IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 15 July 2022, accepted 27 July 2022, date of publication 1 August 2022, date of current version 5 August 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3195901

RESEARCH ARTICLE

Solar Cell Surface Defect Detection Based on Improved YOLO v5

MENG ZHANG AND LIJU YIN
School of Electrical and Electronic Engineering, Shandong University of Technology, Zibo 255000, China
Corresponding author: Liju Yin (55660487@qq.com)

This work was supported in part by the Natural Science Foundation of Shandong Province, China, under Grant ZR2020MF127; and in part by the National Natural Science Foundation of China (NSFC) under Grant 62101310.

ABSTRACT A solar cell defect detection method with an improved YOLO v5 algorithm is proposed for the characteristics of the complex solar cell image background, variable defect morphology, and large-scale differences. First, the deformable convolution is incorporated into the CSP module to achieve an adaptive learning scale and perceptual field size; then, the feature extraction capability of the model is enhanced by introducing the ECA-Net attention mechanism; finally, the model network structure is improved and one tiny defect prediction head is added to improve the accuracy of target detection at different scales. To further optimize and improve the YOLO v5 algorithm, this paper uses Mosaic and MixUp fusion data enhancement, K-means++ clustering anchor box algorithm, and CIOU loss function to enhance the model performance. The experimental results show that the improved YOLO v5 algorithm achieves 89.64% mAP for the model trained on the solar cell EL image dataset, which is 7.85% higher than the mAP of the original algorithm, and the speed reaches 36.24 FPS, which can complete the solar cell defect detection task more accurately while meeting the real-time requirements.

INDEX TERMS Deep learning, YOLO v5, solar cell, defect detection, EL image.

I. INTRODUCTION

At the present stage, under the dual pressure of environmental pollution and the increasingly prominent traditional energy crisis, people have turned their attention to the development and utilization of new energy sources [1]. Due to the advantages of a wide range of applications, low cost, safety, and reliability, solar energy has become one of the mainstream new energy sources with high-speed development. Solar panels are important components of photovoltaic power generation, silicon crystal plates are fragile and fragile, and defects are easily produced by improper operation in production and installation [2], these defects cannot only affect the efficiency of solar cell power generation but also seriously threaten people's life and property safety [3]. Therefore, the study of solar cell defect detection methods is of great significance [4].

Electroluminescence (EL) imaging involves injecting a forward bias current into the PV module to put it in an excited state and then using a silicon charge-coupled device (CCD) or an InGaAs camera to capture the infrared light generated by the solar cell in the excited state for imaging. With the advantages of nondestructive and non-contact, electroluminescence imaging cannot only effectively detect tiny cracks, finger interruption, and other process defects that cannot be observed by conventional imaging systems, but also avoid blurring of imaging caused by lateral thermal propagation [5], [6]. Based on its excellent performance, electroluminescence imaging has become the main way of solar cell defect detection.

Traditional visual inspection requires operation and maintenance engineers to carry instruments to inspect solar cells one by one, which is a high workload, low efficiency, and overly dependent on the subjective experience of O&M engineers, and the inspection accuracy cannot be guaranteed. To automatically and accurately identify defects in images, researchers have proposed traditional

The associate editor coordinating the review of this manuscript and approving it for publication was Chuan Li.

computer vision based on manual feature extraction and classifiers [7]–[10]. Tsai *et al.* proposed a method for detecting defects in polysilicon solar cells based on the Fourier image reconstruction technique, which removes possible defects in EL images by setting the frequency components of line and strip defects to 0 [11]. Demant *et al.* proposed a classification recognition method based on local descriptors and support vector machines, which achieves effective detection of photoluminescence (PL) images and infrared (IR) images of small-grain silicon wafers [7]. However, traditional computer vision relies on manual extraction of descriptors, which requires a large number of parameter adjustments and has poor robustness and generalization capabilities.

In recent years, deep learning models represents by convolutional neural networks have been widely used in the fields of target detection, image classification, and semantic segmentation [12]–[15]. However, most convolutional neural networks are designed for natural scene images, and the direct application of mature deep learning models to detect surface defects in solar cell EL images has inapplicable problems, mainly because (1) solar cell defect detection is susceptible to interference from complex backgrounds (2) solar cells have diversity in the shape of the same class of defects (3) as network training progresses and downsampling continues, micro defect features such as cracks and finger interruption tend to disappear. The above problems make solar cell defect detection challenging. Deep learning-based solar cell defect detection is faced with the above difficulties. Therefore, two-order detection models based on the idea of candidate regions are widely used in the early stage of the research, such as the R-CNN series and R-FCN, etc. These algorithms have high detection accuracy but relatively slow detection speed. With the continuous efforts of scientific researchers, the first-order detection model represented by YOLO series algorithms is constantly improving the accuracy and speed of target detection, and the detection effect is increasing day by day.

The YOLO family of algorithms is a typical first-order target detection algorithm that uses an anchor box to combine classification and target localization. To date, five versions of the YOLO family of algorithms have been released, including YOLO v1, YOLO v2, and YOLO v3, all of which are proposed by the YOLO research team, and YOLO v3 is considered to be a milestone in the performance and speed of the YOLO family of algorithms with significant improvements, while YOLO v4 and YOLO v5 are released by different research teams. YOLO v5 detection model is smaller and faster than the other four generations of models and is fully implemented by Python (Pytorch), which is widely welcomed by the target detection field. It is noteworthy that researchers in different research directions have improved the original YOLO v5 model based on the characteristics of their detection targets, making the improved YOLO v5 algorithm excellent in many research areas. Among them, Li *et al.* proposed an improved YOLO v5 target detector for infrared images by adding the cross-stage-partial-connections (CSP) module to the improved model and introducing an improved attention module in the residual module, making the detection model reduce the network parameters while ensuring the detection accuracy [16]. Luo *et al.* proposed an improved YOLO v5-based aircraft target detection method to achieve a large improvement in detection accuracy and detection speed by incorporating centering and scale calibration, improving the cross-entropy loss function, and adding the CSandGlass module to the residual module [17]. Zhu *et al.* achieved a large improvement in detection accuracy and detection speed by replacing the original prediction head with the Transformer prediction head and increasing the number of prediction heads and incorporating an attention mechanism to form the TPH-YOLO v5 model, the improved model detection performance improved by about 7% over the original performance [18]. Kim *et al.* proposed an online copy-and-paste and hybrid data enhancement method to alleviate the class imbalance of the dataset during training and effectively improve the classification performance of the YOLO v5 detection model [19]. Mseddi *et al.* proposed a lightweight YOLOv5 detection model to detect visited networks and loop closures by introducing a Siamese network for binary classification at the neck [20].

After the above analysis and demonstration, the first-order detector YOLO v5 plays an important role in target detection with powerful real-time processing capability and low hardware requirements, which can be ported to mobile devices for real-time monitoring. Based on this, this paper proposes an improved YOLO v5 model for three different characteristics of solar cell surface defects, namely, cracks, black core, and finger interruption. In the design of the improved YOLO v5 network, deformable convolution is introduced into the CSP module to achieve effective extraction of defects of different sizes and shapes; and the ECA-Net attention module is introduced in the Neck part to achieve improved detection performance through cross-channel interaction; meanwhile, the model structure is optimized and the prediction head is added to achieve four-scale feature defect detection and improve the detection accuracy of tiny defects. Finally, the detection effect of the improved model in this paper is objectively evaluated through experiments such as ablation experiments and a comparison of mainstream methods, and the results show that the improved model improves the detection accuracy of solar cell defects while ensuring the real-time detection.

The main contributions of this paper can be summarized as follows:

1) Part of the conventional convolution in the CSP module is replaced with deformable convolution to realize the detection network adaptive learning of feature point receptive fields and effectively extract defect features of different sizes and shapes;

2) Adding the Neck part to the ECA-Net of the deep convolutional network to achieve considerable performance improvement by adding only a small number of parameters

through a local cross-information interaction strategy without dimensionality reduction.

3) Improving the network structure and optimizing the parameters of the YOLO v5 detection model, and increasing the number of prediction heads from 3 to 4. The new prediction heads use shallow features to achieve the detection of micro defects, making the improved model more applicable to solar cell surface defect detection.

4) Using the K-means++ algorithm for anchor box clustering, the improved clustering anchor box size is more in line with the data set, effectively reducing the impact of initial points on the clustering results and speeding up the convergence of network training; at the same time replacing the loss function of the detection network with the complete loss function (CIOU), making the improved prediction box more in line with the real box.

5) The mosaic data augmentation and Mixup data augmentation are proportionally fused for data expansion, which effectively reduces the memory loss of data augmentation while satisfying the demand for data expansion.

The remainder of this paper is organized as follows, Part II introduces the work related to solar cell defect detection, Part III introduces the improved YOLO v5 model framework and implementation details, Part IV conducts evaluation experiments and analyzes the experimental results accordingly, and Part V elaborates the conclusions.

II. RELATED WORK

A. CLUSTERING ANCHOR BOX ALGORITHM

The original YOLO v5 model uses the K-means algorithm to cluster the detection dataset, using the boundaries of the training set as a benchmark, and setting the feature mappings of three different sizes as three anchor boxes, using the anchor boxes as a priori boxes to assist in predicting the target sizes. However, the clustering centers of the K-means algorithm in the initial clustering are randomly selected, which may result in the initial clustering center being far away from the optimal clustering center location, which will not only affect the convergence speed of the model but also lead to poor detection results [21].

Therefore, to obtain an anchor box with a larger average intersection over union (avg-iou), in this paper, we use the K-means++ algorithm to perform multidimensional clustering of labeled target frames, taking one sample in the dataset as the initial cluster center, then calculate the distance between each sample and the existing cluster center, and categorize the sample into the category corresponding to the cluster center with the smallest distance from it, and calculate the probability of each sample being set as the next cluster center, select the sample with the largest probability value as the next center, and repeat the above process until no object is assigned to other clusters, and finally filter out K cluster centers. Although the K-means++ algorithm takes slightly more time to select the initial cluster centers than the K-means method, the convergence speed after the cluster centers are selected is faster than the original method, and the local optimum problem can be effectively avoided by using the improved method.

The K-means++ algorithm flow is shown in Table 1.

TABLE 1. K-means++ algorithm flow.

Input: width-height set S of all targets in the training set, clustering center K.
Output: Group K anchor box.
Step 1: Take a random value from S as the initial clustering center.
Step 2: Calculate the minimum iou distance $d(x)$ between all samples in S and the existing clustering centers and select the next clustering center C_i.
Step 3: Repeat step 2 until K clustering centers are found.
Step 4: for each sample x_i in the dataset, calculate its iou distance to the K cluster centers and assign it to the class corresponding to the cluster center with the smallest distance.
Step 5: recalculation of K clustering centers based on the division results.
Step 6: Repeat step 4 and step 5 until the cluster center position no longer changes and output the final cluster center.

B. MULTI-MODEL INTEGRATION METHODS IN TARGET DETECTION

Deep learning is a kind of collection of highly complex data modeling through multi-layer nonlinear transformation. Deep neural network can effectively realize multi-layer nonlinear transformation and expand in proportion according to the amount of training data, thus having strong flexibility. However, it will cause that deep neural network is very sensitive to the details of training data set. As a result, the weight sets of each training are different and the prediction results are different, which makes the stability of deep neural network poor. In order to solve this problem, researchers replace the original single-model training with multi-model training, and combine the training results of multiple models to make predictions, which significantly improves the stability of the deep neural network model.

In the post-processing of target detection, the prediction boundary box processing methods of different target detection models mainly include NMS, soft-NMS, GIOU, CIOU, and so on. The original YOLO v5 uses GIOU as a regression loss function, which can effectively solve the situation where the prediction box and the real box do not intersect. However, when the two boxes are contained or the union of length to width is different, the GIOU function cannot accurately judge the position relationship of the two boxes, resulting in a large error in model positioning. The improved YOLO v5 model uses CIOU [22] as a regression loss function. CIOU function can provide movement direction for boundary frames when they do not overlap. Meanwhile, the distance of the center point, overlap area, and aspect ratio of overlapping boundary frames are taken into account. Its performance is higher than other methods. The calculation formula of CIOU is as follows:

$$IoU = IoU - \left(\frac{\mathrm{p}^2\left(\mathrm{b}, \mathrm{b}^{\mathrm{gt}}\right)}{\mathrm{c}^2} + \alpha \mathrm{v} \right) \tag{1}$$

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

$$\alpha = \frac{v}{(1 - IoU) + v} \quad (3)$$

$$Loss_{CIoU} = 1 - IoU + \frac{p^2\left(b, b^{gt}\right)}{c^2} + \alpha v \quad (4)$$

In the above equation, b and b^{gt} denote the centroids of the prediction frame and the real frame, respectively, c denotes the square of the diagonal length of the minimum enclosing frame, p denotes the Euclidean distance between the two centroids of the prediction frame, and the real frame, α is the weight factor, v denotes the similarity of the aspect ratio, w^{gt} denotes the width of the labeled frame, h^{gt} denotes the height of the labeled frame, w denotes the width of the prediction frame, and h denotes the height of the prediction frame.

C. DATA EXPANSION

Deep neural networks need to be trained with a large amount of data to have good performance. Since PV plants are mostly located in the middle of nowhere, in remote and harsh environments, it is difficult to collect images, so data augmentation can effectively solve this problem. Offline data augmentation expands a large amount of data by augmentation factor calculation, but it occupies a large storage space; online data augmentation is to expand data within the deep learning framework during model training, which can effectively improve the training effect of the network by obtaining a large amount of data at the cost of very small memory consumption.

At present, online data augmentation methods are widely used in machine learning for data expansion, among which mosaic data augmentation and Mixup data augmentation are the most widely used. Mosaic data augmentation is performed by randomly selecting four images in the dataset for cropping and scaling operations, and then randomly arranging them into one image; MixUp augmentation is performed by randomly selecting two images in the dataset for weighted summation, and the labels of the images are weighted and summed accordingly. The improved YOLO v5 detection model combines the mosaic data enhancement and MixUp data enhancement methods for data expansion and uses the mosaic data enhancement in 50% probability and the MixUp data method in 25% probability during the training process, and only 280 training generations (70% of the total training generations) are used for data augmentation, which effectively reduces the memory consumption of data augmentation while meeting the data expansion requirements.

III. IMPROVED THE YOLO v5 MODEL

A. YOLO v5 OVERVIEW

YOLO v5 model has four different models, YOLO v5s, YOLO v5m, YOLO v5l, and YOLO v5x. The four models have depth and width parameters set as shown in Table 2. YOLO v5s is the model with the smallest network depth and width, while the other three models are products that deepen and expand based on YOLO v5s. The smaller the network model, the lower the performance requirements on mobile terminals and the easier the deployment. YOLO v5 uses the CSPDarknet53 architecture with the SPP layer as the backbone and PANet as the YOLO v5 prediction head.

TABLE 2. Parameter settings for depth and width of YOLO v5 models of four different models.

Parameters	YOLO v5s	YOLO v5m	YOLO v5l	YOLO v5x
Depth	0.33	0.67	1.00	1.33
Width	0.50	0.75	1.00	1.25

B. IMPROVING THE YOLO v5 MODEL

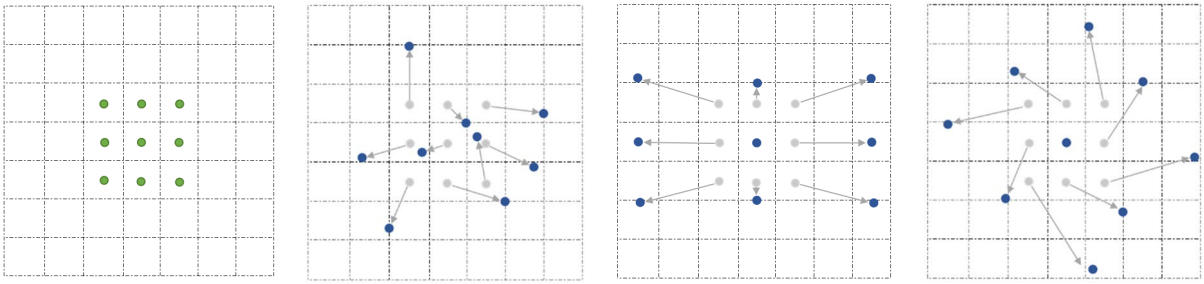
The original YOLO v5 target detection model has high accuracy in many target detection tasks, but it is not ideal for detecting objects with large differences in categories such as solar cell defects. Therefore, in this paper, the following parts of the detection model are improved according to the characteristics of solar cell defects.

1) DEFORMABLE CONVOLUTIONAL CSP MODULE

The original YOLO v5 detection model for solar cells defect inspection of rupture, solid black shapes such as changeable shortcomings because conventional convolution of rectangular structure can only be fixed sampling the input characteristic figure of the fixed position and feature points of the receptive field is fixed, but in different locations in the same feature layer corresponds to the different scale and shape of the target, Therefore, target detection has certain limitations. By introducing deformable convolution, the shortcoming of sampling of fixed rectangular structures can be overcome effectively and adaptive learning of scale and receptive field size can be realized.

Deformable convolution can improve the transformation modeling capability of the target by learning the offset from the previous feature mapping through parallel convolution layers [23], and the sampling points of the convolution kernel are shifted thus the sampling network is freely deformed to achieve sampling points focused on the target or region of interest. Figure 1 shows the schematic diagrams of sampling locations for conventional and deformable convolution. Figure (a) shows that conventional convolution only has a sampling network with a fixed rectangular structure, and figures (b), (c), and (d) show that the sampling points of each convolution kernel are increased with offsets, which can break through the limitations of conventional convolution and achieve random sampling near the current location.

Figure 2 shows the illustration of the 3 × 3 deformable convolutions, using the convolution layer to calculate the offset to the input feature map, the convolution kernel and the current convolution layer have the same spatial distribution and expansion, and the calculated offset and the input features have the same spatial resolution, and the number of input channels is 3 times the number of N convolution kernel sampling points, where N are the sampling point weights and 2N are the offsets in the x, y direction [24].



(a) Regular convolution (b)Deformable Convolution (c) Deformable convolution special case 1 (d) Deformable convolution special case 2

FIGURE 1. Schematic diagram of 3 × 3 conventional and deformable convolution.

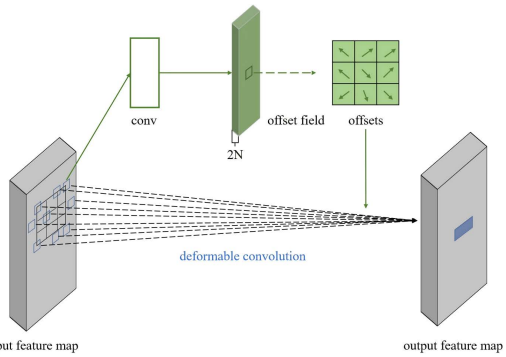


FIGURE 2. 3 × 3 deformable convolution illustration.

The conventional convolution operation is divided into ① sampling the input feature map using the regular grid R ② weighting the sampled points using the convolution kernel. Where R defines the size of the perceptual field and the dilation rate, as shown in (5), which defines a convolution kernel of size 3 × 3 and a dilation rate of 1.

$$R = \{(-1,-1),(-1,0),\cdots,(0,-1),(1,1)\} \tag{5}$$

For each position p_0 on the output feature map, the output value $y(p_0)$ is calculated by (6).

$$y(p_0) = \sum_{p_n \in R} w(p_n) \times x(p_0 + p_n) \tag{6}$$

where p_n denotes the unknown enumeration listed in R.

In the operation of deformable convolution, the input feature map F is convolutionally sampled using a regular network, and the set of sampled positions V is offset by combining the offset Δp_n with a weight Δm_n predicted for each sampled point, where N is the number of pixels in the grid, and the output value $y(p_0)$ for each position p_0 on the output feature map is.

$$y(p_0) = \sum_{p_n \in R} \boldsymbol{w}(p_n) \cdot x(p_0 + p_n + \Delta p_n) \cdot \Delta m_n \tag{7}$$

$$\Delta p_n = 1, 2, \cdots, N \tag{8}$$

Since the sampling point is sampled at $p_n + \Delta p_n$ after the offset, but the offset Δp_n is usually fractional and the pixel value at the location cannot be obtained accurately, the value of x is calculated by bilinear interpolation, as shown in (9)-(11).

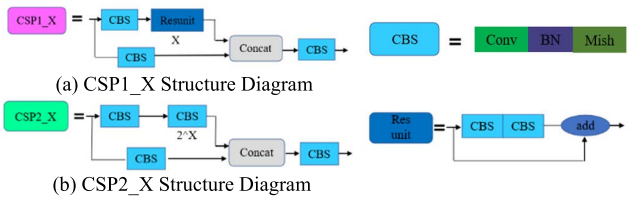


FIGURE 3. Original CSP structure diagram.

$$g(a,b) = \max(0, 1 - |a - b|) \tag{9}$$

$$G(q,p) = g(q_x, p_x) \times g(q_y, p_y) \tag{10}$$

$$x(p) = \sum_{q} G(q,p) \times x(q) \tag{11}$$

where $p = p_0 + p_n + \Delta p_n$ denotes the position after offset, $x(q)$ denotes the pixel values of the four adjacent integer coordinates of the feature map F, and $G(\cdot,\cdot)$ is the weight corresponding to each of the four coordinates. Through the above analysis, it can be obtained that the deformable convolution can realize the adaptive learning perceptual field, and the sampling position is closer to the shape and size of the defect itself, which is more conducive to the extraction of defect features.

Traditional CSP modules divide feature mapping into two branches to extract features [25] and then merge them by cross-stage hierarchy to ensure accuracy while reducing computation. Two CSP structures are designed in the YOLO v5 network, CSP1_X, and CSP2_X. CSP1_X is applied in the backbone network, which contains three convolutional layers and X residual unit modules, and its function is to improve the capability of the convolutional layers while reducing the computation. CSP2_X is applied in the neck network, and the difference between CSP2_X and CSP1 in the backbone is that the residual modules in CSP2_X are replaced by ordinary convolutional modules, whose role is to enhance the capability of network feature fusion.

In this paper, we improve two CSP modules by replacing the conventional convolution in the conventional CSP1_X and CSP2_X lower branch CBS modules with deformable convolution to ensure that the improved CSP module can

achieve accurate sampling of the size and shape of the target with only a small increase in computation.

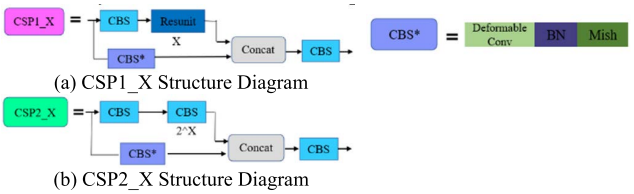


FIGURE 4. Improved CSP structure diagram.

2) EFFICIENT CHANNEL ATTENTION

In addition to the defect information, there is also a large amount of complex background information in the solar cell defect image. During the convolution operation, the iterative accumulation of complex background information forms a large amount of redundant information to overwhelm the defect information, resulting in poor detection accuracy. To solve the above problems, this paper introduces the Efficient channel attention (ECA-Net) in the YOLO v5 model and adds it to Part of the Neck of the YOLO v5s model for feature fusion to make the model's localization as well as target recognition more accurate [26], [27].

Attention mechanisms are used to obtain more critical information by focusing on the important regions of the input object. Mainstream attention mechanisms such as BAM, CBAM, SE-Net, and ECA-Net have been validated to lead to improved detection model performance [28]–[30]. It is worth noting that ECA-Net changes the status quo of obtaining detection performance improvement at the cost of increasing complexity, and only by adding a small number of parameters can achieve a considerable performance improvement. To improve the detection accuracy of the YOLO v5 model while maintaining detection efficiency and better embedding in mobile for engineering applications, the ECA-Net attention module is selected for improving the detection model.

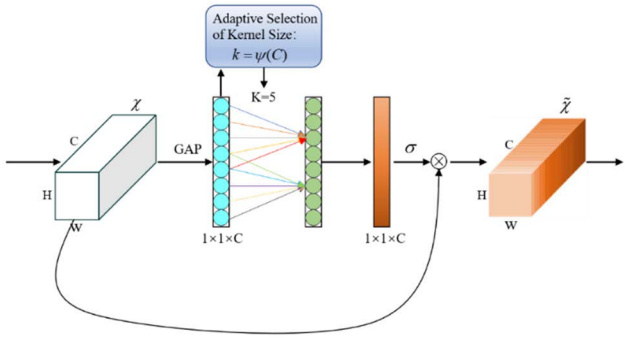


FIGURE 5. Schematic diagram of ECA-Net.

The ECA-Net attention structure is shown in Fig. 5. The ECA-Net attention module first performs global average pooling of the original input feature images, and on this basis, obtains local cross-channel interactions by fast 1D convolution of size k. After that, the channel weights are generated by the sigmoid function, and then the original input features are combined with the channel weights to obtain features with channel attention. The adaptive function of fast one-dimensional convolution of size k is shown in (12).

$$K = \varphi(C) = \left| \frac{lb(C)}{\gamma} + \frac{b}{\gamma} \right|_{odd} \tag{12}$$

In the defective target detection task, extracting different features of the object through different convolutional channels will result in too many training resources being devoted to non-defective regions, resulting in inefficient training of the network. To solve this problem, in this paper, ECA-Net attention is added to the feature fusion layer of the YOLO v5 target detection model, and the specific network structure is shown in Figure 6. By adding the ECA-Net attention module, different weights are assigned to different convolutional channels to highlight solar cell defective features, and the complexity of the model is significantly reduced by appropriate cross-channel interactions to avoid the impact of dimensionality reduction on the learning channels, and objective performance improvement is achieved by adding only a small number of parameters.

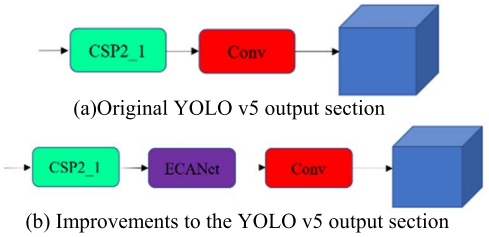


FIGURE 6. Schematic diagram of the output part of the YOLOv5 detection model.

3) MULTI-LEVEL FEATURE FUSION

There are obvious differences between the three types of defects of solar cells: crack, finger interruption, and black core, and the three types of defects are shown in Figure 7. The cracks are mainly fireworks shaped and varied in shape. Finger interruption mainly show stripes with the vertical black distribution. The black core mainly shows an irregular elliptical shape, which is a cluster of darker black areas relative to the background area. The crack and finger interruption are usually small, while the black core are usually large. In the network model, the role of the convolution layer is to extract the feature information in the input image, so the first part of convolution can output some larger feature mappings to capture small-sized defects, and the later convolution can form some smaller feature mappings to capture large-sized defects [31]. The three types of defects contained in the solar cell EL image dataset require different feature levels, but the original YOLO v5 detection model has only deep feature extraction networks, and these deep feature extraction networks are not sufficient to extract all the features of the three types of defects in the data.

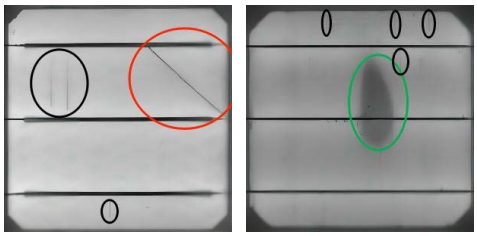


FIGURE 7. Schematic diagram of the three types of defects (black oval mark for finger interruption, red oval mark for crack, green oval mark for black core).

To address the above issues, this paper adds a prediction head for tiny defect detection to the original YOLO v5 target detection model, and the added prediction head is generated with low-level, high-resolution feature maps that are more sensitive to tiny defects. Meanwhile, to make the prediction head better detect target defects, the improved YOLO v5 detection model adds a CSP section and a CBS section, and further up-samples the fused feature map to generate a larger feature map for detecting tiny defects, and the improved detection model network structure is shown in Figure 8.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. DATA SET

In this paper, we use the solar cell EL image dataset for training, which has 2534 images with a size of 300 × 300. In the training, the dataset is randomly divided into a training set of 2281 images and a test set of 253 images according to the ratio of 9:1.

The labeling software LabelImg is used to label the defect location and category of the dataset in YOLO format, and there were three labeling categories, crack, finger interruption, and black core. In the marking process, the defects in the solar cell EL image are surrounded by rectangular boxes, which can reflect the specific location and category of the defects. The annotations are saved as XML files in PASCAL VOC format.

B. EXPERIMENTAL CONDITIONS AND PARAMETER SETTINGS

The experimental environment is Windows 10 operating system, using NVIDIA GeForce RTX 3070 graphic processing unit for computing, GPU size is 8GB, CPU configuration is Intel(R) Core(TM) i7-11800H @ 2.30GHz, CUDNN version is 11.0, Pytorch version is 1.7.1, and the python language environment is 3.6.0.

The hyperparameters of the network for this experiment are configured as follows: in the model training, the parameters are tuned using the Adma optimizer, the category confidence threshold of the target is set to 0.5, the initial learning rate is 0.01, the momentum is 0.937, the cosine learning rate decay is used, and the weight decay coefficient is set to 0.0005 to prevent data overfitting. In addition, the batch size is set to 16, and a total of 400 epochs are trained.

TABLE 3. Indicator parameters of the four models.

Model Type	Depth	Width	mAP(%)
YOLO v5s	0.33	0.50	81.79
YOLO v5m	0.67	0.75	82.64
YOLO v5l	1.00	1.00	82.33
YOLO v5x	1.33	1.25	80.92

C. EVALUATION INDICATORS

In this paper, recall (R), average precision (AP), mean average precision (mAP), and frames per second (FPS) are used to evaluate the performance of the improved detection model. The above evaluation metrics are calculated as follows.

$$AP = \int_0^1 P \mathrm{d}R \tag{13}$$

$$\mathrm{mAP} = \frac{\sum_{i=1}^{N} \mathrm{AP}_i}{N} \tag{14}$$

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{15}$$

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{16}$$

TP(true positive) indicates the number of defects detected in the defective image, TN(true negatives) indicates the number of defects detected in the defect-free image, FN(false negatives) indicates the number of defects detected in the defect-free image, and FP(false positive) indicates the number of defects detected in the defect-free image. The AP value is the area of the P-R curve. The mAP is obtained by averaging the average accuracy of the three defects of crack, finger interruption, and black core, the number of categories of defects in detection N = 3, and the larger the value of mAP, the better the detection of defects by the detection model and the higher the recognition accuracy.

D. EXPERIMENTAL RESULTS AND ANALYSIS

1) PERFORMANCE COMPARISON ANALYSIS OF THE BASE MODEL

This section explores the effects of the depth and width of the model on the mean average precision of solar cell defect detection. In deep learning models, usually the more complex the model structure and the deeper the depth the better the detection effect. However, small sample data may not show optimal detection in the most complex model. In order to design the most cost-effective model, four different models of YOLO v5s, YOLO v5m, YOLO v5l, and YOLO v5x are trained, and the metrics of the models are shown in Table 3.

The experimental results show that both YOLO v5m and YOLO v5l have better detection results than YOLO v5s, but the maximum mAP difference is only 0.85%. Considering the hardware requirements and the detection accuracy, we choose YOLO v5s as the base detection model.

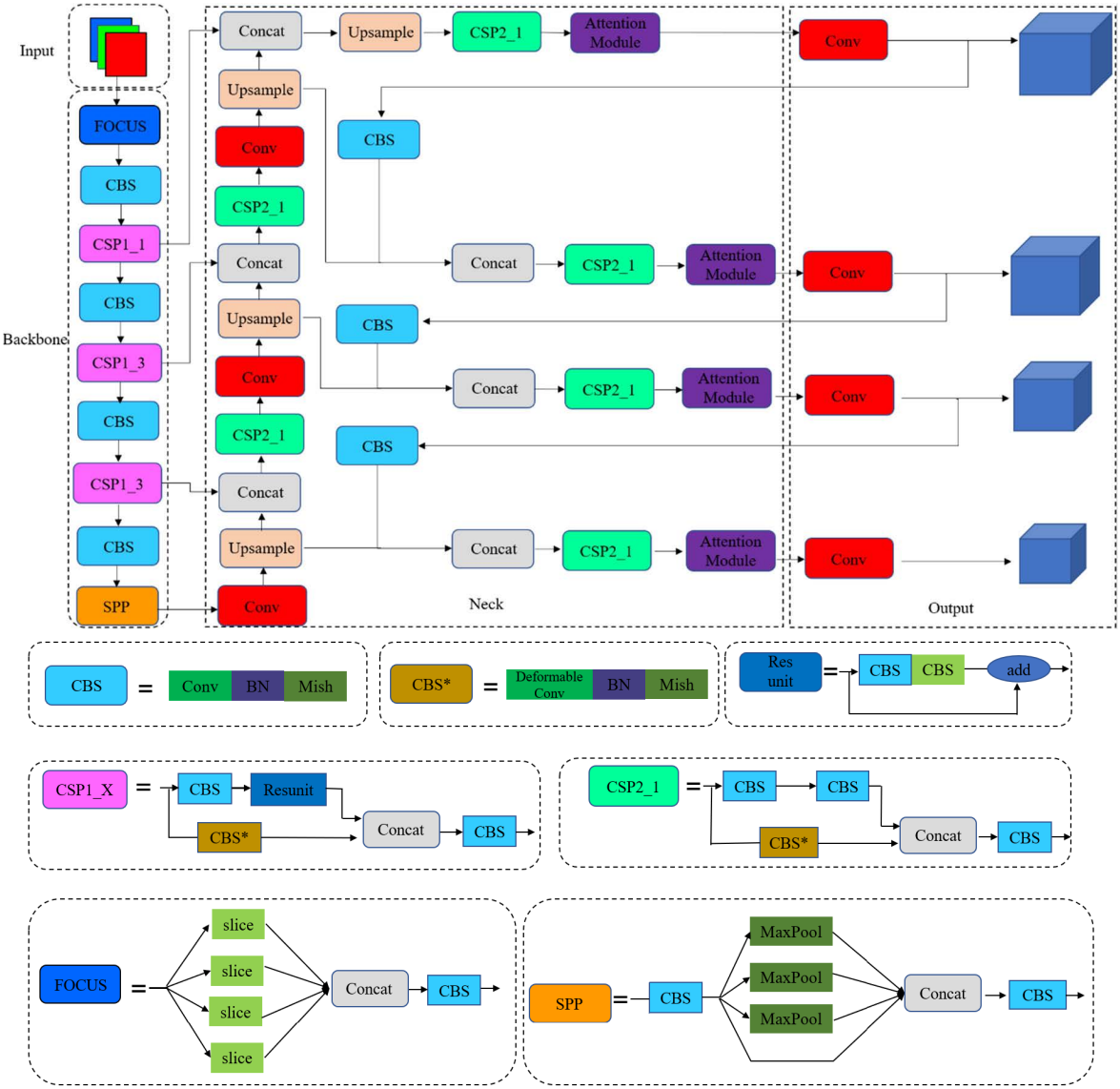


FIGURE 8. Improved YOLO v5s structure schematic.

2) PERFORMANCE COMPARISON ANALYSIS OF CLUSTERING ANCHOR BOX ALGORITHM

This section verifies the effectiveness of the improved clustering anchor box algorithm in improving the size matching of the prior frame and enhancing the accuracy of the network in solar cell defect detection through comparative experiments. The original YOLO v5s network model and the improved YOLO v5s network model are trained on the same dataset with the same number of epochs, and the experimental results are shown in Table 4.

From the analysis in Table 4, it can be obtained that the clustering anchor box algorithm is improved based on the original YOLO v5s detection model, the K-means algorithm is changed to the K-means++ algorithm, and the mAP of the model is improved by 0.28% and the FPS is improved by 0.02. The improved YOLO v5s detection model is optimized based on the original model by improving the network structure and adding the attention model. FPS and detection accuracy improved by 7.85% compared to the original YOLO v5_K-means with mAP. The improved YOLO v5s detection model changed the K-means algorithm to the K-means++ algorithm, and the mAP of the model improved by 0.43%

TABLE 4. Comparison of experimental test results.

Models	Clustering Methods	P%	R%	mAP(%)	FPS
Original YOLO v5s	K-means	87.19	91.34	81.79	44.21
Original YOLO v5s	K-means++	88.02	90.96	82.07	44.23
Improved YOLO v5s	K-means	92.93	97.52	89.21	36.21
Improved YOLO v5s	K-means++	93.53	97.04	89.64	36.24

and the FPS improved by 0.03. In summary, the improved clustering anchor box algorithm generates a more reasonable size of the prior anchor, which effectively improves the detection accuracy rate of the detection model for solar cell defects.

3) COMPARATIVE ANALYSIS OF ATTENTIONAL MECHANISMS

Figure 9 shows the comparison of the detection results of different attention mechanisms embedded in the Neck part of the YOLO v5s detection model, and the solar cell EL image dataset is used for training in this experiment. It can be seen from the table that SE-Net and ECA-Net attention mechanisms can improve the detection accuracy of the network after the introduction of the YOLO v5 model, but the detection accuracy decreases after the introduction of the CBAM attention mechanism compared with the original algorithm. the ECA-Net attention mechanism achieves the best result with a 3.82% improvement compared with the original algorithm. In summary, the ECA-Net attention mechanism is more suitable for target detection tasks with a high cross-merge ratio.

In this experiment, ECA-Net attention mechanism is embedded in different positions of YOLO v5s target detection model, and the network is trained with solar cell defect EL data set. The experimental results are shown in Table 5. The ECA-Net attention mechanism is embedded into YOLO v5s backbone network, and the detection accuracy is improved by 2.46% compared with the original model. When the ECA-Net attention mechanism is embedded in the multi-scale feature fusion, mAP@0.5 is 85.61%, and the number of model parameters increases by 1.17m, which has the most obvious improvement effect on the detection model. When the ECA-Net attention mechanism is embedded into the prediction end, mAP@0.5 is 84.21%, and the number of model parameters increased by 1.35m. By comparing the experimental results, it can be concluded that the ECA-Net attention mechanism embedded in the Neck part of YOLO v5s can locate and identify targets more accurately.

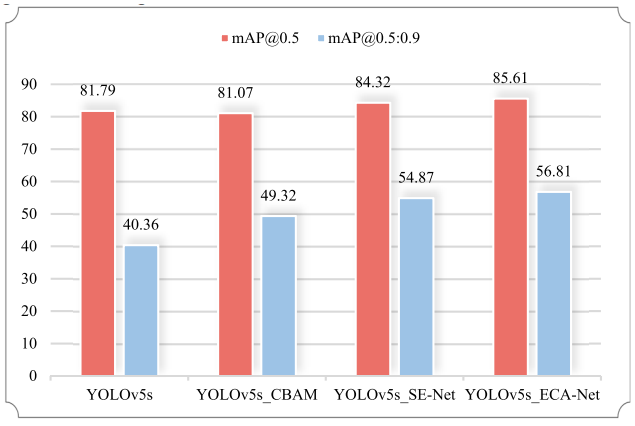


FIGURE 9. Comparison of various attention mechanisms under YOLOv5s.

TABLE 5. Detection results of different regions of the ECA-Net attention mechanism embedded in the network.

Models	Number of parameters	P%	R%	mAP@0.5(%)
YOLOv5s	8.47	87.19	91.34	81.79
YOLOv5s-backbone	9.73	89.65	93.76	84.77
YOLOv5s-neck	9.64	90.35	94.62	85.61
YOLOv5s-prediction	9.82	89.74	93.59	84.23

4) PERFORMANCE COMPARISON EXPERIMENTS BETWEEN THE IMPROVED ALGORITHM AND OTHER ALGORITHMS

In this section, the improved YOLO v5s detection model and four mainstream algorithms are selected to detect solar cell defects (including the first-order detection models YOLO v3, YOLO v4, SSD, and the second-order detection model Faster RCNN), and the detection results are multivariate analyzed, and the obtained data are shown in Table 6.

TABLE 6. Performance comparison results between the improved algorithm and other algorithms.

Models	Parameters/M	FPS	P%	R%	mAP @0.5(%)
YOLOv3	236	18.39	85.72	94.56	77.81
YOLOv4	245	14.55	88.64	95.19	81.39
SSD	100	37.53	79.80	96.42	74.36
Faster RCNN	108	6.54	86.22	96.04	83.42
Ours	10.94	36.24	93.53	97.35	89.64

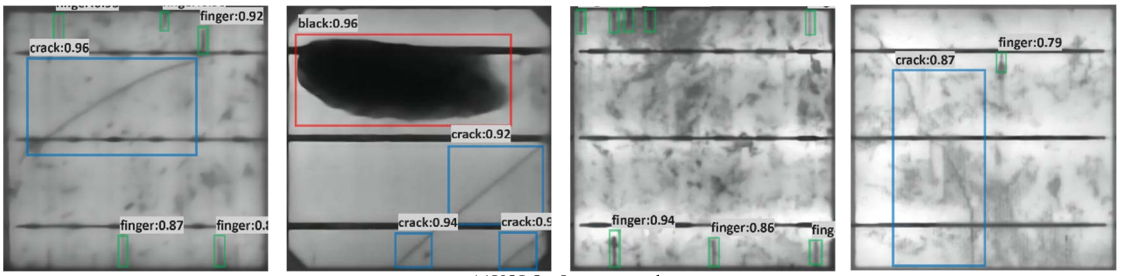
The improved YOLO v5s model has the best performance in the structural complexity of all models, and the model parameters are only 10.94% of SSD model parameters, and 89.64% of mAP@0.5 model parameters. Compared with other detection models, the accuracy is improved by at least 4.89%, and the FPS is 36.24, which can meet the actual engineering application. Compared with other algorithms, the second-order detection model Faster RCNN has a good detection effect, but its detection speed is slow and cannot meet the requirements of engineering practice. Finally, through comparative tests, it can be concluded that the improved YOLO v5s detection model can effectively detect solar cell defects and achieve better detection performance.

5) ABLATION EXPERIMENT

In order to visually observe the impact of different improved modules on the performance of the detection model, this section uses ablation experiments for verification. Specifically, the K-means++ clustering anchor box algorithm, hybrid data enhancement, improved CSP module, ECA-Net, and prediction head is added to the original YOLO v5s model, respectively, to ensure that the detection effects are compared under the same data set and the same number of training generations.

TABLE 7. Statistical results of ablation experiments.

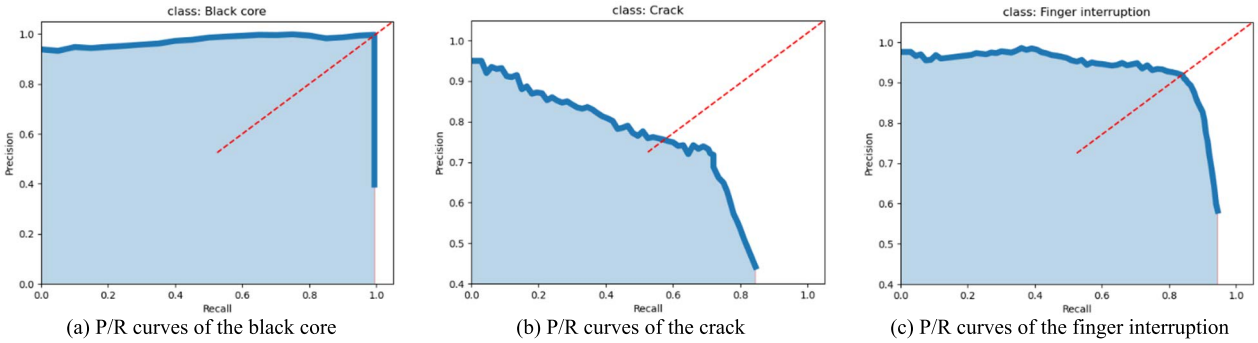
K-means++	Hybrid Data Enhancement	Improved CSP	ECA-Net	Add prediction head	mAP@0.5(%)	mAP Increase
					81.79	
√					82.37	0.28
	√				83.34	1.55
		√			82.69	0.9
			√		85.61	3.82
				√	84.46	2.67
√	√	√	√	√	89.64	7.85



(a)YOLO v5s assay results

(b)Improved YOLO v5s assay results

FIGURE 10. Improved YOLO v5s and the original YOLO v5s detection results comparison.



(a) P/R curves of the black core

(b) P/R curves of the crack

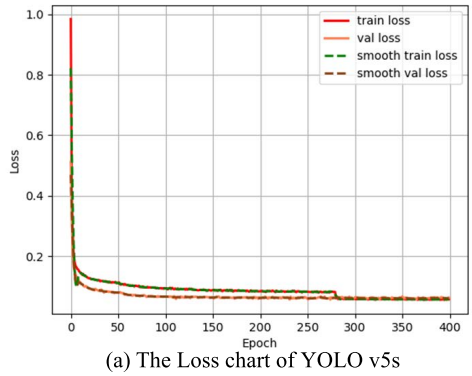
(c) P/R curves of the finger interruption

FIGURE 11. Improved P/R curve of YOLO v5s detection model.

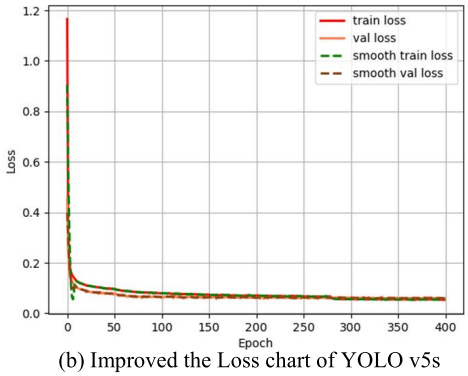
The experimental results are shown in Table 7. By adding the K-means ++ clustering anchor box algorithm, hybrid data enhancement, and improving the CSP module, ECA-Net, and prediction head, the accuracy index of defect detection is improved. When YOLO v5s integrated with five improved modules generated the final defect detection model, the detection accuracy is better than that of the five modules alone, mAP@0.5 increased by 7.85%.

6) ANALYSIS OF DETECTION RESULTS

In this section, different types of solar cell defect images are randomly selected for testing, and the detection results of the original YOLO v5s detection model and the improved YOLO v5s model are shown in Figure 10. Figure 10(a) shows the detection results of the original YOLO v5s model, and Figure 10(b) shows the detection results of the improved YOLO v5 model. From Fig. 10(a), it can be seen that YOLO v5s shows a missed detection with a low confidence level



(a) The Loss chart of YOLO v5s



(b) Improved the Loss chart of YOLO v5s

FIGURE 12. Loss plots of two YOLO v5s detection models.

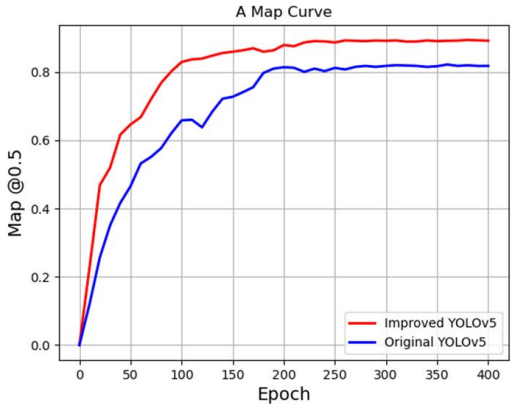


FIGURE 13. Comparison of two YOLO V5S detection models mAP@0.5.

under the interference of complex background. While in Figure 10(b), the improved YOLO v5s model in this paper is not disturbed by the non-uniform, complex background, and 100% of the defects in the picture are detected, the detection box is surrounded by a more accurate position, and has a higher confidence level. The comparison shows that the improved YOLO v5s model has more accurate detection results, can capture the key information of defects, and has excellent generalization performance.

The P/R curves of the improved YOLO v5s inspection model for three common defects in solar cells are shown in Figure 11. The horizontal and vertical axes indicate the recall and precision, respectively, and the mAP value for each type of defect is the area enclosed by the curve and the two axes. The point at which Recall = Precision is the equilibrium point, and the detection effect is proportional to the equilibrium point value. From Figure 11, it can be seen that the mAP value of solid black defects is close to 1 due to their large area and simple texture, etc. The detection effect is also good due to the fixed vertical black line shape of the broken fence, which has a single shape. However, the crack defect shape is diverse, the defect area is small, and the detection effect is relatively poor.

Figure 12 and 13 respectively show the loss curve and mAP@0.5 curve of the YOLO v5s detection model in the training process before and after improvement. In Figure 12, the loss value of the improved YOLO v5s network is lower, the inflection point occurred earlier than the original detection model and became smoother after 200 epochs, indicating that the improved YOLO v5s network could converge faster and more smoothly. It can be seen from Figure 13 that the improved YOLO v5s curve is above the original YOLO v5s curve, which means that the detection accuracy of the improved YOLO v5s network is higher than that of the original YOLO v5s network on the whole, and the learning curve of the improved network is smoother, indicating that the improved model has better stability. In general, the improved YOLO v5s detection model in this paper has a high accuracy and optimized the network detection performance, which can better meet the needs of solar cell defect applications.

V. CONCLUSION

In this paper, an improved YOLO v5 target detection model is proposed for the characteristics of solar cell defects, introducing deformable convolutional CSP module, ECA-Net attention mechanism, improved network structure and adding prediction head to enhance the feature extraction capability to achieve defect detection at different scales. Meanwhile, in order to optimize and improve the model, this paper uses mosaic and MixUp scale fusion data enhancement, K-means++ clustering anchor box algorithm, and invoking multi-model integration methods. The comparison experiments and ablation experiments show that the improved target detection model achieves an average accuracy of 89.64%, an improvement of 7.85% over the mAP of the original detection model, and a speed of 36.24 FPS, with significant enhancement effects. The next work direction is to reduce the complexity of the model and achieve high detection speed by processing the detection model network pruning and distillation to achieve a lighter improvement of the model.

REFERENCES

[1] International Energy Agency. (2021). *Renewables*. [Online]. Available: https://www.iea.org/reports/renewables-2021

[2] M. Köntges, I. Kunze, S. Kajari-Schröder, X. Breitenmoser, and B. Bjørneklett, "The risk of power loss in crystalline silicon based photovoltaic modules due to micro-cracks," *Sol. Energy Mater. Sol. Cells*, vol. 95, no. 4, pp. 1131–1137, Apr. 2011, doi: 10.1016/j.solmat.2010.10.034.

[3] P. Rupnowski and B. Sopori, "Strength of silicon wafers: Fracture mechanics approach," *Int. J. Fract.*, vol. 155, no. 1, pp. 67–74, Mar. 2009, doi: 10.1007/s10704-009-9324-9.

[4] B. Du, R. Yang, F. Wang, S. Huang, and Y. He, "Nondestructive inspection, testing and evaluation for Si-based, thin film and multi-junction solar cells: An overview," *Renew. Sustain. Energy Rev.*, vol. 78, pp. 1117–1151, Oct. 2017, doi: 10.1016/j.rser.2017.05.017.

[5] T. Fuyuki and A. Kitiyanan, "Photographic diagnosis of crystalline silicon solar cells utilizing electroluminescence," *Appl. Phys. A, Solids Surf.*, vol. 96, no. 1, pp. 189–196, Jul. 2009, doi: 10.1007/978-1-4419-9792-0_27.

[6] M. Dhimish and V. Holmes, "Solar cells micro crack detection technique using state-of-the-art electroluminescence imaging," *J. Sci., Adv. Mater. Devices*, vol. 4, no. 4, pp. 499–508, Dec. 2019, doi: 10.1016/j.jsamd.2019.10.004.

[7] M. Demant, T. Welschehold, M. Oswald, S. Bartsch, T. Brox, S. Schoenfelder, and S. Rein, "Microcracks in silicon wafers I: Inline detection and implications of crack morphology on wafer strength," *IEEE J. Photovolt.*, vol. 6, no. 1, pp. 126–135, Jan. 2016, doi: 10.1109/JPHOTOV.2015.2494692.

[8] B. Su, H. Chen, Y. Zhu, W. Liu, and K. Liu, "Classification of manufacturing defects in multicrystalline solar cells with novel feature descriptor," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 12, pp. 4675–4688, Dec. 2019, doi: 10.1109/TIM.2019.2900961.

[9] K. Firuzi, M. Vakilian, B. T. Phung, and T. R. Blackburn, "Partial discharges pattern recognition of transformer defect model by LBP HOG features," *IEEE Trans. Power Del.*, vol. 34, no. 2, pp. 542–550, Apr. 2019, doi: 10.1109/TPWRD.2018.2872820.

[10] Q. Luo, Y. Sun, P. Li, O. Simpson, L. Tian, and Y. He, "Generalized completed local binary patterns for time-efficient steel surface defect classification," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 3, pp. 667–679, Mar. 2019, doi: 10.1109/TIM.2018.2852918.

[11] D.-M. Tsai, S.-C. Wu, and W.-C. Li, "Defect detection of solar cells in electroluminescence images using Fourier image reconstruction," *Sol. Energy Mater. Sol. Cells.*, vol. 99, pp. 250–262, Apr. 2012, doi: 10.1016/j.solmat.2011.12.007.

[12] X. Qian, H. Zhang, C. Yang, Y. Wu, Z. He, Q.-E. Wu, and H. Zhang, "Micro-cracks detection of multicrystalline solar cell surface based on self-learning features and low-rank matrix recovery," *Sensor Rev.*, vol. 38, no. 3, pp. 360–368, Jun. 2018, doi: 10.1108/SR-08-2017-0166.

[13] D.-M. Tsai, G.-N. Li, W.-C. Li, and W.-Y. Chiu, "Defect detection in multi-crystal solar cells using clustering with uniformity measures," *Adv. Eng. Inform.*, vol. 29, no. 3, pp. 419–430, 2015, doi: 10.1016/j.aei.2015.01.014.

[14] H. Chen, H. Zhao, D. Han, and K. Liu, "Accurate and robust crack detection using steerable evidence filtering in electroluminescence images of solar cells," *Opt. Lasers Eng.*, vol. 118, pp. 22–33, Jul. 2019, doi: 10.1016/j.optlaseng.2019.01.016.

[15] S. A. Anwar and M. Z. Abdullah, "Micro-crack detection of multicrystalline solar cells featuring an improved anisotropic diffusion filter and image segmentation technique," *EURASIP J. Image Video Process.*, vol. 2014, no. 1, pp. 1–17, Mar. 2014, doi: 10.1109/ICCSCE.2012.6487131.

[16] S. Li, Y. Li, Y. Li, M. Li, and X. Xu, "YOLO-FIRI: Improved YOLOv5 for infrared image object detection," *IEEE Access*, vol. 9, pp. 141861–141875, 2021, doi: 10.1109/ACCESS.2021.3120870.

[17] S. Luo, J. Yu, Y. Xi, and X. Liao, "Aircraft target detection in remote sensing images based on improved YOLOv5," *IEEE Access*, vol. 10, pp. 5184–5192, 2022, doi: 10.1109/ACCESS.2022.3140876.

[18] X. Zhu, "TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV) Workshops*, 2021, pp. 2778–2788.

[19] J.-H. Kim, N. Kim, Y. W. Park, and C. S. Won, "Object detection and classification based on YOLO-V5 with improved maritime dataset," *J. Mar. Sci. Eng.*, vol. 10, no. 3, p. 377, Mar. 2022, doi: 10.3390/jmse10030377.

[20] W. S. Mseddi, M. A. Sedrine, and R. Attia, "YOLOv5 based visual localization for autonomous vehicles," in *Proc. 29th Eur. Signal Process. Conf. (EUSIPCO)*, 2021, pp. 746–750, doi: 10.23919/EUSIPCO54536.2021.9616354.

[21] Z. Wang, L. Wu, T. Li, and P. Shi, "A smoke detection model based on improved YOLOv5," *Mathematics*, vol. 10, no. 7, p. 1190, Apr. 2022, doi: 10.3390/math10071190.

[22] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, and D. Ren, "Distance-IoU loss: Faster and better learning for bounding box regression," in *Proc. AAAI Conf. Artif. Intell.*, Apr. 2020, vol. 34, no. 7, pp. 12993–13000, doi: 10.1609/aaai.v34i07.6999.

[23] J. Dai *et al.*, "Deformable convolutional networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 764–773. [Online]. Available: https://ieeexplore.ieee.org/document/8237351 doi: 10.1109/ICCV.2017.89.

[24] H. Fengqi, C. Ming, and F. Guofu, "Improved YOLO object detection algorithm based on deformable convolution," *Comput. Eng.*, vol. 47, no. 10, pp. 269–275 and 282, 2021.

[25] L. Zhu, X. Geng, Z. Li, and C. Liu, "Improving YOLOv5 with attention mechanism for detecting boulders from planetary images," *Remote Sens.*, vol. 13, no. 18, p. 3776, Sep. 2021, doi: 10.3390/rs13183776.

[26] Q. Wang, "ECA-Net: Efficient channel attention for deep convolutional neural networks," 2019, *arXiv:1910.03151*.

[27] R. Kadri, M. Tmar, B. Bouaziz, and F. Gargouri, "Alzheimer's disease detection using deep ECA-ResNet101 network with DCGAN," in *Proc. Int. Conf. Hybrid Intell. Syst.* Cham, Switzerland: Springer, 2021, doi: 10.1007/978-3-030-96305-7_35.

[28] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 3–19.

[29] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Sep. 2018, pp. 7132–7141.

[30] J. Park, S. Woo, J.-Y. Lee, and I. S. Kweon, "BAM: Bottleneck attention module," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, 2018, pp. 1–14, doi: 10.48550/arXiv.1807.06514.

[31] L. Cui, R. Ma, P. Lv, X. Jiang, Z. Gao, B. Zhou, and M. Xu, "MDSSD: Multi-scale deconvolutional single shot detector for small objects," 2018, *arXiv:1805.07009*.

MENG ZHANG received the B.E. degree in smart grid information engineering from the Shandong University of Technology, Zibo, China, in 2021, where he is currently pursuing the M.E. degree with the School of Electrical and Electronic Engineering. His research interests include deep learning-based image processing and object detection in computer vision.

LIJU YIN received the master's degree from China Agricultural University, Beijing, China, in 2002, and the Ph.D. degree from the Nanjing University of Science and Technology, Nanjing, China, in 2012. She is currently a Professor with the School of Electrical and Electronic Engineering, Shandong University of Technology, Zibo, China. Her research interests include photoelectric detection technology (low light level detection), photon counting imaging, machine vision, and intelligent information processing.