applied sciences

MDPI

Article

An Improved YOLOv5 Model for Detecting Laser Welding Defects of Lithium Battery Pole

Yatao Yang, Yunhao Zhou, Nasir Ud Din, Junqing Li, Yunjie He and Li Zhang *

College of Electronics and Information Engineering, Shenzhen University, Shenzhen 518000, China
* Correspondence: zhang_li@szu.edu.cn; Tel.: +86-755-22673392

Abstract: Focus on the requirement for detecting laser welding defects of lithium battery pole, a new model based on the improved YOLOv5 algorithm was proposed in this paper. First, all the 3×3 convolutional kernels in the backbone network were replaced by 6×6 convolutional kernels to improve the model's detection capability of a small defect; second, the last layer of the backbone network was replaced by our designed SPPSE module to enhance the detection accuracy of the model; then the improved RepVGG module was introduced in the head network, which can help to improve the inference speed of the model and enhance the feature extraction capability of the network; finally, SIOU was used as the bounding box regression loss function to improve the accuracy and training speed of the model. The experimental results show that our improved YOLOv5 model achieved 97% mAP and 270 fps on our dataset. Compared with conventional methods, ours had the best results. The ablation experiments were conducted on the publicly available datasets PASCAL VOC and MS COCO, and their mAP@0.5 was improved by 2.4% and 3%, respectively. Additionally, our model improved the average detection rate for small targets on the MS COCO dataset by 2.4%, showing that it can effectively detect small target defects.

Keywords: laser welding defect; lithium battery pole; loss function; RepVGG; YOLOv5

check for updates

Citation: Yang, Y.; Zhou, Y.; Din, N.U.; Li, J.; He, Y.; Zhang, L. An Improved YOLOv5 Model for Detecting Laser Welding Defects of Lithium Battery Pole. *Appl. Sci.* **2023**, *13*, 2402. https://doi.org/10.3390/app13042402

Academic Editor: Guijun Bi

Received: 18 January 2023
Revised: 9 February 2023
Accepted: 10 February 2023
Published: 13 February 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

In recent years, new energy electric vehicles have gradually occupied the global automobile manufacturing market with their advantages of low energy consumption and low carbon emission. At the same time, the manufacturing process of new energy power batteries has also attracted widespread attention [1]. Defects will inevitably appear on the pole of new energy power batteries during the process of laser welding [2], which can cause great potential hazards to the production and safe usage of batteries. Therefore, it is of great significance to accurately detect these laser welding defects. Automatic optical inspection (AOI) systems are often used in industrial production to detect and process weld defects [3,4]. An AOI system mainly consists of three modules: the image acquisition module, image processing module, and image analysis module. The image acquisition module primarily consists of a CMOS industrial camera and LED light source [5]. The image analysis module is the most important link in the AOI system, and it is also the core module that determines whether the AOI system can efficiently identify weld defects. Usually, an AOI system mainly utilizes the defect detection algorithm to identify laser welding defects. Similarly, for the safe production of new energy power batteries, it is necessary to design a highly efficient laser welding defect detection algorithm for battery poles. Currently, with the rapid development of convolutional neural network (CNN) research and image processing algorithms, deep learning-based target detection has obtained considerable development and has gradually been used in AOI systems for defect detection. Deep learning-based target detection algorithms are divided into two categories: two-stage and one-stage target detection algorithms. One of the first two-stage target detection algorithms was the R-CNN proposed by Girshick et al. [6] in 2014. Subsequently, to improve the

R-CNN algorithm, Girshick et al. [7] proposed Fast R-CNN in 2015. In the same year, Ren SQ et al. [8] proposed Faster R-CNN, which further improved the detection speed of the model. Regarding one-stage target detection algorithms, Redmon J et al. [9] first proposed You Only Look Once (YOLO) in 2016. Single Shot Multibox Detector (SSD) was proposed by W Liu et al. [10] where the one-stage algorithm reduces the target detection problem to a regression problem without generating candidate frames, greatly increasing the model's speed and making it possible to deploy the target detection algorithm to the industry. The subsequent proposals of some algorithms such as YOLOv2 [11], YOLOv3 [12], YOLOv4 [13], and YOLOX [14] have led to a considerable improvement in the detection accuracy and speed of the model.

Among the aforementioned target detection algorithms, Faster R-CNN is the most mature target detection algorithm in the second stage algorithms, which is used in industrial production in large quantities. Kaihua Zhang et al. [15] proposed an improved Faster R-CNN algorithm by clustering to generate anchor frames and fused migration learning with ResNet-101 to solve the problem of inefficiency and the lack in the safety of lithium battery connector solder joint detection methods. Min-jae Jung et al. [16] proposed an automatic weld defect detection method based on the Faster R-CNN algorithm to address the problem of manual and costly weld quality inspection in the shipbuilding industry. Moyun Liu et al. [17] proposed an improved YOLO algorithm by designing an enhanced multi-scale feature module to enable the extracted feature maps to represent richer information, resulting in a better balance between the accuracy and speed of the model in the X-ray image weld defect dataset. Yu-Ting Li et al. [18] used ResNet-101 as a classifier based on the YOLOv2 algorithm to achieve the automatic detection of solder defects in the printed circuit board (PCB) dual inline package (DIP) process, reducing the cost of manual inspection. Jiexin Zheng et al. [19]. proposed a YOLOv3-based algorithm for steel surface defect detection using MobileNet as the backbone network with good performance on the NEU-DET dataset. Srihari M et al. [20] proposed an effective casting defect detection model based on YOLOv4 for improvement. Although early one-stage inspection algorithms achieved the basic accuracy required for defect detection in industry, the detection speed still needs improvement.

With the introduction of the YOLOv5 algorithm, it is finally possible to achieve a high accuracy and fast rate in the industry field of defect detection. Meng Zhang et al. [21]. solved the problem of complex background and difficult defect detection in solar cell images based on the YOLOv5 algorithm, but the model complexity increased, leading to a decrease in detection speed. Zhuang Li et al. [22] proposed a two-stage industrial defect detection framework based on the improved YOLOv5 model and Optimized-Inception-ResnetV2 to achieve the accurate identification of minor target defects on steel surfaces. However, the detection accuracy still fell short of the industrial production requirements. Dingming Yang et al. [23] proposed an improved pipeline weld defect detection algorithm for YOLOv5, which effectively improved the detection efficiency and basically met the accuracy and speed requirements for defect detection in the industry. Although the above methods have basically achieved defect detection in the industry, some problems remain unsolved. The main problems at present are: how to improve the detection accuracy while ensuring the model detection speed is fast enough and how to solve the problem that small target defects are challenging to detect.

However, with the rapid development of industrial automation manufacturing, the defect detection speed and accuracy have put forward higher requirements. Aiming to solve the rapidly expanding demand for detecting laser welding defects of a lithium battery pole, we developed a YOLOv5-based algorithm as an image analysis module for the AOI system. We did not use the officially provided pre-training weights, and all network models were trained from scratch. We compared the improved model with some influential algorithms such as YOLOv7 [24], YOLOv6 [25], and YOLOX, and the results show that our model has high accuracy and can meet the industrial demand for real-time detection.

2. Laser Welding Defect Detection Model for Lithium Battery Pole

2.1. YOLOv5 Algorithm

The official YOLOv5 source code provides four network models with increasing network depth and feature map width, namely YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x. To meet the industrial requirements for defect detection algorithms such ass real-time detection and easy deployment, we chose YOLOv5s as the infrastructure, which has the smallest model size and the fastest detection speed. Figure 1 shows the network structure of YOLOv5, which mainly includes four parts: Input, Backbone, Neck, and Head.

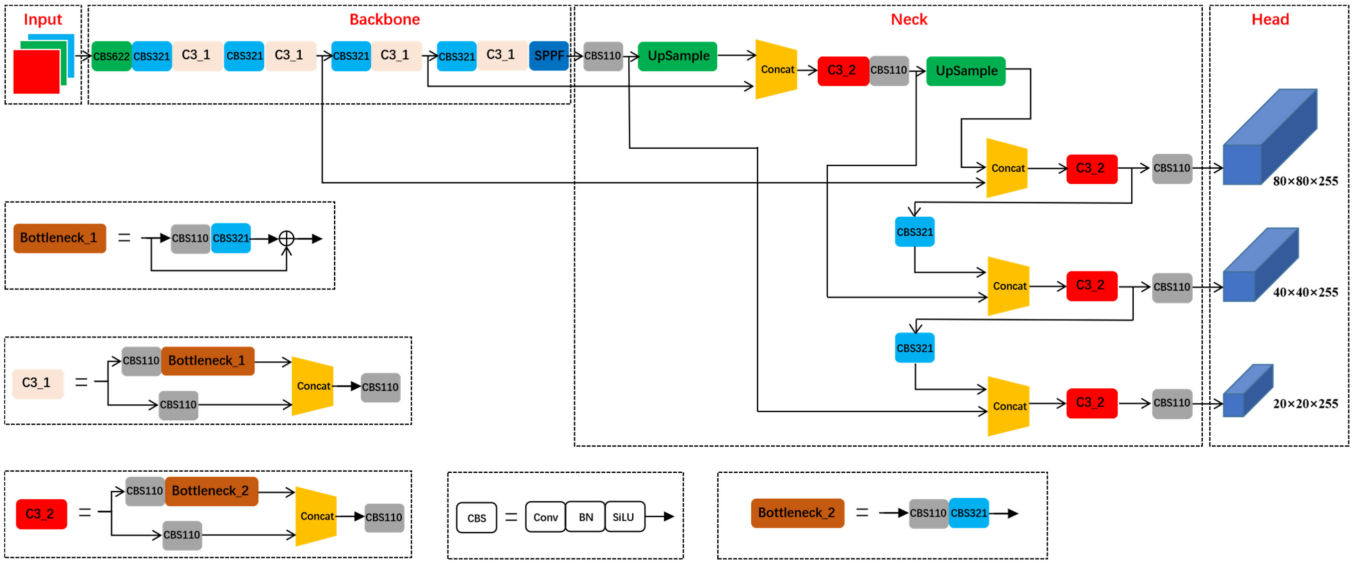


Figure 1. The network structure of the YOLOv5 model.

Input includes three modules: Adaptive image scaling, Mosaic data enhancement, and Adaptive anchor frame calculation. When images of different sizes are entered into Input, the algorithm first normalizes the image size to 640 × 640 by the adaptive image scaling technique. It then randomly selects four images for cropping, scaling, and layout processing by Mosaic data enhancement before finally inputting them into the network for training. The YOLOv5 algorithm provides nine priori anchor frames trained on the MS COCO dataset, calculates the maximum recall between the standard information of the dataset and the priori anchor frames before starting the training, and uses the K-means algorithm to re-cluster the anchor frames that best fit the dataset when the maximum recall is below 0.98.

Backbone is the backbone network used to extract the input image features, which mainly consists of Conv Batch Normalization SiLU (CBS), CSP bottleneck with 3 Conv (C3), and Spatial Pyramid Pooling Fast (SPPF) modules. Compared with the previous algorithms, the latest YOLOv5 algorithm uses CBS with 6 × 6 convolutional kernels to replace the original Focus module at the first layer of the network, reducing the training time cost. The first CBS was followed by four CBS, four C3 modules, and the SPPF module, which is in the last layer of the backbone network. The input image of 640 × 640 × 3 is first halved, and the number of channels is changed to four times by 6 × 6 convolutional kernels to obtain a 320 × 320 × 12 feature map and then fed into the CBS and C3 modules. Finally, the feature map was obtained by SPPF to extract advanced semantic information.

YOLOv5, like YOLOv4, still uses the feature pyramid network and pixel aggregation network (FPN-PAN) structure for its Neck part. The FPN structure transmits semantic information by an up-sampling operation from top to bottom, and the PAN structure transmits location information by a down-sampling process from bottom to top. Combining the two structures enables the network to fuse more feature information, constituting a multi-scale feature fusion module that can retain large-scale and small-scale target feature

information. When the input image is 640 × 640, the Head part outputs a feature map grid of 20 × 20, 40 × 40, and 80 × 80, representing its predicted small, medium, and large targets. At the same time, there are three anchor boxes for prediction at each scale, and finally, the prediction box with the highest confidence is filtered by non-maximum suppression.

2.2. The Improved YOLOv5 Model

The laser welding defects of a battery pole are irregular in shape, random in location, and vary in size. In addition, there are often many small target defects. In this case, the original YOLOv5 model cannot fully meet the detection requirements, and there is a low detection accuracy and high missing inspection rate. The improved YOLOv5 network model is shown in Figure 2. First, to improve the algorithm's ability to detect small target defects, we improved the backbone network by replacing the original 3 × 3 convolutional kernels with 6 × 6 convolutional kernels and replacing the original SPPF module in the last layer of the backbone network with our SPPSE module, then we introduced the lightweight convolutional neural network Re-Parameterization of Visual Geometry Group (RepVGG) [26] in the upper layer of the three detection heads. Finally, the original loss function CIoU [27] was modified to SIoU [28].

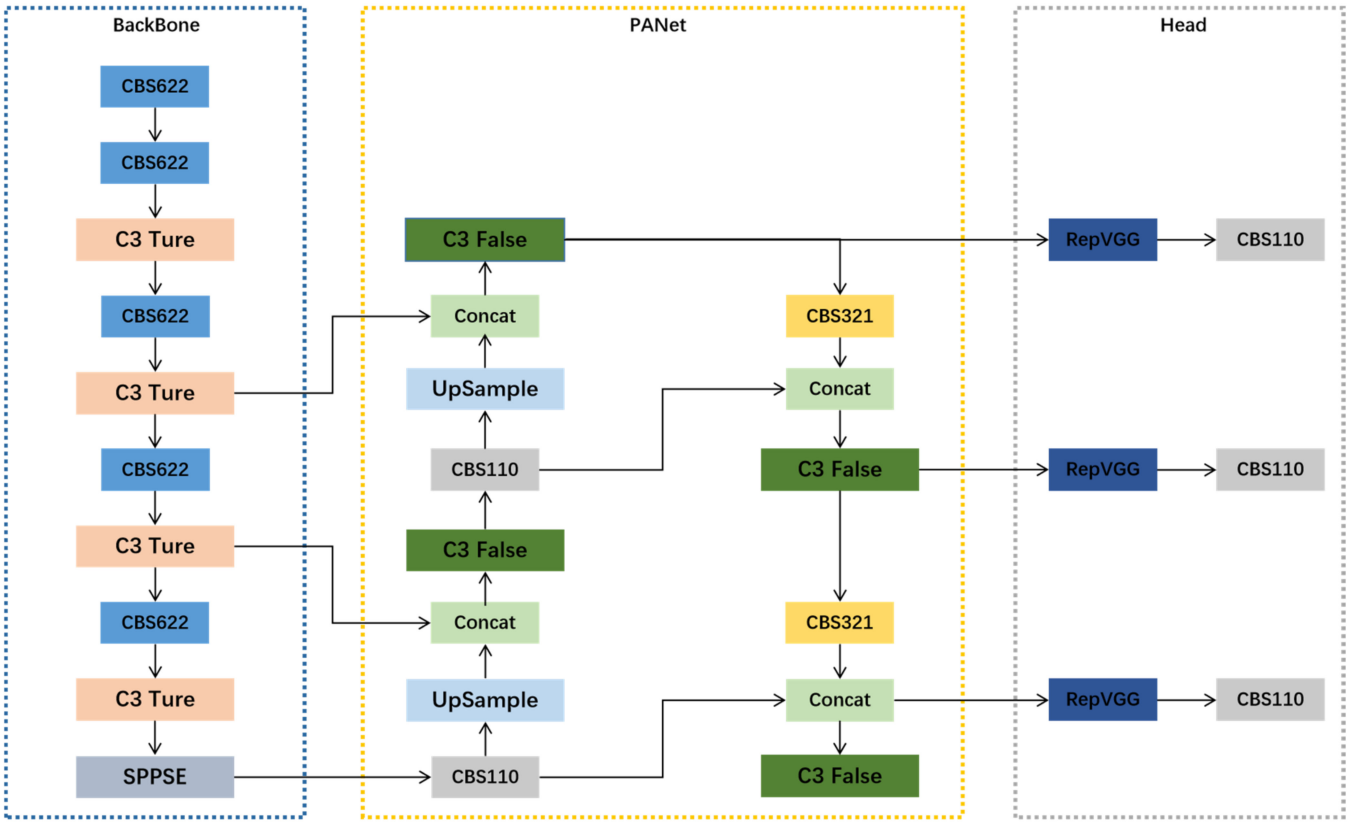


Figure 2. The architecture of the improved YOLOv5 model.

2.2.1. The Improved CBS Module

To improve the feature extraction capability of the backbone network for small targets of welding defects, we increased the convolutional kernels in the CBS module from 3 × 3 to 6 × 6. Using larger convolutional kernels can effectively improve the sensory field perception, obtain more contextual information, enable the backbone network to capture small low-level features more efficiently, and improve the small target detection capability of the model. The improved CBS module is shown in Figure 3, which consists of 6 × 6 convolutional kernels, BN, and activation function layers (SiLU) connected in series. BN significantly improves convergence without needing other forms of regularization; SiLU

has the properties of unbounded upper and lower bounds, smooth and non-monotonic, and outperformed ReLU on deep models.

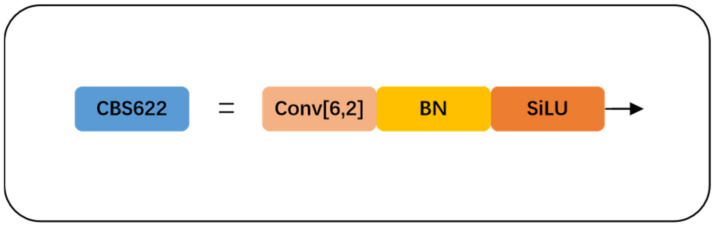


Figure 3. The improved CBS module.

2.2.2. The SPPSE Module

Spatial Pyramid Pooling (SPP) [29] was first proposed to avoid the problems of incomplete cropping and the shape distortion of image objects caused by the R-CNN algorithm for image region cropping and scaling operations, to solve the problem of repeated feature extraction of images by convolutional neural networks, to greatly improve the speed of generating candidate frames, and to save computational costs. To make the model adapt to images with different resolutions, we combined the Spatial Pyramid Pooling Cross Stage Partial Conv (SPPCSPC) module of YOLOv7 with the attention mechanism Squeeze and Excitation Network (SENet) [30], and we named it the SPPSE module, which is shown in Figure 4a. In the SPPSE module, the input features are divided into two parts, one for the convolution operation and the other for the SPP structure. Finally, the two elements are combined using Concat. This design reduces the computation by half and makes the model inference faster, and the accuracy is improved. Meanwhile, to select the key information in the current task and improve the efficiency and accuracy of image information processing, we added the attention mechanism SENet at the top of the SPPSE module, which can suppress other useless information from different channels and enhance the focus on the target region.

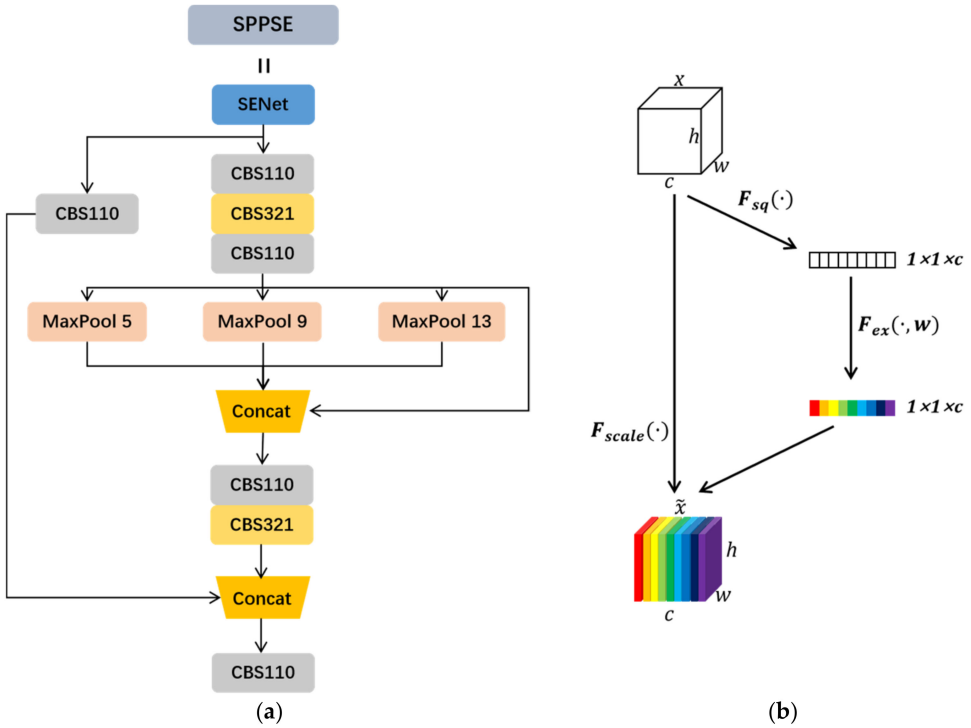


Figure 4. (**a**) The network structure of SPPSE. (**b**) The network structure of SENet.

As shown in Figure 4b, the attention mechanism SENet consists of squeeze and excitation. The model can determine the importance of each channel feature based on the interrelationship between different channels and then assign different weights to each channel to achieve the effect of the channel of interest and more important to the result.

2.2.3. The Improved RepVGG Module

RepVGG was proposed in 2021 based on the idea of re-parameterization. Its core technology is to enhance the feature extraction capability by re-parameterizing the structure and improving the inference speed by using a multi-branch design during training and a single-branch procedure during inference. The training and inference schematic of the RepVGG module is shown in Figure 5.

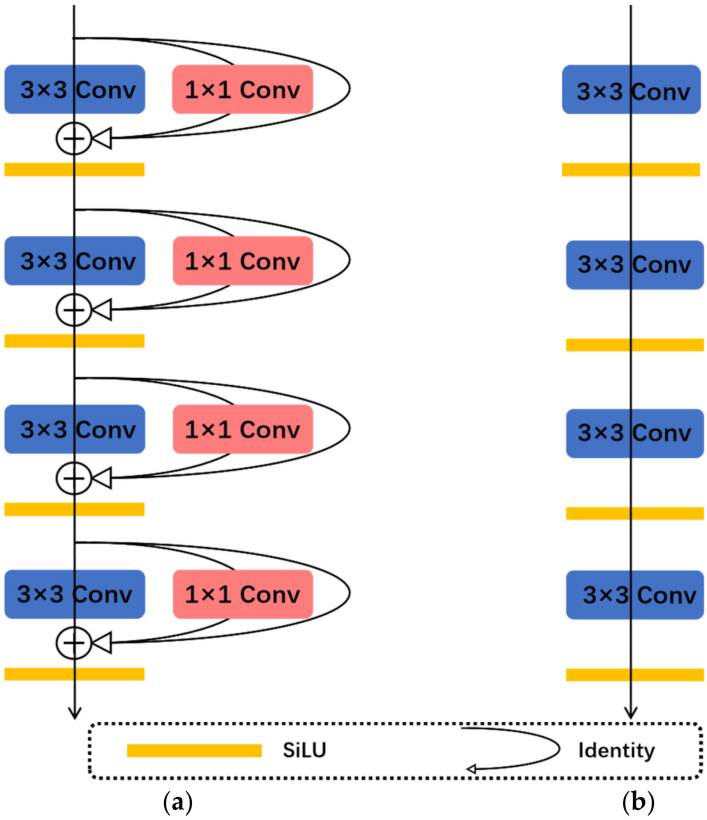


Figure 5. The improved RepVGG structure. (**a**) Stage of training. (**b**) Stage of inference.

As shown in Figure 5a, the RepVGG module uses a multi-branch structure consisting of 3 × 3 convolutional kernels, 1 × 1 convolutional kernel, and identity branches in the training phase. It enhances the model's ability and efficiency of extracting feature information by connecting the different perceptual fields obtained from each branch to an additive block and then extending the network downward and doing the same operation. When inference is made, the RepVGG model transforms the training model into the inference model, as shown in Figure 5b. The inference model is equivalent to the structure of multiple VGGs directly connected, which improves the inference speed of the model. Meanwhile, to improve the detection accuracy of the model, we used SiLU to replace the previous activation function ReLU.

2.2.4. The Improved Loss Function

The loss function of the YOLOv5 model is shown in Equation (1), and it consists of three components: confidence loss function L_{obj}, classification loss function L_{cls}, and bounding box regression loss function L_{box}.

$$LOSS = L_{obj} + L_{cls} + L_{box} \tag{1}$$

The YOLOv5 source code uses CIoU as the bounding box regression loss function. CIoU can take into account the dimensional information of the predicted frame and the real frame such as the overlap area, centroid distance, and aspect ratio, but it cannot truly reflect the difference between the confidence level, width, and height, so we chose to use SIoU instead. The SIoU loss function redefines the penalty measure to solve the problem of mismatch between the true and predicted frames and considers the vector angle between the expected regressions.

The SIoU [28] loss function mainly contains the following four parts: Angle cost Λ, Distance cost Δ, Shape cost Ω, and IoU cost. Angle cost is defined by Equation (2):

$$\Lambda = 1 - 2 \cdot \sin^2\left(\arcsin(x) - \frac{\pi}{4}\right) \tag{2}$$

where $x = \frac{c_h}{\sigma}$, σ is the distance between the center point of the prediction box and the ground truth box; c_h is the vertical distance between the center point of the prediction box and the ground truth box.

Considering the Angle cost defined above, the Distance cost is defined by Equation (3):

$$\Delta = \sum_{t=x,y} \left(1 - e^{-\gamma \rho_t}\right) \tag{3}$$

where $\rho_x = \left(\frac{b_{c_x}^{gt} - b_{c_x}}{c_w}\right)^2$, $\rho_y = \left(\frac{b_{c_y}^{gt} - b_{c_y}}{c_h}\right)^2$, $\gamma = 2 - \Lambda$, $b_{c_x}^{gt}$ and $b_{c_y}^{gt}$ are the horizontal and vertical coordinates of the center point of the ground truth box; b_{c_x} and b_{c_y} are the horizontal and vertical coordinates of the center point of the prediction box; c_w is the horizontal distance between the center point of the prediction box and the ground truth box. Shape cost is defined by Equation (4):

$$\Omega = \sum_{t=w,h} \left(1 - e^{-w_t}\right)^{\theta} \tag{4}$$

where $\omega_w = \frac{|w - w^{gt}|}{\max(w, w^{gt})}$, $\omega_h = \frac{|h - h^{gt}|}{\max(h, h^{gt})}$; w^{gt} and h^{gt} are the width and height of the ground truth box; w and h are the width and height of the prediction box.

The value of θ is calculated by the genetic algorithm, which determines how much attention should be paid to the cost of the shape. The value of θ is different in various datasets and ranges from 2 to 4. IoU cost is defined by Equation (5):

$$IoU = \frac{|B \cap B^{GT}|}{|B \cup B^{GT}|} \tag{5}$$

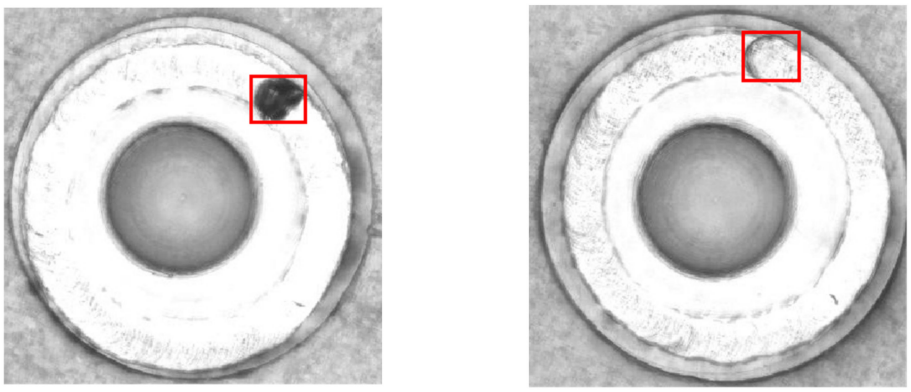
where B^{GT} is the area of the ground truth box; B is the area of the prediction box. The final total defining equation of the loss function SIoU is obtained as Equation (6).

$$SIoU_LOSS = 1 - IoU + \frac{\Delta + \Omega}{2} \tag{6}$$

3. Experiments and Analysis

3.1. Dataset Building and Experiment Design

We used a CMOS digital camera and collected 8439 laser welding defect images of a lithium battery pole of 2448 × 2048 size from the automatic laser welding line as the original dataset for the experiment. The defect images were appropriately cropped and then labeled according to the labeling format of the PASAL VOC dataset, and we used the software Labelimg V1.7 made by Max Tkachenko (New York, NY, USA) to obtain the Battery Pole Laser Welding Defect (BPLWD) dataset. As shown in Figure 6, welding hole (W) and bulge (B) are the most common types of weld defects in actual production.



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

Figure 6. Laser welding defects of the lithium battery pole. (**a**) Welding hole. (**b**) Bulge.

In BPLWD dataset, the ratio of the welding hole and bulge was about 5:3, which is shown in Table 1, and we divided the training set, validation set, and test set according to the ratio of 8:1:1. The environment configuration and training parameter settings for this experiment are shown in Tables 2 and 3.

Table 1. Composition of the BPLWD dataset.

Type	Number
Welding hole	5231
Bulge	3208

Table 2. Experimental setting.

Name	Parameter
Operating system	Ubuntu 20.04
CPU	Intel(R) Xeon(R) Silver 4110 CPU @ 2.10 GHz 16
GPU	NVIDIA GeForce GTX 2080Ti 11 GB
CUDA	Version 11.7
CUDNN	Version 8.5
Deep learning framework	PyTorch 1.12.1
Language	Python 3.7.13

Table 3. Parameter setting.

Name	Value
Size of picture	640 × 640
Epochs	200
Batch size	32
Learning rate	0.01
Momentum	0.937
Weight decay	0.0005

The evaluation metrics used in this experiment include AP, mAP@0.5, mAP@0.5:0.95, precision (P), recall (R), fps, GFLOPs, latency, and params. mAP@0.5 and mAP@0.5:0.95 refer to the average precision when the IoU is 0.5 and 0.5~0.95, respectively. Fps represents the number of images processed per second, GFLOPs refers to the number of floating point operations per second of 1 billion, latency denotes the inference time, and params is the number of parameters. As shown in Equations (7)–(10), accuracy P is the ratio of the number of correctly predicted positive samples to all predicted positive samples; recall R is the ratio of the number of correctly predicted positive samples to all positive samples; precision AP [31] is the area obtained by doing the integration operation on the graph enclosed by the P–R curve and the horizontal axis, and mAP is the average precision of all categories.

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

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

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

$$mAP = \frac{1}{k}\sum_{i=1}^{k} AP(i) \tag{10}$$

3.2. Experimental Results and Discussion

To verify the actual effect of the improved model, we validated the model performance with three datasets including PASCAL VOC [31], MS COCO [32], and our BPLWD dataset. PASCAL VOC contains 20 target categories with 21,503 images and MS COCO includes 80 target categories with a total of 123,287 images.

3.2.1. Experiments on BPLWD Dataset

We experimentally compared the improved YOLOv5 algorithm (YOLOv5-ours) with the more influential one-stage algorithms YOLOv5, YOLOv7, YOLOv6, and YOLOX in recent years, and the results are shown in Table 4 and Figure 7. Our improved YOLOv5 improved the average accuracy by 1.1% compared to the original model. It increased the precision and recall of the welding hole by 1.5% and 2%, respectively, and the recall of bulge by 1.6%. Despite the larger computation of the model, the inference speed remained within 2 ms, and the fps reached 270, which is fully capable of real-time industrial inspection. As shown in Figure 7, YOLOv5-ours had the highest average accuracy compared with the other four YOLO algorithms. This demonstrates that our improved model has good detection performance and can meet the industrial requirements for detecting the laser welding defects of a lithium battery pole.

Table 4. Comparison experiments of the different algorithms.

Algorithm	Category	P (%)	R (%)	AP (%)	mAP@0.5 (%)	fps	GFLOPs	Latency (ms)
YOLOV5-ours	W	95.2	93.0	96.4	97.0	270	37.9	2.0
	B	94.7	95.3	97.6				
YOLOv5	W	93.7	91.0	95.1	95.9	323	15.8	1.5
	B	96.0	93.7	96.7				
YOLOv7	W	93.0	88.9	94.6	95.4	122	103.2	6.7
	B	95.7	94.6	96.1				
YOLOv6	W	95.7	88.7	96.2	95.7	355	11.0	1.66
	B	92.5	93.2	96.0				
YOLOX	W	87.9	87.2	89.1	89.7	59	26.8	8.53
	B	87.4	89.1	90.4				

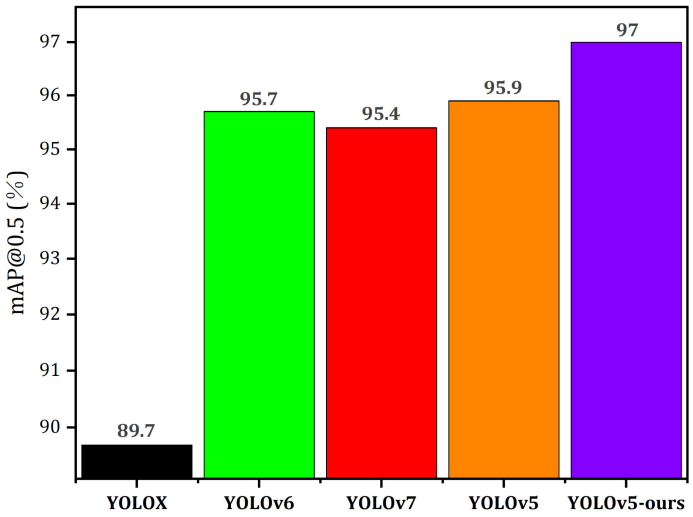


Figure 7. Comparison of the accuracy of different algorithms.

We selected four recently proposed lightweight CNN modules: MobileNetV3 [33], GhostNet [34], EfficientNet [35], ShuffleNetV2 [36], and replaced RepVGG in the YOLOv5-ours network with the above modules to compare the effects of the detection accuracy and performance of different modules. The experimental results are shown in Table 5, and the network using the RepVGG module had the highest detection accuracy and fast inference speed, indicating that the RepVGG module based on structural re-parameterization can increase the model's detection accuracy and shorten the inference time.

Table 5. Comparison of different lightweight models.

Algorithm	mAP@0.5 (%)	mAP@0.5:0.95 (%)	Latency (ms)
YOLOv5-ours(RepVGG)	97	61.2	1.7
YOLOv5-ours(MobileNetV3)	95.5	58.9	1.8
YOLOv5-ours(GhostNet)	95.3	60.1	3.5
YOLOv5-ours(EfficientNet)	95.5	60.6	3.6
YOLOv5-ours(ShuffleNetV2)	96.1	58.5	1.6

Additionally, we compared the defect detecting effects of the original YOLOv5 and our improved YOLOv5 algorithm, and the results are shown in Figure 8. Obviously, YOLOv5 had a low confidence level for the detection welding hole, and missed detecting the bulge

defects. Relatively, the YOLOv5-ours model not only successfully detected all of the defects, but also improved the confidence score. The results indicate that our improved algorithm is good at feature extraction and can better meet the industrial requirement of a low miss inspection rate for detection tasks.

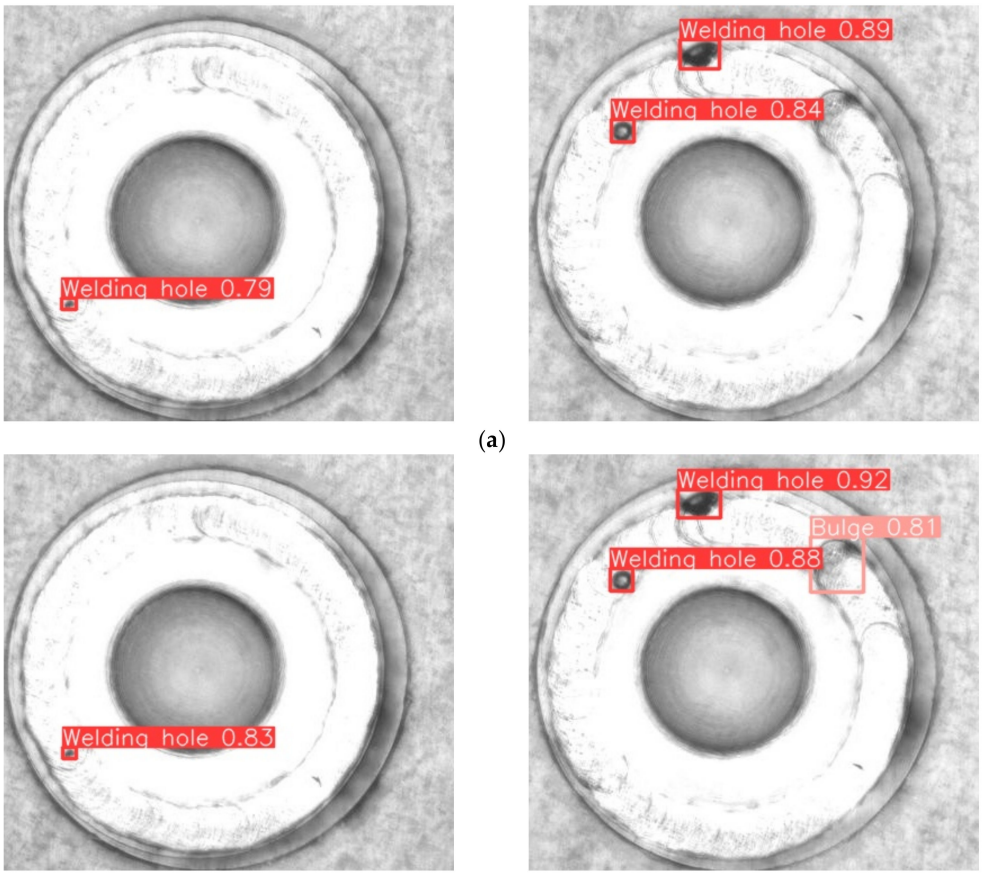


Figure 8. Comparison of the detection results of different models. (**a**) YOLOv5 model. (**b**) YOLOv5-ours model.

3.2.2. Experiments on the MS COCO and PASCAL VOC Datasets

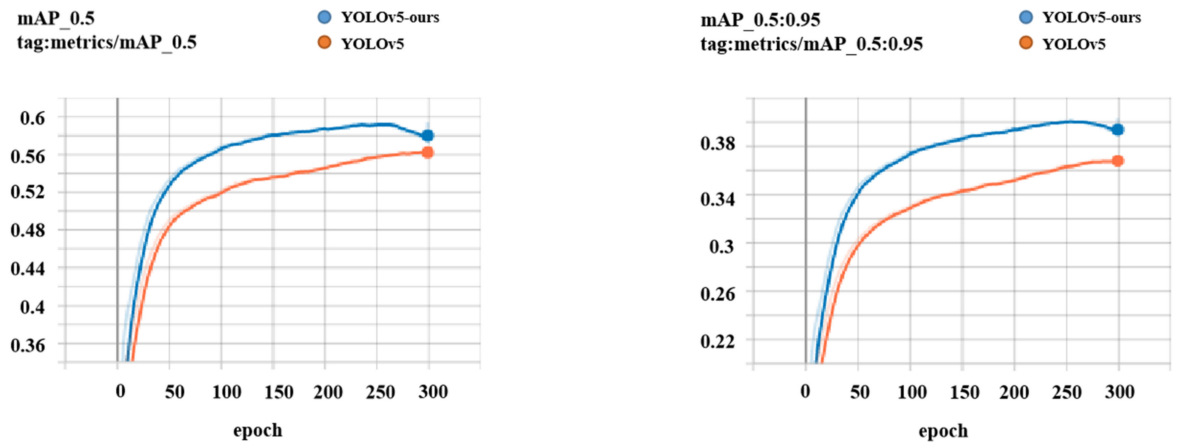
To verify the effectiveness of our improvement of the YOLOv5 algorithm, we conducted ablation experiments on the MS COCO and PASCAL VOC datasets. We set the training period to 300 epochs, and kept the other experimental settings the same as Table 2. The experimental results are shown in Table 6.

Table 6. Ablation experiments on the MS COCO and PASCAL VOC datasets.

Dataset	SPPSE	6 × 6	RepVGG	SIoU	Params (M)	fps	mAP@0.5 (%)	mAP@0.5:0.95 (%)	mAP@small (%)
MS COCO					7.2	357	56.8	37.0	20.4
	√				13.6	323	58.7	38.6	21.0
	√	√			18.4	313	59.1	39.3	21.6
	√	√	√		21.5	303	59.6	40.1	22.2
	√	√	√	√	21.5	303	59.8	40.3	22.8
PASCAL VOC					7.1	370	78.9	54.3	-
	√				13.5	333	80.0	56.1	-
	√	√			18.2	323	80.9	57.9	-
	√	√	√		21.3	286	81.1	58.1	-
	√	√	√	√	21.3	294	81.3	58.3	-

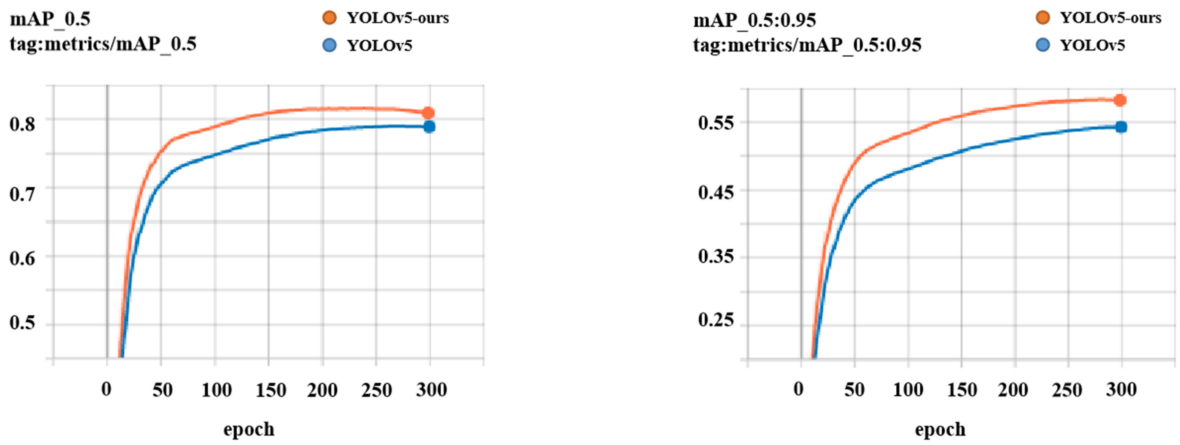
The ablation experiments were conducted with the YOLOv5 model as the baseline, and the marked tick represents the addition of this improvement item. From the experiments on MS COCO, it can be found that when the network model was replaced with the SPPSE module, mAP@0.5 increased by 1.9%, mAP@0.5:0.95 rose by 1.6%, and the average detection accuracy for small targets rose by 0.6%, indicating that SPPSE had the effect of focusing on small targets, and could thus increase the detection accuracy; in addition, the number of model parameters increased by 6.4 million (M) and the fps decreased from 357 to 323. After replacing the convolutional kernel of the backbone network to 6 × 6, mAP@0.5 increased by 0.4% and mAP@0.5:0.95 increased by 0.9%, the detection accuracy of small targets increased by 0.6%, the number of parameters increased by 4.8 M to 18.4 M, and the fps became 313, indicating that after expanding the convolutional kernel of the backbone network, its feature extraction ability was enhanced and the average detection accuracy was also improved; meanwhile, the complexity of the model did not increase significantly, and it could still achieve real-time detection. When the RepVGG module was added to the network model, mAP@0.5 and mAP0.5:0.95 increased by 0.5% and 0.8%, respectively, and the small target detection accuracy improved by 0.4%, the number of parameters reached 21.5 M, and the fps decreased by 10 to 303, indicating that the RepVGG module can increase the feature extraction ability of the detection head and then increase the average detection accuracy while improving the model's inference speed. Finally, after replacing the loss function with SIoU, mAP@0.5 and mAP@0.5:0.95 both improved by 0.2%. Although the average accuracy increase was not as good as other improvements, the small target detection accuracy was improved by 0.6%, the number of parameters and fps of the model remained unchanged, which indicates that SIoU can improve the detection accuracy without increasing the complexity of the model. The ablation experiments on the PASCAL VOC dataset can also demonstrate that each improvement to the original YOLOv5 model can improve the detection accuracy of the algorithm. Figure 9 more starkly shows the improvement in detection accuracy after we improved the model.

To further verify the superior performance of SIoU_Loss compared to CIoU_Loss, we compared their bounding box regression loss curves on the MS COCO dataset, as shown in Figure 10. SIoU_Loss had a faster convergence rate, dropping to below 0.03 after 50 epochs. When the training was over, SIoU_Loss had a lower value in the loss function, and the difference between the predicted and actual values of the model was more negligible, indicating that the model had superior performance and was better able at achieving laser welding defect detection.



(**a**)

Figure 9. *Cont.*



(**b**)

Figure 9. Comparison chart of mAP_0.5 and mAP_0.5:0.95 on different datasets. (**a**) MS COCO. (**b**) PASCAL VOC.

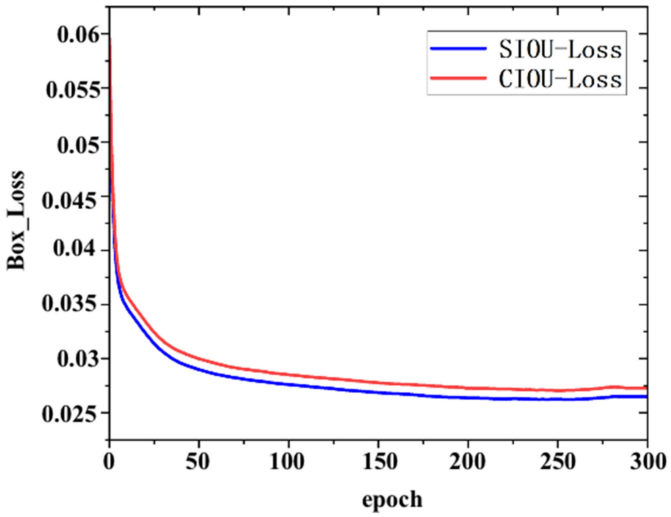


Figure 10. Comparison of the SIoU_Loss and CIoU_Loss curve.

4. Conclusions

To solve the problem of laser welding defect detection for lithium battery poles and to realize the automatic welding defect detection in the industry, this paper proposed an end-to-end improvement algorithm based on the YOLOv5 model. Given the fact that small targets in laser welding defects are difficult to detect, we suggest replacing the original 3×3 convolution kernel with a larger convolutional kernel; replacing the SPPF of the backbone network with our SPPSE module to improve the feature extraction capability; introducing the RepVGG module to enhance the accuracy and detection speed of the model; and finally, using SIoU as the bounding box regression loss function to reduce the negative impact of the mismatch between the predicted and real frames on the network. The experimental results showed that the improved model improved mAP by 1.1% to 97% relative to YOLOv5 on our BPLWD dataset, with a detection speed of 270 fps, and can meet the needs of the industry for high accuracy and speed of defect detection. Meanwhile, the model accuracy was improved by 3% in the MS COCO dataset and 2.4% in the PASCAL VOC dataset. The results prove that our improved model makes a good trade-off between detection accuracy and speed, which can meet the industrial requirements for defect detection when embedded into the image analysis module of the AOI system. In future work, we will focus on model compression and pruning to facilitate its subsequent deployment in industry.

Author Contributions: Conceptualization, Y.Y. and L.Z.; Methodology, Y.Z.; Software, Y.Z.; Validation, N.U.D., J.L. and Y.H.; Formal analysis, N.U.D.; Investigation, Y.Z.; Resources, Y.Y.; Data curation, J.L.; Writing—original draft preparation, Y.Z.; Writing—review and editing, L.Z.; Visualization, Y.H.; Supervision, L.Z.; Project administration, Y.Y.; Funding acquisition, Y.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Shenzhen Science and Technology Program (JCYJ20210324093806017), and Shenzhen-Hong Kong Joint Innovation Foundation (SGDX20190919094401725).

Data Availability Statement: Data is unavailable due to privacy or ethical restrictions.

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

References

1. Voulodimos, A.; Doulamis, N.; Doulamis, A.; Protopapadakis, E. Deep Learning for Computer Vision: A Brief Review. *Comput. Intell. Neurosci.* **2018**, *2018*, 7068349. [CrossRef] [PubMed]
2. Cui, L.; Zhang, Y.; He, D.; Li, X.; Jiang, J. Research progress of high power fiber laser welding. *Laser Technol.* **2012**, *36*, 154–159.
3. Xiao-bin, W.; Ji-hua, G.U.; Yong, Y.; Hao, Z.; Yong-bin, W.; Ya-juan, X.U. Automatic optical inspection for solder joints based on BP neural network. *Opt. Technol.* **2009**, *35*, 905–909.
4. Lu, R.S.; Shi, Y.Q.; Li, Q.; Yu, Q.P. AOI Techniques for Surface Defect Inspection. In Proceedings of the International Conference on Precision Instrumentation and Measurement 2010, Kiryu, Japan, 17–20 March 2010; pp. 297–302.
5. Yang, Y.T.; Pan, L.H.; Ma, J.X.; Yang, R.Z.; Zhu, Y.S.; Yang, Y.Z.; Zhang, L. A High-Performance Deep Learning Algorithm for the Automated Optical Inspection of Laser Welding. *Appl. Sci.* **2020**, *10*, 11. [CrossRef]
6. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In Proceedings of the 27th IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
7. Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 11–18 December 2015; pp. 1440–1448.
8. Ren, S.Q.; He, K.M.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [CrossRef] [PubMed]
9. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 27–30 June 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 779–788.
10. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. Ssd: Single Shot Multibox Detector. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; pp. 21–37.
11. Redmon, J.; Farhadi, A. YOLO9000: Better, Faster, Stronger. In Proceedings of the 30th IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 6517–6525.
12. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
13. Bochkovskiy, A.; Wang, C.-Y.; Liao, H.-Y.M. Yolov4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
14. Ge, Z.; Liu, S.; Wang, F.; Li, Z.; Sun, J. Yolox: Exceeding yolo series in 2021. *arXiv* **2021**, arXiv:2107.08430.
15. Zhang, K.; Shen, H. Solder joint defect detection in the connectors using improved faster-rcnn algorithm. *Appl. Sci.* **2021**, *11*, 576. [CrossRef]
16. Oh, S.-j.; Jung, M.-j.; Lim, C.; Shin, S.-c. Automatic detection of welding defects using faster R-CNN. *Appl. Sci.* **2020**, *10*, 8629. [CrossRef]
17. Liu, M.; Chen, Y.; He, L.; Zhang, Y.; Xie, J. LF-YOLO: A Lighter and Faster YOLO for Weld Defect Detection of X-ray Image. *arXiv* **2021**, arXiv:2110.15045.
18. Li, Y.-T.; Kuo, P.; Guo, J.-I. Automatic Industry PCB Board DIP Process Defect Detection with Deep Ensemble Method. In Proceedings of the 2020 IEEE 29th International Symposium on Industrial Electronics (ISIE), Delft, The Netherlands, 17–19 June 2020; pp. 453–459.
19. Zheng, J.; Zhuang, Z.; Liao, T.; Chen, L. Improved Yolo V3 for Steel Surface Defect Detection. In Proceedings of the International Conference on Computer Engineering and Networks, Hangzhou, China, 22–24 July 2022; pp. 729–739.
20. Raj, V.G.; Srihari, M.; Mohan, A. Casting defect detection using YOLO V4. *Int. Res. J. Mod. Eng. Technol. Sci.* **2021**, *3*, 1581–1585.
21. Zhang, M.; Yin, L. Solar cell surface defect detection based on improved YOLO v5. *IEEE Access* **2022**, *10*, 80804–80815. [CrossRef]
22. Li, Z.; Tian, X.; Liu, X.; Liu, Y.; Shi, X. A two-stage industrial defect detection framework based on improved-yolov5 and optimized-inception-resnetv2 models. *Appl. Sci.* **2022**, *12*, 834. [CrossRef]
23. Yang, D.; Cui, Y.; Yu, Z.; Yuan, H. Deep learning based steel pipe weld defect detection. *Appl. Artif. Intell.* **2021**, *35*, 1237–1249. [CrossRef]
24. Wang, C.-Y.; Bochkovskiy, A.; Liao, H.-Y.M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv* **2022**, arXiv:2207.02696.

25. Li, C.; Li, L.; Jiang, H.; Weng, K.; Geng, Y.; Li, L.; Ke, Z.; Li, Q.; Cheng, M.; Nie, W. YOLOv6: A single-stage object detection framework for industrial applications. *arXiv* **2022**, arXiv:2209.02976.
26. Ding, X.; Zhang, X.; Ma, N.; Han, J.; Ding, G.; Sun, J. Repvgg: Making vgg-Style Convnets Great Again. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 13733–13742.
27. Zheng, Z.; Wang, P.; Ren, D.; Liu, W.; Ye, R.; Hu, Q.; Zuo, W. Enhancing geometric factors in model learning and inference for object detection and instance segmentation. *IEEE Trans. Cybern.* **2021**, *52*, 8574–8586. [CrossRef] [PubMed]
28. Gevorgyan, Z. SIoU Loss: More Powerful Learning for Bounding Box Regression. *arXiv* **2022**, arXiv:2205.12740.
29. He, K.; Zhang, X.; Ren, S.; Sun, J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1904–1916. [CrossRef] [PubMed]
30. Hu, J.; Shen, L.; Sun, G. Squeeze-and-Excitation Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 7132–7141.
31. Everingham, M.; Winn, J. *The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Development Kit*; Technical Report for PASCAL—Pattern Analysis, Statistical Modelling and Computational Learning; European Commission: Brussels, Belgium, 2012; Volume 2007, 45p.
32. Lin, T.-Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft Coco: Common Objects in Context. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; pp. 740–755.
33. Howard, A.; Sandler, M.; Chu, G.; Chen, L.-C.; Chen, B.; Tan, M.; Wang, W.; Zhu, Y.; Pang, R.; Vasudevan, V. Searching for Mobilenetv3. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 1314–1324.
34. Han, K.; Wang, Y.; Tian, Q.; Guo, J.; Xu, C.; Xu, C. Ghostnet: More Features from Cheap Operations. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 1580–1589.
35. Tan, M.; Le, Q. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In Proceedings of the 36th International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019; pp. 6105–6114.
36. Ma, N.; Zhang, X.; Zheng, H.-T.; Sun, J. Shufflenet v2: Practical Guidelines for Efficient cnn Architecture Design. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 116–131.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.