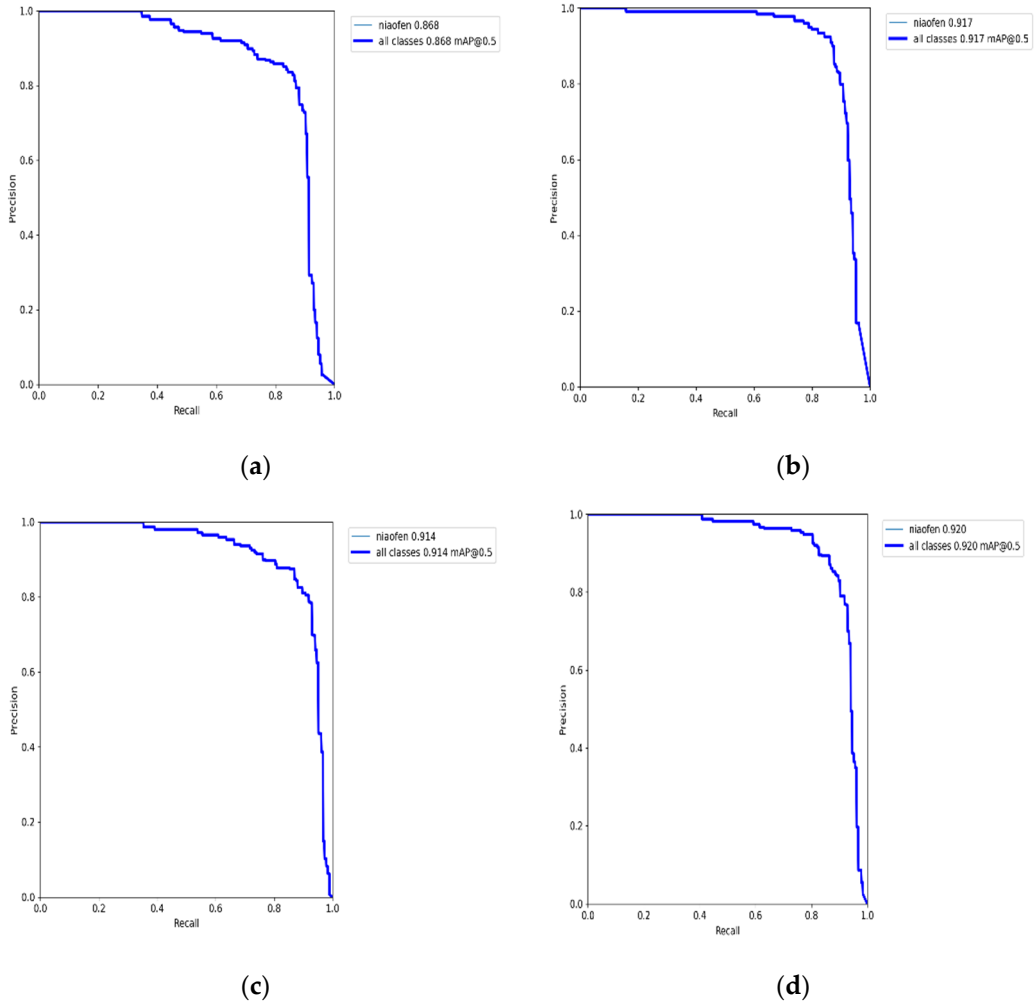


Figure 9. Comparison of mAP values of experiments: (**a**) mAP result of the YOLOv5s algorithm; (**b**) mAP result of YOLOv5s+CA; (**c**) mAP result of YOLOv5s+Detect; (**d**) mAP result of the CA-YOLOv5s algorithm.

6. Discussion

6.1. Comparison Experiment

To better verify the performance of the proposed algorithm, the same data set and the same training conditions were used to test the detection effects of different algorithms, and the comparison experiments were conducted with other mainstream algorithms. The results are shown in Table 2.

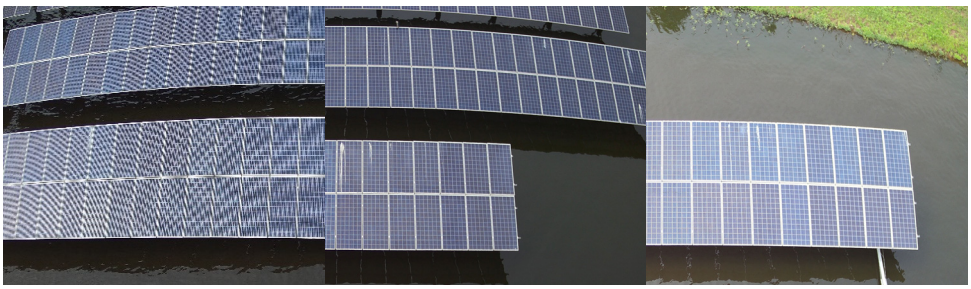
Table 2. The results were compared with other experiments.

Model	Parameter	P	R	mAP@.5	Model Size
YOLOv3	61497430	0.854	0.891	0.905	123.5 MB
YOLOv3-tiny	8666692	0.924	0.858	0.902	17.4 MB
YOLOv3-spp	62546518	0.89	0.837	0.898	125.6 MB
YOLOv5s	7012822	0.836	0.859	0.868	14.4 MB
YOLOv5m	20852934	0.892	0.853	0.905	42.2 MB
YOLOv5l	46108278	0.888	0.864	0.909	92.8 MB
YOLOv5n	1760518	0.895	0.788	0.873	3.9 MB
YOLOv5x	86173414	0.881	0.848	0.902	173.1 MB
CA-YOLOv5s	7599256	0.893	0.864	0.92	16.1 MB

As can be seen from the table, compared with other mainstream algorithms, the mAP value of the improved algorithm based on the YOLOv5s model in this paper is significantly higher than that of other similar YOLO series algorithms, and the number of parameters and model size are also lower than those of most target detection algorithms of the same type. Among them, compared with the YOLOv3 series algorithms, the improved algorithm in this paper has the highest detection accuracy and the smallest model volume. Compared with the highest YOLOv3 algorithm, the mAP value is increased by 1.5%, and the model volume is reduced by 7.47% compared with the lightest YOLOv3-tiny algorithm. Compared with YOLOv5 series algorithms, the improved algorithm in this paper has the highest detection accuracy, and the model volume is second only to YOLOv5s and YOLOv5n. The mAP value is 1.1% higher than that of the highest YOLOv5l algorithm, and the mAP value is 4.7% higher than that of the lightest YOLOv5n algorithm. In general, the improved algorithm proposed in this paper is relatively lightweight but also has high detection accuracy and achieves a good balance between model volume and detection efficiency, which further demonstrates the feasibility of the proposed algorithm for the detection of bird droppings-related defects in photovoltaic modules.

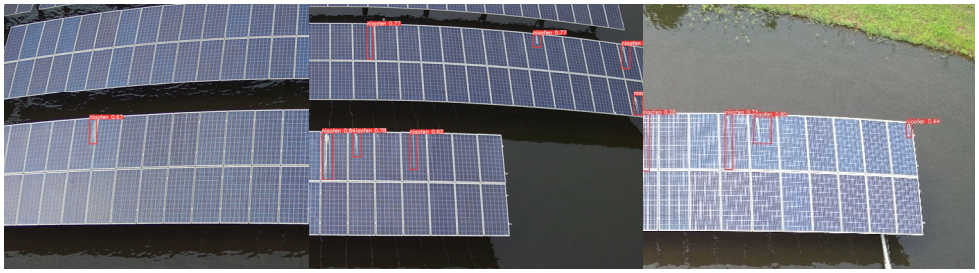
6.2. Algorithm Effect and Analysis

To display the detection effect of the improved algorithm in this paper more directly and further verify the effectiveness of the algorithm in this paper, a unified data set is used to detect the algorithm before and after the improvement. The detection results are shown in Figure 10.

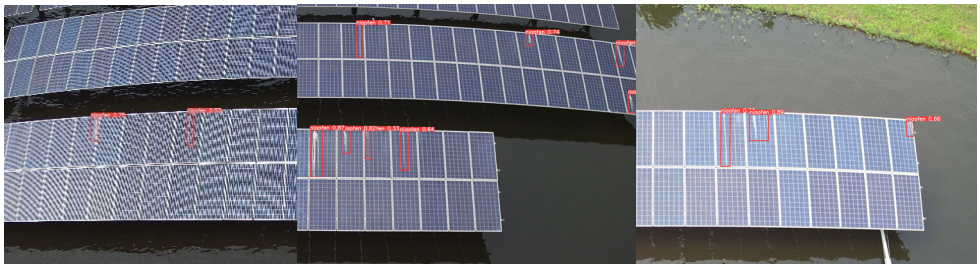


(**a**)

Figure 10. *Cont.*



(**b**)



(**c**)

Figure 10. Comparison of detection effects before and after algorithm improvement (**a**) original image; (**b**) detection result of the YOLOv5s algorithm; (**c**) detection result of the CA-YOLOv5s algorithm.

The figure shows the detection result of the model on the bird droppings-related defects in the photovoltaic module. The detection result contains a rectangular box with the location information of the bird droppings-related defects, the name of the defect class, and the confidence information. From the comparison and analysis of detection results between YOLOv5s and the improved algorithm in this paper, it can be seen that the detection results of the unimproved YOLOv5s algorithm have such problems as missing detection, false detection, low confidence, failure to completely frame targets, etc. After the improvement of the algorithm in this paper, these conditions have been improved to varying degrees.

Considering the influence of different lighting environments on the model detection results, this paper tested the photovoltaic module pictures collected under different lighting conditions, including morning light, noon light, and evening light. In the data set, image samples collected in the morning accounted for 71%, and the mAP value of the detection results was 91.7%; at noon, 12% of the samples were collected when the sunlight was strongest, and the mAP value of the detection results was 89.5%; at evening, 18% of the samples were collected, and the mAP value of the detection results was 91.2%. The test results are shown in Figure 11. In Figure 11, the left picture shows the detection results of the unimproved YOLOv5s algorithm, and the right picture shows the detection results of the improved YOLOv5s algorithm in this paper. Through comparison, it can be clearly seen that even under more extreme lighting conditions, such as direct sunlight, the improved algorithm in this paper can still accurately identify the bird droppings-related defect on the photovoltaic module. In the case of insufficient illumination, the improved algorithm can overcome the interference caused by insufficient illumination and still accurately locate the defect location. Under strong light, the algorithm also showed excellent performance and could accurately identify the bird droppings-related defects in the pictures. On the whole, the improved YOLOv5s algorithm in this paper has shown higher accuracy in detecting bird droppings-related defects of photovoltaic modules under different lighting environments, and its performance has been improved compared with the traditional YOLOv5s algorithm. This result provides a reference for the further development and application of photovoltaic module detection technology.

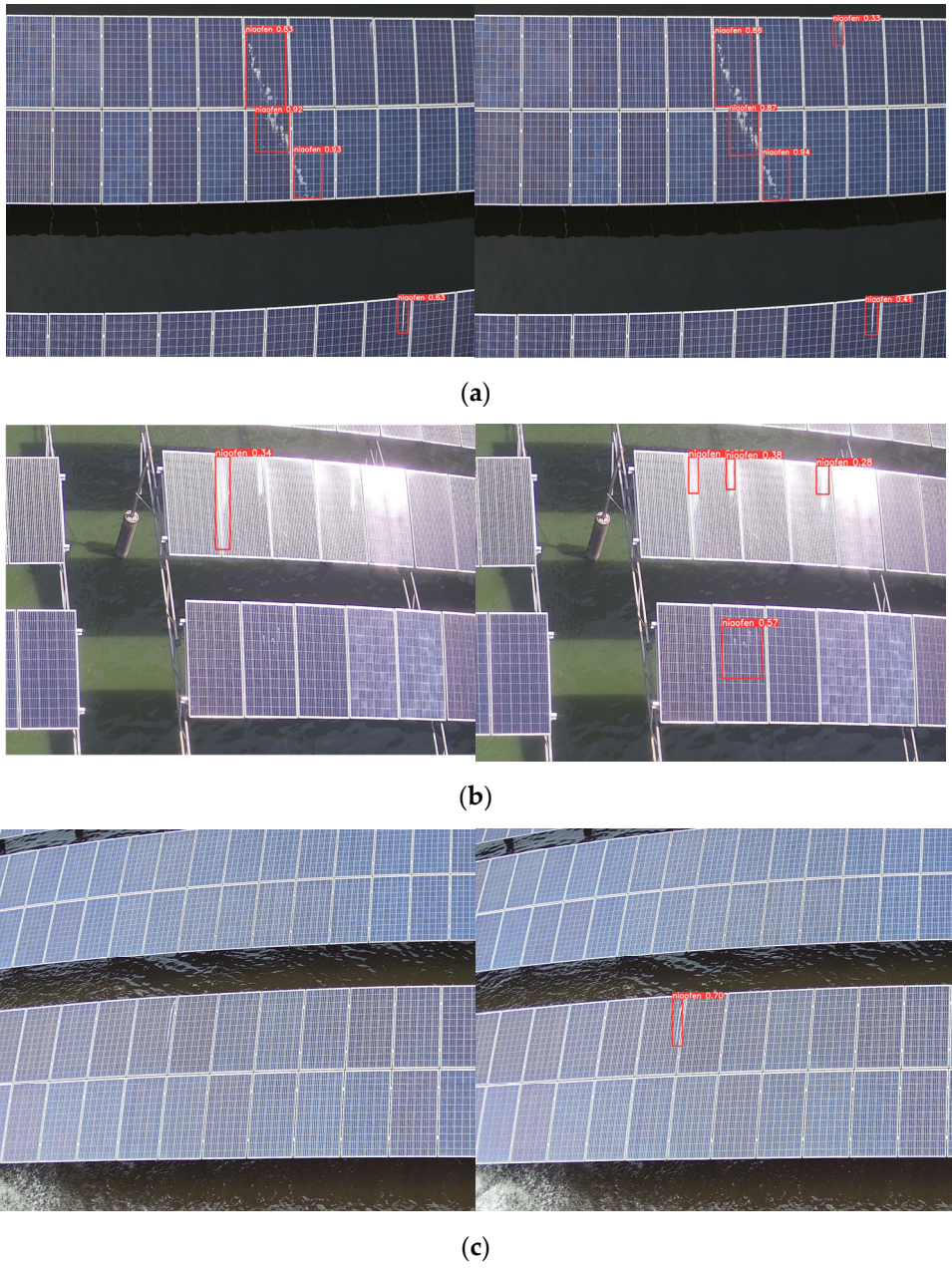


Figure 11. Detection results under different lighting conditions: (**a**) morning light detection results; (**b**) test results in direct sunlight at noon; (**c**) detection of evening light.

7. Conclusions

In this paper, an improved algorithm based on the YOLOv5s algorithm is proposed to improve the detection performance of small target defects of photovoltaic modules, aiming at the problems of missing detection, false detection, and misdetection of small target defects by the YOLOv5 algorithm.

Firstly, Coordinate Attention is inserted between the Backbone and Neck network to enhance the model's ability to accurately identify and extract key information from images. By introducing Coordinate Attention between the two networks, the model can pay more attention to those channels that are more critical to tasks when transmitting features. In the process of feature transmission, these key channels are given higher weights, and the model can locate the key information in the image more accurately, thus significantly improving its overall performance. Then, a small target detection layer is added, and the shallow network location information and deep network semantic information are fused, making

full use of the advantages of the two, which not only retain the location information and detailed features of the small target but also obtain enough semantic information to support its accurate identification, enhance the multi-scale learning ability of the model, and have a more efficient small target detection effect.

Experimental results show that compared with the original algorithm, the detection effect of the improved YOLOv5s algorithm has been improved, with mAP reaching 92% and an increase of 5.2%. Compared with other YOLO series algorithms, the improved algorithm can not only improve the detection accuracy but also effectively reduce the model volume, achieve a good balance between the model volume and detection effect, and have better detection results. This algorithm can be applied to our intelligent diagnosis platform for photovoltaic module image defects. By collecting image data from drones and processing them on the platform, we can identify photovoltaic module defects and achieve better detection results.

In future work, more in-depth research will be carried out in the aspects of network structure, volume, and parameter quantity to improve the detection speed while ensuring the detection accuracy of the algorithm. Future studies will also consider the detection effect of the model on other defects in PV modules and the detection effect under different weather conditions.

Author Contributions: L.L. and Q.L. designed and accomplished the study and wrote the paper. X.L. and W.W. gave advice. All authors have read and agreed to the published version of the manuscript.

Funding: National Natural Science Foundation of China (No. 52267008), and "Wind Farm Dynamic Operation and Maintenance System Development based on AR Smart Glasses", Jiangxi Province Science and Technology Department project.

Data Availability Statement: The data sets used and analyzed in the current study are available from the corresponding author upon reasonable request.

Conflicts of Interest: Author Wenbao Wu was employed by the company China Power Construction Group Jiangxi Power Construction Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Shi, Y.; Dai, F.; Yang, C. Defect detection of solar photovoltaic cells. *J. Electron. Meas. Instrum.* **2020**, *34*, 157–164.
2. 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; IEEE: New York, NY, USA, 2016; pp. 779–788.
3. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. SSD: Single shot multibox detector. In *Computer Vision–ECCV 2016: Proceedings of the 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016*; Springer: Cham, Switzerland, 2016; pp. 21–37.
4. 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, Mandi, India, 16–19 December 2017; IEEE: New York, NY, USA, 2014; pp. 580–587.
5. Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; IEEE: New York, NY, USA, 2015; pp. 1440–1448.
6. Ren, S.; He, K.; Girshick, R.; Sun, J. R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern. Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [CrossRef] [PubMed]
7. Kong, S.; Xu, Z.; Lin, X.; Zhang, C.; Jiang, G.; Zhang, C.; Wang, K. Infrared thermal imaging defect detection of photovoltaic modules based on improved YOLO v5 algorithm. *Infrared Technol.* **2023**, *45*, 974–981.
8. Chen, W.; Jia, X.; Zhu, X.; Ran, E.; Xie, H. Object detection in UAV aerial image based on DSM-YOLO v5. *Comput. Eng. Appl.* **2023**, *59*, 226–233.
9. Xie, L.; Zhu, W.; Xie, K.; Xiao, S. EL defect image detection model of photovoltaic cell based on improved YOLOv5s. *Foreign Electron. Meas. Technol.* **2023**, *42*, 93–102.
10. Yu, J.; Jia, Y. Improved small target Detection Algorithm of YOLOv5. *Comput. Eng. Appl.* **2019**, *59*, 201–207.
11. Guo, L.; Liu, Z. Defect detection of photovoltaic modules based on improved YOLOv5. *Adv. Laser Opto Electron.* **2019**, *60*, 148–156.
12. Wang, Y.; Gao, R.; Li, M.; Sun, Q.; Li, X.; Hu, X. Defects detection of photovoltaic modules based on YOLOv5 LiteX algorithm. *Acta Sol. Energy Sin.* **2023**, *44*, 101–108.

13. Zhou, Y.; Ye, H.; Wang, T.; Chang, M. Defects identification of photovoltaic modules based on multi-scale CNN. *Acta Sol. Energy Sin.* **2022**, *43*, 211–216.
14. Yan, H.; Dai, J.; Gong, X.; Wu, Y.; Wang, J. Defect detection of photovoltaic panel based on multi-source image fusion. *Infrared Technol.* **2023**, *45*, 488–497.
15. Tian, H.; Zhou, Q.; He, C. Photovoltaic modules defect detection based on multi-scale feature fusion. *Comput. Eng. Appl.* **2023**, *60*, 340–347.
16. Kisantal, M.; Wojna, Z.; Murawski, J.; Naruniec, J.; Cho, K. Augmentation for small object detection. *arXiv* **2019**, arXiv:1902.07296.
17. Chen, J.; Wang, X. Improved YOLOv5 UAV Aerial Image Dense small target Detection Algorithm. *Comput. Eng. Appl.* **2019**, *60*, 100–108.
18. Song, P.; Chen, H.; Gou, H. Improved UAV target Detection Algorithm based on YOLOv5s. *Comput. Eng. Appl.* **2019**, *59*, 108–116.
19. Luo, X.; Liu, Y.; Chu, G.; Pu, H. UAV Image Object Detection Algorithm Based on Improved YOLOv5. *Radio Eng.* **2023**, *53*, 1528–1535.
20. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature Pyramid Net works for Object Detection. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; IEEE: New York, NY, USA, 2017; pp. 936–944.
21. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path Aggregation Network for Instance Segmentation. In Proceedings of the 2018 IEEE/CVF IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; IEEE: New York, NY, USA, 2018; pp. 8759–8768.
22. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; IEEE: New York, NY, USA, 2018; pp. 7132–7141.
23. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European conference on computer vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
24. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient channel attention for deep convolutional neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 11534–11542.
25. Hou, Q.; Zhou, D.; Feng, J. Coordinate attention for efficient mobile network design. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 13713–13722.
26. Zhou, M.; Tang, Q.; Shi, T.; Luo, T.; Zhang, Z.; Xue, Y. Rail Surface Crack detection Algorithm based on improved YOLOv5s. *Chin. J. Liq. Cryst. Disp.* **2019**, *38*, 666–679. [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.