Lightweight Chip Pad Real-Time Alignment Detection Method and Application Based on Improved YOLOv5s

Yanli Zou *, Zongjian Zhang, Chiyang Zhou, Yufei Tan

Posted Date: 14 August 2024

doi: 10.20944/preprints202408.1058.v1

Keywords: deep learning; chip detection; small target detection; lightweighting

Article

Lightweight Chip Pad Real-Time Alignment Detection Method and Application Based on Improved YOLOv5s

Yanli Zou [1,2,*], Zongjian Zhang [1,2], Chiyang Zhou [1,2] and Yufei Tan [1,2]

[1] Guangxi Key Laboratory of Brain-inspired Computing and Intelligent Chips, School of Electronic and Information Engineering, Guangxi Normal University, Guilin, China
[2] Key Laboratory of Nonlinear Circuits and Optical Communications (Guangxi Normal University), Education Department of Guangxi Zhuang Autonomous Region, Guilin, China
* Correspondence: zouyanli72@163.com

Abstract: Chip pad alignment inspection is of great importance in the industrial field. However, due to the fact that chip pads are usually small, problems such as misdetection and missed detection often occur. When applying deep learning methods for chip pad detection, it is necessary to ensure accurate detection of small target chips while meeting the requirements of lightweight detection models for industrial needs. To solve the above problems, this paper proposes a lightweight model based on improved YOLOv5s. Firstly, the feature extraction part is improved to increase the network's focus on the target. Secondly, the feature fusion layer is improved to double the resolution of the prediction head, and the context-aware network is designed to enhance the context-capture ability of key features of small targets. Finally, SIoU is adopted as the loss function to improve the speed and accuracy of the regression frame. The experimental results show that the improved YOLOv5s algorithm improves the detection accuracy by 2.3% and reduces the network parameters by 81.8% compared to the original algorithm. The improved algorithm is combined with image processing techniques to design correction methods for alignment anomalies and realize real-time alignment anomaly correction in industry.

Keywords: deep learning; chip detection; small target detection; lightweighting

1. Introduction

Chip pad inspection is the basis of chip pad alignment inspection, which is a very critical step in the semiconductor manufacturing process. It ensures the accuracy and immediacy of chip alignment detection and alignment correction, and has a significant impact on subsequent decisions.

In the actual alignment detection work, the traditional manual detection of alignment usually requires manual measurement under a microscope, which cannot meet the demand for high precision and high efficiency in industrial assembly lines. In 2010, Chen [1] et al. used pattern recognition and image processing techniques for fast positioning of graphic tracking for automatic wafer alignment. In 2012, Xiao [2] proposed a simplified algorithm of template matching to extract wafers from the edge detection processed image to extract the wafer cut channel, and the wafer cut channel center line is obtained by straight line fitting to complete the positioning of the wafer. In 2013, Wu H [3] et al. proposed feature selection and two-stage classifier for weld joint detection, which improves the recognition rate of weld joints by extracting the color features, average grey level, and template-matching features. In 2017, Xu[4] et al. proposed a Fourier transform based direction alignment and least squares regression for positional pre-alignment, which improves the pre-alignment accuracy. In 2022, Wang [5] et al. proposed an adaptive Kalman filter with a dual-rate structure for uncalibrated visual localisation of wafer chips in LED packages by designing an adaptive Kalman filter for estimating the varying calibration parameters, and an introduced dual-rate structure for compensating the visual latency and achieving multi-rate sensor The dual-rate structure is

 doi:10.20944/preprints202408.1058.v1

introduced to compensate the visual delay and achieve the time synchronisation of multi-rate sensors.

With the development of computer vision in recent years, various deep learning based methods have been proposed to be applied in inspection. In 2019, Yu [6] et al. proposed a convolutional neural network based method for pattern recognition and analysis of constable defects, which inspects wafer defects by building an 8-layer CNN model. In 2020, Chien [7] et al. proposed a deep learning convolutional neural network based method that provides a reliable machine vision method instead of manual inspection by using Faster-RCNN model for training. In 2021, Bian [8] et al. propose a method based on improved YOLOv5s, which improves the detection accuracy by building an infrared image database for model training, and introduces an ECA module to enhance the feature extraction capability of the network. In 2022, Xu [9] et al. constructed an attention mechanism with long dependencies to enhance the correlation between features and proposed a design guideline for a single attention layer, which reduces the requirements for hardware devices in real scenarios. The target detection algorithm can quickly detect the chip pads and thus indirectly determine the alignment of the chip. Target detection uses techniques such as image processing and convolutional neural networks to classify and locate targets in images or videos.

Although some research progress has been made by previous researchers in the detection of chip pads, most of the research backgrounds are relatively homogeneous and differ greatly from the environment in actual industrial production. In reality, chip pads tend to be more numerous, densely arranged, and smaller in size. Although traditional CNN models can obtain a relatively good accuracy by stacking layers, their large number of parameters and complex structure lead to the inability of inference and deployment in edge devices with limited computational resources. Therefore, the requirements for chip pad detection networks are to achieve lightweight network models while ensuring detection performance. In 2015, He [10] et al. proposed the residual connection method, which effectively solves the problem of gradient disappearance or gradient explosion due to the deepening of the network layers. In 2017, Huang [11] et al. proposed the dense connection method, which solves the problem of parameter redundancy of the deeper network, and further reduces the model size and network size. further reducing the model size and network parameters. Subsequently, Howard [12–14] et al. proposed deep separable convolution, which divides the convolution process into two parts: channel-by-channel convolution and point-by-point convolution, and reduces the computation amount of convolution to 1/3 of the ordinary convolution; Zhang [15,16] et al. carry out channel disruption during channel-by-channel convolution, which makes the information that was originally not interoperable between the groups flow and interact, and enhances model expression.

There are two main categories of target detection algorithms, one is region-based second-order detection algorithms, such as Faster R-CNN [17] and R-CNN [18], etc., which first generate multiple candidate regions from an image, and then extract features and perform classification and regression for each region, so as to improve the detection accuracy. However, the disadvantages of this type of algorithms are many network parameters, complex models, slow detection speed, which are not suitable for real-time detection scenarios. The other category is the single-order detection algorithms represented by SSD [19] and YOLO [20–23], which predict and classify candidate frames directly on the picture, with the advantages of fast speed and simple model, which are more suitable for real-time detection needs. Currently widely used is the fifth generation algorithm of YOLO series, YOLOv5 [24], of which YOLOv5s version is the YOLOv5 in which a good balance between detection accuracy and model size is achieved. Therefore, in this paper, YOLOv5s is used as the baseline network for chip pad alignment detection. When we apply the YOLOv5s network directly on the chip pad dataset, the detection of small targets is not satisfactory. Zhu [25] et al. introduced Transformer [26] into the YOLO network for the first time, and the self-attention mechanism captures the contextual information and improves the detection accuracy of small targets by means of global composition. Literature [27–31] combines Swin-Transformer [32] into YOLO networks to reduce network parameters in global composition by using a moving sliding window. However, both

 doi:10.20944/preprints202408.1058.v1

Transformer and Swin-Transformer, the huge consumption of parameter computation and the highly complex model structure make the network impossible to be deployed into embedded devices.

In existing work on target detection, the introduction of more efficient convolutional and attentional modules effectively improves the detection performance of the network, but most of the work does not take into account the relationship between the image resolution and the feature receptive field in a small target detection environment. Stacking convolutional kernels to obtain a larger receptive field can capture richer semantic information about the target, but too deep a network will increase the network parameters and computation, and increasing the receptive field also leads to a decrease in resolution and a reduction in the ability to perceive the details of the image, thus affecting the detection of small targets. And the surrounding of small targets can often provide useful contextual information to help detect small targets.

To address the chip pad detection problem, this paper proposes a lightweight real-time detection network based on YOLOv5s, which not only ensures the detection accuracy of small target chip pads, but also effectively reduces the network parameters to meet the requirements of automated production. The main contributions of our work are as follows:

1. Using lightweight convolutional module (GhostNet) and attention module (CBAM), we improve the feature extraction module of the backbone network, which effectively reduces the parameter redundancy and computational complexity in feature extraction, enhances the network's attention to the target, and improves the detection effect of the network.

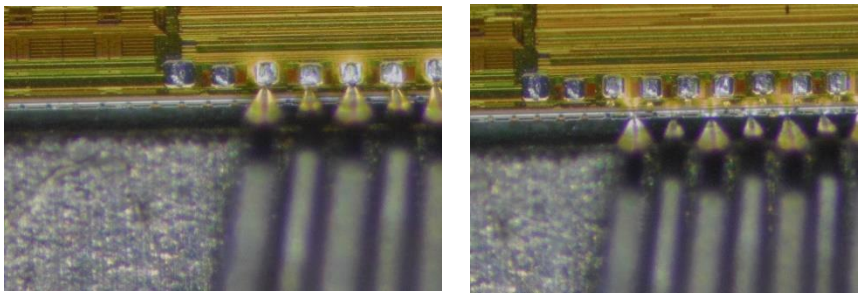
2. Propose a lightweight improvement method for small target detection on chip pads. Starting from the contradictory relationship between resolution and sensory field, the resolution of the customised prediction head is doubled by fusing shallower backbone network feature layers and cropping the last extraction layer. The sensing field is improved by introducing the cavity convolution in the spatial pyramid to enhance the contextual information perception of the key features of the small targets, so as to improve the detection performance of small targets with chip pads.

3. A correction method for real-time detection is designed, and the improved network is deployed in embedded devices to achieve real-time alignment detection and anomaly correction on industrial assembly lines by combining deep learning target detection algorithms with image processing techniques.

2. Materials and Methods

2.1. Datasets

The chip pad datasets originated from a semiconductor company in Guangxi. According to the actual inspection of the chip pads, an industrial camera with a pixel resolution of 6112*3440 was used to collect images under different lighting conditions with 200 times magnification, so as to construct the image data set of chip pads. The acquired data are cropped and enhanced, and the 1464 images are annotated using LabelImg annotation software after filtering and sorting, and finally consist of 6772 pads that require probe alignment (rig), 1610 pads that do not require probe alignment (wro), and 4237 solder joints (Probe). The chip pad dataset is divided into training set and validation set in the ratio of 9:1. Figure 1 shows an example of some images of the chip pad datasets.



(**a**) (**b**)

 doi:10.20944/preprints202408.1058.v1

Figure 1. Chip pad datasets.

2.2. Improved YOLOv5 algorithm

In the chip pad dataset, which contains a large number of small targets to be detected, the size of the feature map decreases as the network becomes deeper as feature extraction is continuously performed in the network, and this change can have a significant impact on the detection of small objects. In order to achieve a lightweight algorithm and reduce redundant computations and parameters. To address the problem of low accuracy of small object detection, the network proposed in this paper makes the following improvements based on the YOLOv5 algorithm to enhance the detection of chip pads and reduce the network parameters and model size to better suit the automated detection of chip pads in industry.

In the backbone layer, we use GhostNet [33] to replace the original convolutional module in C3, and embed the lightweight and efficient CBAM [34] attention mechanism. GhostNet can reduce the redundancy generated by feature extraction, which makes the network lighter, and the CBAM attention can make the network more focused on the small targets of chip pads from both the channel and the spatial dimensions to obtain higher detection accuracy, which is more suitable for industrial automation. It is more suitable for industrial automated production.

In the feature fusion and prediction section, the last layer of the C3 module is trimmed, and a customised prediction head is used to double the resolution, reducing the computational effort of the network and effectively detecting smaller targets. Meanwhile, hollow convolution [35] is introduced in the spatial pyramid to improve the feature sensing field and capture the rich contextual information around salient features. The SIoU [36] loss function is used instead of CIoU [37], and the SIoU considers the angle, distance, and shape of the bounding box, which is more consistent with the actual detection work. The improved YOLOv5 network is shown in Figure 2.

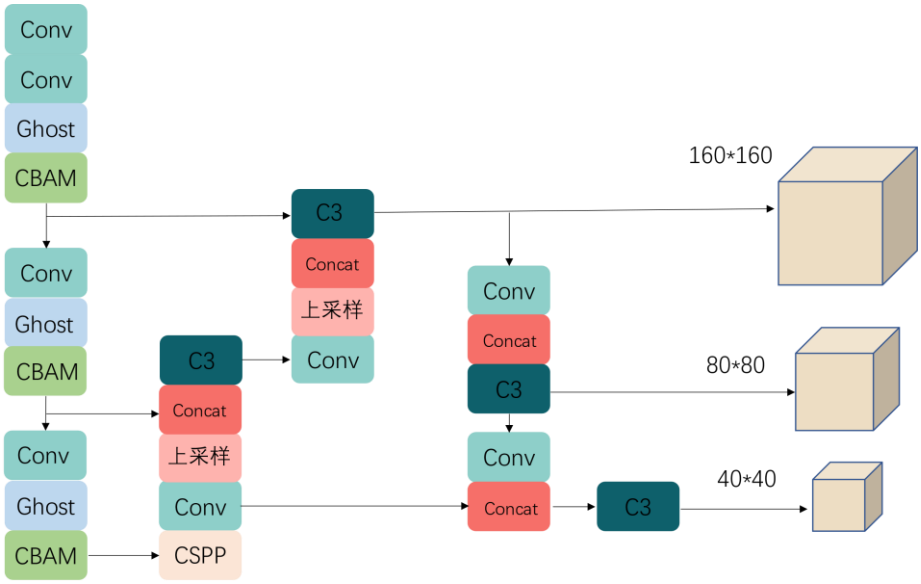


Figure 2. Improved YOLOv5 network structure diagram.

2.3. Feature Extraction Module Improvement

2.3.1. GhosetNet

In target detection, usually only a small part of the region contains the target to be detected, and there is a lot of redundant information in the whole image. In order to extract a more comprehensive feature map, it is usually necessary to use more convolutional kernels for the feature extraction work, but this can lead to redundancy of the convolutional kernels, especially when a large number of convolutional kernels are used as well as a too deep number of channels. Therefore, in this paper, GhostNet is used to replace the original convolutional layer in the backbone network. As shown in

 doi:10.20944/preprints202408.1058.v1

Figure 3, firstly, some feature maps are generated using ordinary convolution, and then the generated feature maps are processed by applying deep separable convolution to get more feature maps, and finally the original feature maps are spliced with Ghost feature maps. In this way, the feature expression can be enhanced with less computation.

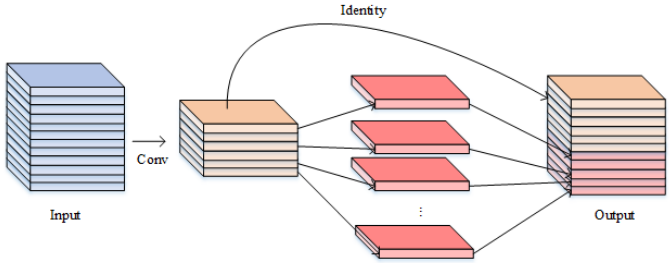


Figure 3. GhostNet.

2.3.2. CBAM Attention Mechanism

Just as when humans see an image, the brain usually pays attention to the whole image in its entirety, and when there is too much information in the image, the brain tends to selectively focus its attention on certain parts of the image. It is from the study of human vision that the attention mechanism originated and has been widely used in the field of computer vision to process information in images. It is based on assigning weights to different parts of the feature map to select useful information and ignore most irrelevant information. Attention mechanisms can be classified into channel attention, spatial attention, and hybrid attention, in which the channel attention represented by SE (Squeeze-and-Excitation Networks) [38] and ECA (Efficient Channel Attention) [39] attention focuses only on the channel information of the image, thus ignores the detail information in the spatial dimension; and a single spatial attention cannot meet the demand for channel feature extraction.

In the chip pad detection task, there are feature information of pads and probes in terms of color and position. The CBAM attention module cited in this paper is a hybrid lightweight attention module that combines channel and spatial attention, which makes the detection network more focused on small targets, thus obtaining higher detection accuracy. The overall flow structure of the CBAM attention module is shown in Figure 4, and the module consists of two independent modules, namely, channel attention (CAM) and spatial attention (SAM). Firstly, the channel feature map is generated by channel attention, which is multiplied with the residual input features for weighted refinement to strengthen the useful channel information; similarly the output results enter the spatial attention to get the final results. Adding CBAM attention after feature extraction can effectively aggregate the network's attention to the target and improve the detection of small targets.

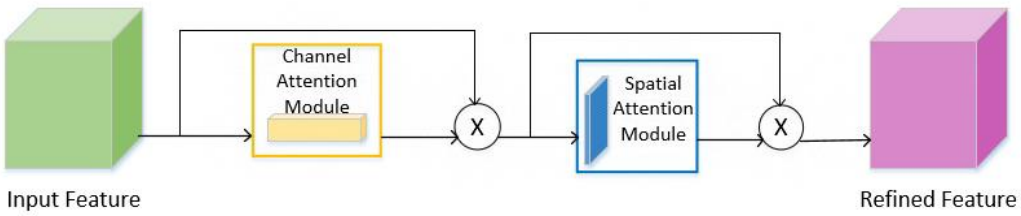


Figure 4. CBAM attention mechanism

2.4. Lightweight Improvement Methods

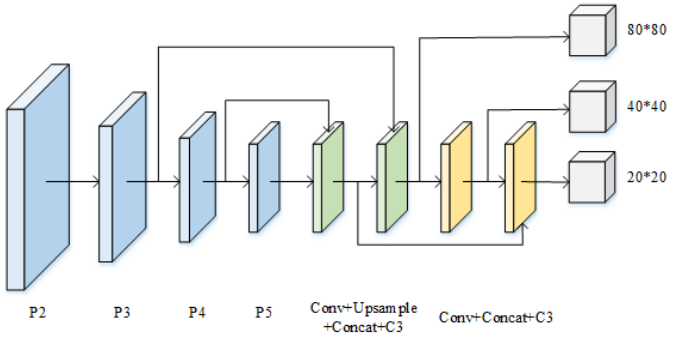
2.4.1. Lightweight High-Resolution Prediction Network

In the chip pad dataset, the target mainly occupies 2% to 8% of the image proportion, and the original YOLOv5's 80x80 resolution prediction head is difficult to complete the work of accurate

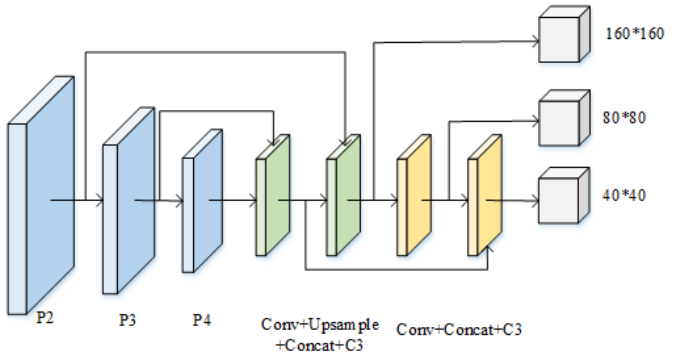
 doi:10.20944/preprints202408.1058.v1

detection of this smaller target. Small target detection has always been a difficult problem in the field of target detection, in the current mainstream detection algorithms, small targets are often detected in high-resolution feature maps, while low-resolution feature maps may reduce the network's ability to perceive image details.

In order to achieve the lightweight of the network while avoiding the influence of low-resolution feature maps, as shown in Figure 5, this paper proposes a lightweight high-resolution prediction network for chip pad detection, which firstly crops the default P5 of the YOLOv5 backbone network in Figure. 5(a), and then doubles the resolution of the prediction header by fusing the information of the shallower feature layer to obtain a larger resolution of 160x160 size feature maps, the network is shown in Figure 5(b). This strategy achieves the detection of higher resolution images while mitigating the model parameters, allowing smaller sized targets to be effectively detected.



(**a**) Origin YOLOv5



(**b**) Lightweight high-resolution prediction networks

Figure 5. Lightweight and high-resolution improvements.

2.4.2. Context-Aware Networks

However, the resolution of an image and the feature receptive field are contradictory existences. Although a larger receptive field can be obtained by stacking convolutional kernels to capture richer semantic information, this also leads to a reduction in resolution, which in turn affects the detection of small targets. The P5 of the cropping backbone network ensures that the image resolution is no longer degraded by the downsampling operation, but at the same time it results in insufficient extraction of semantic information.

In the chip pad detection task, pads and probes have obvious color and edge features relative to the background. In order to make full use of these features, this paper constructs a context-aware network CSPP (Context Spatial Pyramid Pooling) by introducing void convolution after the maximum pooling layer of SPPF. The maximum pooling layer is responsible for extracting the most significant features in the chip pads, while the cavity convolution uses the retained feature information to expand the sensory field by introducing the expansion rate in the convolution to obtain more contextual information from the local area. The context-aware network is shown in Figure. 6. CSPP expands the receptive field while keeping the resolution constant, which effectively retains more feature information, reduces information loss, and improves the expressive ability of the network.

 doi:10.20944/preprints202408.1058.v1

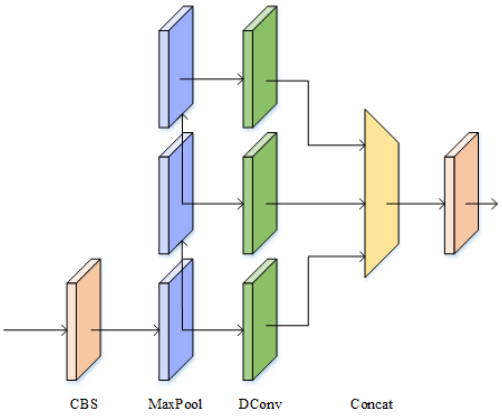


Figure 6. CSPP.

2.5. SIoU Loss Function

The loss function for target detection consists of two parts, Classification Loss and Bounding Box Regression Loss.YOLOv5 uses the binary cross-entropy loss function to calculate the probability of the category and the loss of the confidence score of the target, and in the regression loss calculation, the CIoU serves as the current stage of the commonly used form of YOLOv5 regression loss, which is calculated as shown in Equations (1) and (2).

$$CIoU = \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v \tag{1}$$

$$Loss_{CIoU} = 1 - IoU + CIoU = \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v \tag{2}$$

In Equations. (2) IoU represents the intersection and concurrency ratio of the real and predicted frames, b,b^gt represents the centroids of the predicted and real frames, respectively, ϱ^2 (b,b^gt) computes the Euclidean distance of the two centroids, and c represents the diagonal distance of the smallest closed region that can contain both the predicted and the real frames. α is a weight parameter, and v is used to measure the similarity of the width to height ratios, and the computation of α and v is shown in Equations (3) and (4) are shown.

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

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

In Equations. (4), (w,h) and (w^gt,h^gt) are the width and height of the predicted and real frames, respectively. CIoU takes into account the overlap area, the distance from the centroid, and the aspect ratio, but the difference in the aspect ratio that it responds to is not the difference between the real width and height and the confidence level, and this shortcoming leads to a slower convergence during training. The SIoU chosen in this paper takes the angle of the vector to be regressed into consideration, and consists of four loss functions: angle loss, distance loss, shape loss, and IoU loss, and the SIoU is defined as shown in Equation (5).

$$Loss_{SIoU} = 1 - IoU + \frac{\Delta + \Omega}{2} \tag{5}$$

In Equations. (5), Δ and Ω denote the distance and shape loss, respectively, which are defined as shown in Equations. (6) and (7).

$$\Delta = \sum_{t=x,y} (1 - e^{-\gamma p_t}) \tag{6}$$

 doi:10.20944/preprints202408.1058.v1

$$\Omega = \sum_{t=w,h} (1 - e^{-\omega_t})^{\theta} \tag{7}$$

In Equations. (6), $\rho_x = \left(\frac{b_{c_x}^{gt} - b_{c_x}}{c_w}\right)^2$, $\rho_y = \left(\frac{b_{c_y}^{gt} - b_{c_y}}{c_h}\right)^2$, and $\gamma = 2 - \Lambda$. Λ is the angular loss, which is defined in Equation. (8).

$$\Lambda = 1 - 2 * \sin^2\left(\sin^{-1}\frac{c_h}{\sigma} - \frac{\pi}{4}\right) \tag{8}$$

In Equation (8),

$$\sigma = \sqrt{\left(b_{c_x}^{gt} - b_{c_x}\right)^2 + \left(b_{c_y}^{gt} - b_{c_y}\right)^2} \tag{9}$$

$$c_h = \max\left(b_{c_y}^{gt}, b_{c_y}\right) - min\left(b_{c_y}^{gt}, b_{c_y}\right) \tag{10}$$

In Equation. (7), $w_w = \frac{|w - w^{gt}|}{max(w, w^{gt})}$ and $w_h = \frac{|h - h^{gt}|}{max(h, h^{gt})}$. In Equation. (9), σ is the distance between the centroid of the true frame and the prediction frame, and c_h in Equation. (10) is the height difference between the centroid of the true frame and the prediction frame, where $b_{c_x}^{gt}$、$b_{c_y}^{gt}$、b_{c_x}、b_{c_y} denote the x, y coordinates of the centroid of the true frame and the prediction frame, respectively. SIoU redefines the regression angle of the prediction frame by calculating the distance loss and accelerates network convergence.

3. Experiments and Results

This experiment was conducted on a graphics processor workstation equipped with an Intel Core i7-12700K processor and NVIDIA GeForce RTX 3090 with 24G of video memory. We used a 64-bit Ubuntu operating system with version number 18.04.6LTS and Linux kernel version 5.4.0-126-generic.The experiments were conducted using the PyTorch deep learning framework configured with CUDA with version number 11.1, and Python version 3.6.10.The experiments were conducted using SGD (Stochastic Gradient Descent) as the optimisation algorithm, with weight decay set to 0.0005, learning rate initial value of 0.01, image fixed input size of 640x640, batch size of 32 and training batch epoch of 300.

3.1. Feature extraction module comparison experiment

In order to verify the effectiveness of the improved feature extraction module to improve the detection effect, this paper introduces GhostNet to replace the C3 module of the backbone network on the basis of the YOLOv5s model, GhostNet can effectively reduce the redundancy of feature maps in the process of feature extraction, and the improved model is named YOLOv5s-g. Comparison experiments are conducted on the chip pad dataset. set to experiment on the two algorithms, and the experimental results are shown in Table 1.

Table 1. Comparison experiment of feature extraction module.

Method	Model	Parameters	FLOPs	mAP@0.5
YOLOv5s	13.75M	7.0M	16.0G	0.867
YOLOv5s-g	11.57M	5.8M	12.6	0.872

As can be seen in Table 1, the detection accuracy of the network for chip pads is improved by 0.5%, and the network parameters are reduced by 1.2M, which verifies that the introduction of GhostNet in the feature extraction part can effectively reduce the computational cost of the network and improve the expressive ability of the model. In order to further improve the feature extraction ability of the model, on the basis of YOLOv5s-g, the attention design scheme in the literature is used to introduce the CBAM attention mechanism, and a side-by-side comparison is made with the mainstream attention methods, and the comparison experiments are shown in Table 2.

 doi:10.20944/preprints202408.1058.v1

Table 2. Attention cross-sectional comparison experiment.

Method	rig	wro	Probe	mAP@0.5
YOLOv5s-g	0.913	0.853	0.851	0.872
+SE	0.917	0.849	0.858	0.875
+ECA	0.918	0.855	0.856	0.876
+CBAM	0.914	0.873	0.851	0.879

From the experimental results, it can be seen that the embedded attention mechanism can have a positive gain for chip pad alignment detection, in which the embedded CBAM attention has the largest improvement of 87.9% on the detection performance, in which the accuracy is 91.4% for the need to align the solder joints rig, 87.3% for the need not to align the solder joints wro, and 85.1% for the detection accuracy of the probes Probe. The side-by-side comparison experiments verify the effectiveness of embedded CBAM attention for improving the detection accuracy.

3.2. Comparison Experiment of Lightweight Improvement Methods

In order to verify the effectiveness of the proposed lightweight improvement method on the detection effect as well as the lightweight, based on Section 3.1, the deep feature extraction layer is firstly cropped, the shallower features are fused, and the resolution of the prediction head is doubled, and the final outputs are detected on the high-resolution feature maps. The improved model is named YOLOv5s-HR(high resolution). Comparative experiments are taken to test the two algorithms on the chip pad dataset, and the experimental results are shown in Table 3.

Table 3. Lightweight high-resolution prediction network.

Method	rig	wro	Probe	mAP@0.5	Model	Parameters	FLOPs
YOLOv5s-CBAM	0.914	0.873	0.851	0.879	7.2M	3.3M	12.6G
YOLOv5s-HR	0.918	0.892	0.844	0.884	3.25M	1.27M	10.2G

As can be seen from Table 3, the size of the lightweight high-resolution prediction network model obtained by trimming the deep feature extraction layer and fusing the shallower features is reduced by 3.95M, the amount of parameters is reduced by 2.03M, and the detection accuracy is improved by 0.5%, of which the detection accuracy is 91.8% for the pad rig that needs to be aligned, and the detection accuracy for the pad wro that does not need to be aligned is 89.2%, which is higher than the improvement of the former algorithm. The detection accuracy for probe Probe is 84.4%, which is slightly lower than that of the previous algorithm. This is because the improved high-resolution prediction network method reduces the ability to capture rich semantic information, and by adopting the context-aware network approach, richer semantic information can be captured to improve the detection performance of probes as well as targets to be inspected, and the experimental results are shown in Table 4.

Table 4. CSPP comparison experiments.

Method	rig	wro	Probe	mAP@0.5
YOLOv5s-HR	0.918	0.892	0.844	0.884
+CSPP	0.919	0.893	0.852	0.888

As can be seen in Table 4, the CSPP module can effectively improve the detection performance of the network, verifying the effectiveness of the lightweight improvement method in the chip pad detection task.

 doi:10.20944/preprints202408.1058.v1

3.3. Ablation Experiment

Based on the chip pad detection task, the model size, parameters, the number of floating-point operations and the average accuracy, etc., as an indicator to assess the model performance, the ablation experiment on each module has verified the effectiveness of the algorithm proposed in this paper, the experimental results are shown in Table 5.

Table 5. Ablation experiment.

Ghost+CBAM	HR	CSPP	SIoU	Model	Parameters	FLOPs	mAP@0.5
				13.75M	7.0M	16.0G	0.867
√				7.2M	3.3M	12.6G	0.879
√	√			3.25M	1.27M	10.2G	0.884
√	√	√		3.46M	1.3M	10.6G	0.888
√	√	√	√	3.46M	1.3M	10.6G	0.89

According to the experimental results in Table 5, it can be found that after the improvement of the feature extraction part, compared with the original network, the detection accuracy is improved by 1.2%, and the network parameters, model weights, and model complexity are reduced by 3.7M, 6.55M, and 3.4G, respectively, which proves that the Ghost and the CBAM can effectively mitigate the redundant features and focus the attention. Next, we propose a lightweight improvement method for chip pad detection, which improves the detection accuracy by 0.5% through a lightweight high-resolution prediction network, and reduces the network parameters, model weights, and model complexity by 2.03M, 3.95M, and 2.4G, respectively, which demonstrates that the proposed lightweight improvement method is effective in improving the performance of the detection of chip pads, and at the same time makes the network more lightweight. Introducing the CSPP composed of cavity convolution in SPPF, the detection accuracy reaches 88.8% despite the slight increase in network weights, model parameters and model complexity, proving that the proposed CSPP is capable of extracting the semantic information of key features; finally, the SIoU Loss is used as the regression loss function to form the final model, and compared with the initial model, the detection accuracy of the proposed algorithm increases by 2.3%, the network weights increase by 2.3%, and the network weights increase by 2.3%, and the network weights increase by 2.3%. improves by 2.3%, the network weights are reduced by 74.8%, the model parameters are reduced by 81.4%, and the model complexity is reduced by 5.4 G. This indicates that the algorithm proposed in this paper achieves a good balance between the detection accuracy and the model size, and the improved network reduces the cost of the hardware and is easy to be deployed on the edge devices, which ensures the practical use in the industry. In summary, the improved algorithm proposed in this paper has very high practical value.

3.4. Mainstream algorithm comparison experiment

In order to evaluate the performance of the improved algorithm proposed in this study, the network was compared with the classical mainstream algorithms, mainly SSD, Efficientdet-d0 [40], YoloX-s [41] and Yolo-lite-g [42], as shown in Table 6.

Table 6. Mainstream algorithm comparison experiment.

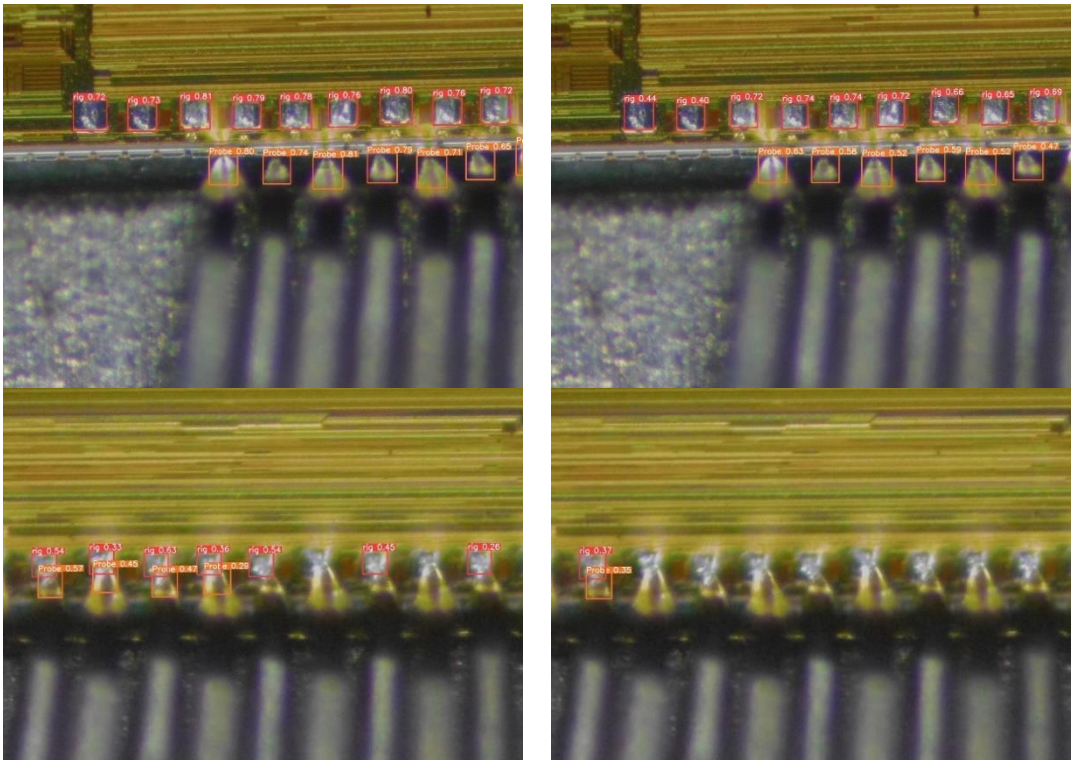
Method	rig	wro	Probe	mAP@0.5	Parameters	Model
SSD-mobile	0.6	0.35	0.49	0.482	25.06M	15.32M
Efficientdet-d0	0.76	0.48	0.688	0.641	3.7M	15.08M
YOLOX-s	0.854	0.847	0.752	0.818	9.1M	34.36M
YOLOv5-lite-g	0.908	0.855	0.826	0.863	5.5M	10.76M
YOLOv5s	0.906	0.853	0.84	0.867	7.0M	13.7M
YOLOR	0.903	0.851	0.793	0.849	9.0M	17.46M
YOLOv3-tiny	0.879	0.791	0.807	0.826	8.7M	16.63M

 doi:10.20944/preprints202408.1058.v1

YOLOv7-tiny	0.911	0.87	0.855	0.879	6.0M	11.72M
Ours	0.922	0.887	0.863	0.89	1.3M	3.46M

As can be seen from the indicators in the table, the improved model in this paper improves the average accuracy by 2.3% compared to the original Yolov5s and outperforms the new popular algorithms of YOLO series, YOLOX-s and YOLOR, with a performance improvement of 7.2% and 0.41%, respectively; compared to the same lightweight algorithms, Efficientdet-d0, YOLOv5-lite-g, YOLOv3-tiny and YOLOv7-tiny[44], the proposed algorithm in this paper improves detection accuracy by 34.9%, 2.7%, 6.4% and 1.1%, respectively, and the model parameters and network weights are reduced. In summary, the improved method in this paper has higher accuracy in the alignment detection of chip pads while achieving a lighter model, which proves its superiority and is more suitable for the deployment of reasoning in the real industry.

The actual detection of chip pads is shown in Figure. 7, where (a) shows the detection effect of the improved algorithm and (b) shows the detection effect of the YOLOv5 algorithm. It can be seen that compared with the original YOLOv5 network, the improved algorithm in this paper has a higher detection rate on the same solder pad detection, and at the same time, the algorithm proposed in this paper can detect more targets. This implies that in practice, our method detects better and is more robust.



(**a**) (**b**)

Figure 7. Chip pad detection effect. (a) ours. (b) YOLOv5s.

3.5. Model Deployment and Calibration Detection Methods

3.5.1. Real-Time Detection Processing

Digital images in the process of acquisition, transmission and processing, often subject to the shooting equipment and the external environment, the inevitable impact on the image, the identification of the target in the image will have a greater impact.

In this section, we combine the chip pad detection video screen, select any frame and greyscale it to get the image shown in Figure 8, and use the fast Fourier transform to get the frequency domain spectrum of the image, after centering as shown in Figure 9 (a). After centring, the low frequency

 doi:10.20944/preprints202408.1058.v1

signals are distributed in the middle part of the frequency domain spectrum and the high frequency signals are distributed around.

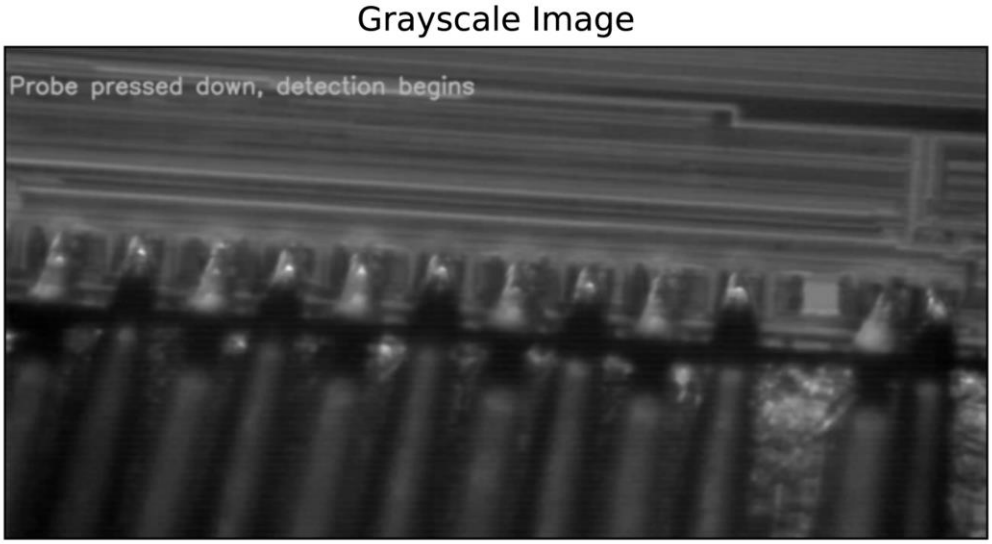
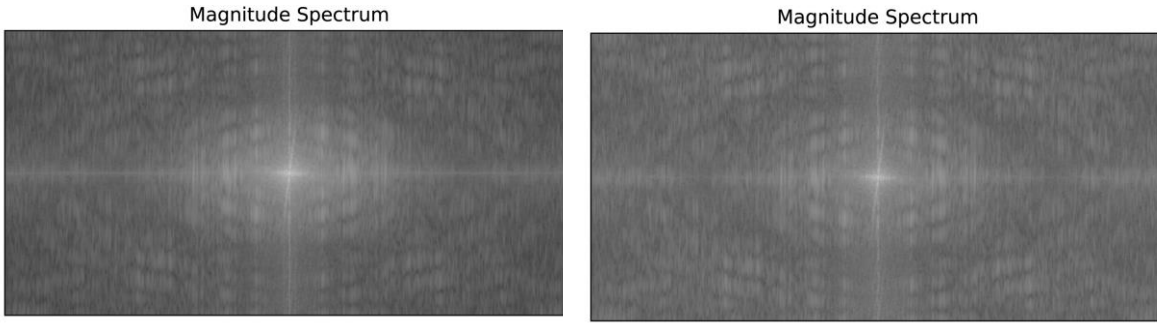


Figure 8. Grayscaled image.



(a) **(b)**

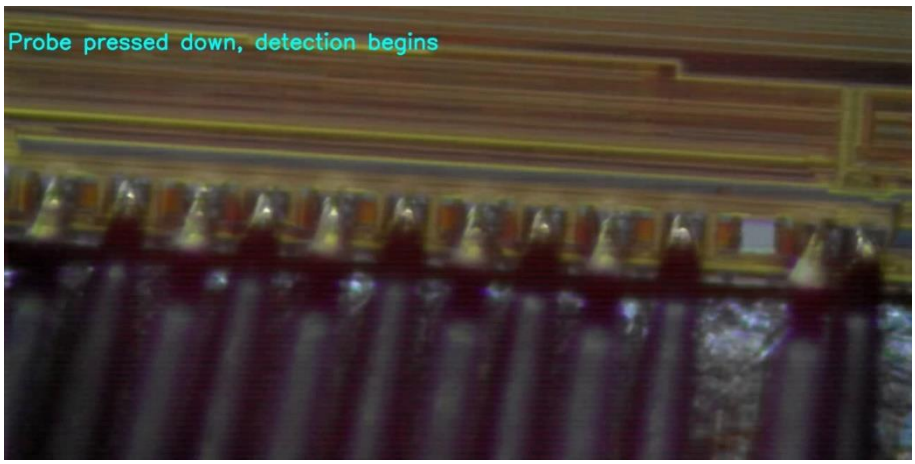
Figure 9. Image frequency domain spectrum.

As can be seen in Figure. 9 (a), the low-frequency portion of the frequency domain spectrum of the original image is obvious and the high-frequency portion is not prominent enough. The low-frequency part of the image corresponds to the overall brightness and color respectively, smooth changes in large areas, etc., while the high-frequency part corresponds to the edges and details of the object as well as the noise in the image. When the chip pad is detected, the overall image becomes smoothed due to the distortion of the image caused by the data line transmission, which also causes the loss of the target edges and details of the chip pad and the probe, which is required. In order to enhance the high frequency component of the image screen, the enhancement of the high frequency component is achieved by using high pass filtering in the spatial domain.

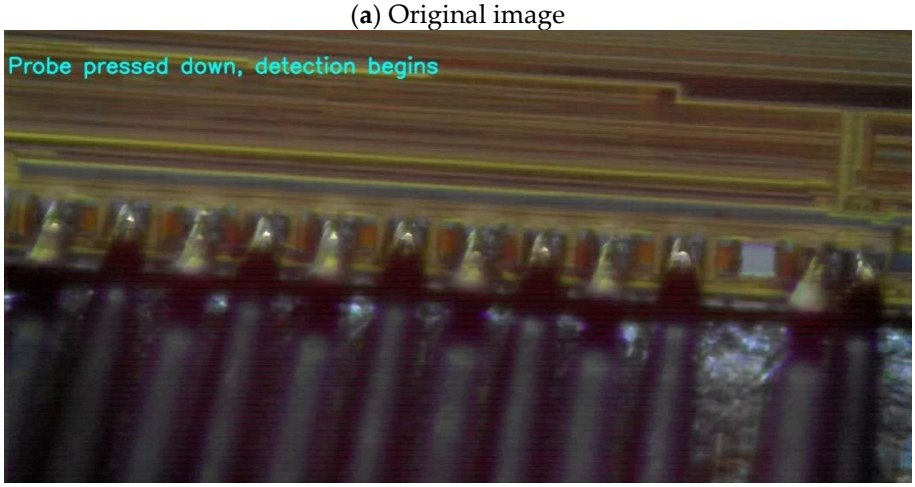
High-pass filtering enhances the edge information of the target in the image and improves the clarity of the image, so it is also often referred to as image sharpening. In this paper, a 3×3 sharpening convolution kernel is used to enhance the video at high frequencies, in order to try to avoid the interference of noise brought by sharpening on detection, the sharpening convolution kernel is designed in the form as shown in Equation. (11), and the frequency spectrum of the image after high-pass filtering is shown in Figure. 9 (b). It can be seen that the high-frequency enhancement of the image after high-pass filtering, the low-frequency a little weakened, in line with requirements.

$$H = \begin{bmatrix} -0.25 & -0.5 & -0.25 \\ -0.5 & 4 & -0.5 \\ -0.25 & -0.5 & -0.25 \end{bmatrix} \tag{11}$$

 doi:10.20944/preprints202408.1058.v1



(**a**) Original image



(**b**) Image after high pass filtering

Figure 10. Comparison effect of filtering effect.

After comparative analysis, in the chip pad alignment detection, high-pass filtering, although it will introduce a small amount of noise, but due to the data transmission after the picture will be with a small amount of distortion and cause image smoothing, high-pass filtering method is clearer than the original picture, the target characteristics of the information is more obvious. The experimental comparison results of high-pass filtering are shown in Figure 10.

3.5.2. Real-Time Anomaly Correction Method

Previous work has successfully deployed the improved lightweight network to edge devices for chip pad alignment detection. However, in practice, the alignment detection of chip pads and probes often has anomalies, and the detection can only get the coordinate information and category of the target to be inspected, while the offset of the detected pads and probes is unknown, so the real-time anomaly calibration task is difficult to complete. In order to solve this problem and improve the efficiency and accuracy of chip pad alignment testing, this paper proposes a matching scheme aimed at real-time anomaly calibration.

The matching scheme for real-time detection and calibration is divided into sequential marking and determining the matching relationship. It is as follows: firstly, the video stream is acquired by an industrial camera, and the real-time detection of each frame is performed using the improved network in this paper to obtain the detection results of each frame. The targets to be inspected in the inspection frame are labelled according to the horizontal coordinates, and the targets with horizontal coordinates from small to large are obtained. Subsequently, we need to ensure the correspondence between pads and probes.

In the actual alignment detection process, the pads and probes may have certain angle and distance deviations, so we adopt a more flexible method to determine their matching relationship.

 doi:10.20944/preprints202408.1058.v1

Specifically, the reference target is determined by determining the offset direction of the probe. When the probes as a whole are offset to the right, we select the pad as the category benchmark and match the nearest probe within a certain offset range. And when the probes are shifted to the left as a whole, we select the probes as the category datum and match the nearest pads within a certain offset range.

For the matched pads and probes in each frame, the average pixel distance and average angle are calculated and the data is published in real time. This enables the robotic arm module to subscribe and resolve the abnormal distance and angle of the alignment situation for real-time calibration. This approach not only improves the flexibility of alignment detection, but also provides the basis for real-time calibration of the robotic arm.

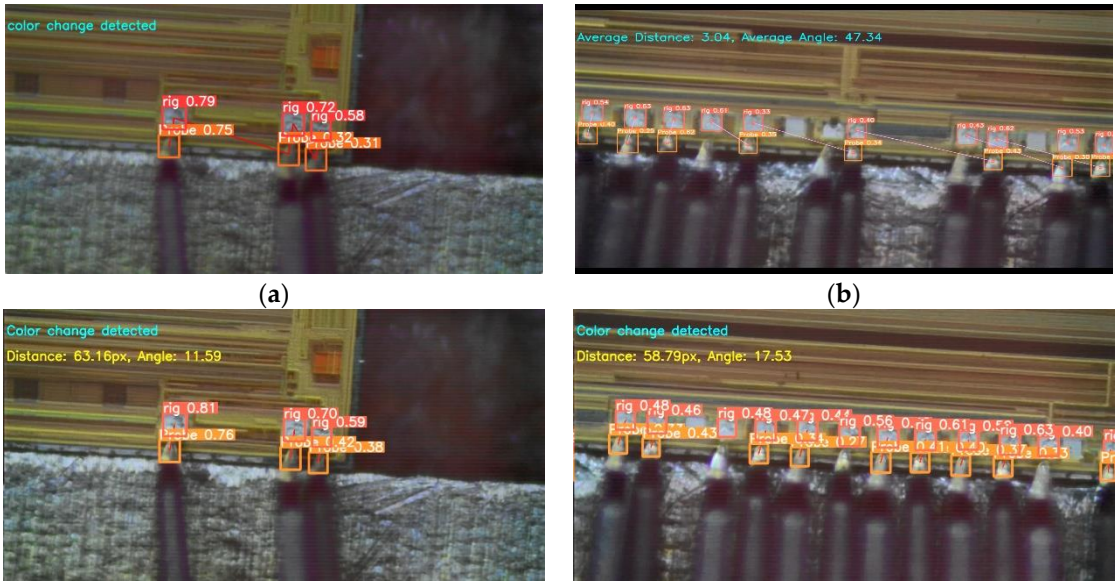


Figure 11. Real-time anomaly calibration matching comparison results.

In Figure. 11, (a) shows the results without pad-probe matching, and it can be seen that it is not possible to determine the offset when the one-to-one correspondence is not determined. (b) is the result of determining the matching relationship based only on the coordinates, and it can be seen that there is a large detection error. (c) and (d) are the results after carrying out the matching scheme, and it can be seen that after designing the anomaly calibration matching scheme can effectively improve the accuracy of calibration。

3.5.3. Actual Detection Effect

Subsequently, the designed calibration detection method was deployed into the edge device to perform chip pad alignment detection for the ongoing chip test work The actual detection is shown in Figure 12.

 doi:10.20944/preprints202408.1058.v1

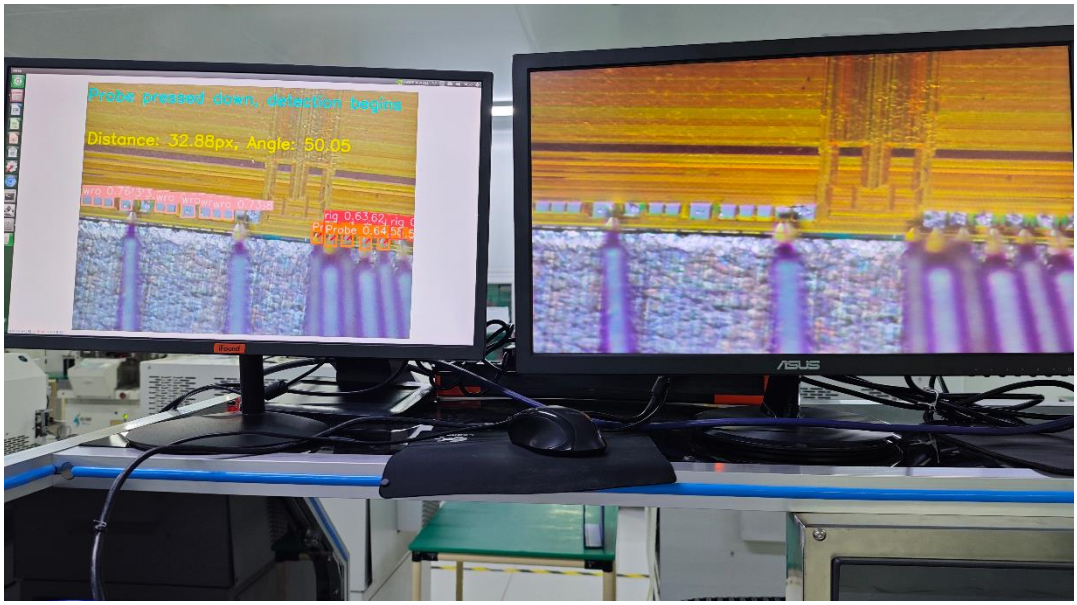


Figure 12. Actual detection effect.

In Figure 12, the display on the right shows the output screen obtained by the industrial camera, and the display on the left shows the alignment detection results of the real-time screen after passing the calibration detection method, with an average offset distance of 32.88px and an average offset angle of 50.05 degrees. The proposed chip pad calibration detection method performs well and meets the demand for real-time and high accuracy of chip pad alignment detection. Therefore, the method can be applied to the environment of actual industrial inspection and provide a data base for subsequent automatic calibration.

4. Discussion

In this section, we will further discuss the capability of the proposed algorithm on other datasets to evaluate whether the proposed algorithm is generalizable for small target detection, and then determine whether the improvements therein have a wide range of application scenarios. In this section, the VisDrone [45], WIDER FACE [46] public datasets are selected for testing. the VisDrone and WIDER FACE datasets contain a large number of small targets, which are suitable for the validation of the proposed algorithm in this paper. The comparison results are shown in Table 7 and Table 8.

Table 7. Performance comparison of VisDrone dataset.

Method	mAP	FLOPs	Parameters
YOLOv5s	0.35	16.0G	7.0M
Ours	0.387	10.6G	1.3M
Improve	+3.7%	-5.4G	-81.4%

Table 8. Performance comparison of WIDER FACE dataset.

Method	mAP	FLOPs	Parameters
YOLOv5s	0.736	16.0G	7.0M
Ours	0.75	10.6G	1.3M
Improve	+1.4%	-5.4G	-81.4%

Through the experimental validation on VisDrone dataset and WIDER FACE dataset, it can be seen that the improved algorithm proposed in this paper still improves in detection accuracy and comprehensive performance. This indicates that the improved method has obvious improvement

 doi:10.20944/preprints202408.1058.v1

effect in the direction of small target detection and has certain universality. The experimental results are shown in Figures 13 and 14.

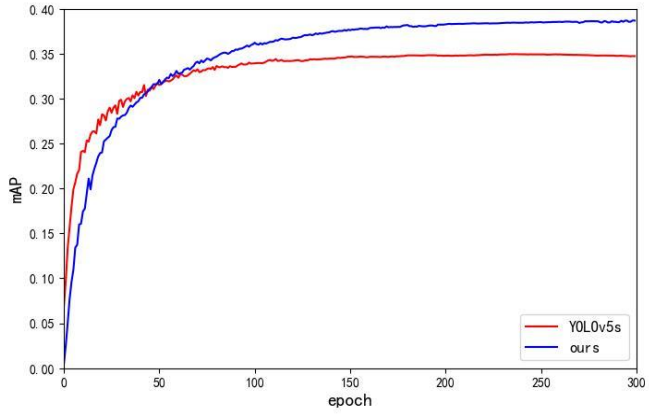


Figure 13. Comparison of detection accuracy on the VisDrone dataset.

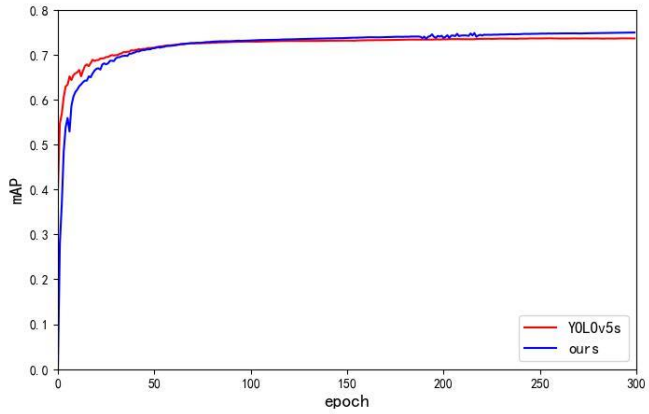


Figure 14. Comparison of detection accuracy on the WIDER FACE dataset.

The above experiments verify the effectiveness and universality of the method proposed in this paper. In addition, the lightweight design of the model can still be handled by model compression for lightweighting, such as knowledge distillation, in addition to the network structure design. Therefore, our future work will explore in the knowledge distillation technique.

5. Conclusions

This paper is dedicated to solving the problem of chip pad alignment detection in industry and proposes a lightweight detection algorithm based on the improved YOLOv5s. In order to solve the problem of poor alignment detection in the case of dense distribution and small percentage of chips in industrial production, we have optimized the YOLOv5s network in many aspects. First, the feature extraction part is improved by introducing the lightweight and efficient Ghost convolution with CBAM attention mechanism, which effectively reduces the redundancy of feature extraction in the convolution process and improves the network's ability to pay attention to the target. Starting from the contradictory relationship between resolution and sensory field, we double the resolution of the prediction header and introduce a context-aware network to improve the detailed grasp of key information and achieve a balance between resolution and sensory field. Finally, we choose SIoU Loss as the loss function to accelerate the model convergence and improve the accuracy. In order to verify the effectiveness of the improved model, we conduct a large number of ablation experiments and comparison experiments. The experimental results show that the average accuracy is improved by 2.3% on the chip pad dataset. The detection accuracy is also improved on the public datasets

 doi:10.20944/preprints202408.1058.v1

VisDrone and WIDER Face. The improved network has fewer parameters and is more suitable for industrial applications.

6. Patents

一种基于 YOLOv5 的轻量级芯片焊盘对准检测方法 202310891662.1

Author Contributions: Y.Z. and Z.Z. performed the analysis.Z.Z. validated the analysis and drafted the manuscript. Y.Z. reviewed the manuscript. Z.Z. designed the research. Y.T. reviewed the manuscript. C.Z. reviewed the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: We acknowledge support from Guangxi Innovation Driven Development Project Guike AA21077015 and National Natural Science Foundation of China Grants 12162005.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: We sincerely thank the anonymous reviewers for their critical comments and suggestions for improving the manuscript.

Conflicts of Interest: We sincerely thank the anonymous reviewers for their critical comments and suggestions for improving the manuscript.

References

1. Chen, M.-F., Y.-S. Ho, and S.-M. Wang. A fast positioning method with pattern tracking for automatic wafer alignment. in 2010 3rd International Congress on Image and Signal Processing. 2010. IEEE.
2. Yang W and Xiao M, Research on Automatic Alignment System of Scribing Machine Based on Matlab. Mechanical Design and Manufacturing, 2012(05): p. 96-98.
3. Wu, H., et al., Classification of solder joint using feature selection based on Bayes and support vector machine. IEEE Transactions on Components, Packaging and Manufacturing Technology, 2013. 3(3): p. 516-522.
4. Xu, J., et al., A wafer prealignment algorithm based on Fourier transform and least square regression. IEEE Transactions on Automation Science and Engineering, 2017. 14(4): p. 1771-1777.
5. Wang, Z., D. Zhou, and S. Gong, Uncalibrated visual positioning using adaptive Kalman Filter with dual rate structure for wafer chip in LED packaging. Measurement, 2022. 191: p. 110829.
6. Yu, N., Q. Xu, and H. Wang, Wafer defect pattern recognition and analysis based on convolutional neural network. IEEE Transactions on Semiconductor Manufacturing, 2019. 32(4): p. 566-573.
7. Chien, J.-C., M.-T. Wu, and J.-D. Lee, Inspection and classification of semiconductor wafer surface defects using CNN deep learning networks. Applied Sciences, 2020. 10(15): p. 5340。
8. Bian, Y.-C., et al. Using improved YOLOv5s for defect detection of thermistor wire solder joints based on infrared thermography. in 2021 5th International Conference on Automation, Control and Robots (ICACR). 2021. IEEE
9. Xu, J., et al., Chip Pad Inspection Method Based on an Improved YOLOv5 Algorithm. Sensors, 2022. 22(17): p. 6685.
10. He, K., et al. Deep residual learning for image recognition. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
11. Huang, G., et al. Densely connected convolutional networks. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.
12. Howard, A.G., et al., Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861, 2017.
13. Sandler, M., et al. Mobilenetv2: Inverted residuals and linear bottlenecks. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.
14. Howard, A., et al. Searching for mobilenetv3. in Proceedings of the IEEE/CVF international conference on computer vision. 2019.
15. Zhang, X., et al. Shufflenet: An extremely efficient convolutional neural network for mobile devices. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.
16. Ma, N., et al. Shufflenet v2: Practical guidelines for efficient cnn architecture design. in Proceedings of the European conference on computer vision (ECCV). 2018.
17. Ren, S., et al., Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems, 2015. 28.

 doi:10.20944/preprints202408.1058.v1

18. Girshick, R., et al. Rich feature hierarchies for accurate object detection and semantic segmentation. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2014.
19. Liu, W., et al. Ssd: Single shot multibox detector. in Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14. 2016. Springer.
20. Redmon, J., et al. You only look once: Unified, real-time object detection. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
21. Redmon, J. and A. Farhadi. YOLO9000: better, faster, stronger. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.
22. Redmon, J. and A. Farhadi, Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767, 2018.
23. Bochkovskiy, A., C.-Y. Wang, and H.-Y.M. Liao, Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934, 2020.
24. Ultralytics, Yolov5. Available online: https://github.com/ultralytics/yolov5 (accessed on 22 February 2022).
25. Zhu, X., et al. TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios. in Proceedings of the IEEE/CVF international conference on computer vision. 2021.
26. Vaswani, A., et al., Attention is all you need. Advances in neural information processing systems, 2017. 30.
27. Gong, H., et al., Swin-transformer-enabled YOLOv5 with attention mechanism for small object detection on satellite images. Remote Sensing, 2022. 14(12): p. 2861.
28. Lei, F., F. Tang, and S. Li, Underwater target detection algorithm based on improved YOLOv5. Journal of Marine Science and Engineering, 2022. 10(3): p. 310.
29. Lu, S., et al., Swin-transformer-YOLOv5 for real-time wine grape bunch detection. Remote Sens 14: 5853. 2022.
30. Ling, Q., et al., Insulated Gate Bipolar Transistor Solder Layer Defect Detection Research Based on Improved YOLOv5. Applied Sciences, 2022. 12(22): p. 11469.
31. Zhang, S.G., et al., Swin-YOLOv5: Research and Application of Fire and Smoke Detection Algorithm Based on YOLOv5. Computational Intelligence and Neuroscience, 2022. 2022.
32. Liu, Z., et al. Swin transformer: Hierarchical vision transformer using shifted windows. in Proceedings of the IEEE/CVF international conference on computer vision. 2021.
33. Han, K., et al. Ghostnet: More features from cheap operations. in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020.
34. Woo, S., et al. Cbam: Convolutional block attention module. in Proceedings of the European conference on computer vision (ECCV). 2018.
35. Yu, F. and V. Koltun, Multi-scale context aggregation by dilated convolutions. arXiv preprint arXiv:1511.07122, 2015.
36. Gevorgyan, Z., SIoU loss: More powerful learning for bounding box regression. arXiv preprint arXiv:2205.12740, 2022.
37. Zheng, Z., et al. Distance-IoU loss: Faster and better learning for bounding box regression. in Proceedings of the AAAI conference on artificial intelligence. 2020.
38. Hu, J., L. Shen, and G. Sun. Squeeze-and-excitation networks. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.
39. Wang, Q., et al. ECA-Net: Efficient channel attention for deep convolutional neural networks. in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020.
40. Tan, M., R. Pang, and Q.V. Le. Efficientdet: Scalable and efficient object detection. in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020.
41. Ge, Z., et al., Yolox: Exceeding yolo series in 2021. arXiv preprint arXiv:2107.08430, 2021.
42. Yolov5-lite. Available online: https://github.com/ppogg/YOLOv5-Lite.
43. Wang, C.-Y., I.-H. Yeh, and H.-Y.M. Liao, You only learn one representation: Unified network for multiple tasks. arXiv preprint arXiv:2105.04206, 2021.
44. Wang, C.-Y., A. Bochkovskiy, and H.-Y.M. Liao. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023.
45. Zhu, P., et al., Detection and tracking meet drones challenge. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021. 44(11): p. 7380-7399.
46. Yang, S., et al. Wider face: A face detection benchmark. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.