

Article

Photometric-Stereo-Based Defect Detection System for Metal Parts

Yanlong Cao ^{1,2}, Binjie Ding ^{1,2}, Jingxi Chen ^{1,2}, Wenyuan Liu ^{1,2}, Pengning Guo ^{1,2}, Liuyi Huang ^{3,4} and Jiangxin Yang ^{1,2,*}

¹ State Key Laboratory of Fluid Power and Mechatronic Systems, School of Mechanical Engineering, Zhejiang University, Hangzhou 310058, China

² Key Laboratory of Advanced Manufacturing Technology of Zhejiang Province, School of Mechanical Engineering, Zhejiang University, Hangzhou 310058, China

³ Zhejiang Academy of Special Equipment Science, Zhejiang University, Hangzhou 310023, China

⁴ Key Laboratory of Special Equipment Safety Testing Technology of Zhejiang Province, Hangzhou 310023, China

* Correspondence: yangjx@zju.edu.cn

Abstract: Automated inspection technology based on computer vision is now widely used in the manufacturing industry with high speed and accuracy. However, metal parts always appear in high gloss or shadow on the surface, resulting in the overexposure of the captured images. It is necessary to adjust the light direction and view to keep defects out of overexposure and shadow areas. However, it is too tedious to adjust the position of the light direction and view the variety of parts' geometries. To address this problem, we design a photometric-stereo-based defect detection system (PSBDDS), which combines the photometric stereo with defect detection to eliminate the interference of highlights and shadows. Based on the PSBDDS, we introduce a photometric-stereo-based defect detection framework, which takes images captured in multiple directional lights as input and obtains the normal map through the photometric stereo model. Then, the detection model uses the normal map as input to locate and classify defects. Existing learning-based photometric stereo methods and defect detection methods have achieved good performance in their respective fields. However, photometric stereo datasets and defect detection datasets are not sufficient for training and testing photometric-stereo-based defect detection methods, thus we create a photometric stereo defect detection (PSDD) dataset using our PSBDDS to eliminate gaps between learning-based photometric stereo and defect detection methods. Furthermore, experimental results prove the effectiveness of the proposed PSBDD and PSDD dataset.

Keywords: photometric stereo; defect detection; convolutional neural network



Citation: Cao, Y.; Ding, B.; Chen, J.; Liu, W.; Guo, P.; Huang L.; Yang, J. Photometric-Stereo-Based Defect Detection System for Metal Parts. *Sensors* **2022**, *22*, 8374. <https://doi.org/10.3390/s22218374>

Academic Editor: Steve Vanlanduit

Received: 6 October 2022

Accepted: 27 October 2022

Published: 1 November 2022

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



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Quality control is an essential part of the manufacturing industry, which helps standardize both production and reactions to quality issues. Most of the current inspection work is mainly done manually by visual inspection. The results are heavily influenced by subjective human factors, leading to unstable production quality control. The development of computer vision has greatly alleviated this problem. Automated inspection technology based on computer vision is one of the popular ways for efficient and high-precision defect detection through processing and analysing the images. Traditional defect detection methods based on computer vision mainly use techniques such as edge detection [1,2], grey-scale thresholding [3,4], and image segmentation [5,6]. However, traditional ones have limitations such as their reliance on artificially defined features. In recent years, numerous learning-based techniques have been proposed for defect detection, which transform data into complex and abstract representations that enable the features to be learned during

the training phase to overcome the requirement of predefined features. A necessary condition for defect detection using computer vision algorithms is the ability to extract the characteristics of the defects in the image. However, the high reflective properties of the metal surfaces cause overexposure in some areas of the captured image and override the characteristics of the defect, as shown in Figure 1.

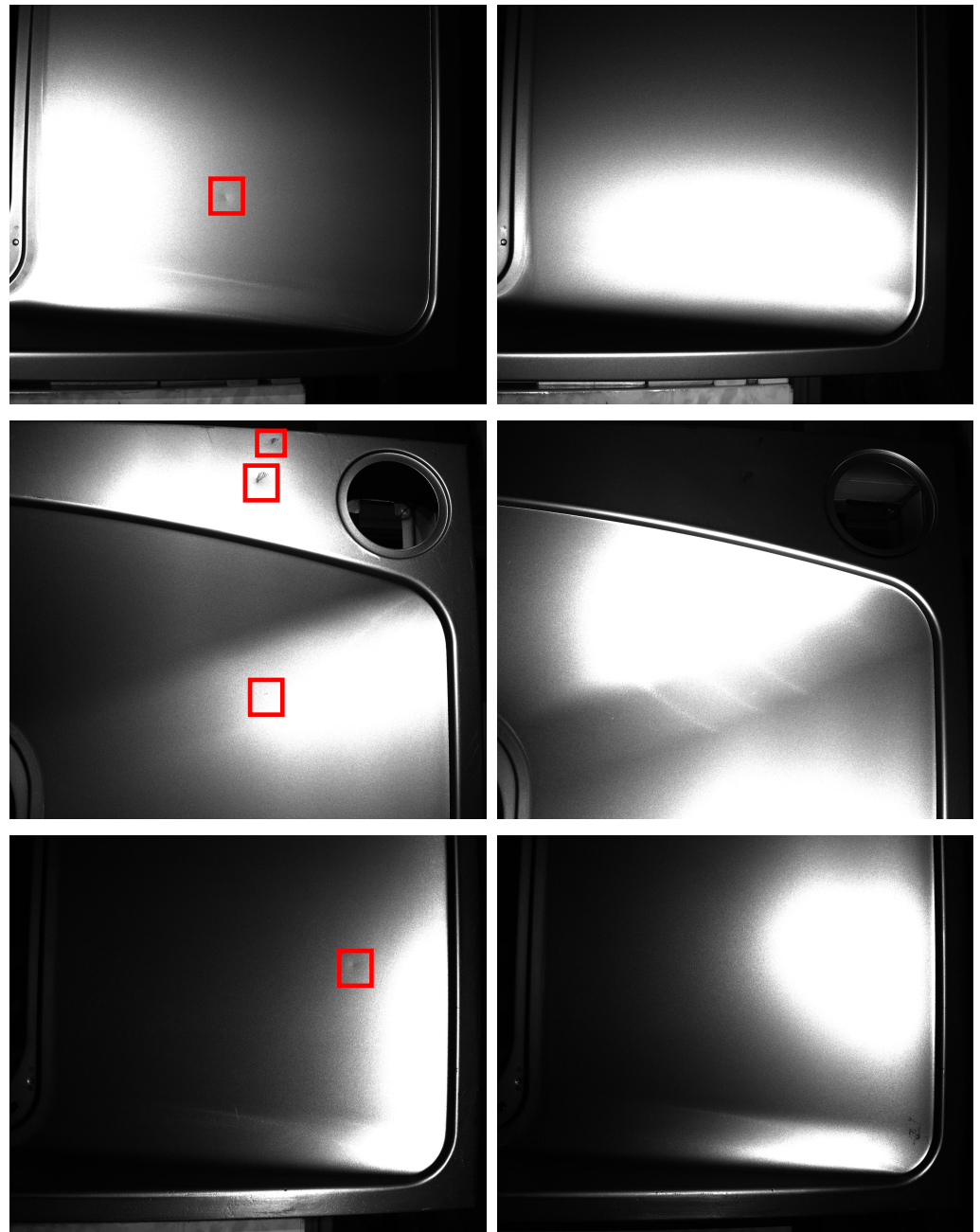


Figure 1. Images of stationary metal objects under different lighting conditions. Each row represents images of an object captured under different directional light. The red box indicates the location of the defect. (**Left**): defect visible. (**Right**): defect in overexposed or shadow area.

When the angular bisector of the incident light and the view coincide with the normal vector, the observation points appear glossy. The surface prevents the light from falling on itself, which leads to the attached shadow in the observation area. When the surface prevents the light from falling on another surface, the observation are appears with a cast shadow [7]. However, as the shape of the object is unknown, the way to avoid defects

being in the highlight or shadow areas is to adjust the angle between the direction of light incidence and the direction of view. However, it is too tedious to adjust the light direction and the view with the variety of shapes of metal parts. It is therefore effective and convenient to reduce the probability of defects remaining in the highlight or shadow areas by illuminating the object in different directions. Furthermore, we introduce photometric stereo in the metal defect detection task, which removes the interference of highlights by using the information from multiple illuminations.

In recent years, photometric stereo has received significant attention in the field of advanced manufacturing [8–12], which estimates the accurate and highly detailed surface normal of a target object based on a set of images captured in different light directions using a fixed-viewpoint camera. The basic theory of photometric stereo was first proposed by Woodham et al. [13] based on the assumption of ideal Lambertian reflectance in the 1980s. However, there are few objects in reality that fit the ideal Lambertian assumption. Therefore, over the last forty years, numerous methods aimed at non-Lambertian objects have been proposed to improve the applicability of photometric stereo. Before the development of deep learning, most traditional photometric stereo methods treated the non-Lambertian pixels as outliers [14–18]. However, these methods only worked for objects with sparse non-Lambertian regions. To handle dense non-Lambertian reflections, some methods used analytical or empirical models to simulate reflection on the surface models [19–25]. However, this type of method is only applicable to a limited range of materials. In recent years, numerous deep-learning-based photometric stereo methods have been proposed to handle non-Lambertian objects in reality. According to how learning-based methods process the input image, methods can be divided into per-pixel [26–29] and all-pixel methods [30–33]. Per-pixel methods take the observed intensities as the input and output a surface normal for a single pixel. Santo et al. [34] first introduced deep learning into photometric stereo and propose a deep photometric stereo network (DPSN), which assumed that the light direction stayed the same during the testing and training. To generalize the learning-based per-pixel methods, Ikehata projected the observation intensities onto a 2D space and proposed an observation map, which rearranged the observation intensities according to light directions. The operation of the observation map enabled per-pixel methods to handle the arbitrary number of directions and order-agnostic lighting. Numerous methods, such as CNN-PS [26], LMPS [28], and SPLINET-Net [29], used this data preprocessing method in their neural networks and achieved great performance on the DiLiGenT benchmark dataset. All-pixel methods take whole images or patches as well as their corresponding light directions as input and output a surface map with the same resolution as the input. To handle order-agnostic light, Chen et al. [30] proposed an all-pixel photometric stereo network named PS-FCN, which used a sharing-weight feature extractor to handle per-input image and used max pooling to aggregate the extracted features from multiple images. Based on this strategy, SDPS-Net [32] and GC-Net [35] were proposed to handle the uncalibrated photometric stereo task. According to the summary by Zheng et al. [36], per-pixel methods are robust to nonuniform distributions of surface materials but cannot handle well regions with global illuminations effects (shadows or inter-reflection) since shape information is not explicitly considered. Because all-pixels methods are generally trained using various shapes with a uniform material for each shape, they perform well for regions with a global illumination effect, while they are ineffective for objects with nonuniform materials. There are several methods to improve performance by combining the per-pixel and all-pixel methods. SPS-Net used the attention mechanism to extract photometric information, then applied convolution layers to extract spatial information. Yao et al. [37] proposed a graph-based photometric stereo network (GPS-Net) that first introduced structure-aware graph convolution filters to explore per-pixel information and classical convolution to explore spatial information. Ikehata [38] applied Transformer [39] to extract per-pixel and spatial information and proposed PS-Transformer, which aimed to handle the sparse photometric stereo task. Both defect detection methods and photometric stereo methods achieve outstanding performance on their respective benchmark. There are also methods [8,9,40,41]

that combine photometric stereo with defect detection. Ren et al. [42] used a data-driven PS method to extract the surface normal and separate defects from the background through filters. The method could only locate defects but not classify them and could not be applied when the defects and the background had relatively similar characteristics in the frequency domain. Feyza et al. [43] used an L2 photometric stereo [13] method to estimate the albedo map and normal map of objects. Furthermore, they took the combination of the albedo map and the normal map as input and used a CNN to detect the defects. The L2 photometric method is based on the Lambertian assumption and does not apply to most objects in reality. Defect detection and photometric stereo methods have achieved outstanding performance on their respective benchmarks with the development of deep learning. However, there are few methods that combine learning-based photometric stereo methods with learning-based defect detection methods. This is because most of the current defect detection data are captured under a single illumination and cannot be applied to photometric stereo. Meanwhile, existing photometric stereo datasets lack defect samples.

To address this problem, we design a **photometric-stereo-based defect detection system (PSBDDS)** to detect defects using photometric information under multilight illumination and create a **photometric stereo defect detection dataset (PSDD)** for combining existing defect detection networks with photometric stereo networks. Compared to existing methods, our method can estimate the most realistic objects with non-Lambertian surfaces while being effective in locating and classifying multiple types of defects. We first estimate the normal map from images, captured by the PSBDDS, through the non-Lambertian photometric stereo network and feed the normal map into the defect detection network. In summary, the contributions of this paper are as follows. (1) We build a photometric stereo defect detection dataset for the integration of existing photometric stereo methods with defect detection methods. (2) We design a photometric-stereo-based defect detection system to collect photometric data on the surface of metal parts and apply the photometric-stereo-based defect detection method. (3) We experimentally validate that our dataset can be used to combine the state-of-the-art photometric stereo methods with the advanced defect detection methods for highly accurate and robust defect detection performance.

2. Method

2.1. Photometric-Stereo-Based Defect Detection System (PSBDDS)

PSBDDS consists of hardware and control modules.

Hardware. Our hardware module consisted of photometric stereo data capture equipment, a manipulator, and a profile frame, as shown in Figure 2. We mounted the PS data capture equipment at the end of a six-degree-of-freedom manipulator, which could adjust the posture and trajectory to capture data at target points. As shown in Figure 3, the photometric stereo data acquisition device consisted of 18 LED beads and an industrial camera. The inspection system of photometric stereo needs to take several pictures in a short time. We thus chose industrial cameras equipped with CMOS sensors with their advantages in high-speed imaging. We set the inspection area size to $v_x = 150$ mm, $v_y = 120$ mm and the inspection accuracy to 0.5 mm. Based on our experience in the design of computer-vision-based defect detection systems, 5~10 pixels were required to characterize the accuracy of the detection. Therefore, the corresponding pixel accuracy was approximately 0.05~0.1 mm. We took the middle value of the pixel accuracy $a_{pixel} = 0.08$ mm for calculating the resolution of the camera.

$$p_x = \frac{v_x}{a_{pixel}} = \frac{150}{0.08} = 1875 \quad (1)$$

$$p_y = \frac{v_y}{a_{pixel}} = \frac{120}{0.08} = 1500 \quad (2)$$

where p_x and p_y represent the minimum length and width of the resolution. According to the calculation result, we chose an industrial camera of 5 megapixels and a 2448×2048 resolution. We chose a fixed-focus lens with a focal length of 25 mm and an aperture range of f2.8~16.

The light source was also a key part of the data acquisition system. Following the general settings on the photometric stereo, we used the power of 3 W LED beads as the light source and installed them evenly in a rectangle frame; the aspect ratio of the frame was the same as the camera's field of view.

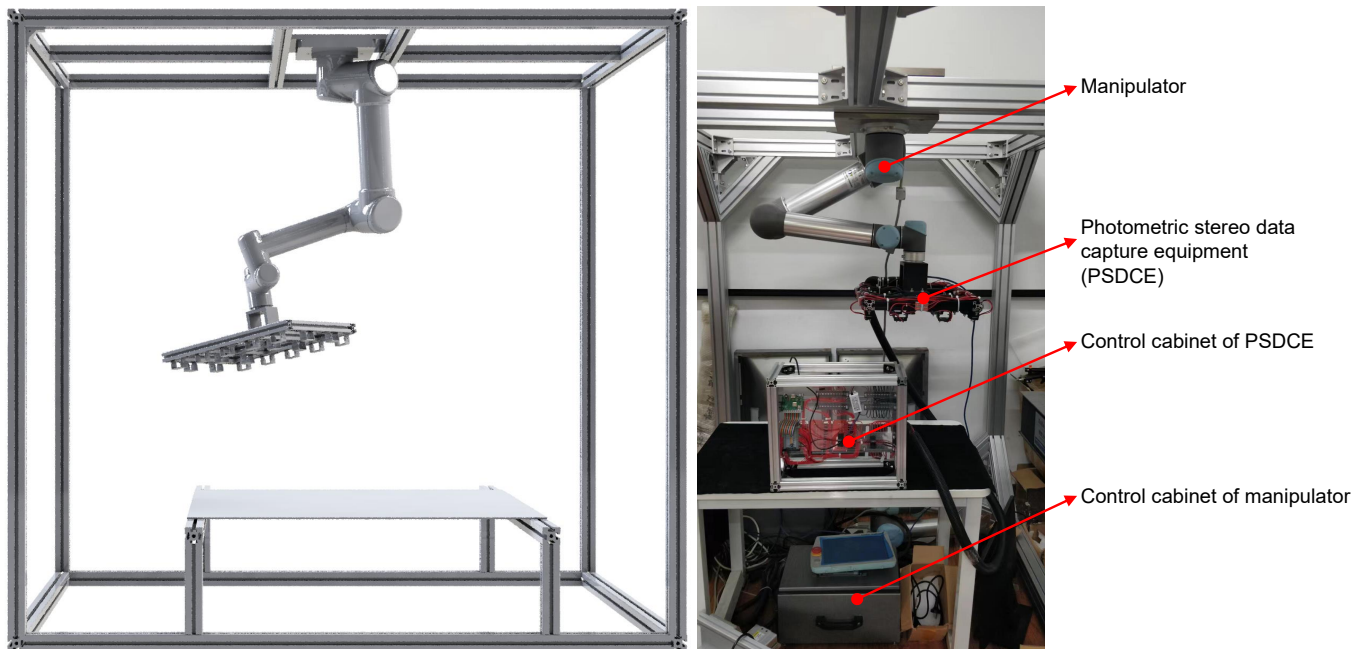


Figure 2. The diagram of our PSBDDS. The image on the left is a 3D model and the image on the right is a real device. The manipulator is mounted on a frame made of aluminium profiles and the PSDCE is attached to the end of the manipulator for capturing photometric stereo. The PSDCE and the manipulator are controlled by their corresponding control cabinets.

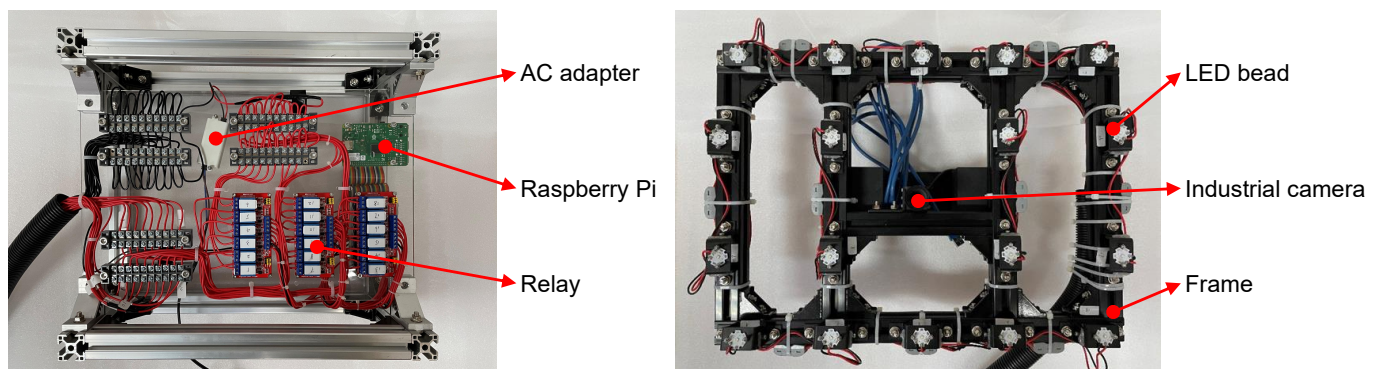


Figure 3. Diagram of the PSDCE and its control cabinet. The image on the left is a control cabinet. We used the Raspberry Pi to control the LED. As the IO port voltage of the Raspberry Pi cannot meet the needs of all LED beads, we used an AC adapter in conjunction with a relay to power LED. The image on the right is the PSDCE. The frame has been blackened to prevent reflective interference. The LED beads are mounted on the frame and the industrial camera is mounted at the centre of the frame.

Control modules. The control system of the defect detection system was divided into two parts: the control system of the PSDCE, as shown in Figure 3 and the control of the manipulator. The manipulator had existing control boxes. After establishing a connection with the host computer, they received manipulator control commands sent from the host computer and controlled the manipulator to achieve fast and accurate movement. Figure 4 shows the control scheme for hardware. We chose to use a Raspberry Pi as the controller between the PC and the photometric stereo data capture equipment, using the GPIO pins to control the switch of the LEDs and send the data capture command to the camera. The industrial camera had an external trigger line that could be connected directly to the GPIO pins of the Raspberry Pi for image capture triggering. The Raspberry Pi used an Ethernet interface for data interaction with the host PC. A TCP/IP connection between the host computer and the device was established via the socket interface to enable information transfer. However, the output voltage of the GPIO pins of the Raspberry Pi could not meet the demand for LED power. We thus used a constant current supply to provide a steady current and GPIO control relays to switch the circuit.

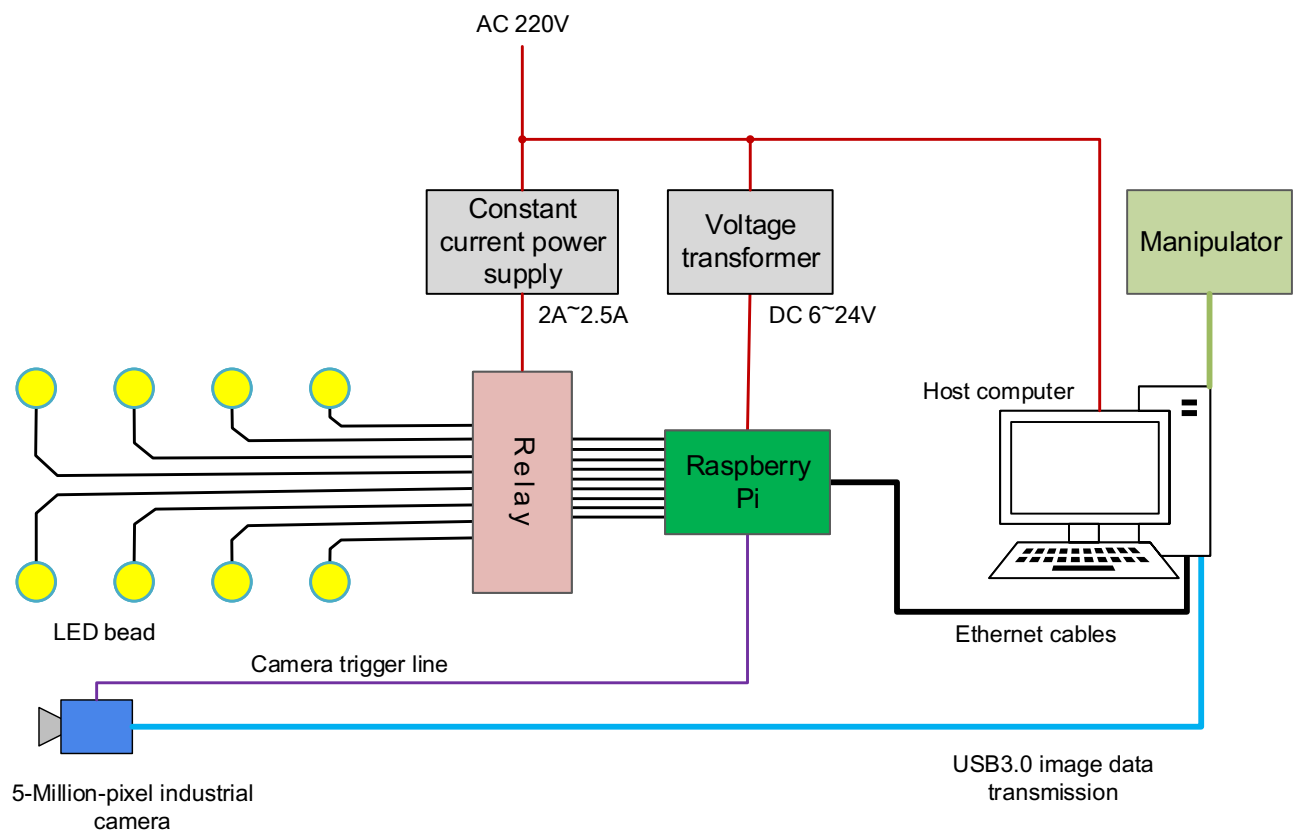


Figure 4. The control scheme for hardware. The host computer controls the communication among the manipulator, the Raspberry Pi, and the industrial camera. The Raspberry Pi controls the operation of the LED beads.

2.2. Dataset

We used four types of objects to produce our dataset, namely *yellow hemisphere*, *silver hemisphere*, *yellow cone*, and *silver cone*, as shown in Figure 5. The objects contained both curved and flat samples, made of 304 stainless steel and had a wide range of high-gloss areas on the surface. We created defects on the surfaces of objects by simulating the impact of the production process. We used 18 different directional lights to illuminate the object and captured the observed images with a resolution of 400×400 . The photometric stereo model used images or pixels under different illuminations as input to estimate the normal map, thus processing 18 high-resolution images together would consume huge computational resources as well as significantly slow down the computation. Most of the existing defect detection methods generate region proposals through enumeration or sliding windows or split the image into patches for processing, thus the high-resolution images would also consume a lot of time for the detection method.

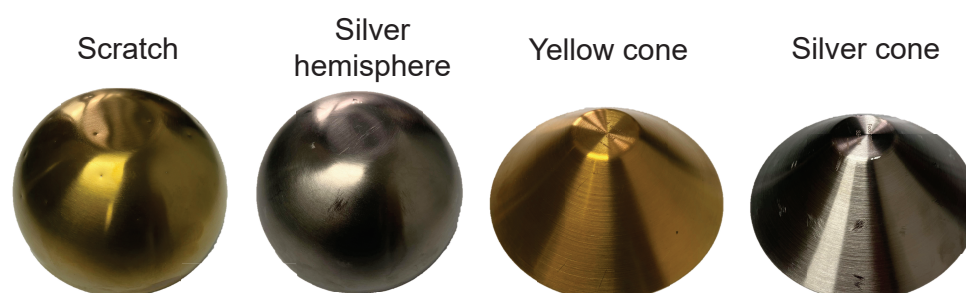


Figure 5. Objects for datasets. The dataset contained two colours (yellow and silver) and two shapes (hemisphere and cone) of objects. The material of the objects was 304 stainless steel. The two objects on the left are hemispheres with a diameter of 16 cm and a height of 5 cm. The two on the right are cones with a diameter of 15 cm and a height of 6 cm.

To improve the efficiency of the detection, we used sliding windows to split the whole image into multiple low-resolution images. The process of making the dataset is shown in Figure 6. The split images are shown in Figure 7. We obtained a total of 2017 defective samples, including 924 **stamp** samples, 542 **scratch** samples, and 551 **abraded** samples. The minimum **stamp** size was 0.45×0.87 mm, the minimum **scratch** size was 0.39×2.21 mm, and the minimum **abraded** size was 1.73×2.78 mm. Each sample contained 18 images with a corresponding light direction and intensity. We divide the dataset into a training set, a validation set, and a test set based on a 3:1:1 split. The number of samples in each category is shown in Table 1.

Table 1. The number of samples for each type of defect in the dataset.

	Stamp	Scratch	Abrade	Total
Training set	548	325	329	1202
Validation set	190	112	110	412
Test set	186	105	112	403

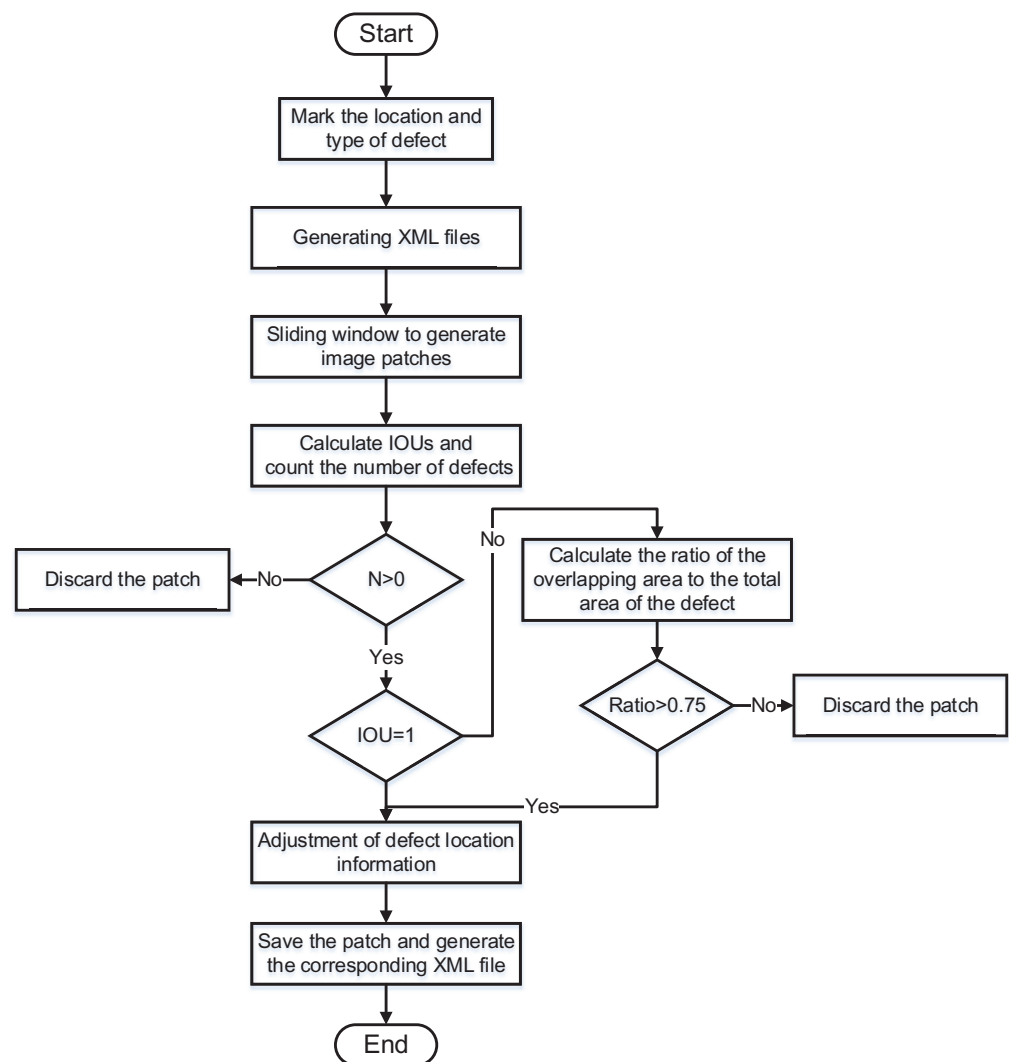


Figure 6. Flowchart of dataset processing. First, we use the labelling tool LabellImg to label the rectangular box in the target area of the defect. We experimentally set the initial size of the sliding window to 400×400 and the step size to 200. The algorithm reads the annotation file of each image, gets the location and category information for each defect in the image, and traverses the image starting from the top left corner of the image. The algorithm first determines whether there is a defective target in the area of the window. If there is no defect, the cropped image is discarded. Otherwise, the defect area information is saved. Then, the algorithm determines whether there is a window containing an incomplete defect. If not, the defect location information is converted to the image coordinate system of the window and saved to generate a new annotation file. When there is an incomplete defect area, the algorithm calculates the ratio of that defect area in the window to the area of the actually marked defect. When the ratio is greater than 0.75, the annotation information of the defect is adjusted and the window boundary is used as the new boundary. When the ratio is less than 0.75, the defect is discarded, and the window image is then saved and a new annotation file is generated.

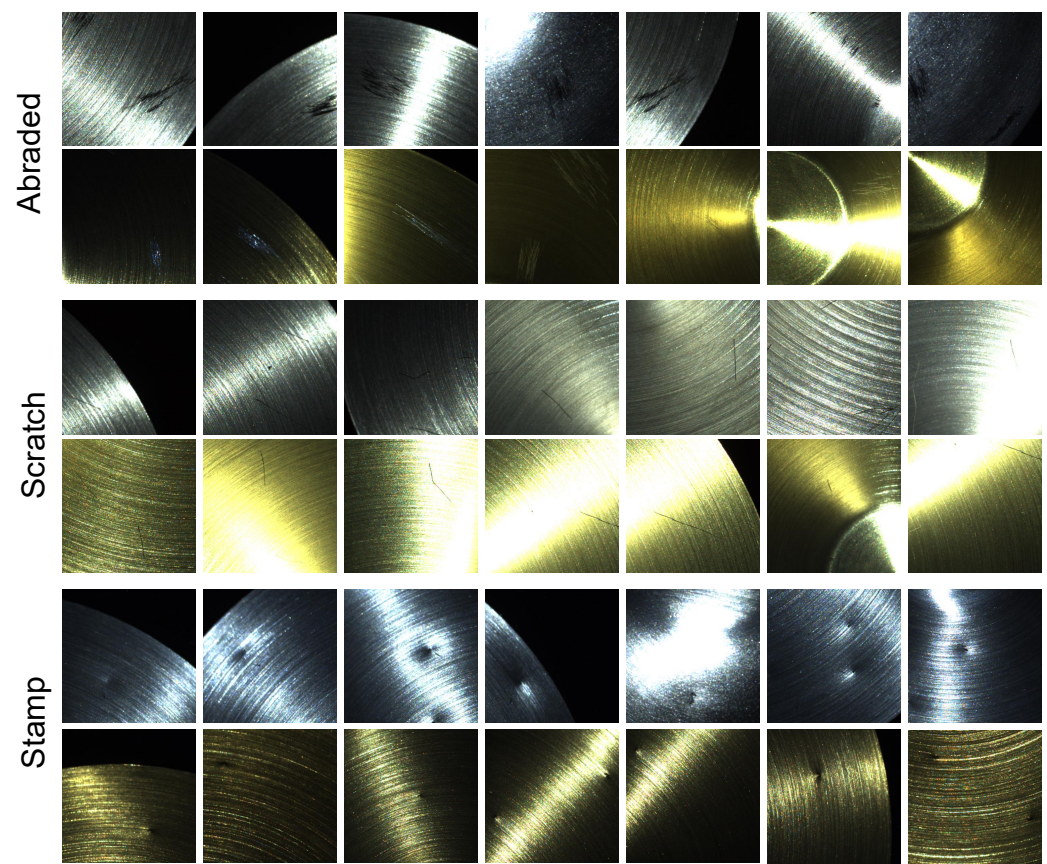


Figure 7. Diagram of the defects. Our dataset contains three types of defects. Rows 1 and 2 are samples of abraded defects. The abraded defects occur in patches and are irregular in shape. Rows 3 and 4 are samples of scratch. The scratches appear individually and are narrow and long. Rows 5 and 6 are samples of the stamp. Stamps usually appear to be circular, with a significant variation in depth.

2.3. Detection Framework

As illustrated in Figure 8, our detection framework consisted of two components, the photometric stereo module and the defect detection module. The photometric stereo module took captured images and corresponding light maps as input and estimated the normal map. Then, the defect detection module used the normal map as input to locate and classify defects. We chose MT-PS-CNN [44] as our photometric stereo model for estimating normal maps. Compared to all-pixel methods (i.e., PS-FCN [30], SDPS-Net [32]) and per-pixel methods (i.e., DPSN [34], CNN-PS [26]), MT-PS-CNN combines the advantages of both types of methods through a 3-D convolution, effectively eliminating the interference highlights, and it performs well with sparse inputs. For a detailed description of MT-PS-CNN, please refer to [44].

The number of LED beads has a huge impact on the estimation of the normal map. More LED beads mean more valid clues. However, a large number of image inputs reduces the efficiency of the model computation. Therefore, we analysed the performance of MT-PS-CNN with different numbers of inputs to find a trade-off between the accuracy of the normal estimation and computational speed, as shown in Figure 9. We found that the normal estimation error rose rapidly when the number of inputs was less than 18. The model prediction time grew linearly with the number of inputs. Combining photography time, computational resource consumption, normal solution accuracy, and model prediction speed, we finally chose 18 images as our inputs and set up 18 LED beads.

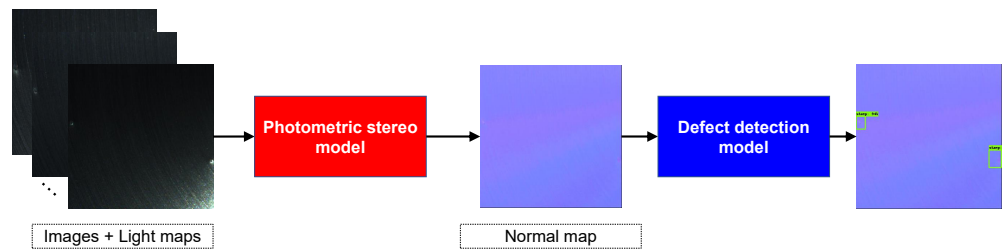


Figure 8. The framework of our detection model. The images captured under different illumination and the corresponding light information are used as input and the normal map is estimated through the photometric stereo model. The defect detection locates and classifies defects with the input of the normal map.

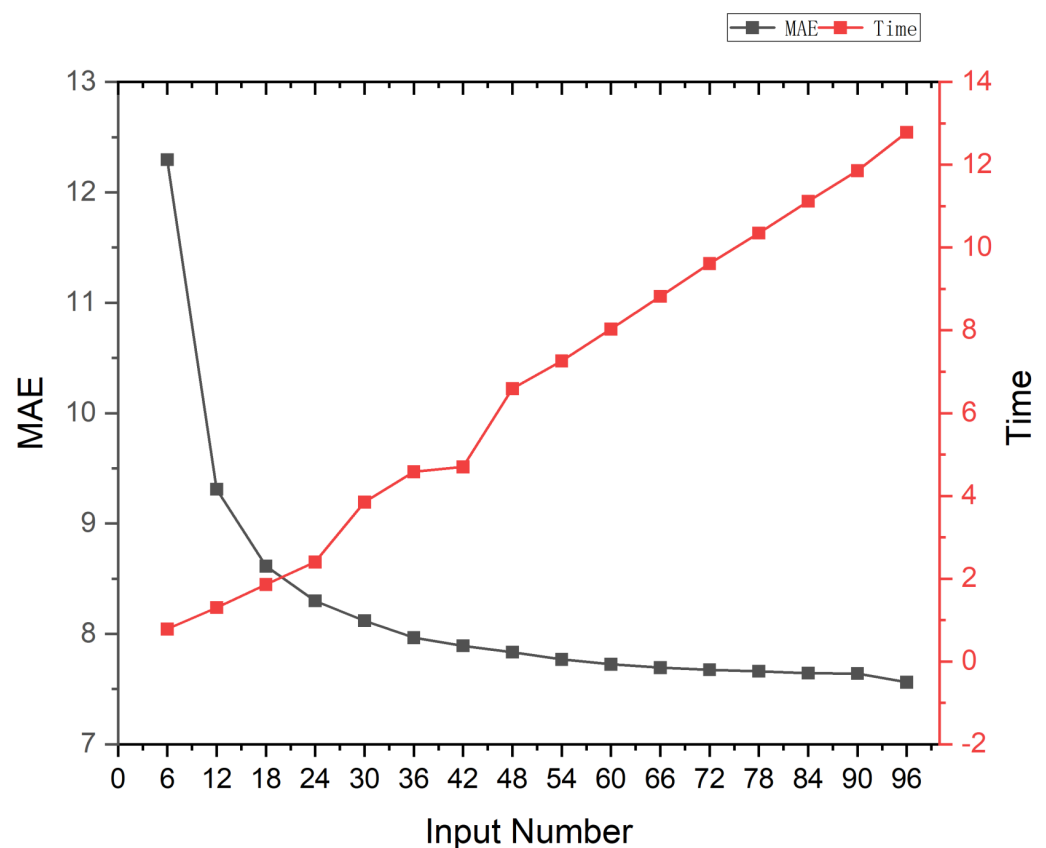


Figure 9. Performance and prediction time of MT-PS-CNN with the different number of inputs on the DiLiGenT dataset. The horizontal axis represents the number of inputs. The left vertical axis represents the mean angular error (MAE °). The right vertical axis represents the prediction time (s). Beginning with 96 inputs, we reduced the number of inputs by 6 at a time and calculated the average MAE over 10 trials.

Existing defect detection methods are mainly divided into one-stage (i.e., YOLO [45–47], SSD [48], RON [49]) and two-stage methods (i.e., SPP-Net [50], Fast R-CNN [51], Faster R-CNN [52]). The two-stage approach divides the detection problem into two stages, first generating region proposals, then adjusting the position of the generated region and using the information within the region for classification. The two-stage method is highly accurate but slower than one-stage methods. One-stage methods directly predict the bounding boxes and object classification through a single network. Compared to two-stage methods, one-stage methods are faster but less accurate. In our detection task, the method needs to

detect tiny defects in a large background area, and the requirement for detection accuracy is higher than the speed.

The detection speed is similar for both types of methods running on a high-performance computer. For balance, we chose the representative and effective two-stage method, Faster R-CNN, as the basis detection network. The Faster R-CNN model is divided into four parts: the feature extraction module, the region proposal network (RPN), the ROI pooling layer, and the target prediction module. Taking the normal map as input, the Faster R-CNN acquires the feature map using a feature extractor. The RPN generates the region proposals: generating anchor boxes with the pixels of the feature map and selecting and correcting the position of anchor boxes to obtain region proposals. The ROI pooling layer acquires the feature maps according to the region proposals and resizes the features maps into fixed-size feature vectors through the pooling operation. The target prediction module performs class classification and location regression based on feature vectors. For a detailed description of Faster R-CNN, please refer to [52].

3. Results

All experiments were performed on a GeForce RTX 2080Ti and 64GB RAM. The SGD optimizer was used to optimize our network with a learning rate of 2.5×10^{-3} , momentum of 0.9, and weight decay of 1×10^{-4} . The training process took 20 epochs with a batch size of 32.

3.1. Performance Criteria

We used the Intersection over Union (IoU) to evaluate the overlap between the ground-truth bounding box and the predicted bounding box, which can be expressed as

$$\text{IoU} = \frac{P \cap G}{P \cup G}, \quad (3)$$

where P is the predicted bounding box and G represents the ground-truth bounding box. The range of IoU is from zero to one where zero means that there is no overlap between the predicted and ground-truth regions and one means both regions overlap perfectly. According to common practice [53–55], we set the IoU threshold to 0.5. The sample with a threshold greater than 0.5 was positive, while the opposite was a negative sample. We applied the mean average precision (mAP) to evaluate the detection accuracy of the model. The confusion matrix was used to summarize the classification results; taking binary classification as an example, the target sample was positive, and the nontarget sample was negative, as shown in Table 2.

Table 2. Binary classification confusion matrix.

	Estimated	Positive	Negative
Ground Truth			
Positive		True positive (TP)	False negative (FN)
Negative		False positive (FP)	True negative (TN)

The ratio of positive samples correctly predicted to total samples is called the precision.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (4)$$

The ratio of positive samples correctly predicted to total positive samples is called the recall.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (5)$$

The mAP is the mean of the average precision (AP) for each category. The mAP for the N categories can be expressed as:

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (6)$$

where AP indicates the average precision of one category, which is the area under the precision–recall curve.

3.2. Detection Performance Analysis

The backbone of Fast R-CNN was used to extract features from input images. The performance of Fast R-CNN with different backbones varied considerably. We compared the performance of three mainstream backbones: VGG-16 [56], ResNet-50 [57], ResNet-50+FPN [58], as shown in Table 3.

Table 3. Fast R-CNN performance comparison of different backbones on our PSDD datasets.

Backbone	mAP/%
VGG-16	16.4
ResNet-50	50.1
ResNet-50+FPN	75.5

The introduction of the FPN (feature pyramid network) structure in the backbone can effectively retain shallow detail information and improve performance. Thus, we chose ResNet-50+FPN as the backbone of Fast R-CNN. Taking the normal map as input, we also trained representative one-stage defect detection methods (SSD [48] and a YOLO-V3 [55]) network. Table 4 shows the performance of one-stage and two-stage approaches on the test dataset.

Table 4. Performance comparison of different learning-based defect detection methods on our PSDD dataset. Numbers represent mAP (%).

Model	Stamp	Abraded	Scratch	Avg.
YOLOv3	35.5	13.6	33.3	29.5
SSD	54.9	45.1	48.9	50.9
Faster R-CNN	74.0	75.1	79.0	75.5

Faster R-CNN achieved good performance through two-stage operations and FPN that made full use of shallow local information and deep global information. We compared the number of parameters and the speed of the above methods, as shown in Table 5.

Table 5. Comparison of parameters and detection speed for different learning-based defect detection methods. The size of the test images was 400×400 .

Model	Parameters	FPS
YOLOv3	62.6 MB	14
SSD	13.5 MB	4
Faster R-CNN	41.4 MB	11

It can be found that the Faster R-CNN had advantages in terms of speed and detection accuracy. Figure 10 shows a diagram of the detection results for different defects. We can see that our method was effective in detecting defects in the captured images that were disturbed by shadows or highlights. However, due to the information consumption in the photometric stereo, some defect areas were incompletely marked, such as sample 8 in Figure 10.

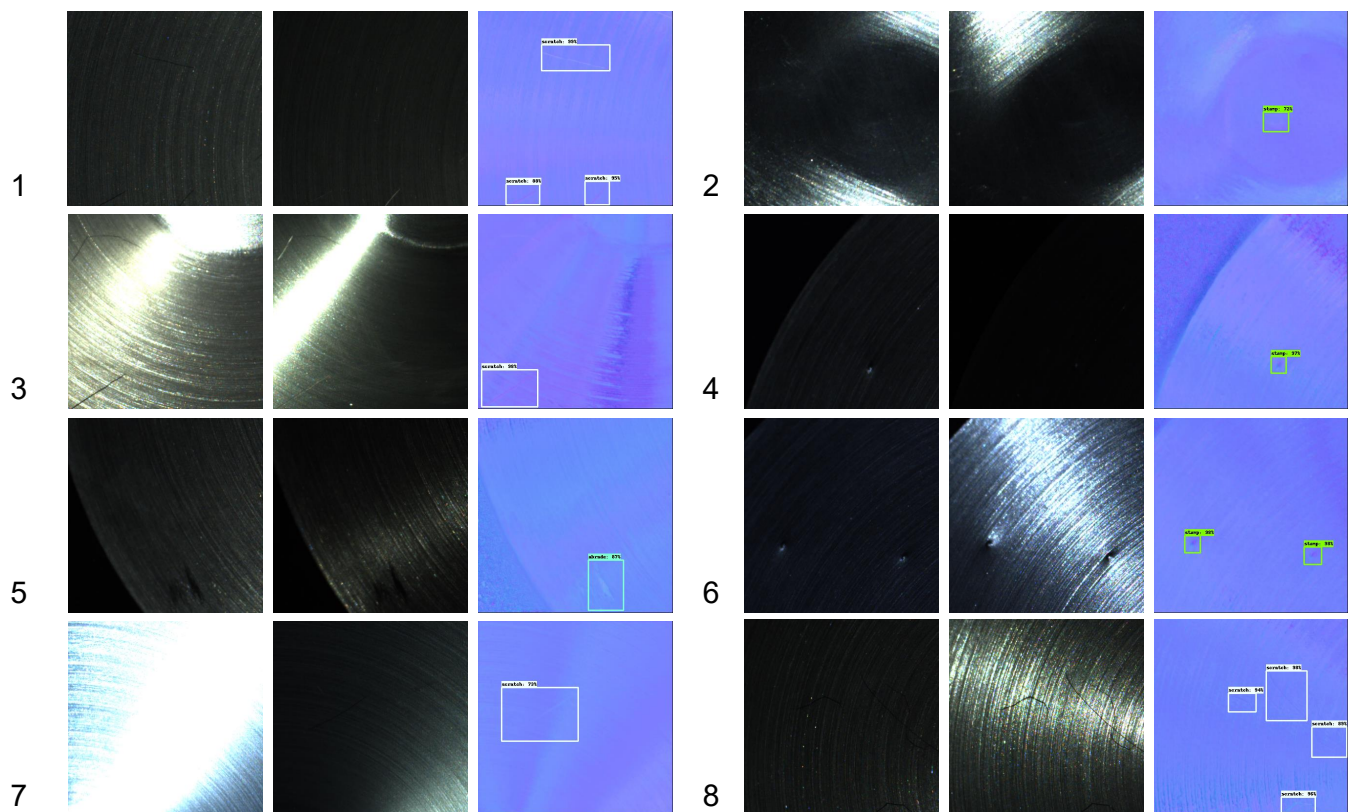


Figure 10. Diagram of detection results. We show the results of 8 sample. The numbers in the images represent the serial numbers of the samples. For each sample, we display 2 input images with the detection result.

3.3. Detection Time Analysis

We took a stainless steel part with a size of 600×450 mm as the test object. Due to the large size of the part, it needed to be inspected in pieces. The PSDCE was driven by a manipulator and the detection took place in eight different workstations.

The detection times are shown in Table 6.

Table 6. Time for each step of the system. Each step runs in tandem.

Manipulator Movement	PSDCE Work	Model Prediction	Total
10.5 s	8.8 s	24.8 s	36.1 s

All processes were carried out in tandem, and it took about half a minute to complete the detection of a part. According to the result, our method is suitable for spot testing or offline testing. In future work, we will improve the method by parallelising the processes and shortening the prediction speed of the model to meet the real-time requirements of an industrial site.

4. Discussion

As demonstrated by the experimentation shown previously, our system and dataset are effective for defect detection on a glossy surface. This is because the photometric stereo removes the effects of highlights and shadows by fusing the effective information from multiple directions of light and retaining only the normal information that is effective for defect detection. However, there are still some aspects which need improvement. We captured a small variety of objects in our dataset. In future work, we will capture metal

parts in a variety of shapes and materials to expand the dataset and prevent overfitting of the model. We combined an existing photometric stereo model directly with the defect detection model during testing. Photometric stereo is not only able to obtain the normal map of the object but also to solve for the surface albedo. The defects mentioned above all vary in the depth direction and the albedo contributes less to defect detection, so we input the normal map directly into the detection model. However, when the photometric stereo model removes highlights, some texture information is inevitably lost at the same time, leading to a failure of the defect detection, as shown in Figure 11.

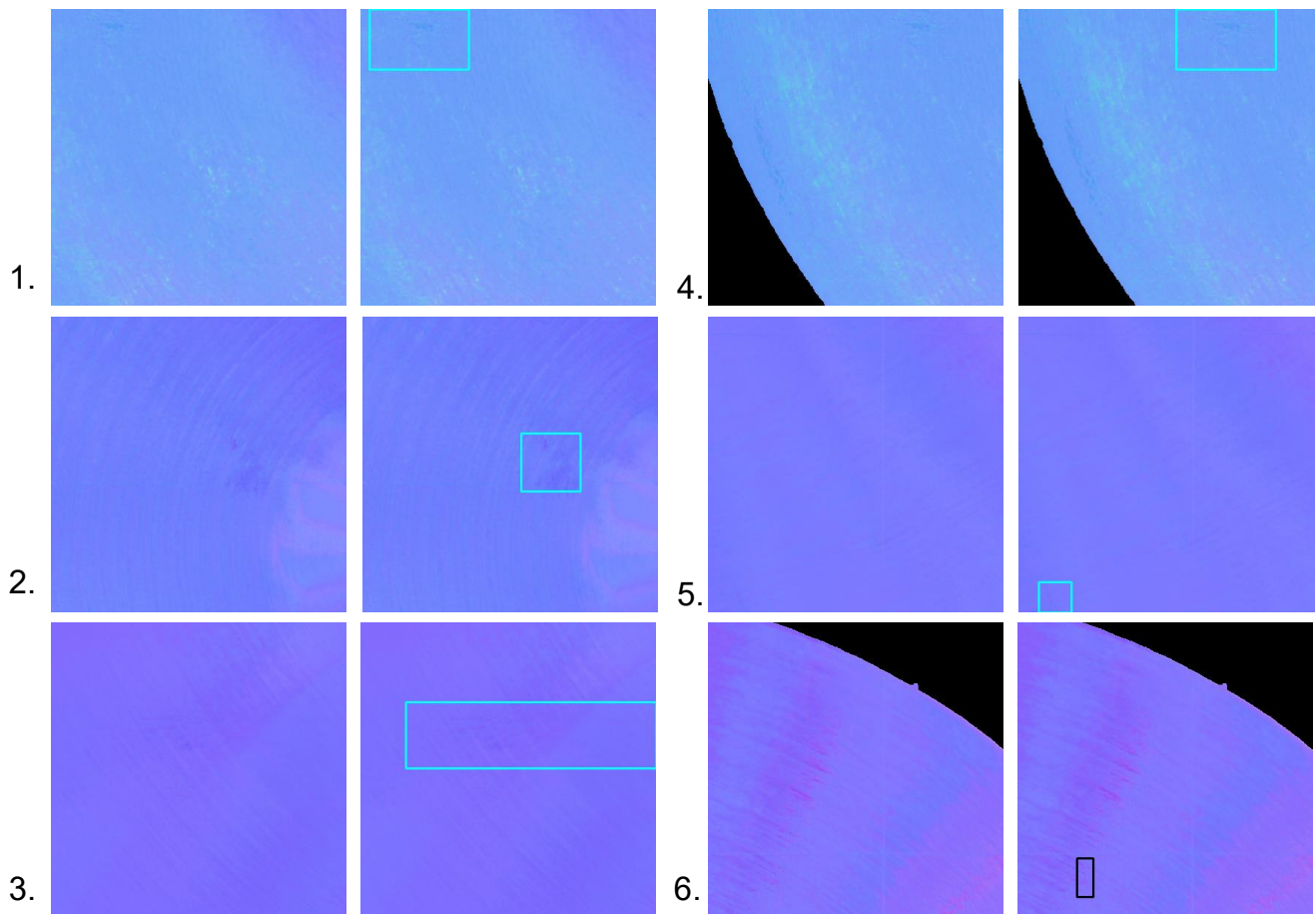


Figure 11. Detection failure cases. A total of 6 groups, of which groups 1–5 are abraded samples and group 6 is a scratch sample. The left image of each group is the estimated result and the right one is the ground truth. Boxes indicate the location of defects. For each group, the defect detection model did not identify the defect successfully.

We can see that the smoothing operation of the convolution made some of the defects too indistinguishable from the background. In future work, we will improve the existing photometric stereo model to retain sufficient detail information when removing highlights and provide a priori information on the background to the defect detection methods. Meanwhile photometric stereo is also capable of obtaining an albedo map. Furthermore, for defects with insignificant changes in depth, the albedo is still valid information. Therefore, the integration of albedo information into the defect detection models is also an avenue for future research.

5. Conclusions

In this paper, we designed a photometric-stereo-based defect detection system (PSB-DDS) to locate and classify the defects on metal parts and eliminate gloss interference. We

used the PSBDDS to make a photometric defect detection (PSDD) dataset, which contained images under different illumination and corresponding light information. Based on the PSDD dataset, we proposed a framework for photometric-stereo-based defect detection, which could combine the existing state-of-the-art photometric stereo methods with a defect detection method. Experimental results showed that our system could effectively fuse valid information under multidirectional illumination, remove highlights and shadow interference, and improve the efficiency of defect detection on high-gloss surface objects such as metal parts.

Author Contributions: Conceptualization, Y.C. and B.D.; methodology, Y.C., B.D., and J.C.; software, W.L.; validation, P.G.; writing—original draft preparation, B.D. and J.Y.; writing—review and editing, L.H., B.D., and J.Y.; supervision, Y.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key Research and Development Program of China (2020YFB1711400), National Natural Science Foundation of China (52075485) and Zhejiang Provincial Key Research and Development Program (2021C03112).

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare that they have no known competing financial interest or personal relationships that could have appeared to influence the work reported in this paper.

References

- Shi, T.; Kong, J.Y.; Wang, X.D.; Liu, Z.; Zheng, G. Improved Sobel algorithm for defect detection of rail surfaces with enhanced efficiency and accuracy. *J. Cent. South Univ.* **2016**, *23*, 2867–2875. [\[CrossRef\]](#)
- Robert, M.; Paindavoine, M.; Gorria, P. An edge detection ASIC for real time defect detection. In Proceedings of the Fifth Annual IEEE International ASIC Conference and Exhibit, Rochester, NY, USA, 21–25 September 1992; pp. 193–196.
- Ma, Y.; Li, Q.; Zhou, Y.; He, F.; Xi, S. A surface defects inspection method based on multidirectional gray-level fluctuation. *Int. J. Adv. Robot. Syst.* **2017**, *14*, 1729881417703114. [\[CrossRef\]](#)
- Ng, H.F. Automatic thresholding for defect detection. *Pattern Recognit. Lett.* **2006**, *27*, 1644–1649. [\[CrossRef\]](#)
- Yazdchi, M.; Yazdi, M.; Mahyari, A.G. Steel surface defect detection using texture segmentation based on multifractal dimension. In Proceedings of the 2009 International Conference on Digital Image Processing, Bangkok, Thailand, 7–9 March 2009; pp. 346–350.
- Funck, J.; Zhong, Y.; Butler, D.; Brunner, C.; Forrer, J. Image segmentation algorithms applied to wood defect detection. *Comput. Electron. Agric.* **2003**, *41*, 157–179. [\[CrossRef\]](#)
- Mamassian, P.; Knill, D.C.; Kersten, D. The perception of cast shadows. *Trends Cogn. Sci.* **1998**, *2*, 288–295. [\[CrossRef\]](#)
- Soukup, D.; Huber-Mörk, R. Convolutional neural networks for steel surface defect detection from photometric stereo images. In Proceedings of the International Symposium on Visual Computing, Las Vegas, NV, USA, 8–10 December 2014; Springer: Berlin/Heidelberg, Germany, 2014; pp. 668–677.
- Lee, J.H.; Oh, H.M.; Kim, M.Y. Deep learning based 3D defect detection system using photometric stereo illumination. In Proceedings of the 2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC), Okinawa, Japan, 11–13 February 2019; pp. 484–487.
- Kang, D.; Jang, Y.J.; Won, S. Development of an inspection system for planar steel surface using multispectral photometric stereo. *Opt. Eng.* **2013**, *52*, 039701. [\[CrossRef\]](#)
- Smith, M.L.; Smith, G.; Hill, T. Gradient space analysis of surface defects using a photometric stereo derived bump map. *Image Vis. Comput.* **1999**, *17*, 321–332. [\[CrossRef\]](#)
- Landstrom, A.; Thurley, M.J.; Jonsson, H. Sub-millimeter crack detection in casted steel using color photometric stereo. In Proceedings of the 2013 International Conference on Digital Image Computing: Techniques and Applications (DICTA), Hobart, Australia, 26–28 November 2013; pp. 1–7.
- Woodham, R.J. Photometric method for determining surface orientation from multiple images. *Opt. Eng.* **1980**, *19*, 191139. [\[CrossRef\]](#)
- Solomon, F.; Ikeuchi, K. Extracting the shape and roughness of specular lobe objects using four light photometric stereo. In Proceedings of the 1992 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, hampaign, IL, USA, 15–18 June 1992; pp. 466–471. doi: 10.1109/CVPR.1992.223149. [\[CrossRef\]](#)
- Coleman, E.N.; Jain, R. Obtaining 3-dimensional shape of textured and specular surfaces using four-source photometry. *Comput. Graph. Image Process.* **1982**, *18*, 309–328. doi: 10.1016/0146-664X(82)90001-6. [\[CrossRef\]](#)

16. Wu, L.; Ganesh, A.; Shi, B.; Matsushita, Y.; Wang, Y.; Ma, Y. Robust Photometric Stereo via Low-Rank Matrix Completion and Recovery. In *Lecture Notes in Computer Science (including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 703–717. doi: 10.1007/978-3-642-19318-7_55. [\[CrossRef\]](#)
17. Mukaigawa, Y.; Ishii, Y.; Shakunaga, T. Analysis of photometric factors based on photometric linearization. *JOSA A* **2007**, *24*, 3326–3334. [\[CrossRef\]](#)
18. Miyazaki, D.; Hara, K.; Ikeuchi, K. Median Photometric Stereo as Applied to the Segonko Tumulus and Museum Objects. *Int. J. Comput. Vis.* **2010**, *86*, 229–242. doi: 10.1007/s11263-009-0262-9. [\[CrossRef\]](#)
19. Phong, B.T. Illumination for computer generated pictures. *Commun. ACM* **1975**, *18*, 311–317. doi: 10.1145/360825.360839. [\[CrossRef\]](#)
20. Blinn, J.F. Models of light reflection for computer synthesized pictures. In Proceedings of the 4th Annual Conference on Computer Graphics and Interactive Techniques—SIGGRAPH '77, San Jose, CA, USA, 20–22 July 1977; ACM Press: New York, NY, USA, 1977; pp. 192–198. doi: 10.1145/563858.563893. [\[CrossRef\]](#)
21. Torrance, K.E.; Sparrow, E.M. Theory for Off-Specular Reflection From Roughened Surfaces*. *J. Opt. Soc. Am.* **1967**, *57*, 1105. doi: 10.1364/JOSA.57.001105. [\[CrossRef\]](#)
22. Kay, G.; Caelli, T. Estimating the parameters of an illumination model using photometric stereo. *Graph. Model. Image Process.* **1995**, *57*, 365–388. doi: 10.1006/gmip.1995.1032. [\[CrossRef\]](#)
23. Georgiades. Incorporating the Torrance and Sparrow model of reflectance in uncalibrated photometric stereo. In Proceedings of the Ninth IEEE International Conference on Computer Vision, Washington, DC, USA, 13–16 October 2003; IEEE: Piscataway, NJ, USA, 2003; Volume 2, pp. 816–823. doi: 10.1109/ICCV.2003.1238432. [\[CrossRef\]](#)
24. Chung, H.S.; Jia, J. Efficient photometric stereo on glossy surfaces with wide specular lobes. In Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, Alaska, 23–28 June 2008; pp. 1–8. doi: 10.1109/CVPR.2008.4587771. [\[CrossRef\]](#)
25. Shape and Spatially-Varying BRDFs from Photometric Stereo. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 1060–1071. doi: 10.1109/TPAMI.2009.102. [\[CrossRef\]](#)
26. Ikehata, S. CNN-PS: CNN-based photometric stereo for general non-convex surfaces. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–18.
27. Santo, H.; Samejima, M.; Sugano, Y.; Shi, B.; Matsushita, Y. Deep photometric stereo networks for determining surface normal and reflectances. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *44*, 114–128. [\[CrossRef\]](#)
28. Li, J.; Robles-Kelly, A.; You, S.; Matsushita, Y. Learning to minify photometric stereo. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–17 June 2019; pp. 7568–7576.
29. Zheng, Q.; Jia, Y.; Shi, B.; Jiang, X.; Duan, L.Y.; Kot, A.C. SPLINE-Net: Sparse photometric stereo through lighting interpolation and normal estimation networks. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019; pp. 8549–8558.
30. Chen, G.; Han, K.; Wong, K.Y.K. PS-FCN: A flexible learning framework for photometric stereo. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–18.
31. Taniai, T.; Maehara, T. Neural inverse rendering for general reflectance photometric stereo. In Proceedings of the International Conference on Machine Learning, PMLR, Stockholm Sweden, 10–15 July 2018; pp. 4857–4866.
32. Chen, G.; Han, K.; Shi, B.; Matsushita, Y.; Wong, K.Y.K. Self-calibrating deep photometric stereo networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 8739–8747.
33. Hold-Geoffroy, Y.; Gotardo, P.; Lalonde, J.F. Single day outdoor photometric stereo. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *43*, 2062–2074. [\[CrossRef\]](#)
34. Santo, H.; Samejima, M.; Sugano, Y.; Shi, B.; Matsushita, Y. Deep Photometric Stereo Network. In Proceedings of the 2017 IEEE International Conference on Computer Vision Workshops (ICCVW), Venice, Italy, 22–29 October 2017; pp. 501–509.
35. Chen, G.; Waechter, M.; Shi, B.; Wong, K.Y.K.; Matsushita, Y. What is learned in deep uncalibrated photometric stereo? In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; Springer: Berlin/Heidelberg, Germany, 2020; pp. 745–762.
36. Zheng, Q.; Shi, B.; Pan, G. Summary study of data-driven photometric stereo methods. *Virtual Real. Intell. Hardw.* **2020**, *2*, 213–221. [\[CrossRef\]](#)
37. Yao, Z.; Li, K.; Fu, Y.; Hu, H.; Shi, B. Gps-net: Graph-based photometric stereo network. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 10306–10316.
38. Ikehata, S. Ps-transformer: Learning sparse photometric stereo network using self-attention mechanism. In Proceedings of the BMVC, Online, 22–25 November 2021; Volume 2, p. 11.
39. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is all you need. In Proceedings of the 31st Conference on Neural Information Processing Systems 30 (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017.
40. Huang, S.; Xu, K.; Li, M.; Wu, M. Improved visual inspection through 3D image reconstruction of defects based on the photometric stereo technique. *Sensors* **2019**, *19*, 4970. [\[CrossRef\]](#)

41. Wang, L.; Xu, K.; Zhou, P. Online detection technique of 3D defects for steel strips based on photometric stereo. In Proceedings of the 2016 Eighth International Conference on Measuring Technology and Mechatronics Automation (ICMTMA), Macau, China, 11–12 March 2016; pp. 428–432.
42. Ren, M.; Wang, X.; Xiao, G.; Chen, M.; Fu, L. Fast defect inspection based on data-driven photometric stereo. *IEEE Trans. Instrum. Meas.* **2018**, *68*, 1148–1156. [\[CrossRef\]](#)
43. Saiz, F.A.; Barandiaran, I.; Arbelaiz, A.; Graña, M. Photometric Stereo-Based Defect Detection System for Steel Components Manufacturing Using a Deep Segmentation Network. *Sensors* **2022**, *22*, 882. [\[CrossRef\]](#)
44. Cao, Y.; Ding, B.; He, Z.; Yang, J.; Chen, J.; Cao, Y.; Li, X. Learning inter-and intraframe representations for non-Lambertian photometric stereo. *Opt. Lasers Eng.* **2022**, *150*, 106838. [\[CrossRef\]](#)
45. Redmon, J.; Divvala, S.K.; Girshick, R.B.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016.
46. Redmon, J.; Farhadi, A. YOLO9000: Better, Faster, Stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.
47. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. *arXiv* **2018**, arXiv:1804.02767
48. 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; Springer: Berlin/Heidelberg, Germany, 2016; pp. 21–37.
49. Kong, T.; Sun, F.; Yao, A.; Liu, H.; Lu, M.; Chen, Y. Ron: Reverse connection with objectness prior networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 5936–5944.
50. 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. [\[CrossRef\]](#)
51. Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
52. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process. Syst.* **2015**, *28*. [\[CrossRef\]](#)
53. Shou, Z.; Gao, H.; Zhang, L.; Miyazawa, K.; Chang, S.F. Autoloc: Weakly-supervised temporal action localization in untrimmed videos. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 154–171.
54. Cai, Z.; Vasconcelos, N. Cascade r-cnn: Delving into high quality object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 6154–6162.
55. Solovyev, R.; Wang, W.; Gabruseva, T. Weighted boxes fusion: Ensembling boxes from different object detection models. *Image Vis. Comput.* **2021**, *107*, 104117. [\[CrossRef\]](#)
56. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
57. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
58. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.