applied sciences

MDPI

Article

CCA-YOLO: An Improved Glove Defect Detection Algorithm Based on YOLOv5

Huilong Jin [1,2], Ruiyan Du [1], Liyong Qiao [1], Lingru Cao [1], Jian Yao [3] and Shuang Zhang [1,*]

[1] College of Engineering, Hebei Normal University, Shijiazhuang 050024, China; jhl981@hebtu.edu.cn (H.J.); 13903216975@163.com (R.D.); qiaoly@hebtu.edu.cn (L.Q.); caolr0708@163.com (L.C.)
[2] Vocational and Technical College, Hebei Normal University, Shijiazhuang 050024, China
[3] School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430072, China; jian.yao@whu.edu.cn
* Correspondence: zshuang@hebtu.edu.cn

Abstract: Aiming to address the issue of low efficiency and a high false-detection rate in artificial defect detection in nitrile medical gloves, CCA-YOLO was proposed on the basis of YOLOv5 to realize the detection of tear and scratch defects. CCA-YOLO added a small-target detection layer to the YOLOv5 network backbone and further proposed an innovative channel coordinate attention mechanism. According to the different characteristics of tears and scratches, focal and efficient IoU loss and α-IoU loss functions were introduced to further improve the positioning accuracy. The data enhancement method was used to generate a dataset of nitrile gloves, which was divided into datasets for horizontal angular tear detection, vertical angular tear detection, and scratch detection. The problem of class imbalance with few defect samples was solved. Our experiments show that CCA-YOLO can effectively identify tear and scratch defects in nitrile medical gloves in the self-made datasets. Compared with YOLOv5, the mean average precision (mAP) of the three models for horizontal angular tear detection, vertical angular tear detection, and scratch detection can reach 99.3%, 99.8%, and 99.6%, showing increments of 4.2%, 5.3%, and 12.4%, respectively, thereby meeting the performance requirements of glove defect detection.

Keywords: α-IoU; coordinate attention; data enhancement; defect detection; focal and efficient IoU; YOLOv5

Citation: Jin, H.; Du, R.; Qiao, L.; Cao, L.; Yao, J.; Zhang, S. CCA-YOLO: An Improved Glove Defect Detection Algorithm Based on YOLOv5. *Appl. Sci.* **2023**, *13*, 10173. https://doi.org/10.3390/app131810173

Academic Editor: Chilukuri K. Mohan

Received: 11 August 2023
Revised: 5 September 2023
Accepted: 8 September 2023
Published: 10 September 2023

1. Introduction

The rapid digitization of the global manufacturing industry and the impact of the epidemic are increasing the global annual demand for nitrile gloves. With the continuous expansion of the nitrile glove production scale, surface defects frequently occur during the production process, and defect detection has become a major difficulty in the production process. Common surface defects of nitrile gloves include tears and scratches [1].

Currently, defect detection in nitrile glove production has issues such as low efficiency, a high false detection rate, and a high missed detection rate. In the early research on surface defect detection in gloves, physical detection methods were mainly used for defect detection [2]. In 2000, Sohn et al., tested pinhole defects in gloves, using brine to expand gloves and soak them in brine through water load experiments and conductance tests [3]. In 2003, Murray identified pinhole defects in rubber gloves through air expansion and water immersion experiments [4]. In 2016, Thang et al., used image processing techniques based on the region of interest and integrated grayscale, morphology, threshold, hole filtering, and noise removal for glove defect detection; the overall accuracy was only 81% [5]. The machine-learning-based technology for detecting glove defects has developed gradually. In 2016, Sun and Chen proposed a glove defect detection method based on machine vision. The color space of image detection was changed from RGB to HSV, and a Canny edge detector was used to extract the glove contour and detect the tearing condition [6]. This

method is slow and requires cumbersome image processing; therefore, it is unsuitable for defect detection in high-efficiency production. With the continuous development of deep learning, two-stage target detection methods are becoming increasingly computationally intensive owing to the complexity of the underlying network, number of candidate frames, and complexity of the classification and regression subnetworks.

With the widespread application of deep learning, the accuracy and efficiency of object detection have been considerably improved; moreover, future defect detection in nitrile medical gloves is expected to primarily focus on defect target detection. On the basis of many studies, object detection algorithms can be approximately categorized into two-stage and one-stage object detection algorithms. The classical algorithms for two-stage target detection include R-CNN [7], SPP-Net [8], Fast R-CNN [9], and Faster R-CNN [10]. In 2015, Liu et al. performed threshold segmentation to extract glove defects on the basis of gray images using the different characteristics of gray color values of defects and non-defects; however, this method could not be implemented during actual production [11]. In 2023, Mohd Anul Haq used deep-learning-based super-resolution to improve the spatial resolution of HSI, broadening the idea for the research of computer vision to distinguish materials [12].

In 2016, Redmon et al. proposed the "you only look once" (YOLO, also known as YOLOv1) approach to overcome the inefficiency of the two-stage target detection algorithm [13]. YOLO discards the candidate frame extraction branch of the algorithm and directly implements feature extraction, candidate frame classification, and regression in the same branchless deep convolutional network. This simplified the network structure, enabling deep-learning-based target detection algorithms to meet the demands of real-time detection tasks, given the computing power available at the time. With the emergence of YOLO, deep-learning-based target detection algorithms started to have dual and single stages. Between 2016 and 2020, the YOLO algorithm gradually evolved into YOLOv2 [14], YOLOv3 [15], YOLOv4 [16], and YOLOv5, and since 2020, the YOLO family of algorithms has gradually become a research focus for target detection. In 2021, Ge et al. proposed YOLOX [17], which uses YOLOv3 as the base network for improvements, with three decoupled heads added to the output layer. In 2022, YOLOv6 was introduced, bringing the re-param VGG structure to YOLO to increase its suitability for the use of GPU (Graphics Processing Unit) devices. Almost simultaneously, YOLOv7 was proposed as a sequel to YOLOv4; it mainly focused on model structure referencing and dynamic label assignment issues. In the same year, YOLOv8 was further developed by the team that developed YOLOv5; the main additions were structural algorithms, a command line interface, and Python API. YOLOv8 is a step up in accuracy compared with YOLOv5 but a slight step down in speed.

Compared with previous YOLO-series algorithms, the YOLOv5 algorithm offers improvements in model size. The input terminal uses adaptive anchor frame calculation and adaptive picture scaling methods, considerably improving detection performance. Experimental proof indicates that YOLOv5 is the more classical algorithm in the current YOLO series and is more suitable for defect detection in nitrile medical gloves. In 2022, Jawaharlalnehru improved the YOLO algorithm by using pre-trained network classification and multi-scale detection training and changing the screening rules of candidate boxes, which can be effectively used for multi-scale target detection [18]. In 2022, Thang used the YOLO model for training in Google Colab and successfully distinguished torn gloves from normal ones [19] but failed to detect other defects. In the same year, H. Wang proposed the glove defect detection algorithm YOLO-G [20], which used Ghostnet to replace part of the structure of YOLOv5 but did not realize the defect detection of scratches and tears.

Since the YOLO algorithm was proposed, various network structures have been continuously integrated into it to improve detection performance. However, only a few studies have been conducted on defect detection in nitrile medical gloves. Most of the defect detection methods involve physical detection and object detection realized through a simple convolutional neural network. However, problems such as tedious processing,

high requirements for hardware and the environment, inconvenient deployment, false detection, and leakage detection still exist. To solve these problems, herein, CCA-YOLO is proposed; this model can meet the performance requirements of glove defect detection. The achievements of this study can be briefly summarized as follows.

(1) The receipt of nitrile medical gloves was created using 39,941 images after data enhancement, effectively solving the problem of class imbalance with fewer defective samples.
(2) The CCA-YOLO algorithm was used for defect detection, and a small-target detection layer was added to the YOLOv5 network backbone [21].
(3) On the basis of the spatial squeeze and channel excitation block [22] and coordinate attention mechanism, an innovative channel coordinate attention (CCA) mechanism was proposed to improve the detection of tears and scratch defects. This mechanism can focus on the channel relationship in addition to the spatial relationship of the network.
(4) On the basis of the different characteristics of tears and scratches, EIoU was introduced to detect glove tear defects and α-IoU was introduced to detect glove scratch defects, thus improving defect detection accuracy in one step.

2. Related Work

YOLOv5 is a fully convolutional network comprising a convolutional layer and batch normalization layer, without a full connection layer. YOLOv5 provides four network structures of different sizes: YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x, in ascending order. Among them, the YOLOv5s network has the smallest volume and fast reasoning speed but low detection accuracy, which is suitable for devices with low computing power.

The network structure of YOLOv5s comprises a backbone network, bottleneck layer network, and detection layer [23]. The backbone network comprises a focus module to extract the features of the input image data. The neck layer uses a path aggregation network [24] for feature fusion, combining superficial graphic features with deep semantic features to obtain more complete features. The CioU [25] function is used as the loss function of bbox regression. The original network structure of YOLOv5s is shown in Figure 1.

The YOLOv5 algorithm has been successfully improved and applied to defect detection. Owing to the large downsampling multiple of YOLOv5, it struggles to learn the feature information of small glove defect targets from deep feature maps, resulting in the poor detection of small-target defects. Therefore, a small-target detection layer was proposed to detect shallow and deep feature images after concatenation. This layer can enhance the network detection of small glove defect targets and improve detection performance. To overcome the problem of small defect targets, based on the original output layer, Yu et al., added an output layer specifically for small-target detection using a cascading network [26].

Attention mechanisms are resource allocation schemes that allocate computing resources to more important tasks and solve the problem of information overload in cases of limited computing power. In the process of the development of attention mechanisms, channel attention (CA) mechanism coordinate attention has been proposed; it aims to enhance the expression ability of the learning features of the mobile network [27]. The CA mechanism can not only obtain long-range dependencies in the spatial direction but can also enhance the expression of the location information of the features and increase the global receptive field of the network. The CA mechanism can consider attention in both channel and spatial dimensions to better pay attention to the defect features of glove tears and scratches [28]. In 2023, Zhu et al., introduced coordinate attention in the backbone network to enhance information interaction among all channels and make the network focus on high-weight areas [29].

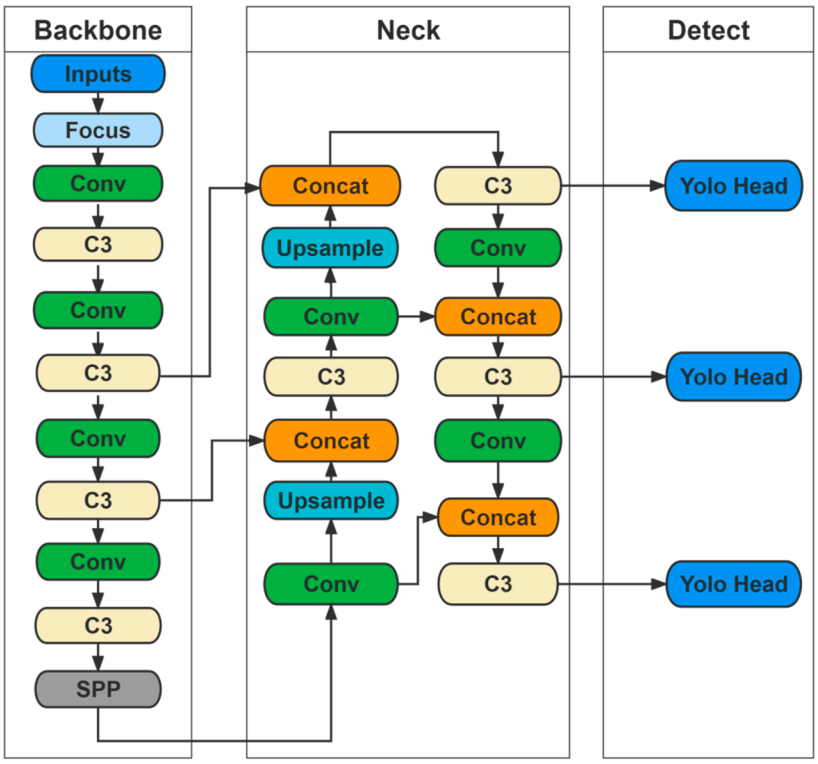


Figure 1. YOLOv5s network structure.

Bounding box (bbox) regression is the basis of target detection, location, and tracking. In recent studies related to bbox, IoU loss and its related variants were directly adopted as location loss. The related variants of IoU loss are GioU, DioU, CioU, EioU, and α-IoU. GioU could not directly reflect the distance between the predicted boundary box and real boundary box [30]. In the case of DioU, the diagonal distance normalizes the distance between the center points of the detection box and prediction box, but when the IoU value is the same as the distance between the center points of the two boxes, it cannot distinguish the distance between these boxes [31]. CioU consumes a certain amount of computational power in the process of calculation. EioU calculates the differences in width and height instead of the aspect ratio and introduces focal loss to solve the problem of the imbalance between difficult and easy samples [32]. α-IoU allows more flexibility in achieving different levels of bbox regression accuracy by adjusting the α to weight losses and gradients; EioU and α-IoU can solve the problem of high surface noise in glove defect detection. Yixuan et al. replaced the original CioU loss function with an EioU loss function to improve the ability to extract surface defect features. Furthermore, Yinsheng et al. changed the loss function of frame regression to α-IoU loss, further improving the accuracy of bbox regression.

3. CCA-YOLO Network Model

3.1. CCA-YOLO Structure

The feature extraction network of YOLOv5s was redesigned by adding the small-target feature detection layer and the CCA mechanism proposed in this study. CioU was replaced with the EioU loss function to improve the detection of tear defects. The α-IoU loss function was used to improve the scratch defects detection accuracy. The structure CCA-YOLO is shown in Figure 2.

3.2. Improvement of Small-Target Detection Layer

The main reason for the poor detection of small-target defects is the size of the target. The original YOLOv5s model has only three target detection layers of different scales. In the case of the 608 × 608 network input, the sizes of the three features are 19 × 19, 38 × 38,

and 76 × 76. As shown in the feature mapping schematic in Figure 3, the largest feature is responsible for detecting small targets, corresponding to 608 × 608; thus, the receptive field of each cell feature graph is 608/76 = 8 × 8.

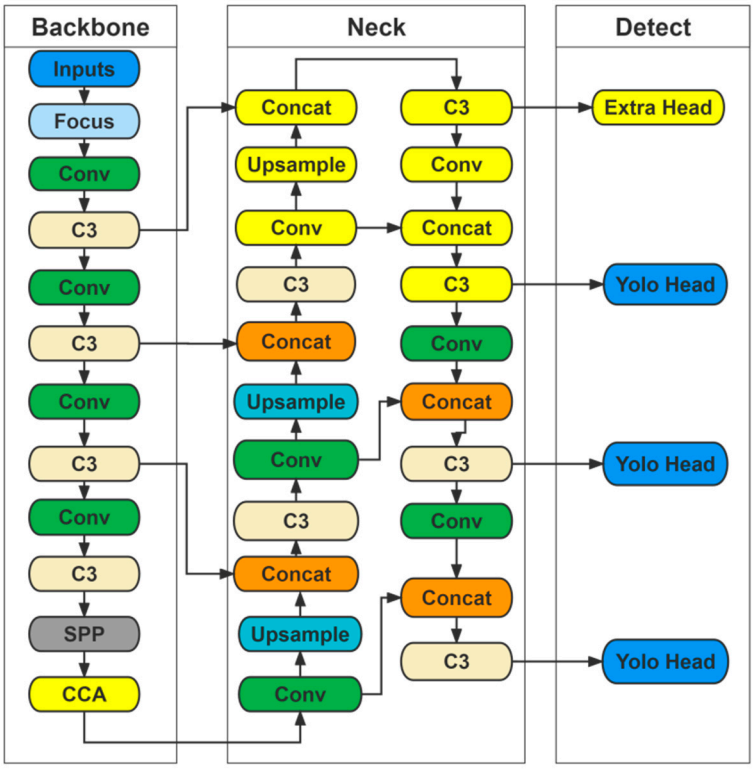


Figure 2. CCA-YOLO structure.

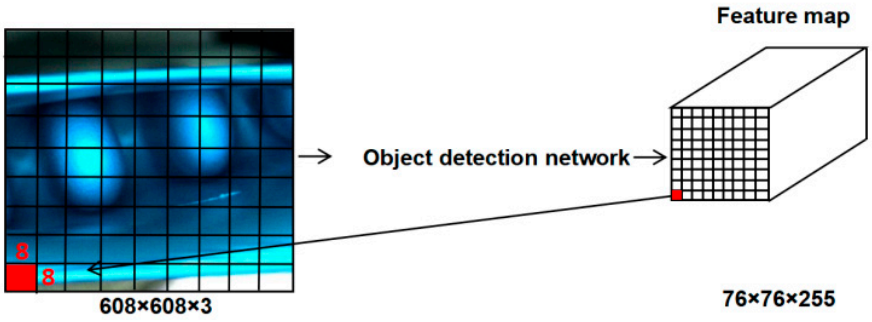


Figure 3. Feature mapping.

If the width or height of the target in the original image is less than 8 pixels, it becomes challenging for the network to learn feature information. Moreover, many images have high resolutions, and simple subsampling can result in a large loss of data information if performed using a high subsampling multiple. However, if the multiple is substantially small, many feature graphs must be stored in the memory for network forward propagation, which consumes considerable GPU resources and easily causes video memory explosion, making normal training and reasoning impossible. The feature extraction layer for small targets added in this study continues to perform upsampling and other types of processing on the feature map after the 17th layer, so the feature map continues to expand. Meanwhile, at the 20th layer, the obtained feature graph with a size of 160 × 160 is concatenated with the second layer feature graph in the backbone network to obtain a larger feature graph for small-target detection. At the 31st layer, a segmentation detection module is added and four layers are used for detection, improving the accuracy of small scratch detection.

3.3. Channel Coordinate Attention Mechanism

Coordinate attention encodes channel relationships and long-term dependencies through precise position information. This encoding is divided into two steps: the coordination of information embedding and coordinate attention generation [32]. As shown in Figure 4, one-dimensional adaptive average pooling of input features in the x-axis and y-axis directions was performed to obtain independent directional perception features with x-axis and y-axis information retained, respectively. One spatial direction captures long-range dependencies, whereas the other retains accurate position information.

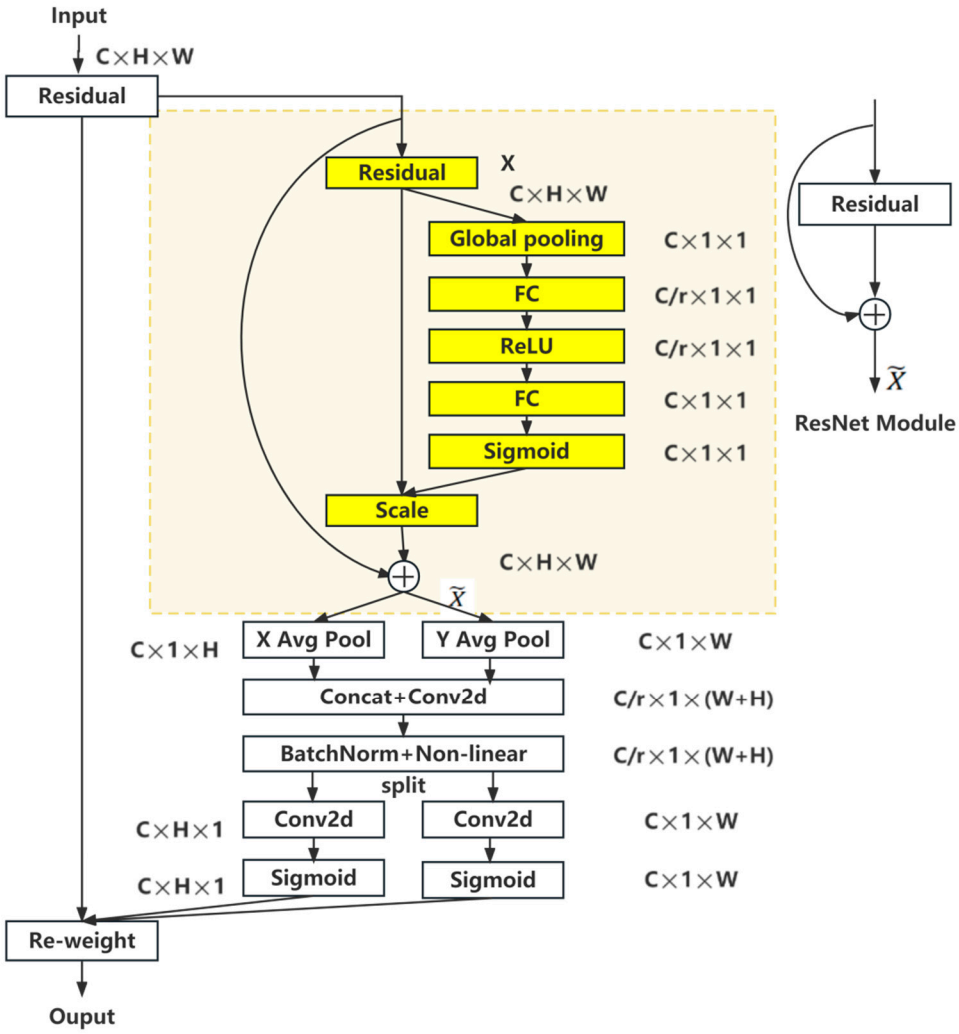


Figure 4. Network diagram of the CCA mechanism.

As shown in Figure 4 and Formula (1), the feature map is changed from [C, H, W] to [C, 1, 1] using the global average pooling method, and the information is subsequently processed using two $1 \times 1 \times 1$ convolutions, resulting in a C-dimensional vector. The addition join does not add new dimensions without preserving the characteristics of the previous layer. It is generally believed that the input mapping features **U** as a combination and embedding of the global spatial information in vector z, encoding the channel-related dependencies:

$$z_k = \frac{1}{H \times W} \sum_i^H \sum_j^W \mathbf{u}_k \tag{1}$$

Using a sigmoid function to pass $\sigma(\hat{z})$ for normalization, $\hat{z}$ is brought to the interval [0, 1] to obtain the corresponding mask, and channel-wise multiplication is finally per-

formed to obtain an information-calibrated feature map, emphasizing the useful features to suppress the useless ones. The resulting vector is used to recalibrate or excite **U** to

$$\hat{\mathbf{U}}_{\mathrm{CCA}} = \mathrm{F}_{\mathrm{CCA}}(\mathbf{U}) = [\boldsymbol{\sigma}(\hat{z}_1)\mathbf{u}_1, \boldsymbol{\sigma}(\hat{z}_2)\mathbf{u}_2, \ldots, \boldsymbol{\sigma}(\hat{z}_{\mathbf{C}})\mathbf{u}_{\mathrm{C}}] \tag{2}$$

$\sigma(\hat{z}_i)$ indicates the importance of the channel being rescaled. As the network learns, this activation is adaptively adjusted to ignore less important channels and emphasize the important ones, improving the network performance by focusing on the key locations of the image.

After optimizing the channel dimension, the two obtained one-dimensional features are spliced in the *W* dimension and a convolution and nonlinear activation function are passed; subsequently, the features are split in the channel dimension [31]. Two feature graphs with long-range dependencies of specific spatial direction are obtained through convolution and the sigmoid activation function. These two feature maps can be complementarily applied to the input feature map to enhance the target of interest. A feature graph with attention weight in the width and height directions is obtained through feature fusion. The experimental results show that CCA can effectively improve the accuracy of the model while slightly increasing the number of computations.

3.4. Improvement of Loss Function

On the basis of multiple studies, in this experiment, the introduction of the EioU loss function led to the better identification and detection of tear defects taken at vertical and horizontal angles. The EioU loss function is defined as follows:

$$\begin{aligned} \mathrm{L}_{\mathrm{EIoU}} &= \mathrm{L}_{\mathrm{IoU}} + \mathrm{L}_{\mathrm{dis}} + \mathrm{L}_{\mathrm{asp}} \\ &= 1 - \mathrm{IoU} + \frac{\rho^2(\mathrm{b},\mathrm{b}^{\mathrm{gt}})}{\mathrm{c}^2} + \frac{\rho^2(\mathrm{w},\mathrm{w}^{\mathrm{gt}})}{\mathrm{c}_{\mathrm{w}}^2} + \frac{\rho^2(\mathrm{h},\mathrm{h}^{\mathrm{gt}})}{\mathrm{c}_{\mathrm{h}}^2} \end{aligned} \tag{3}$$

This function includes overlap loss $\mathrm{L}_{\mathrm{IoU}}$, distance loss $\mathrm{L}_{\mathrm{dis}}$, and width and height loss $\mathrm{L}_{\mathrm{asp}}$. In the aforementioned equation, c represents the diagonal distance that can wrap the minimum rectangles of two frames; (b, b^{gt}) represents the distance between the center points of two frames; c_{w} and c_{h} represent the width and height that can wrap the minimum rectangles of two frames, respectively; and (w, w^{gt}) and (h, h^{gt}) represent the difference between the width and height of the two boxes. EioU loss not only considers the distance factor of the two rectangular boxes but also minimizes the difference between the width and height of the two boxes. Therefore, the detection accuracy for smaller targets is improved to a certain extent and the convergence stability of the model is enhanced.

From experimental research, it can be found that although the EioU loss function can effectively improve tear detection accuracy, it cannot accurately identify small defects such as scratches. Therefore, to address the characteristics of small scratch targets that are difficult to identify and the high level of noise interference in the plant environment, the experiment improved the loss function of the YOLOv5s model using α-IoU. Unlike other variants, α-IoU is a new IoU loss function proposed for the existing IoU loss based on the introduction of power transformation, with a power IoU term, an additional power regular term, and a single power parameter α, which can considerably exceed the existing IoU loss [32].

Common IoU losses are defined as follows:

$$\mathrm{L}_{\mathrm{IoU}} = 1 - \mathrm{IoU} \tag{4}$$

The general form of the α-IoU is as follows:

$$\mathrm{L}_{\alpha\text{-IoU}} = 1 - \mathrm{IoU}_{\alpha} \tag{5}$$

Ordinary IoU losses can be reduced to α-IoU losses through Box–Cox transformation:

$$L_{\alpha-IOU} = \frac{1}{\alpha}(1 - IOU^{\alpha}), \alpha > 0 \quad (6)$$

Compared with L_{IoU}, $L_{\alpha-IoU}$ increases the loss and gradient of high-IoU targets when $\alpha > 1$, thus improving the accuracy of bbox regression. The selection of α is crucial to the loss of α-IoU. In most cases, the best effect is attained at $\alpha = 3$. In this study, it was found that $\alpha = 2$ accelerated the learning of all positive IoU targets at AP_{50}.

The calculation formula is as follows:

$$L_{2-IOU} = \frac{1}{2}\left(1 - IOU^2\right) = \frac{1}{2}L_{IOU}{}^2 \quad (7)$$

Common IoU loss is defined as the effect of penalty conditions on the properties of vanilla α-IoU. Therefore, based on the analysis of vanilla α-IoU (4), it can be concluded that the power transformation of $L_{\alpha\text{-IoU}}$ retains the key properties of L_{IoU}, including non-negativity, indistinguishable identity, symmetry, and triangle inequality. Additionally, $L_{\alpha\text{-IoU}}$ also includes the following five characteristics: Order preservation ensures that L_{IoU} and $L_{\alpha\text{-IoU}}$ are both monotonic decreasing functions. Relative loss reweighting increases the weighting factor monotonically (from 1 to α) with increasing IoU when $\alpha > 1$ and can help the model focus more on high-IoU targets to improve location and detection performance [33]. Relative gradient reweighting monotonically increases the aforementioned reweighting factor with increasing IoU when $\alpha > 1$, which allows the model to learn targets with adaptive velocities depending on the IoU of the target.

4. Experimental Results and Discussion

4.1. Creating a User-Defined Glove Dataset

Currently, there are no publicly available datasets on nitrile medical gloves. The detection effect in real nitrile glove defect detection scenarios can be affected by complex conditions such as light and angle. In this paper, the data collection work was carried out on actual defect samples in the production workshop to ensure that the user-defined glove dataset represents real variation in defects. To better detect small scratch and tear targets and enhance the model generalization ability, defect data were enhanced before model training. The image was preprocessed for defect data, including clipping, noise, dimness, brightness, and rotation. The image obtained after data enhancement is shown in Figure 5.

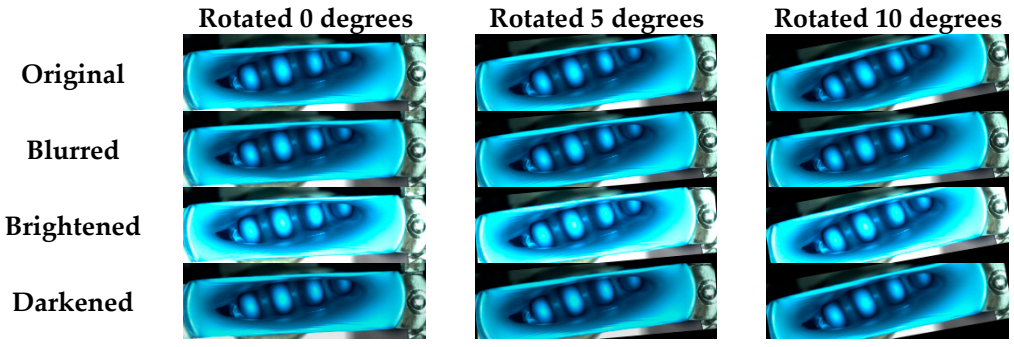


Figure 5. Partial dataset after data enhancement.

The customized nitrile glove dataset includes frontal and vertical shooting, each with tear and scratch defect types, and contains 39,941 images. From the data, 31,953 training sets and 7988 test sets were randomly selected. The sample dataset is shown in Figure 6.

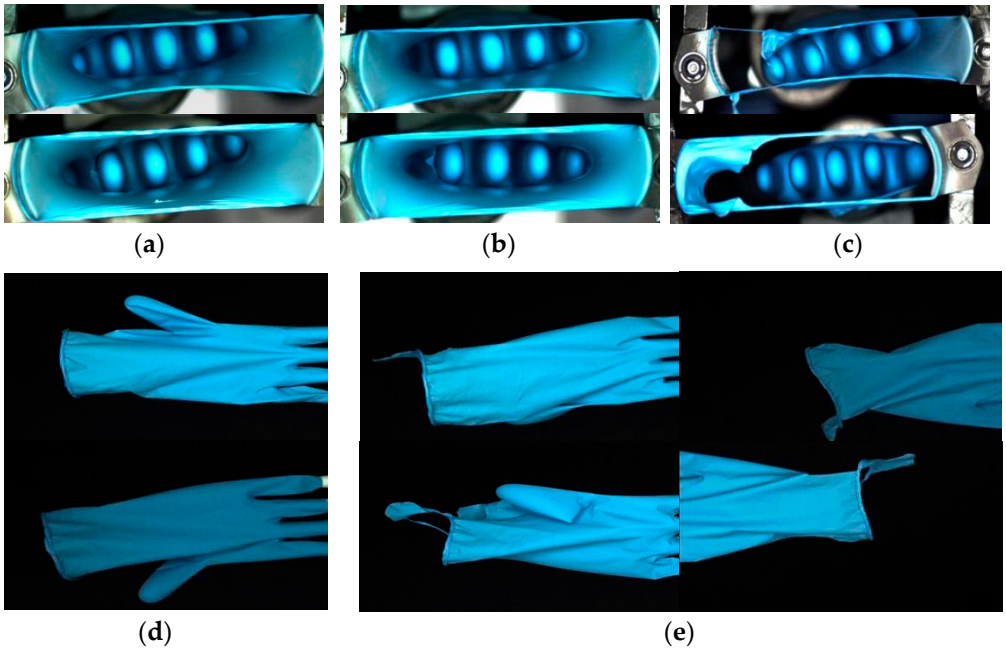


Figure 6. Example dataset diagram. (**a**) Scratches (vertical angle). (**b**) Qualified (vertical angle). (**c**) Tears (vertical angle). (**d**) Qualified (horizontal angle). (**e**) Tears (horizontal angle).

Table 1 shows that after 1000 iterations of training, the data-enhanced dataset improved the generalization ability and robustness of the model. Upon noise, rotation, and clipping data enhancements based on geometric transformation, the differences in scale, position, and perspective between the training set and test set could be eliminated. By adjusting the data enhancement based on color space transformation, such as brightness, the differences in illumination, color, and brightness between the training set and test set could be eliminated. The data enhancement method adopted in this study made the amplified training data as close as possible to the real distributed data, thus improving the detection accuracy.

Table 1. The data sheet of ablation experiment 1.

Dataset	Data Volume	P (%)	R (%)	mAP@0.5 (%)	mAP@0.5:0.95 (%)
Original—horizontal	2235	90.6	92.6	95.1	68.7
Enhanced—horizontal	19,815	99.8	95.4	98.3	82.7
Original—vertical	1155	91.8	91.8	94.5	75.5
Enhanced—vertical	18,434	97.4	98.8	98.8	90.0
Original—scratch	497	85.1	83.3	87.2	33.6
Enhanced—scratch	2322	92.8	95.0	96.6	57.6

4.2. Evaluation Indexes

In this experiment, the mean average precision (mAP@0.5), parameters (Param), Giga Floating-point Operations Per Second (GFLOPs), precision (P), recall (R), and Frames Per Second (FPS) were used as the evaluation indexes of model performance. P and R were found as follows:

$$P = \frac{T_P}{T_P + F_P} \tag{8}$$

$$R = \frac{T_P}{T_P + F_N} \tag{9}$$

where T_P (true positive) indicates that the object to be detected in the image is correctly identified and the IoU is greater than the threshold; F_p (false positive) indicates that the

detection object is not correctly identified and the IoU is less than the threshold; and F_N (false negative) indicates that the target is not detected.

The calculation process of mAP is as follows:

$$mAP = \frac{1}{n}\sum_{0}^{N}\int_{0}^{1} P_n(r)dr. \quad (10)$$

N is the number of target categories detected in the dataset and the mAP value of a certain category. In particular, mAP@0.5 and mAP@0.5:0.95 are commonly used to evaluate the model performance; mAP@0.5 focuses on the variation trend of model accuracy with the recall rate, and mAP@0.5:0.95 pays more attention to the comprehensive performance of the model under different IoU thresholds, reflecting the fitting range between the detection and real frames. Unless otherwise specified, mAP in subsequent sections refers to mAP@0.5.

4.3. Experimental Results and Analysis

4.3.1. Experimental Comparison of Optimized Loss Functions

To verify its contribution to model improvement, the loss function used herein was compared with other common loss functions. Figure 7a shows a comparison of the mAP curves of YOLOv5s with EioU and other loss function models for tears, and Figure 7b shows a comparison of YOLOv5s with α-IoU for scratches. The detection accuracy and network performance improved with the increasing mAP value. With the same dataset, the improved model after 100 iterations exhibited a higher recall value and higher mAP than the model with the mainstream attention mechanism, representing a 2.9% improvement compared with the YOLOv5s network.

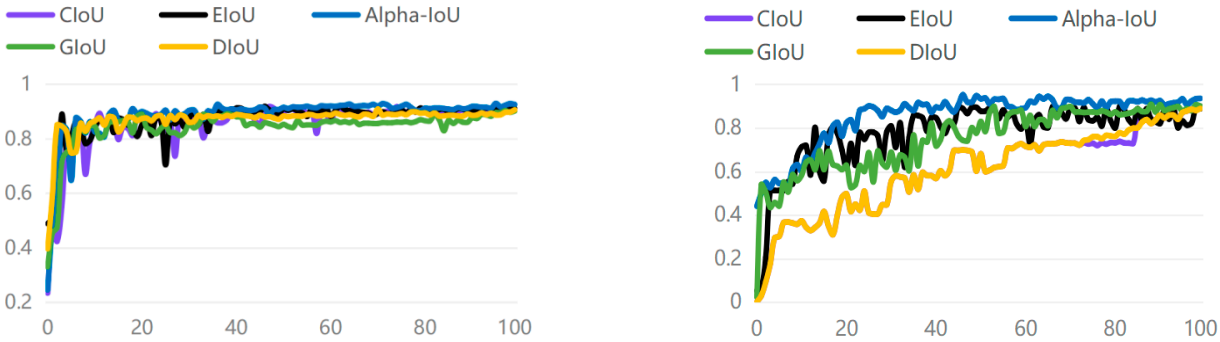


Figure 7. Comparison of loss curves of different loss functions. (**a**) Tear detection. (**b**) Scratch detection.

4.3.2. Experimental Comparison of Optimized CCA Mechanism

On the basis of the proposed CCA mechanism, the validity and rationality of the CCA mechanism are verified in two cases. In the first case, the CCA mechanism was used to replace the last C3 module in the backbone for the experiment, and in the second case, the CCA mechanism was added after the SPP module for the experiment. Meanwhile, to further verify its effectiveness in defect detection in nitrile medical gloves, this article compared the proposed CCA mechanism with the mainstream attention mechanisms: SE, CBAM, original CA, and ECA. As shown in Table 2, the CCA mechanism added after SPP exhibits the best recall value and mAP and meets the requirements of defect detection in nitrile medical gloves with a small increase in the values of the parameters.

4.3.3. Experimental Comparison of the Optimization Mechanism

To effectively analyze the performance of the improved model, the original and CCA-YOLO networks were trained using the same training parameters and methods. To strengthen the ability of the YOLOv5s model to extract spatial information features, this study focused on redesigning the feature extraction network of YOLOv5s. Module A adds

small-target feature detection, and module B introduces the CCA mechanism. Module C replaces CIoU with the α-IoU loss function to improve the scratch detection accuracy, and module D replaces CIoU with EIoU to improve the tear detection accuracy.

Table 2. The data sheet of ablation experiment 2.

Model	P (%)	R (%)	mAP@0.5 (%)
YOLOv5s	93.8	90.1	90.1
SE	94.0	91.1	91.7
CBAM	92.8	91.1	90.9
ECA	92.5	91.2	90.2
CA	97.3	90.0	91.7
CCA (replace)	92.7	93.7	92.0
CCA (add)	94.3	96.1	92.6

The aim of this ablation comparison experiment is to verify the optimization of each improvement module. From the experimental data in Table 3, it can be seen that under the same experimental parameters, if the three types of datasets, vertical angle, horizontal angle, and scratches, are placed in the same model for training, the mAP value is lower than 80%, and the detection effect is poor. If the three models are trained separately and independently, the mAP values are higher than 90%. Meanwhile, according to the experimental results, it is found that, based on the different features of tear and scratch, the introduction of EIoU to detect glove tear defects and the introduction of α-IoU to detect glove scratch defects can improve the two-week defect detection accuracy, and the detection effect is better.

Table 3. The data sheet of ablation experiment 3.

Datasets	A	B	C	D	E	P (%)	R (%)	mAP
Vertical + positive + scratch	√	√	√	√		80.8	81.9	78.3
Vertical + positive + scratch	√	√	√		√	81.3	83.7	79.8
Vertical	√	√	√	√		93.0	92.1	92.6
Vertical	√	√	√		√	90.4	93.6	90.7
Positive	√	√	√	√		94.3	93.7	92.8
Positive	√	√	√		√	94.1	91.1	91.6
Scratch	√	√	√	√		92.7	93.7	91.1
Scratch	√	√	√		√	94.3	93.7	92.8

Symbol "√" indicates that the module is used during the experiment.

Table 4 shows the dataset acquired at a vertical angle and generated under the condition of the same training parameters for 100 epochs.

Table 4. The data sheet of ablation experiment 4.

Model	P (%)	R (%)	mAP	Param	GFLOPs
YOLOv5s	93.8	90.0	90.1	70.6	16.4
YOLOv5s + A	92.8	94.9	93.6	72.5	16.8
YOLOv5s + B	94.1	94.8	91.7	74.3	16.9
YOLOv5s + C	94.1	93.5	92.6	70.6	16.4
YOLOv5s + D	93.9	91.1	91.1	70.6	16.4

As shown in Table 4, the aforementioned changes improved the detection of small scratch targets in nitrile gloves. In particular, mAP increased by 3.5% after the addition of the small-target detection layer, by 1.6 percentage points after the addition of the CCA modules, and by 2.5 percentage points after the use of the α-IoU loss function. After using the EIoU loss function, mAP increased by 1 percentage point. According to the data analysis, the improved module positively affects the feature extraction capability with a small

number of parameters, which further verifies that the improved model is more suitable for defect detection in nitrile gloves.

To more objectively and accurately evaluate the detection effect of the CCA-YOLO network on the vertical defect dataset, a confusion matrix comparison between YOLOv5s and CCA-YOLO is shown in Figure 8. It can be seen from the figure that the true value is close to the predicted value, and the ratio of damaged positive samples has increased by 25%, indicating that the accuracy of the model has been greatly improved.

After verifying the effectiveness and performance of each module, the proposed algorithm was compared with the current mainstream target detection algorithms YOLOv3, YOLOX, YOLOv5s, YOLOv7, and YOLOv8. Table 5 presents the comparison results, indicating the training data generated by the dataset shot at a vertical angle and under the condition of the same training parameters for 100 epochs.

Table 5. The data sheet of ablation experiment 5.

Model	P (%)	R (%)	mAP@0.5 (%)	FPS/fps
YOLOv3	84.7	97.5	90.8	31.9
YOLOv5s	93.8	90.0	90.1	37.1
YOLOX	89.7	91.9	92.1	36.2
YOLOv7	92.5	95.4	91.8	36.0
YOLOv8	94.3	92.1	92.5	38.7
CCA-YOLO	94.3	93.7	92.8	36.8

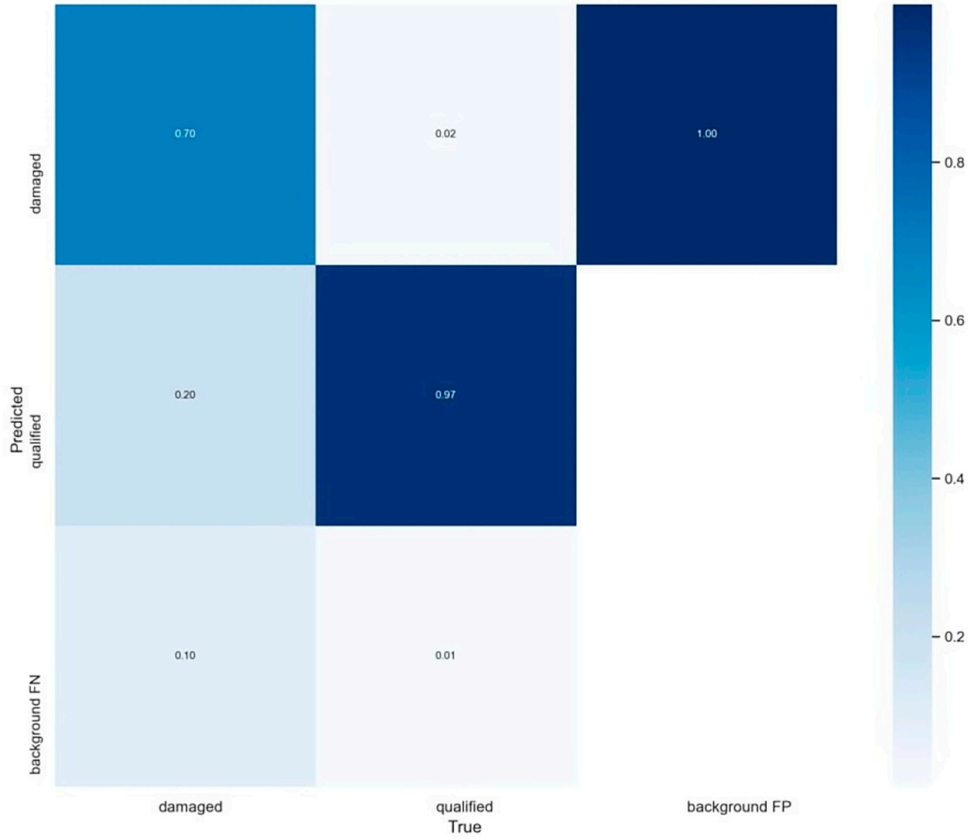
The experiments show that the algorithm achieves an mAP of 92.8% in the target detection task for nitrile medical gloves. Compared with the unimproved YOLOv5s algorithm, this algorithm achieves a 2.8-percentage-point improvement in mAP, a 1.5-percentage-point improvement in P, and a 3.7-percentage-point improvement in R. Compared with the more stable and frequently used YOLOv7 algorithm, this algorithm exhibits higher P and mAP values but lower recall. Compared with the YOLOv8 algorithm, CCA-YOLO performs better in the P, mAP, and FPS values, while reducing the training time by more than half.

To further improve the training, the best weight feeding into the detection network is obtained, and higher precision and a faster monitoring rate are achieved; high-quality data samples were selected from 39,941 data enhancement images for training, the learning rate of the initial training was set to 0.005, the number of picture batches was 32, and the number of training epochs for each dataset was 1000. The training results are shown in Table 6.

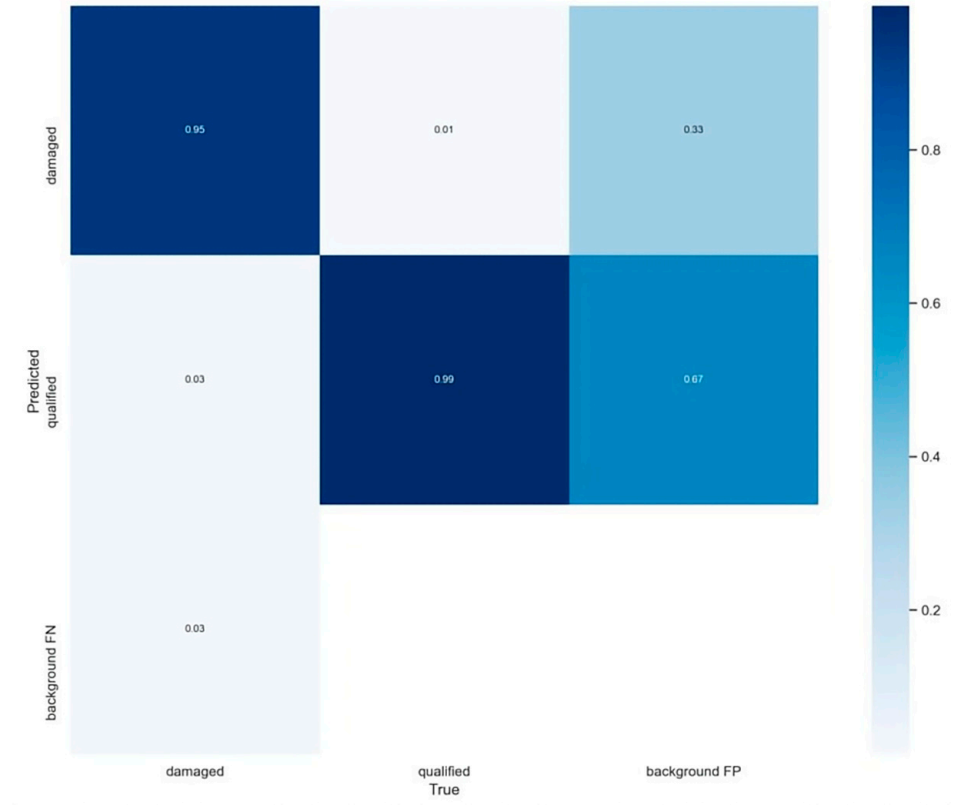
Table 6. The data sheet of ablation experiment 6.

Datasets	Data Volume	P (%)	R (%)	mAP@0.5 (%)
Horizontal tears	18,000	99.3	95.4	99.3
Vertical tears	18,000	99.7	99.7	99.8
Vertical scratches	2300	99.4	99.4	99.6

To better verify the detection performance of the proposed algorithm in cases of small defect targets, the original YOLOv5s network and the CCA-YOLO network were used herein to select some of the image data from the test set for testing, as shown in Figure 9. Figure 9a,b compare three detection results, namely, horizontal tear detection, vertical tear detection, and vertical scratch detection, respectively. Based on these figures, the detection accuracy of the proposed algorithm is significantly higher than YOLOv5l. In Figure 9, YOLOv5s misdetected horizontal tears and failed to accurately detect the target, whereas the proposed algorithm accurately detected tear and scratch targets. This was mainly because the detection layer for small targets, which was added to the proposed model, improved its resolution and the coordinate attention mechanism enlarged the receptive field. Moreover, the EIoU and α-IoU loss functions used for defect types improved the prediction positioning accuracy and detection accuracy for small targets.



(**a**)



(**b**)

Figure 8. Confusion matrix contrast. (**a**) YOLOv5s confusion matrix. (**b**) CCA-YOLO confusion matrix.

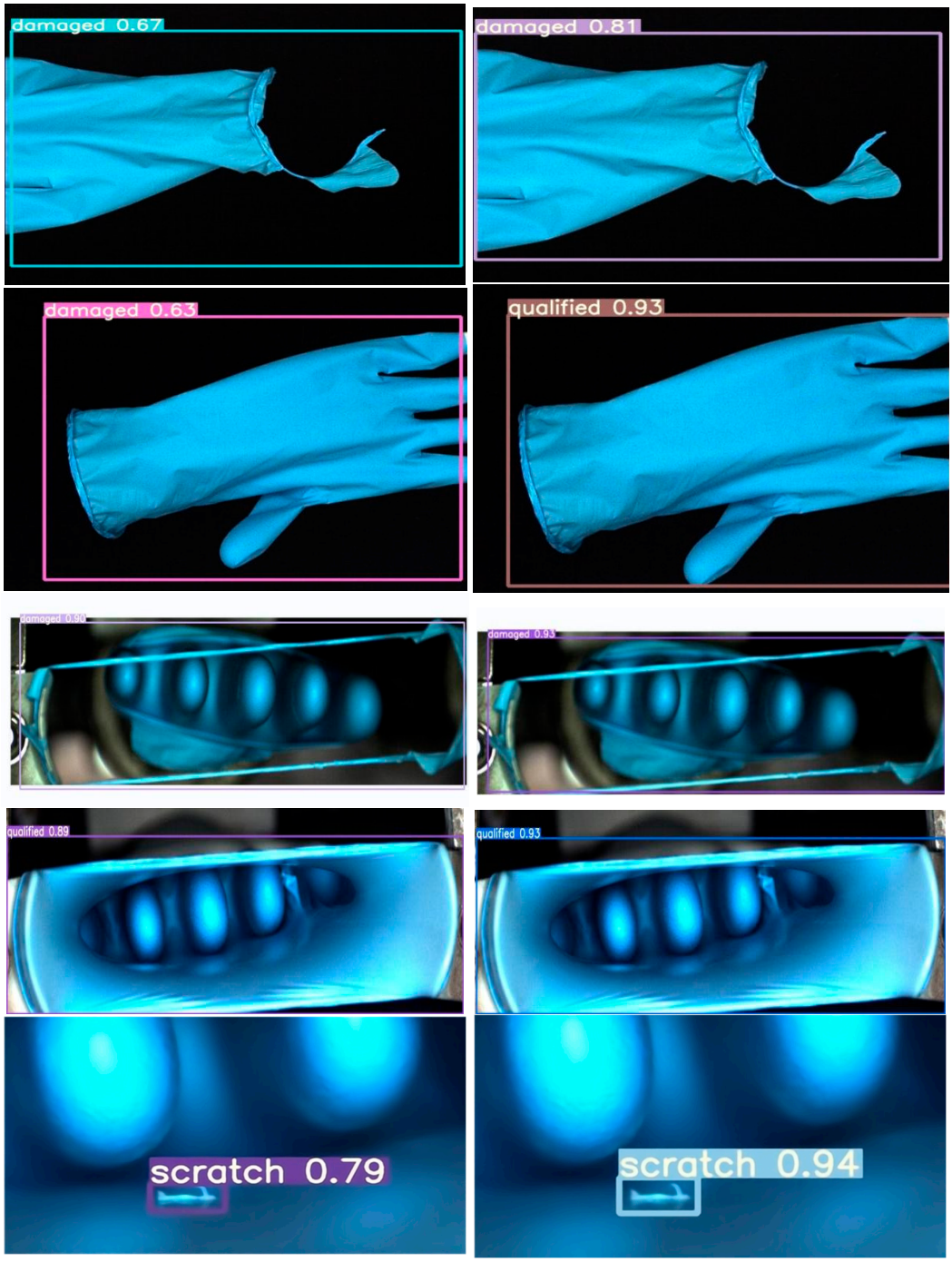


Figure 9. Comparison of the monitoring effect. (**a**) YOLOv5 detection results. (**b**) CCA-YOLO detection results.

5. Conclusions

Despite its remarkable achievements in defect monitoring, deep learning technology cannot successfully be fully utilized for defect detection in nitrile medical gloves. Therefore, herein, a CCA-YOLO algorithm for glove defect detection is proposed. A detection layer for small targets was added to improve the detection accuracy for glove scratches. CCA mechanisms were introduced to obtain a larger receptive field and further improve the feature extraction ability of the network for tear and scratch defects. Furthermore, to improve the positioning accuracy and reduce the target omission rate, the EIoU was introduced to detect glove tear defects and α-IoU was introduced to detect glove scratch defects. With the

experimental equipment unchanged, the mAP values of the three defect detection models for the detection of horizontal tears, vertical tears, and scratches reached 99.3%, 99.8%, and 99.6%, respectively, exhibiting increments of 4.2, 5.3, and 12.4%, respectively, compared with those of the original YOLOv5s. On the basis of the considerably improved mAP, the frame rate reached 36.8 FPS in the experimental environment, meeting the requirements of the real-time detection of glove defects. CCA-YOLO is very applicable to small-target detection and can solve most of the defect detection problems, and thus, can be extended to the industrial surface defect detection field.

Our experiments show that the proposed algorithm exhibits a high accuracy and recall rate as well as a small model size; furthermore, the algorithm can be easily deployed and meets the requirements for detecting glove tears, scratches, and other defects. However, there are a few limitations, mainly in the following two aspects:

(1) The performance of the model reaches a bottleneck using the current hardware; however, in the actual application environment, the hardware requirements are higher and there are more restrictions on the physical size or use environment. The complex network structure increases the number of parameters and calculation complexity, which is not conducive to the deployment of the model in practice.

(2) In practical applications, the defects of nitrile medical gloves are not limited to tears and scratches. There are many types of defects with different characteristics; therefore, the model needs to be further improved to detect more types of defects. Future studies will aim to decrease the number of model parameters, increase the types of detected defects, further reduce the consumption of labor costs, and accelerate the development of the intelligent detection of defects in nitrile medical gloves.

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

Funding: This work was supported by the Industry-University-Research Innovation Foundation of the Chinese University (2021LDA06003); the Provincial Graduate Student Innovation Ability Training Funding Project of Hebei Provincial Education Department (CXZZSS2023058) and Science Foundation of Hebei Normal University (L2021B31).

Data Availability Statement: Not applicable.

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

References

1. Yew, G.Y.; Tham, T.C.; Law, C.L.; Chu, D.-T.; Ogino, C.; Show, P.L. Emerging crosslinking techniques for glove manufacturers with improved nitrile glove properties and reduced allergic risks. *Mater. Today Commun.* **2019**, *19*, 39–50. [CrossRef]
2. Brito, K.; Vasconcelos, S.; Neto, G.F.; Damasceno, A.; Figueirêdo, M.; Ramos, W.; Brito, R. Semi-batch industrial process of nitriles production: Dynamic simulation and validation. *Comput. Chem. Eng.* **2018**, *119*, 38–45. [CrossRef]
3. Sohn, R.L.; Murray, M.T.; Franko, A.; Hwang, P.K.; Dulchavsky, S.A.; Grimm, M.J. Detection of surgical glove integrity. *Am. Surg.* **2000**, *66*, 302–306. [CrossRef]
4. Murray, C.A. Pinhole defects in nitrile gloves. *Br. Dent. J.* **2003**, *195*, 505. [CrossRef]
5. Thang, K.; Lai, N.S. Automated detection of glove defects using vision control. *Int. J. Eng. Technol.* **2016**, *16*, 18–23.
6. Sun, X.; Chen, Q. Defects detecting of gloves based on machine vision. In Proceedings of the IEEE International Conference on Real-Time Computing & Robotics (RCAR), Angkor Wat, Cambodia, 6–10 June 2016; pp. 169–173. [CrossRef]
7. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587. [CrossRef]
8. 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]
9. Girshick, R. Fast R-CNN. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 1440–1448.

10. Ren, S.; He, K.; 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]
11. Liu, H.; Ren, Y.Z.; Wang, L.; Zeng, Q.; Yang, Y.W.; Xu, R.J. The Research of Nitrile Gloves Visual On-Line Automatic Surface Defect Inspection System. *Appl. Mech. Mater.* **2014**, *701–702*, 560–564. [CrossRef]
12. Haq, M.A.; Hassine, S.B.H.; Malebary, S.J.; Othman, H.A.; Tag-Eldin, E.M. 3D-CNNHSR: A 3-Dimensional Convolutional Neural Network for Hyperspectral Super-Resolution. *Comput. Syst. Sci. Eng.* **2023**, *47*, 2689–2705.
13. 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, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
14. Redmon, J.; Farhadi, A. YOLO9000: Better, Faster, Stronger. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 6517–6525.
15. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
16. Bochkovskiy, A.; Wang, C.-Y.; Mark Liao, H.-Y. Yolov4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
17. Zheng, G.; Liu, S.; Wang, F.; Li, Z.; Sun, J. Yolox: Exceeding Yolo series in 2021. *arXiv* **2021**, arXiv:2107.08430.
18. Jawaharlalnehru, A.; Sambandham, T.; Sekar, V.; Ravikumar, D.; Loganathan, V.; Kannadasan, R.; Khan, A.A.; Wechtaisong, C.; Haq, M.A.; Alhussen, A.; et al. Target Object Detection from Unmanned Aerial Vehicle (UAV) Images Based on Improved YOLO Algorithm. *Electronics* **2022**, *11*, 2343. [CrossRef]
19. How, Y.C.; Nasir, A.F.A.; Muhammad, K.F.; Majeed, A.P.P.A.; Razman, M.A.M.; Zakaria, M.A. Glove defect detection via YOLO V5. *Mekatronika* **2022**, *3*, 25–30. [CrossRef]
20. Wang, H.; Wang, Y. Improved glove defect detection algorithm based on YOLOv5 framework. In Proceedings of the 2022 IEEE 6th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Beijing, China, 3–5 October 2022; pp. 1192–1197. [CrossRef]
21. Fan, Y.; Qiu, Q.; Hou, S.; Li, Y.; Xie, J.; Qin, M.; Chu, F. Application of improved YOLOv5 algorithm in parking lot fire detection. *J. Zhengzhou Univ. (Eng. Sci. Ed.)* **2022**, *11*, 2344. [CrossRef]
22. Roy, A.G.; Navab, N.; Wachinger, C. Concurrent spatial and channel 'squeeze & excitation' in fully convolutional networks. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Granada, Spain, 16–20 September 2018; pp. 421–429. [CrossRef]
23. Li, Y.; Fan, Y.; Wang, S.; Bai, J.; Li, K. Application of YOLOv5 Based on Attention Mechanism and Receptive Field in Identifying Defects of Thangka Images. *IEEE Access* **2022**, *10*, 81597–81611. [CrossRef]
24. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path aggregation network for instance segmentation. In Proceedings of the IEEE Conference on Compute Vision & Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 8759–8768. [CrossRef]
25. Zheng, Z.; Wang, P.; Ren, D.; Liu, W.; Ye, R.; Hu, Q.; Zuo, W. Enhancing geometric factors in model learning & inference for object detection & instance segmentation. *IEEE Trans. Cybern.* **2022**, *52*, 8574–8586. [CrossRef]
26. Jun, Y.; Yinshan, J. Improved small target detection algorithm for YOLOv5. *Comput. Eng.* **2023**, *14*, 387–394.
27. He, J.; Erfani, S.; Ma, X.; Bailey, J.; Chi, Y.; Hua, X.-S. Alpha-IoU: A Family of power intersection over union losses for bounding box regression. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 20230–20242. [CrossRef]
28. Hou, Q.; Zhou, D.; Feng, J. Coordinate Attention for Efficient Mobile Network Design. In Proceedings of the 2021 IEEE Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 19–25 June 2021; IEEE Publications: Piscataway, NJ, USA, 2021; pp. 13713–13722.
29. Hao, Z.; Shunyong, Z.; Yalan, Z.; Sicheng, L.; Xue, L. Wood Surface defects based on improved YOLOv5s Detection algorithm. *Wood Sci. Technol.* **2023**, *37*, 8–15.
30. Rezatofighi, H.; Tsoi, N.; Gwak, J.; Sadeghian, A.; Reid, I.; Savarese, S. Generalized intersection over union: A metric and a loss for bounding box regression. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 658–666. [CrossRef]
31. Zheng, Z.; Wang, P.; Liu, W.; Li, J.; Ye, R.; Ren, D. Distance-IoU loss: Faster and better learning for bounding box regression. *AAAI* **2020**, *34*, 12993–13000. [CrossRef]
32. Zhang, Y.-F.; Ren, W.; Zhang, Z.; Jia, Z.; Wang, L.; Tan, T. Focal and efficient IoU loss for accurate bounding box regression. *Neurocomputing* **2022**, *506*, 146–157. [CrossRef]
33. Ma, J.; Zhang, Z.; Xiao, W.; Zhang, X.; Xiao, S. Flame and Smoke Detection Algorithm Based on ODConvBS-YOLOv5s. *IEEE Access* **2023**, *11*, 34005–34014. [CrossRef]