

Article



Research on Surface Defect Detection of Camera Module Lens Based on YOLOv5s-Small-Target

Gang He * D, Jianyun Zhou, Hu Yang D, Yuan Ning and Huatao Zou

Department of Mechanics, College of Mechanical & Electrical Engineering, Hohai University, Changzhou 213022, China

* Correspondence: hegang@hhu.edu.cn

Abstract: For the problem of low resolution of camera module lens surface defect image, small target and blurred defect details leading to low detection accuracy, a camera module lens surface defect detection algorithm YOLOv5s-Defect based on improved YOLOv5s is proposed. Firstly, to solve the problems arising from the anchor frame generated by the network through K-means clustering, the dynamic anchor frame structure DAFS is introduced in the input stage. Secondly, the SPP-D (Spatial Pyramid Pooling-Defect) improved from the SPP module is proposed. The SPP-D module is used to enhance the reuse rate of feature information in order to reduce the loss of feature information due to the maximum pooling of SPP modules. Then, the convolutional attention module is introduced to the network model of YOLOv5s, which is used to enhance the defective region features and suppress the background region features, thus improving the detection accuracy of small targets. Finally, the post-processing method of non-extreme value suppression is improved, and the improved method DIoU-NMS improves the detection accuracy of small targets in complex backgrounds. The experimental results show that the mean average precision mAP@0.5 of the YOLOv5s-Small-Target algorithm is 99.6%, 8.1% higher than that of the original YOLOv5s algorithm, the detection speed FPS is 80 f/s, and the model size is 18.7M. Compared with the traditional camera module lens surface defect detection methods, YOLOv5s-Small-Target can detect the type and location of lens surface defects more accurately and quickly, and has a smaller model volume, which is convenient for deployment in mobile terminals, meeting the demand for real-time and accuracy of camera module lens surface defect detection.

Keywords: deep learning; camera module; surface small target defect detection; YOLOv5; attention mechanism

1. Introduction

The camera module is an important component of digital products such as smartphones and personal computers [1]. In order to produce high quality and high resolution cameras, defect detection on the lens surface of camera modules is an essential process in the production process. Due to the large gap between the camera module lens surface defect features and the target features of the mainstream dataset, which are small target features, the detection accuracy of traditional machine vision algorithms on the camera module lens surface defect detection is not high. In the actual production process, the camera module factory needs to go through a series of processing procedures, such as FPC board cleaning, baking FPC board, wafer fixing, baking wafers, binding, focusing and other processes [2]. As the industrial production workshop is not the ideal dust-free environment, resulting in the module in the processing and installation process, there are often dust, lint and other foreign objects falling on the surface of the camera lens, resulting in the camera module lens surface white spots, white dots, scratches, hair filaments, foreign objects and dirt and other defects, seriously affecting the imaging quality. At present, in the actual production of enterprise, camera module defect detection mainly relies on manual



Citation: He, G.; Zhou, J.; Yang, H.; Ning, Y.; Zou, H. Research on Surface Defect Detection of Camera Module Lens Based on YOLOv5s-Small-Target. *Electronics* **2022**, *11*, 3189. https://doi.org/10.3390/ electronics11193189

Academic Editor: José L. Abellán

Received: 12 August 2022 Accepted: 22 September 2022 Published: 5 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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/). inspection or traditional machine vision inspection technology [3]. For manual inspection, it often produces low efficiency, low detection accuracy and high labor cost. The manual inspection facilities commonly used are shown as Figure 1, where Figure 1a shows an eight times magnifying glass and Figure 1b shows a chromaticity meter. As for the traditional machine vision inspection technology, it can meet the requirements of industrial reality in terms of accuracy and real-time. However, its adaptability to different features is less than satisfactory, and the feature extraction ability for deep features is limited, which makes it difficult to adapt to the complex and diverse defect requirements on the surface of camera module lenses [4]. Therefore, how to quickly and accurately detect the defects on the lens surface of camera modules is an urgent problem in the camera module production line.



Figure 1. Instruments for manual detection. (a) Eight times magnifying glass. (b) Chromaticity measuring instrument.

From the above analysis, it can be seen that camera module lens surface defect detection has high engineering significance and has become an important research topic in the field of defect detection. Traditional machine vision detection algorithms can only extract shallow features of the image, resulting in limited detection accuracy. Deep learning technology has excellent feature learning and feature expression capabilities, and can extract features layer by layer, taking the strengths of the other to complement the weaknesses of the other, effectively improving the accuracy of defect detection [5–7]. From the perspective of scientific research, there is still no universal automatic detection algorithm for camera module lens surface defect detection based on deep learning. Therefore, it is of high practical significance to study an algorithm for camera module lens surface defect detection both in the engineering field and in the academic field of research. In view of this, this paper carries out a study of a deep learning-based surface defect detection method for camera module lenses.

2. Related Works

Defect Detection based on Machine Learning: Chang, C.F. [4] proposed an automatic detection method for compact camera lenses using circular Hough transform, weighted Sobel filter and polar transform, and used a machine learning support vector machine method to obtain accurate detection results. To improve the accuracy and speed of optical lens image thresholding segmentation in optical lens defect detection, Cao Yu et al. [8] proposed a new particle swarm algorithm (PSO) and Otsu thresholding segmentation algorithm, which improves the PSO weight factor update strategy and the global search capability, and assigns the optimal position calculated to the Otsu algorithm, and finally achieves the threshold segmentation of optical lens images. In order to improve the defect detection accuracy of small size curved optical lenses, Pan, J.D. et al. [9] proposed a comprehensive defect detection system based on transmission streak deflection method, dark field illumination and light transmission, and the experimental results show that the proposed system can be applied to the actual mass production of small size curved optical lenses. For defect detection in electronic screens, Gao Yan et al. [10] designed an image processing-based screen defect detection algorithm. Based on the new edge detection algorithm, the defective part is detected by comparing the grayscale difference between the normal and defective regions, thus different types of defects in the screen can be located efficiently and accurately. Although the method basically meets the requirements

of industrial sites in terms of detection speed and accuracy, the setting of many parameters in the algorithm is highly dependent on manual experience, so it is difficult to be widely promoted in the industrial inspection field. To improve the detection of fabric defects, Deng Chao et al. [11] proposed a new algorithm based on edge detection. Fabric defects are detected as the edges of normal texture by using the texture edges generated by the defects and normal texture in the fabric image. Using the directionality of the Sobel operator, the horizontal and vertical gradients of the fabric defects are enhanced, respectively, and the horizontal and vertical gradients of the RGB image are computed for edge detection, and the final detection is performed by image fusion and binarization. However, when the fabric is wrinkled or the sample is not placed correctly, the detection accuracy will be greatly decreased.

The above analysis shows that most of the traditional machine vision algorithms have the following common problems: (1) the setting of parameters is highly dependent on manual definition, and the detection algorithm cannot extract deep semantic information of the image, which in turn limits the improvement of detection accuracy. (2) Traditional machine vision algorithms lack a common, unified detection framework, and it is often needed to combine multiple image processing algorithms to achieve accurate detection of the target. (3) If the defect type is changed, the detection algorithm needs to be redesigned, and the algorithm is poorly reusable, consuming too much manpower and material resources.

Defect Detection based on Deep Learning: With the industry's increasingly stringent requirements for defect detection accuracy and speed, more and more deep learning algorithms are being applied to the field of industrial product surface defect detection. Daniel W. et al. [12] conducted an earlier study on the use of convolutional neural networks for defect classification and recognition. This method passes the acquired image feature information into the backbone feature extraction network for processing to determine whether the image to be detected contains defects. To improve the surface quality of tiles, Xie, L.F. et al. [13] proposed an end-to-end CNN architecture called fused feature CNN (FFCNN). In addition, an attention mechanism is introduced to focus on the more representative parts and suppress the less important information. Experimental results show that the developed system is effective and efficient for magnetic tile surface defect detection. Aiming at the problems of low recognition rate and inaccurate localization of small defects on the surface of industrial aluminum products with traditional detection algorithms, Xiang Kuan et al. [14] proposed an improved deep learning network, Faster RCNN, to detect surface defects on 10 types of aluminum products. Experiments show that the average accuracy (mAP50) of the improved network for detecting surface defects of aluminum products is 91.20%, which is 16% better than the original Faster RCNN network, and its detection ability of small defects of aluminum products is stronger. However, it needs to be further improved in the detection's real-time performance.

Single-stage target detection algorithms are gradually being applied to the field of industrial product inspection to improve production efficiency even further. For example, Wu Tao et al. [15] used the K-means++ algorithm to determine the prior frame, and then built an improved lightweight network based on the YOLOV3 detection architecture to address the problems of low accuracy and slow detection rate of transmission line insulator defects. The experimental results show that the method improves the image detection speed of high-definition insulators and can complete insulator localization and defect detection. Fan, CS et al. [16] proposed a real-time detection algorithm based on improved YOLOv4 to address the problems of low detection accuracy and slow detection rate speed in cell phone lens surface defect detection. YOLOv4's cross-stage partial block and convolutional block attention modules are combined to introduce channel attention and spatial attention to learn the discriminative features of defects. Meanwhile, a new feature fusion network is being designed to combine shallow details with deep semantics. Finally, the proposed model is refined using a structural clipping strategy to improve detection speed without sacrificing accuracy. In comparison to the YOLOv4, this algorithm significantly improves the accuracy of defect detection and achieves real-time performance for industrial production. Guo

Lei et al. [17] proposed a small target detection algorithm based on improved YOLOv5 to address the problems of false detection, missed detection and insufficient feature extraction capability of small targets in target detection. The algorithm applies the Mosaic-8 data augmentation technique, which increases the network's capacity for small target detection by introducing a shallow feature map and modifying the loss function. In comparison to the original YOLOv5 method, the experimental results demonstrate that the algorithm has greater feature extraction ability and higher detection accuracy in small target detection. Zhang, R. [18] suggested a high-precision WTB surface defect detection model SOD-YOLO based on the UAV image analysis of YOLOv5 to address the problems of low accuracy of wind turbine blade surface defect detection layer, and the anchor was re-clustered. In order to reduce the loss of feature information for defects such as small target defects, the K-means algorithm and the CBAM attention mechanism are applied to each feature fusion layer. The experimental results demonstrate that the improved algorithm SOD-YOLO can detect the wind turbine blade surface defects quickly and effectively.

At present, the research of applying deep learning detection algorithm to the field of camera module lens surface defect detection has not been carried out deeply enough, and there are mainly problems in the following aspects.

(1) The problem of limited number of training samples and uneven distribution of sample data.

To obtain detection models with excellent performance, we need sufficient sample data as a driver [19]. However, in engineering practice, the acquisition of defective samples is not easy. In the actual production line, the images acquired by the inspection cameras are mostly qualified products, while the proportion of defective images valid for training is small, and the number of various types of sample data is unevenly distributed.

(2) Small target detection accuracy problem.

Current deep learning models perform well in mainstream datasets such as MS COCO dataset, Pascal VOC, ImageNet, [20] etc., but often fail to meet detection standards in industrial applications. Because most of the objects to be detected in mainstream datasets are large and medium targets, official network models can detect them more easily and achieve high detection accuracy. In practical industrial applications, the targets to be detected are mostly small objects, and the detection accuracy of the official network model is not ideal, so the network needs to be improved and optimized in a targeted way.

Since the detection targets of this topic are all small targets, which have certain requirements on detection accuracy, speed and industrial site deployment, this paper adopts YOLOv5s network model as the base network model. By improving and optimizing it, the algorithm improves the detection and recognition ability of small target defects.

3. Methodology

3.1. Construction of Camera Module Lens Surface Defect Dataset

The images of camera module lens surface defects dataset collected in this paper are mainly from industrial production workshops. Due to the production environment, processing process and other factors, the camera module lens surface produces six kinds of defects: white spots, white dots, scratches, hairy filaments, foreign matter and dirt. Among them, white spots and dirt belong to block defects, white spots belong to point defects, scratches and hairy filaments belong to strip defects, and foreign matter belongs to irregular shape defects. The appearance of each defect is shown in Figure 2.



Figure 2. Camera module lens surface defect appearance. (a) White spots. (b) White dots. (c) Scratch. (d) Hair filaments. (e) Foreign objects. (f) Dirt.

To solve the problem of small and unevenly distributed defect sample data, this paper adopts the image augmentation method to expand the defect samples. Image augmentation refers to a series of random changes to the training images to produce similar but different training samples, thus expanding the size of the dataset and improving the robustness and generalizability of the model. Commonly used image augmentation methods include cropping, flipping, panning, color gamut transformation, adding noise, rotation, etc.

PyQt5 was used to create an image expansion tool to expand the data of the defective sample images, and some of the defective sample data after image expansion are shown in Figure 3.



Figure 3. Cont.



Figure 3. Effect of image enhancement. (a) Original image. (b) Cutting. (c) Flip. (d) Transformation of color gamut. (e) Rotation. (f) Adding noise.

3.2. Network Model

For YOLOv5 target detection network, four different network models are officially given, namely YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x. Since the defect detection target in this paper has higher contrast and more obvious features, and the computing power of computer hardware equipment is limited, this paper uses the YOLOv5s network model to perform camera module lens surface defect detection.

The YOLOv5s network structure mainly consists of four parts: input, backbone network, neck network, and output [21–23], as shown in Figure 4, and the * in the figure represents the meaning of the product.



Figure 4. Network structure diagram of YOLOv5s.

3.3. Optimization of Lens Surface Defect Detection Network for YOLOv5s-Small-Target Camera Module

3.3.1. Improvement of YOLOv5s Input Stage: Introduction of Dynamic Anchor Boxes

YOLOv5s uses K-means clustering to generate anchor frames, using the bounding box of the training dataset as a benchmark, and setting three anchor frames under three different sizes of feature maps by FPN networks. However, the anchor frames generated by using K-means method have the following problems: First, the K-means algorithm has its own limitations; the K-Means algorithm is easily affected by the initial set values and outliers, which can cause the instability of clustering results. Second, the three different sizes of anchor frames are artificially set, and the targets to be detected in reality will not be uniformly distributed according to these three sizes. This can lead to some degree of error, for example, the target which is originally small object is included in the anchor box for detecting medium object, or the target which is originally medium object is included in the anchor box for detecting large object. This can lead to a large loss in the subsequent training process of the bounding box regression and increase the difficulty of model optimization. Third, the anchor frame is generated before the training phase, and the anchor frame is treated as a constant during the training phase. If the anchor box is not set reasonably, it will increase the model loss and convergence difficulty during the training phase.

In order to reduce model loss and improve the accuracy of small target detection, the concept of dynamic anchor frames is introduced in this paper. Specifically, by incorporating the DAFS (Dynamic Anchor Feature Selection) module to solve the problems arising from the generation of anchor frames by K-means clustering in the original YOLOv5s network.

The concept of DAFS model (Dynamic Anchor Feature Selection) was proposed by Li et al. [24] in 2019, which is based on the ARM module. The authors point out that in RefineDet, for any point on the feature map, using the optimized anchor frame of the ARM module as input to the ODM leads to a mismatch problem between the perceptual field of the point and the anchor frame, which instead reduces the detection capability of the model in some cases. Therefore, the authors propose to dynamically adjust the points on the feature map according to the shape size of the refined anchor frame on the detection part of the network (ODM) to reduce this mismatch problem. Additionally, the authors propose to replace TCB with Bidirectional Feature Fusion (BFF), which can fuse top-down and bottom-up bidirectional paths, enabling each layer to receive information from different layers above and below. The DAFS model is shown in Figure 5.



Figure 5. Structure of the DAFS model.

In this paper, the dynamic anchor frame module is added on the basis of YOLOv5 network structure. The initial anchor frame is firstly generated by clustering with K-means algorithm, then the ARM module is connected to the YOLOv5 backbone feature extraction network through the BFF structure, then the feature map information is fine-tuned, and finally the obtained dynamic refined anchor frames (Refined Anchors) are used as the a priori frames for defective sample training. The structure of the model after adding the dynamic anchor frames is shown in Figure 6.

3.3.2. Improving Spatial Pyramidal Pooling SPP Module

In order to reduce the feature information loss and improve the accuracy of small target detection, this paper improves the SPP module and proposes the SPP-D module, thus reducing the feature information loss caused by the maximum pooling of the SPP module. The structure of the module before and after the improvement is shown in Figure 7.



Figure 6. Model structure diagram of YOLOv5+DAFS.



Figure 7. Improved before and after module structure. (a) SPP module. (b) Improved SPP-D module.

Unlike the SPP module, the improved SPP-D module requires intensive feature pooling for each perceptual field before concatenating the features of four different perceptual fields. Among them, the feature fusion operation enhances the feature information reuse rate and reduces the feature information loss due to the maximum pooling operation. In YOLOv5s, the SPP module is embedded only after the feature layer with the smallest feature map size. When the size of the input image is the size, the SPP module is embedded after the special middle layer with the feature map size in order to improve the expressiveness and feature reuse of the features of size and size at the same time. In this paper, the SPP-D module is embedded in layers 13 and 17 of the YOLOv5s network, i.e., after the first and second concat operations in the network. The network structure diagram after adding the SPP-D module is shown in Figure 8.



Figure 8. Diagram of SPP-D module adding position.

3.3.3. Improvement of YOLOv5s Backbone Network: Incorporating the Attention Mechanism Module

(1) The idea of attention mechanism in deep learning

Attention mechanism is a survival mechanism in biology. By adding an attention mechanism module to a machine learning target detection task, the feature extraction ability of a specific target is enhanced, which in turn improves the detection performance of the network [25]. Commonly used attention mechanism modules are SENet [26] (Squeeze-and-Excitation Networks), CBAM [27] (Convolutional Block Attention Module), and CA [28] (Coordinate Attention).

(2) CBAM and CA Attention Mechanism Module

SENet is the basis of CBAM and CA attention mechanism modules. Before introducing CBAM and CA, it is necessary to elaborate on the principle of SENet attention mechanism module.

Figure 9 shows the network structure of SENet attention mechanism module. SENet mainly processes in the channel dimension, firstly, the input feature information is pooled globally averaged in two dimensions, width and height, to obtain a 1×1 spatial feature matrix, and then the weights of different channels are obtained through the processing of fully connected layer and ReLU function. The specific calculation process is shown in Equations (1) and (2):

$$z_{c} = F_{sq}(x_{c}) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} x_{c}(i, j)$$
(1)

$$\hat{z} = T_2(ReLU(T_1(z))) \tag{2}$$



Figure 9. Schematic diagram of SE attention mechanism module.

The SENet module implements the processing of global information through two steps: squeezing and excitation. Equation (1) represents the squeezing operation, and Equation (2) represents the processing of the ReLU activation function.

Finally, the weights obtained are normalized using the Sigmoid activation function, and each channel value in the original feature map is recalculated by weighting the product. As shown in Equation (3):

$$\hat{X} = X \cdot \sigma(\hat{z}) \tag{3}$$

In essence, the SENet attention mechanism module mainly processes the information of image channel dimensions and does not consider the location information. After the action of the SENet module, the network pays more attention to the channel features that are richer in semantic information and suppresses the less important channel features, thus improving the detection performance of the model.

The SENet attention mechanism mainly processes image feature information in the channel dimension. Unlike the SENet attention mechanism, the CA attention mechanism module embeds the location information into the channel attention based on the SENet module. the CA attention mechanism module can improve the feature representation capability of the lightweight network by transforming any intermediate feature tensor in the network and then outputting a feature tensor with the same size and enhanced representation. The execution of the CA attention mechanism module is shown in Figure 10.





As can be seen from Figure 10, the CA attention mechanism module adds attention to the location information on top of considering the channel attention information. First, the global average pooling of the input image in both width and height directions is performed to realize the feature aggregation in both width and height directions and obtain the feature maps in these two directions. The specific calculation method is shown in Equations (4) and (5):

$$z_c^h(h) = \frac{1}{W} \sum_{0 \le j \le W} x_c(h, j) \tag{4}$$

$$z_{c}^{w}(w) = \frac{1}{H} \sum_{0 \le j \le H} x_{c}(j, w)$$
(5)

Then, the obtained feature maps are stitched in both width and height directions, and then processed by the 2D convolution module to obtain the reduced-dimensional feature maps. Then, the feature map is further reduced by the 2D convolution operation after the batch normalization and nonlinear activation function processing. Finally, the reduced-dimensional feature map is fed into the Sigmoid function to obtain the feature map shaped as shown in Equation (6):

$$f = \delta(F_1(z^h, z^w)) \tag{6}$$

A convolution kernel of size 1×1 is used to convolve the feature map f to obtain the feature maps F_h and F_w with the same number of channels as the original one. Then, after the Sigmoid activation function, the attention weight g^h in the height and the attention

11 of 22

weight g^w in the width direction are obtained. The specific calculation formula is shown in Equations (7) and (8):

$$g^h = \sigma(F_h(f^h)) \tag{7}$$

$$g^w = \sigma(F_w(f^w)) \tag{8}$$

Finally, the weights g^h on the height and g^w on the width are used for the generation of the new feature maps. Specifically, the new feature map with attention weights is obtained by using the weighted product calculation. The specific calculation method is shown in Equation (9):

$$y_c(i,j) = x_c(i,j) \times g_c^h(i) \times g_c^w(j)$$
(9)

(3) CBAM attention mechanism

The specific structure of the CBAM convolutional attention mechanism module is shown in Figure 11. As can be seen from the figure, CBAM performs feature aggregation in both channel and spatial dimensions, and mainly consists of two sub-modules, the channel attention module and the spatial attention module, in which the red box is the channel attention mechanism module and the blue box is the spatial attention mechanism module [29].



Figure 11. Schematic diagram of CBAM attention mechanism module.

The execution process of CBAM convolutional attention mechanism is as follows: firstly, the input image information enters the channel attention mechanism module for convolutional processing to reduce the dimensionality of the feature map. Then, the normalized weights are obtained using batch normalization and Sigmoid function. Finally, the original feature map is recalibrated by weighted summation [30]. The specific calculation of the channel attention mechanism module is shown in Equation (10):

$$M_{c}(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) = \sigma(W_{1}(W_{0}(F_{avg}^{c})) + W_{1}(W_{0}(F_{max}^{c})))$$
(10)

After completing the feature determination on the channel dimension, the feature information goes to the spatial attention mechanism module for global maximum pooling (GMP) and global average pooling (GAP) processing. Then the feature map of dimension $1 \times 1 \times C/r$ is obtained by convolution and ReLU activation function module. Then, after a convolution process with dimension 1×1 , the feature map dimension returns to the original dimension. Finally, the feature maps in the channel dimension and the spatial dimension are weighted and combined to realize the recalibration of the feature maps in both channel and spatial dimensions. The specific calculation of the spatial attention mechanism module is shown in Equation (11):

$$M_{s}(F) = \sigma(f^{7 \times 7}([AvgPool(F); MaxPool(F)])) = \sigma(f^{7 \times 7}([F^{s}_{avg}; F^{s}_{max}]))$$
(11)

After processing by the CBAM convolutional attention mechanism module, the recalibrated feature map acquires weights in both channel and space dimensions, which significantly improves the association of features in both channel and space dimensions and is more conducive to extracting effective features of the target to be detected.

(4) Attention Module Add Location

In this paper, the Attention mechanism module is added to BottleNeck, the backbone feature extraction network of YOLOv5s, as shown in Figure 12, and the * in the figure represents the meaning of the product. After the Attention mechanism module is added, all C3-n modules in the backbone feature extraction network will perform the weighted product operation, which leads to a significant improvement in the detection performance of the final trained network model for the target of interest.



Figure 12. Schematic diagram of the location of attentional mechanisms.

3.3.4. Improved Non-Extreme Value Suppression Post-Processing Method DIoU-NMS

In the post-processing process of target detection, Non-Maximum-Suppression (NMS) post-processing method is usually used for target frame screening [31,32]. However, this method has a drawback that NMS only analyzes by overlapping regions, so it is easy to produce false suppression for targets of small size, resulting in missed detection. In order to solve this problem, the post-processing method of non-extreme suppression is improved by borrowing the idea of DIoU loss function. The traditional NMS post-processing method is replaced by the improved DIoU-NMS post-processing method, where DIoU is the function Distance-IoU.

The DIoU loss function is defined by the following equation:

$$R_{DIoU} = \frac{\rho^2(b, b^{gt})}{l^2}$$
(12)

where *b* is the prediction frame; b^{gt} is the true frame. $\rho^2(b, b^{gt})$ is the distance between the centroids of the predicted and real boxes. *l* is the diagonal distance of the minimum outer rectangle of the 2 boxes.

Assuming that the model detects a candidate frame set as H_i , the definition of p_i for the DIoU-NMS method update formula for the prediction frame M with the highest category confidence is shown in Equation (13):

$$p_{i} = \begin{cases} p_{i}, & IoU - R_{DIoU}(M, H_{i}) < \varepsilon \\ 0, & IoU - R_{DIoU}(M, H_{i}) \ge \varepsilon \end{cases}$$
(13)

where $R_{DIoU}(M, H_i)$ is the value of R_{DIoU} with respect to M and H_i . ε is the threshold value set manually by the NMS operation. p_i is the classification score for different defect categories. *IoU* is the intersection ratio and *i* is the number of anchor boxes for each grid.

The DIoU-NMS method takes into account the distance, overlap area and aspect ratio between the prediction frame and the real frame, determines that the two prediction frames with more distant centroid distance may be located on different detection objects, and then combines the intersection ratio of the prediction frame and the real frame with the centroid distance. On the one hand, the IoU loss can be optimized, and on the other hand, the learning of guideline centroids can return the prediction frames more accurately.

The structure diagram of the improved YOLOv5s-Small-Target network model is shown in Figure 13, and the * in the figure represents the meaning of the product.



Figure 13. Structure diagram of the improved YOLOv5s-Small-Target network model.

4. Experimental Results and Analysis

The camera module lens surface defect detection dataset produced in this paper contains 12,000 defect sample images, and these defect sample images are randomly divided into training set, validation set and test set in the ratio of 7:2:1, including 8400 images in the training set, 2400 images in the validation set and 1200 images in the test set. In order to comprehensively verify the testing effectiveness of the three YOLOv5 improvement strategies used in this paper, ablation experiments are conducted on the produced camera module lens surface defect dataset to judge the actual effectiveness of each improvement point. In order to make the model converge as much as possible, the number of iterations was set to 2000. In order to increase the training speed as much as possible and also combine with the computer hardware configuration, after several attempts, the batch size was set to 8. In order to balance the training speed and the quality of the training sample images, the input image size was set to 640×640 . After repeated debugging with several training sessions, the initial learning rate Ir was set to 0.01 and the momentum was set to 0.937 in order to obtain the optimal network model. The results of the ablation experiments are shown in Table 1.

CBAM	DAFS	Precision	Recall	mAP@0.5	mAP@[0.5:0.95]
/	/	92.2%	92.8%	91.5%	89.2%
1	/	95.4%	95.6%	94.8%	90.8%
/	1	94.8%	93.7%	93.4%	89.9%
1	1	95.8%	95.9%	96.2%	94.8%

Table 1. YOLOv5s ablation experiment.

Row 1 of Table 1 indicates the base performance of the original YOLOv5s on the dataset, and the average detection accuracy is 91.5%. After the introduction of CBAM and DAFS, respectively, it can be seen that CBAM improves the detection results more significantly, with significant improvements in Precision, Recall, and AP, while the improvement performance of DAFS is slightly weaker. The analysis suggests that this is related to the different functions of the two modules. The attention mechanism aims to improve the network's ability to extract important features, which is expressed in the improvement of accuracy, while the DAFS dynamic anchor frame module speeds up the regression of the prediction frame and improves the regression accuracy, so there is only a small improvement in the detection accuracy. After introducing both CBAM and DAFS modules, the detection network achieves the best results, with an average accuracy mAP improvement of 4.7% compared to the original network.

To analyze the impact of two improvement points of DAFS and DIoU-NMS on the detection results, some of the test results are visualized as shown in Figure 14. Figure 14a shows some of the detection results of the original YOLOv5s+DAFS, and it can be seen that the network has missed detection when the target is too small, and some of the detection frames are not very accurate. Figure 14b is the DIoU-NMS post-processing method obtained by improving the non-extremely suppressed post-processing method on the basis of (a), and the regression accuracy of the detection frame is significantly improved in the same image compared with Figure 14a, but the same problem of missed detection exists. Figure 14c shows the detection results with the CBAM module, and compared to Figure 14b, the missed white point defects are detected while maintaining a high detection accuracy.



Figure 14. Comparison of ablation experiment results. (a) YOLOv5s+DAFS. (b) YOLOv5s+DAFS+ DIoUNMS. (c) YOLOv5s-Small-Target.

In this paper, we use YOLOv5s as the basic defect detection algorithm. In order to solve the problem of low accuracy of small target defect detection, we add the CBAM convolutional attention mechanism module to the original YOLOv5s to improve it. The evaluation results of the improved YOLOv5s-Small-Target network model are shown in Figure 15a–e.



Figure 15. Performance evaluation results of the improved YOLOv5s-Small-Target model. (**a**) P (precision) curves. (**b**) R (recall) curve. (**c**) PR (Precision-recall) curve. (**d**) F1 score curve. (**e**) Improved YOLOv5s-Small-Target algorithm training results (loss function, precision, recall, and mAP transform).

Figure 16 shows the results of each evaluation of the improved YOLOv5 model. After 2000 epochs, the model reaches the convergence state. During the training of the model, the precision and recall rate are improved very steadily. After the model reaches saturation, its precision rate, which is shown in Figure 16a, can be maintained steadily around 99%; the recall rate, shown in Figure 16b, can be maintained steadily around 100%. The mean average precision and the reconciled mean average precision also remain at a high level, and the mean average precision, as shown in Figure 16c, can be stably maintained around 99%; the reconciled mean average precision, as shown in Figure 16d, can be stably maintained around 99%.



Figure 16. Evaluation results of the improved YOLOv5s-Small-Target model. (**a**) Precision. (**b**) Recall. (**c**) Mean average precision. (**d**) Reconciliated mean average precision.

The best detection model obtained from the training have been used for the defect detection of camera module lens, and some of the detection results are shown in Figure 17. The detection accuracy of white spot defects is basically maintained at about 93%, and the detection accuracy of white spots, scratches and dirty defects is as high as 96%, and there is no leakage and false detection. Therefore, the test experiment has achieved relatively satisfactory results.



Figure 17. Examples of detection test results for different types of defects. (a) Dirt + white dots.(b) Dirt + white dots. (c) White dots + scratch. (d) White spots.

To further validate the testing effectiveness of the YOLOv5s-Small-Target algorithm used in this paper on camera module lens surface defect detection, it is compared with the single-stage SSD algorithm and the YOLOv3+MobileNet lightweight network. The number of model training is set to 300, the batch size is set to 8, and all three algorithms do not use pre-trained models. The specific experimental results are shown in Figures 18–21.



Figure 18. Comparison of defect detection loss values for different detection algorithms.







Figure 20. Comparison of defect detection Recall of different detection algorithms.



Figure 21. Comparison of mean average precision mAP for defect detection with different detection algorithms.

From the experimental results in Figures 18–21, we can see that the accuracy P of the YOLOv5s-Small-Target algorithm designed in this paper is as high as 99%, which is better than the SSD and YOLOv3 algorithms in terms of model fitting speed and accuracy. The main reason for this result is that the detection objects in this paper are mostly small targets, and the improved algorithm contains a CBAM attention module, which improves the ability to detect small targets, so the comprehensive performance of the improved algorithm is improved.

In order to further analyze the defect detection performance of the improved YOLOv5s-Small-Target algorithm in this paper, a longitudinal comparison test and a cross-sectional comparison test are conducted for the camera module lens surface defect dataset.

(i) Longitudinal comparison experiment

The test results of YOLOv5s-Small-Target algorithm were compared with Faster R-CNN, SSD 300 (VGG16), SSD 512 (VGG16), and RetinaNet algorithms in terms of accuracy P, recall R, average precision and inference time while keeping all parameters set consistently, and the specific experimental data are shown in Table 2.

Model Name	Accuracy P	Recall Rate R	mAP@0.5	Reasoning Time (ms)
SSD 300 (VGG16)	0.835	0.867	0.840	9.8
SSD 512 (VGG16)	0.862	0.883	0.851	27
RetinaNet	0.912	0.924	0.893	20.6
Faster R-CNN	0.983	0.991	0.985	106
YOLOv5s-Small-Target	0.960	1.00	0.996	10.5

Table 2. Comparison of detection results of each algorithm.

From Table 2, the improved YOLOv5 algorithm has the highest average precision and the fastest inference speed among the single-stage detection algorithms. Compared with the two-stage Faster R-CNN, the improved recall and average precision are not much different from them, but in terms of inference speed, the improved YOLOv5 algorithm inferred each image 95.6 ms faster than the Faster R-CNN, which also meets the requirement in terms of speed.

(ii) Cross-sectional comparison experiment

In order to further verify the performance of the improved algorithm for lens surface defect detection in this paper, it is compared and analyzed with the mainstream YOLO series target detection algorithm, and the specific experimental results are shown in Table 3.

Model Name	Accuracy P	Recall Rate R	mAP@.5	Model Size/M	Reasoning Time (ms)
YOLOv3	0.866	0.879	0.854	67.2	17.5
YOLOv3+SPP	0.873	0.891	0.860	75.8	16
YOLOv5s	0.922	0.928	0.915	14.4	9.6
YOLOv5m	0.942	0.949	0.945	42.5	14.4
YOLOv3+MobileNetV2	0.812	0.824	0.806	44.0	11.3
YOLOv5s-Small-Target	0.960	1.000	0.996	18.7	10.5

Table 3. Comparison of detection results of each algorithm.

As can be seen from Table 3, the improved YOLOv5s-Small-Target has the best overall performance, with an accuracy of 96.0%, 1.8% higher than YOLOv5m; a recall rate of 100%, 5.1% higher than YOLOv5m; an average precision of 99.6%, 5.1% higher than YOLOv5m, maintaining a high detection accuracy; and an average inference time per image is 10.5 ms, which is 7.0 ms, 5.5 ms, 3.9 ms, and 0.8 ms faster than YOLOv3, YOLOv3+SPP, YOLOv5m, and YOLOv3+MobileNetV2, respectively. In terms of model size, YOLOv5s model has the smallest size, but its accuracy, recall, and average precision values are lower. The improved model after incorporating the attention mechanism has an increase in model size and inference time, but the accuracy, recall, and average precision have increased by 3.8%, 7.2%, and 8.1%, respectively.

Figure 22 shows the comparison of YOLOv5s-Small-Target model with other YOLO models in terms of number of parameters. The number of parameters of the improved algorithm has increased, but the detection accuracy has been improved substantially. Therefore, in comprehensive industrial field camera module lens surface defect detection accuracy and detection of real-time requirements, while considering the mobile deployment of the algorithm at a later stage, the YOLOv5s-Small-Target improved algorithm performs better.



Figure 22. Comparison of the number of parameters of different models.

Overall, although the YOLOv5s-Small-Target model shows some growth in volume and inference time, the accuracy, recall and mean average precision mAP values of the YOLOv5s-Small-Target algorithm are 3.8%, 7.2% and 8.1% higher than the original YOLOv5s algorithm. Therefore, the YOLOv5s-Small-Target algorithm has the best overall performance compared to other algorithms when considering the requirements of camera module lens surface defect detection accuracy and real-time, as well as the difficulty of mobile deployment of the model.

A camera module lens surface defect detection visualization system is designed and implemented, in which the YOLOv5s-Small-Target network model based on the LibTorch framework is deployed. An example graph of the operation results of the camera module lens surface defect detection system under the GPU RTX 3070 configuration is shown in Figure 23.

System for	camera	module lens surface de	fect detection		
	/		defect: 0.97	Initializ	zation
1				Model parameter	5:
				Model Name:	YOLOv5s-5
		defect: 0.94		IoU threshold:	0.6
				Confidence	0.6
		defect: 0.94 □	defect: 0.9	Defect detection Number of Defect size:	results: 5 ~ 257
	>		defect: 0.94	Area/R2:	0.18
.oading in	nage	Start running	Interface r 0 Number of defe	5 Time 5.35	ms
atch process	sing	Degumente) 01 degumente			dline ouhfold
Set path	D:\01My	Documents\01documents	<pre>>\U1MasterGraduationDesign\U2programming\}</pre>		aling subtold
		Start processing	Stop processing Number of	of	rval tips

Figure 23. Example diagram of the software runtime interface under GPU RTX 3070 configuration.

5. Conclusions and Future Works

To solve the problem of low accuracy of the current network model for small target defect detection, a camera module lens surface defect detection algorithm YOLOv5s-Defect based on improved YOLOv5s is proposed. Compared with the original YOLOv5s, the YOLOv5s-Small-Target algorithm makes four improvements: (1) introducing the dynamic anchor frame structure DAFS in the input stage; (2) improving the spatial pyramid SPP module to obtain the SPP-D module; (3) incorporating the CBAM attention mechanism module in the backbone network; (4) using the improved DIoU-NMS non-maximal suppression instead of non-maximal suppression (NMS) as the bounding box regression method in the regression box screening stage. Meanwhile, in order to alleviate the problem of uneven distribution of the number of defective sample pictures on the surface of the camera module lens, a PyQt5-based image broadening tool is produced to perform data enhancement on the defective samples.

The experimental results show that the mean average precision mAP@0.5 of the YOLOv5s-Small-Target algorithm is 99.6%, 8.1% higher than that of the original YOLOv5s algorithm, the detection speed FPS is 80 f/s, and the model size is 18.7M. Compared with YOLOv5s algorithm, YOLOv5s-Small-Target can detect the type and location of lens surface defects more accurately and quickly, which can meet the requirements of real-time and accuracy of camera module lens surface defect detection. Although the performance improvement is achieved, there is still much room for improvement in detection accuracy for detection scenarios where multiple defects overlap each other, which can be considered in terms of obtaining richer datasets, such as using generative adversarial networks to obtain more image samples containing overlapping defects.

Author Contributions: Conceptualization, G.H., J.Z. and H.Y.; methodology, G.H. and J.Z.; software, J.Z. and H.Y.; validation, G.H., J.Z., H.Y. and Y.N.; formal analysis, G.H., J.Z., H.Y. and Y.N.; investigation, G.H., J.Z. and H.Y.; resources, J.Z., Y.N. and H.Z.; data curation, J.Z., Y.N. and H.Z.; writing—original draft preparation, G.H., J.Z. and H.Y.; writing—review and editing, G.H., J.Z. and H.Y.; visualization, J.Z., Y.N. and H.Z.; supervision, G.H.; project administration, G.H.; funding acquisition, G.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (Grant No. 52175223) and the Key Research and Development Program of Changzhou (Grant No. CE20225044).

Data Availability Statement: The sample image data of lens surface defect of camera module used in this study are available from the email: jianyunchow@163.com upon request.

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

References

- Song, W.; Liu, X.; Cheng, Q.; Huang, Y.; Zheng, Y.; Liu, Y.; Wang, Y. Design of a 360-deg panoramic capture system based on a smart phone. *Opt. Eng.* 2020, 59, 015101. [CrossRef]
- Chang, Y.H.; Lu, C.J.; Liu, C.S.; Liu, D.S.; Chen, S.H.; Liao, T.W.; Peng, W.Y.; Lin, C.H. Design of Miniaturized Optical Image Stabilization and Autofocusing Camera Module for Cellphones. *Sens. Mater.* 2017, 29, 989–995.
- Huang, H.; Hu, C.; Wang, T.; Zhang, L.; Li, F.; Guo, P. Surface Defects Detection for Mobilephone Panel workpieces based on Machine Vision and Machine Learning. In Proceedings of the IEEE International Conference on Information and Automation (ICIA), Macao, China, 18–20 July 2017; pp. 370–375.
- Chang, C.F.; Wu, J.L.; Chen, K.J.; Hsu, M.C. A hybrid defect detection method for compact camera lens. *Adv. Mech. Eng.* 2017, 9, 1687814017722949. [CrossRef]
- Yang, J.; Li, S.; Wang, Z.; Yang, G. Real-Time Tiny Part Defect Detection System in Manufacturing Using Deep Learning. *IEEE Access* 2019, 7, 89278–89291. [CrossRef]
- Ma, L.; Lu, Y.; Jiang, H.Q.; Liu, Y.M. An automatic small sample learning-based detection method for LCD product defects. CAAI Trans. Intell. Syst. 2020, 15, 560–567.
- Lv, Y.; Ma, L.; Jiang, H. A Mobile Phone Screen Cover Glass Defect Detection MODEL Based on Small Samples Learning. In Proceedings of the 4th IEEE International Conference on Signal and Image Processing (ICSIP), Wuxi, China, 19–21 July 2019; pp. 1055–1059.
- Cao, Y.; Xu, C. Application of an Improved Threshold Segmentation Algorithm in Lens Defect Detection. *Laser Optoelectron. Prog.* 2021, 58, 1610007.
- 9. Pan, J.; Yan, N.; Zhu, L.; Zhang, X.; Fang, F. Comprehensive defect-detection method for a small-sized curved optical lens. *Appl. Opt.* **2020**, *59*, 234–243. [CrossRef]
- 10. Gao, Y.; Shen, L. Screen Defect Detection Based on Edge Detection. J. Syst. Sci. Math. Sci. 2021, 41, 1442–1454.
- 11. Deng, C.; Liu, Y. Twill Fabric Defect Detection Based on Edge Detection. Meas. Control. Technol. 2018, 37, 110–113.
- 12. Weimer, D.; Scholz-Reiter, B.; Shpitalni, M. Design of deep convolutional neural network architectures for automated feature extraction in industrial inspection. *CIRP Ann.* **2016**, *65*, 417–420. [CrossRef]
- Xie, L.; Xiang, X.; Xu, H.; Wang, L.; Lin, L.; Yin, G. FFCNN: A Deep Neural Network for Surface Defect Detection of Magnetic Tile. IEEE Trans. Ind. Electron. 2021, 68, 3506–3516. [CrossRef]
- 14. Xiang, K.; Li, S.S.; Luan, M.H.; Yang, Y.; He, H.M. Aluminum product surface defect detection method based on improved Faster RCNN. *Chin. J. Sci. Instrum.* **2021**, 42, 191–198.
- 15. Wu, T.; Wang, W.; Yu, L.; Xie, Y.; Yin, W.; Wang, H. Insulator Defect Detection Method for Lightweight YOLOV3. *Comput. Eng.* **2019**, 45, 275–280.
- 16. Fan, C.; Ma, L.; Jian, L.; Jiang, H. A Real-Time Detection Network for Surface Defects of Mobile Phone Lens. In Proceedings of the 13th International Conference on Graphics and Image Processing (ICGIP), Kunming, China, 18–20 August 2021.

- 17. Guo, L.; Wang, Q.; Xue, W.; Guo, J. A Small Object Detection Algorithm Based on Improved YOLOv5. J. Univ. Electron. Sci. Technol. China 2022, 51, 251–258.
- Zhang, R.; Wen, C.B. SOD-YOLO: A Small Target Defect Detection Algorithm for Wind Turbine Blades Based on Improved YOLOv5. Advanced Theory and Simulations. 2022, 5, 2100631. [CrossRef]
- 19. Yao, H.; Ma, G.; Shen, B.; Gu, J.; Zeng, X.; Zheng, X.; Tang, F. Flaws Detection System for Resin Lenses Based on Machine Vision. *Laser Optoelectron. Prog.* **2013**, *50*, 111201.
- Tao, X.; Hou, W.; Xu, D. A Survey of Surface Defect Detection Methods Based on Deep Learning. *Acta Autom. Sin.* 2021, 47, 1017–1034.
- Li, X.; Jeon, S.M.; Maytin, A.; Eadara, S.; Robinson, J.; Qu, L.; Caterina, M.; Meffert, M. Investigating Post-transcriptional Mechanism of Neuropathic Pain. J. Pain 2022, 23, 7. [CrossRef]
- Huang, T.; Cheng, M.; Yang, Y.; Lv, X.; Xu, J. Tiny Object Detection based on YOLOv5. In Proceedings of the 5th Interna-tional Conference on Image and Graphics Processing, ICIGP 2022, Beijing China, 7–9 January 2022; Volume 2021, pp. 45–50.
- Jubayer, F.; Soeb, J.A.; Mojumder, A.N.; Paul, M.K.; Barua, P.; Kayshar, S.; Akter, S.S.; Rahman, M.; Islam, A. Detection of mold on the food surface using YOLOv5. *Curr. Res. Food Sci.* 2021, 4, 724–728. [CrossRef]
- Li, S.; Yang, L.; Huang, J.; Hua, X.S.; Zhang, L. Dynamic Anchor Feature Selection for Single-Shot Object Detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea, 27 October–2 November 2019; pp. 6608–6617.
- 25. Liu, J.; Liu, J.; Luo, X. Research progress in attention mechanism in deep learning. Chin. J. Eng. 2021, 43, 1499–1511.
- 26. Hu, J.; Shen, L.; Albanie, S.; Sun, G.; Wu, E. Squeeze-and-Excitation Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 2020, 42, 2011–2023. [CrossRef] [PubMed]
- 27. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.N. CBAM: Convolutional block attention module. arXiv 2018, arXiv:1807.06521.
- 28. Hou, Q.; Zhou, D.; Feng, J. Coordinate attention for efficient mobile network design. arXiv 2021, arXiv:2103.02907.
- Guo, L.; Wang, Q.; Xue, W.; Guo, J. Detection of Mask Wearing in Dim Light Based on Attention Mechanism. J. Univ. Electron. Sci. Technol. China 2022, 51, 123–129.
- 30. Zhang, W.; Wang, J.; Ren, G. Optical elements defect online detection method based on camera array. *High Power Laser Part. Beams* **2020**, *32*, 051001.
- 31. Zhao, Y.Q.; Rao, Y.; Dong, S.P.; Zhang, J.Y. Survey on deep learning object detection. J. Image Graph. 2020, 25, 629–654.
- Yang, J.; Li, S.; Wang, Z.; Dong, H.; Wang, J.; Tang, S. Using Deep Learning to Detect Defects in Manufacturing: A Comprehensive Survey and Current Challenges. *Materials* 2020, 13, 5755. [CrossRef]