axioms

MDPI

*Article*

# Improved Method for Oriented Waste Detection

**Weizhi Yang [1,*], Yi Xie [2] and Peng Gao [3]**

[1] School of Information and Intelligent Engineering, Guangzhou Xinhua University, Dongguan 523133, China
[2] School of Computer Science and Engineering, Sun Yat-Sen University, Guangzhou 510006, China
[3] College of Electronic Engineering, South China Agricultural University, Guangzhou 510642, China
* Correspondence: weizhiyangq@163.com

**Abstract:** Waste detection is one of the main problems preventing the realization of automated waste classification, which is a basic function for robotic arms. In addition to object identification in general image analysis, a waste-sorting robotic arm not only needs to identify a target object but also needs to accurately judge its placement angle so that it can determine an appropriate angle for grasping. In order to solve the problem of low-accuracy image detection caused by irregular placement angles, in this work, we propose an improved oriented waste detection method based on YOLOv5. By optimizing the detection head of the YOLOv5 model, this method can generate an oriented detection box for a waste object that is placed at any angle. Based on the proposed scheme, we further improved three aspects of the performance of YOLOv5 in the detection of waste objects: the angular loss function was derived based on dynamic smoothing to enhance the model's angular prediction ability, the backbone network was optimized with enhanced shallow features and attention mechanisms, and the feature aggregation network was improved to enhance the effects of feature multi-scale fusion. The experimental results showed that the detection performance of the proposed method for waste targets was better than other deep learning methods. Its average accuracy and recall were 93.9% and 94.8%, respectively, which were 11.6% and 7.6% higher than those of the original network, respectively.

**Keywords:** waste classification; angle detection box; dynamic smoothing; YOLOv5

**MSC:** 68T20; 68T45; 68U10



**Citation:** Yang, W.; Xie, Y.; Gao, P. Improved Method for Oriented Waste Detection. *Axioms* **2023**, *12*, 18. https://doi.org/10.3390/axioms12010018

Academic Editor: Oscar Humberto Montiel Ross

Received: 14 November 2022
Revised: 11 December 2022
Accepted: 21 December 2022
Published: 24 December 2022



## 1. Introduction

Waste disposal is an important problem worldwide that must be addressed. Classifying waste and implementing differentiated treatments can help to improve resource recycling and promote environmental protection. However, many countries and regions still rely on manual waste classification. The main drawbacks of this are twofold. First, the health of operators can be seriously threatened by the large number of bacteria carried by waste [1]. Second, manual sorting is not only costly but also inefficient. Consequently, automated waste management and classification approaches have received extensive attention [2].

Using a robotic arm is a common method for replacing the manual mode with automated waste sorting [3]. In order to enable the robot arm to correctly classify and grasp the target object, each robot arm needs to have the functions of object recognition and placement angle judgment.

Wu et al. [4] proposed a plastic waste classification method based on FV-DCNN. They extracted classification features from original spectral images of plastic waste and constructed a deep CNN classification model. Their experiments showed that the model could recognize and classify five categories of polymers. Chen et al. [5] proposed a lightweight feature extraction network based on MobileNetv2 and used it to achieve image classification of waste. Their experiments showed that the average accuracy of the classification with their dataset was 94.6%. Liu et al. [6] proposed a lightweight neural network based

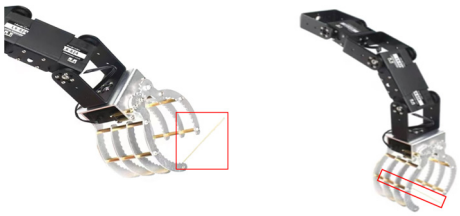
on MobileNet that can reduce the cost of industrial processes. Kang et al. [7] proposed an automated waste classification system based on the ResNet-34 algorithm. The experimental results showed that the classification had high accuracy, and the classification speed of the system was as quick as 0.95 s.

These models can recognize categories of waste intelligently based on convolutional neural networks, but they do not generate location boxes for the waste targets. In addition, when there are multiple categories of waste in an image, such models cannot achieve effective identification. Therefore, they cannot be applied directly to tasks such as automated waste sorting.

Cui et al. [8] used the YOLOv3 algorithm to detect domestic waste, decoration waste and large waste on the street. Xu et al. [9] proposed a five-category waste detection model based on the YOLOv3 algorithm and achieved 90.5% average detection accuracy with a self-made dataset. The dataset included paper waste, plastic products, glass products, metal products and fabrics. Chen et al. [10] proposed a deep learning detection network for the classification of scattered waste regions and achieved good detection results. Majchrowska et al. [11] proposed deep learning-based waste detection method for natural and urban environments. Meng et al. [12] proposed a MobileNet-SSD with FPN for waste detection.

These methods can achieve image-based waste detection, but they do not provide the grasping angle information for the target object. For a target object placed at any angle, these methods only provide a horizontal identification box. Therefore, the robotic arm cannot determine the optimal grasp mode for the shape and placement angle of a waste object, which may easily lead to the object falling or to grabbing failure, especially in cases involving a large aspect ratio, as a small angle deviation can lead to a large deviation in the intersection over union (IoU).

In addition to object identification in general image analysis, a waste-sorting robotic arm not only needs to identify a target object but also needs to accurately judge its placement angle so that the robotic arm can determine the appropriate grasping angle. YOLOv5 has a strong feature extraction structure and feature aggregation network, allowing it to achieve higher detection recall and accuracy. It also provides a series of methods that can be used to achieve data enhancement. YOLOv5 is a good choice for many common identification and classification problems due to its fast detection speed, high detection accuracy and easy deployment, making it popular in many practical engineering applications. Li et al. [13] and Chen et al. [14] proposed improved algorithms for vegetable disease and plant disease detection based on YOLOv5. Their experiments showed that the detection rates reached 93.1% and 70%, respectively, which were better than other methods. Ling et al. [15] and Wang et al. [16] proposed gesture recognition and smoke detection models, respectively, based on YOLOv5. Gao et al. [17] proposed a beehive detection model based on YOLOv5. However, the original YOLOv5 does not provide the grasping angle information required for a target object. For a target object placed at any angle, it only provides a horizontal identification box, as shown in Figure 1a. Therefore, the robotic arm cannot determine the optimal grasp mode for the shape and placement angle of a target object, which may easily lead to the object falling or to grabbing failure, especially in cases involving a large aspect ratio.



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

**Figure 1.** Grabbing with horizontal and oriented detection box. (**a**) Horizontal detection box; (**b**) oriented detection box.

In this work, we made two modifications to YOLOv5 to improve its suitability for automated waste sorting application scenarios. First, we added an angular prediction network in the detection head to provide grasping angle information for the waste object and developed a dynamic smoothing label for angle loss to enhance the angular prediction ability of the model. Second, we optimized the structure of the feature extraction and aggregation by enhancing the multi-scale feature fusion.

The contributions of this work are threefold:

(1) An optimized waste detection approach was designed based on YOLOv5 that provides higher detection accuracy for both general-sized waste and waste with a large aspect ratio;

(2) An angular prediction method is proposed for YOLOv5 that enables the rotation detection box to obtain the actual position of oriented waste;

(3) New optimization schemes are introduced for YOLOv5, including a loss function, feature extraction and aggregation.

## 2. Detection Method for Oriented Waste

### *2.1. Detection Scheme*

As shown in Figure 2, the framework of the proposed waste detection scheme consists of five parts: the input layer, feature extraction backbone network, feature aggregation network, detection head and dynamic smoothing module. In this study, the backbone network mainly consisted of the focus module, the convolution module and an optimized HDBottleneckCSP module based on BottleneckCSP. The focus module reduces the number of computations and improves the speed in accordance with the slicing operation. The BottleneckCSP module is a convolution structure that demonstrates good performance in model learning. The backbone was used to extract the features from waste images and generate feature maps with three different sizes. The feature aggregation network converges and fuses multi-scale features generated from the backbone network to improve the representation learning ability for rotating waste angle features. The detection head generates the category, location and rotation angle for waste based on the multi-scale feature maps. Finally, the dynamic smoothing module partially densifies the "one-hot label encoding" of the angle labels for model training.

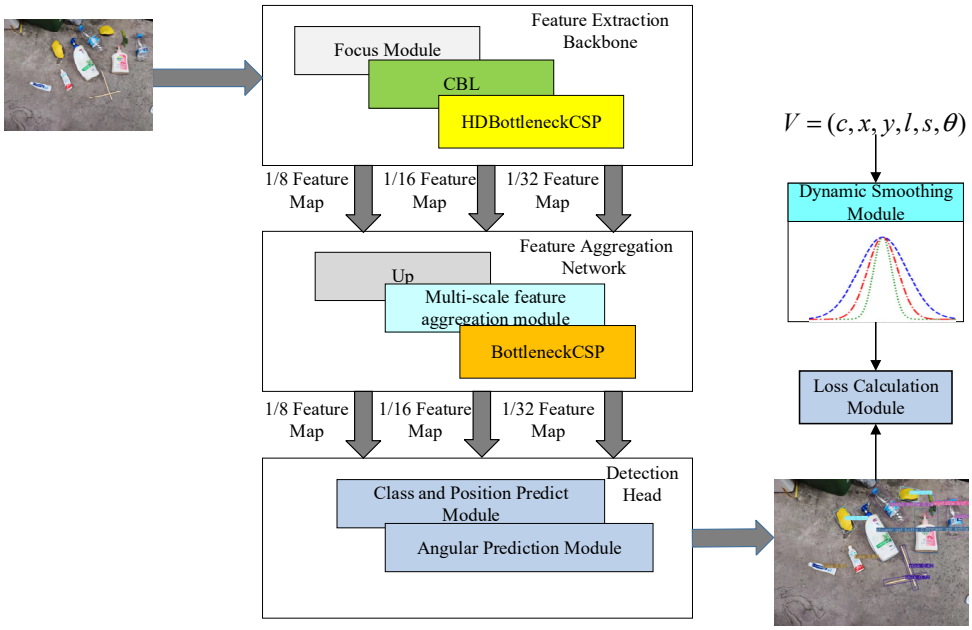



**Figure 2.** Rotation angle waste detection scheme.

### *2.2. Improvement of Detection Model*

The original YOLOv5 model has the following limitations: (i) It can only generate a target detection box with a horizontal angle and not a rotation angle. (ii) The stack of bottleneck modules in BottleneckCSP is serial, which causes the middle-layer features to be lost. (iii) The feature aggregation network lacks end-to-end connection between the input and output feature maps.

To solve these problems, we optimized three aspects of YOLOv5: (i) We added an angular prediction network and loss function, as well as a dynamic angle smoothing algorithm for angular classification, to improve the angular prediction ability. (ii) We optimized the BottleneckCSP module of the backbone network to enhance the model's ability to extract the features of oriented waste. (iii) We optimized the feature aggregation network to improve the effect of multi-scale feature fusion.

#### 2.2.1. Improvement of the Detection Head Network

The original YOLOv5 detector lacks a network structure for angular prediction and cannot provide the grasping angle information for waste objects. Therefore, the robotic arm cannot set the optimal grasp mode according to the placement angle of the waste, which easily leads to the object falling or to grabbing failure. Thus, we optimized the structure of the detection head.

Angular prediction can be realized as regression or classification. The regression mode produces a continuous prediction value for the angle but there is a periodic boundary problem, which leads to a sudden increase in the value of the loss function at the boundary of periodic changes, increasing the difficulty of learning [18]. For example, in the 180° long-side definition method, the defined label range is $(-90°, 90°)$. When the true angle of waste is 89° and the prediction is $-90°$, the error learned by the model is 179°, but the actual error should be 1°, which affects the learning of the model.

Therefore, we added convolution network branches in the detection head and defined the angle label with 180 categories obtained by rotating the long side of the target box clockwise around the center. The angle convolution network generates the angle prediction using information extracted from the multi-scale features obtained by the feature aggregation network.

In the detection head, the angle convolution network and the original network share the output of the feature aggregation network as the input feature graph. The output of the angle prediction network and the original network are merged as follows:

$$V = (\hat{c}, \hat{x}, \hat{y}, \hat{l}, \hat{s}, \hat{\theta}) \tag{1}$$

where $\hat{c}$ is the predicted category of the waste, $\hat{x}$ and $\hat{y}$ are the predicted central coordinates of the object box, $\hat{l}$ and $\hat{s}$ are the predicted lengths of the longer side and shorter side of the object box and $\hat{\theta}$ is the predicted angle of oriented waste.

#### 2.2.2. Angle Smoothing and Loss Function

The realization of angular prediction as classification can avoid the periodic boundary problem caused by regression, but there are still some limitations. The loss function of traditional category tasks is calculated as cross-entropy loss, and the form of the labels is "one-hot label encoding", as shown in Equations (2) and (3):

$$y_{ic} = \begin{cases} 1, c = \theta \\ 0, c \neq \theta \& c \in \{0, 1, \ldots, 179\} \end{cases} \tag{2}$$

$$L = -\frac{1}{N} \sum_{i} \sum_{c=0}^{179} y_{ic} \log(p_{ic}) \tag{3}$$

where $y_{ic}$ is the "one-hot label encoding" for the angle of sample $i$, $\theta$ is the angle of the oriented waste and $p_{ic}$ is the prediction of the detection model.

Equation (8) shows that, for different incorrect predictions of the angle, the same loss value is obtained and the distance of the mistake cannot be quantified, which makes it difficult for model training to determine the angle of the oriented waste.

To solve this problem, we propose a dynamic smoothing label algorithm based on the circular smooth label (CSL) algorithm [18] to optimize the "one-hot label encoding" label of the angle.

The circular smooth label algorithm is shown in Equation (4):

$$\mathrm{CSL}(x) = \begin{cases} g(x), \theta - r < x < \theta + r \ \& \ x \in \{0, 1, \ldots, 179\} \\ 0, \text{others} \end{cases} \tag{4}$$

where $\theta$ is the rotation angle value, $r$ is the range of smoothness and $g(x)$ is the smoothing function. The angle label vector manifests as a "dense" distribution because $g(x)$ is within the range of smoothness.

The value of the smoothing function is shown in Equation (5):

$$0 < g(\theta - \varepsilon) = g(\theta + \varepsilon) \leq 1, |\varepsilon| \leq r \tag{5}$$

where, when $\varepsilon = 0$, the function has a maximum value of 1, and when $\varepsilon = r$, it is 0.

The CSL algorithm partially densifies the "one-hot label encoding". When the angular prediction of the model is in the range of smoothness, different loss values for different predicted degrees are obtained; thus, it can quantify the mistake in the angle category prediction. However, the performance of CSL is sensitive to the range of smoothness. If the range of smoothness is too small, the smoothing label will degenerate into "one-hot label encoding" and lose its effect, and it will be difficult to learn the information from the angle. If the range is too large, the deviation in the angle prediction will be large, which will lead to it missing the object, especially for waste with a large aspect ratio.

Therefore, we propose a dynamic smoothing function for the angle label to adjust the smoothing amplitude and range.

The dynamic smoothing function uses the dynamic Gaussian function to smooth the angle labels. It can be seen from Figure 3 that the smoothing amplitude and the range of the Gaussian function are controlled by the root mean square (RMS) value: the larger

the RMS, the flatter the curve; the smaller the RMS, the steeper the curve and the smaller the smoothing range. Therefore, the RMS of the Gaussian function is gradually shrunk to achieve dynamic smoothing, as shown in Equation (6).

$$\mathrm{DSM}(x) = \exp\left(-\frac{d^2(x,\theta)}{2 \times b^2}\right), x \in \{0, 1, \ldots, 179\} \tag{6}$$

We provide two efficient functions—linear annealing and cosine annealing—to adjust the RMS, as follows:

$$b = c + e \times \cos\left(\frac{0.5 \times \pi \times epoch}{epochs}\right)$$

$$b = c - e \times \frac{epoch}{epochs}$$

where $\theta$ is the value of the rotation angle for the waste, which corresponds to the peak position of the function; $x$ is the encoding range of the waste angle; $b$ is the value of the RMS; and $d(x, \theta)$ is the circular distance between the encoding position and the angle values. For example, if $\theta$ is 179, $d(x, \theta)$ is 1 when $x$ is 0; $epoch$ and $epochs$ represent the current number of training rounds and the maximum number of rounds of the model, respectively, and $c$ and $e$ are hyper-parameters.

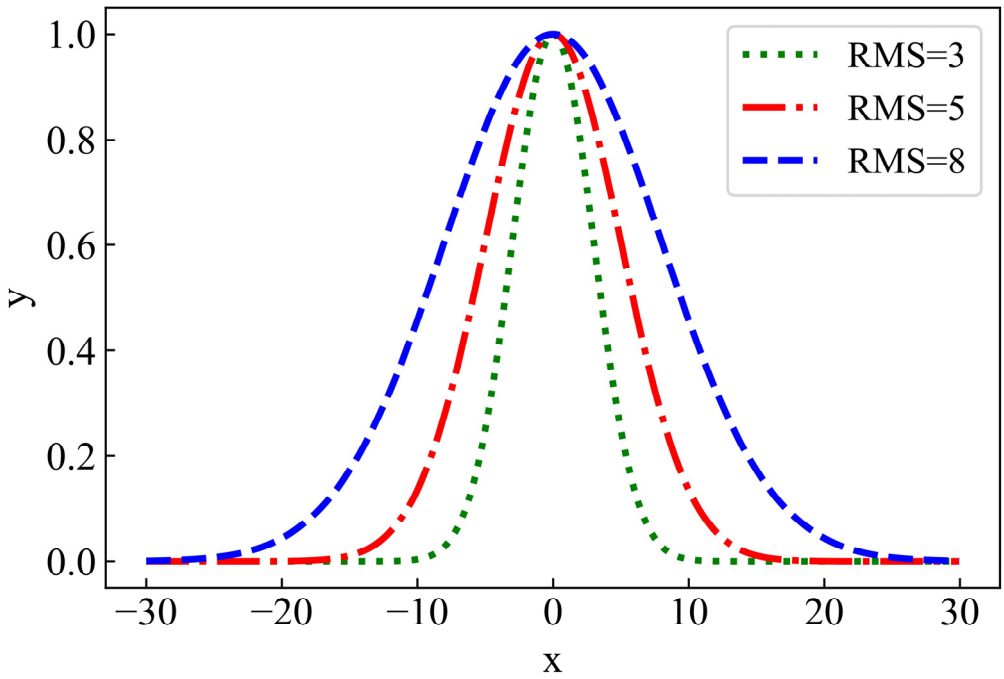



**Figure 3.** Gaussian function curves with different RMS.

It can be seen from Equation (6) that the DSM densifies the angle label according to the distance between the encoding position and the angle value dynamically. In the early stage of model training, $b$ obtained large values because of the small epoch. At this time, the range of smoothing was large, and the model's learning of angles was reflected in the window area. When the smoothing range was more "loose", the model came closer to the neighborhood area of the optimal point; thus, it reduced the difficulty of angle learning and improved the recall rate in image waste detection. The range of angle smoothing decreased with the increase in the *epoch* value. The objective of the model was changed from the optimal region to the learning of the optimal point so that the deviation in the angular prediction would be smaller. The higher accuracy of the angle prediction improved the recall rate for the oriented waste, especially in cases with a large aspect ratio.

The angular loss of waste was calculated using the cross-entropy loss function based on the dynamic smoothing algorithm:

$$loss(a) = -\sum_{i=0}^{s^2} I_{ij}^{obj} \sum_{t=0}^{179} \{\hat{p}_i(t) \log[p_i(t)] + [1 - \hat{p}_i(t)] \log[1 - p_i(t)]\} \tag{7}$$

where $\hat{p}_i(t) = \mathrm{DSM}(t)$. $p_i(t)$ is the prediction of the angle and $s^2$ is the quantification of the subdomain of the picture, and the model provides the prediction of the target for each subdomain. $I_{ij}^{obj}$ is 0 or 1, which indicates whether there is a target. When the prediction is close to the true value, the cross-entropy has a smaller value.

In addition, the GIoU loss function [19] was used to calculate the regression loss of the detection boundary box. In Figure 4, A and B are the real box and the prediction box of the detection target, respectively. C is the smallest rectangle surrounding A and B. The green area is $|\mathrm{C}| - |\mathrm{A} \cup \mathrm{B}|$.

The specific calculation is shown in Equations (8)–(10). GIoU not only pays attention to the overlap of the real box and the prediction box but also to the non-overlapping area, which allows it to solve the problem of the gradient not being calculated caused by A and B not intersecting.

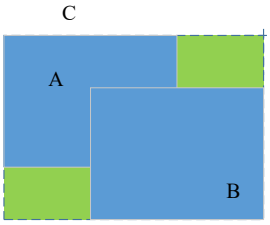



**Figure 4.** Illustration of GIoU.

$$\mathrm{IoU}(\mathrm{A}, \mathrm{B}) = \frac{|\mathrm{A} \cap \mathrm{B}|}{|\mathrm{A} \cup \mathrm{B}|} \tag{8}$$

$$\mathrm{GIoU}(\mathrm{A}, \mathrm{B}) = \mathrm{IoU}(\mathrm{A}, \mathrm{B}) - \frac{|\mathrm{C}| - |\mathrm{A} \cup \mathrm{B}|}{|\mathrm{C}|} \tag{9}$$

$$loss(r) = 1 - \mathrm{GIoU}(\mathrm{A}, \mathrm{B}) \tag{10}$$

In the equations, A and B are the real box and the prediction box of the detection target, respectively. C is the smallest rectangle surrounding A and B. The confidence loss function and category loss function are as shown by Equations (11) and (12):

$$\begin{aligned} loss(o) = \; & -\sum_{i=0}^{s^2} \sum_{j=0}^{B} I_{ij}^{obj} [\hat{c}_i \log(c_i) + (1 - \hat{c}_i) \log(1 - c_i)] \\ & -l_{noobj} \sum_{i=0}^{s^2} \sum_{j=0}^{B} I_{ij}^{noobj} [\hat{c}_i \log(c_i) + (1 - \hat{c}_i) \log(1 - c_i)] \end{aligned} \tag{11}$$

$$loss(c) = -\sum_{i=0}^{s^2} I_{ij}^{obj} \sum_{c \in class} \{\hat{p}_i(c) \log[p_i(c)] + [1 - \hat{p}_i(c)] \log[1 - p_i(c)]\} \tag{12}$$

where $I_{ij}^{obj}$ and $I_{ij}^{noobj}$ indicate whether the prediction box $j$ of the grid $i$ is the target box, and $\lambda_{\mathrm{noobj}}$ indicates the weight coefficients.

The overall loss function of the improved model is a weighted combination of the above loss functions, as shown in Equation (13):

$$Loss = loss(r) + loss(o) + loss(c) + loss(a) \tag{13}$$

2.2.3. Improvement of Feature Extraction Backbone Network

The feature extraction backbone network was used to extract the features of the waste in the image. Due to the addition of angular prediction in the detection of oriented waste, there is a higher demand on the feature extraction to realize effective recognition, especially in cases involving a large aspect ratio due to a narrow area.

BottleneckCSP is the main module in the backbone of YOLOv5. The BottleneckCSP module is stacked using a bottleneck architecture. As shown in Figure 5a, the stacking of the bottleneck modules is serial. With the deepening of the network, the feature abstraction capability is gradually enhanced, but shallow features are generally lost [20]. Shallow features have lower semantics and can be more detailed due to the fewer convolution operations. Utilizing multi-level features in CNNs through skip connections has been found to be effective for various vision tasks [21–23]. The bypassing paths are presumed to be the key factor for easing the training of deep networks. Concatenating feature maps learned by different layers can increase the variation in the input of subsequent layers and improve efficiency [24,25]. In addition, attention mechanisms, which are methods used to assign different weights to different features according to their importance, have been found to be effective for the recognition of an image [26,27]. The coordinate attention mechanism (CA) [28] is one such mechanism that shows good performance. Therefore, as shown in Equation (14), we concentrated and merged the middle features of BottleneckCSP and added the CA module to enhance the feature extraction capability. The attention mechanism is optional in the module at different levels.

$$\boldsymbol{Z}^{out} = g\left(\boldsymbol{Z}^{c}_{h\times w\times(c\times t)}\right) \tag{14}$$

where

$$\begin{cases} \boldsymbol{Z}^{1} = f_1(x) \\ \boldsymbol{Z}^{t} = f_t\left(\boldsymbol{Z}^{t-1}\right) \\ \boldsymbol{Z}^{c} = \left[\boldsymbol{Z}^{1}_{h\times w\times(c\times t)}, \boldsymbol{Z}^{2}_{h\times w\times(c\times t)}, \ldots, \boldsymbol{Z}^{t}_{h\times w\times(c\times t)}\right] \end{cases}$$

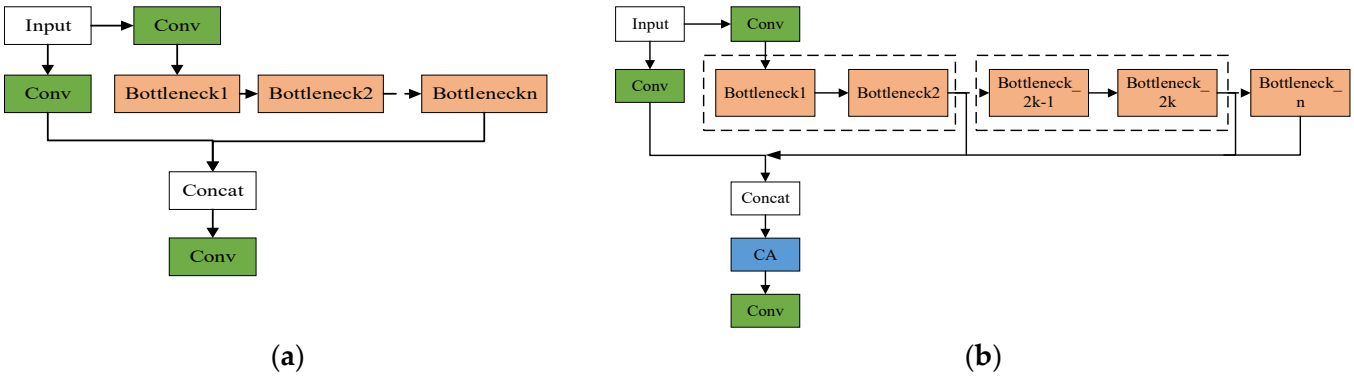



**Figure 5.** Comparison of BottleneckCSP before and after improvement. (**a**) BottleneckCSP module. (**b**) HDBottleneckCSP module.

$\boldsymbol{Z}$ is the feature map, $x$ is the input of the BottleneckCSP module, $f$ is the function mapping of the bottleneck module, and $g$ represents the CA attention operation.

Due to the "residual block" connection in the bottleneck architecture, excessive feature merging between bottlenecks leads to feature redundancy, which is not suitable for model training, and the increased number of parameters means that more resources are consumed. Therefore, the characteristic layers were connected using "interlayer merging", as shown in Figure 5b. The optimized module was named HDBottleneckCSP.

The CA module structure in HDBottleneckCSP is shown in Figure 6. The input feature maps are coded along the horizontal and vertical coordinates to obtain the global field

and to encode position information, respectively, which helps the network to detect the locations of targets more accurately.

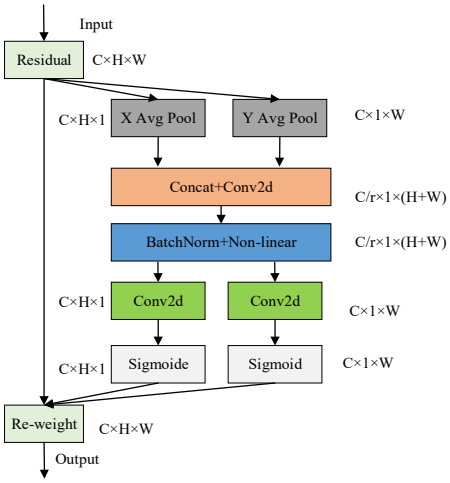



**Figure 6.** CA attention module.

As shown in Equation (15), the CA module generates vertical and horizontal feature maps for the input feature map and then transforms them through a $1 \times 1$ convolution. The generated $\boldsymbol{A} \in \mathbb{R}^{C/r\times(H+W)}$ is the intermediate feature map for the spatial information in the horizontal and numerical directions, $r$ is the down sampling scale and $F_1$ represents the convolution operation.

$$\boldsymbol{A} = \delta\left(F_1\left(\left[\boldsymbol{Z}^h, \boldsymbol{Z}^w\right]\right)\right) \tag{15}$$

where $\boldsymbol{A}$ is divided into $\boldsymbol{A}^h \in \mathbb{R}^{C/r\times H}$ and $\boldsymbol{A}^w \in \mathbb{R}^{C/r\times W}$ in the spatial dimension. As shown in Equations (16) and (17), it is transformed into the same number of channels as the input feature map through the convolution operation, while $g^h$ and $g^w$ are used as the attention weight and participate in the feature map operation. The output result of the CA module is shown in Equation (18).

$$g^h = \delta\left(F_h\left(\left[\boldsymbol{A}^h\right]\right)\right) \tag{16}$$

$$g^w = \delta(F_w([\boldsymbol{A}^w])) \tag{17}$$

$$\boldsymbol{y}_c(i,j) = \boldsymbol{x}_c(i,j) \times \boldsymbol{g}_c^h(i) \times \boldsymbol{g}_c^w(j) \tag{18}$$

The optimized feature extraction backbone network structure is shown in Figure 7. It extracts features through the convolution module and the HDBottleneckCSP module and generates feature maps with three sizes by downsampling (1/8, 1/16 and 1/32).

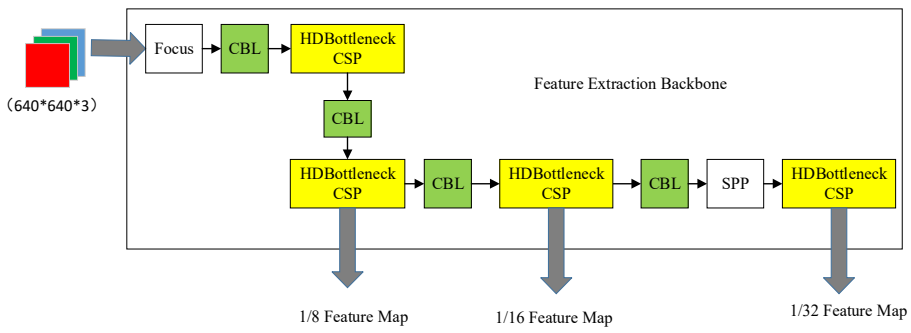



**Figure 7.** Backbone network structure for feature extraction.

2.2.4. Improvement of Feature Aggregation Network

The YOLOv5 feature aggregation network consists of feature pyramid networks [29] (FPNs) and path aggregation networks [30] (PANets). The structure of a PANet is shown in Figure 8a. The PANet aggregates features along two paths: top-down and bottom-up. However, the aggregated features are deep features with high semantics, and the shallow features with high resolution are not fused. In order to make use of the input features more effectively, we used P2P-PANet to replace the PANet based on BiFPN [31], as shown in Figure 8b.

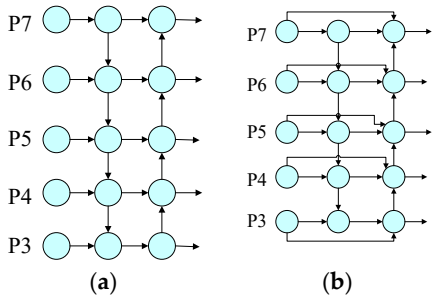



**Figure 8.** Network structures of PANet and P2P-PANet. (**a**) PANet network structure. (**b**) P2P-PANet network structure.

Compared to PANet, P2P-PANet adds end-to-end connection for the input-feature and output-feature maps, which establishes a "point-to-point" horizontal connection path from the low level to the high level, and it can realize the fusion of high-resolution and complex semantic features in an image without adding much cost. Through the extraction and induction of semantic information for the high-resolution and low-resolution feature maps, the angular feature information of rotating waste is further enhanced, and the detection ability of the model is improved.

The method for oriented waste detection after all the optimizations was named YOLOv5m-DSM and is shown in Figure 9. When a picture is input into the model, YOLOv5m-DSM extracts features using the backbone and generates downsampling feature maps with three different sizes for the detection of waste. The feature aggregation network undertakes feature aggregation and fusion to enhance the model's ability to learn features. The detection head generates the prediction information for waste targets based on the multi-scale features. In the model's training stage, the label of the training set is smoothed using the dynamic smoothing module, and the loss in the prediction, including class, angle and position, is calculated using the loss calculation module for iterative learning.

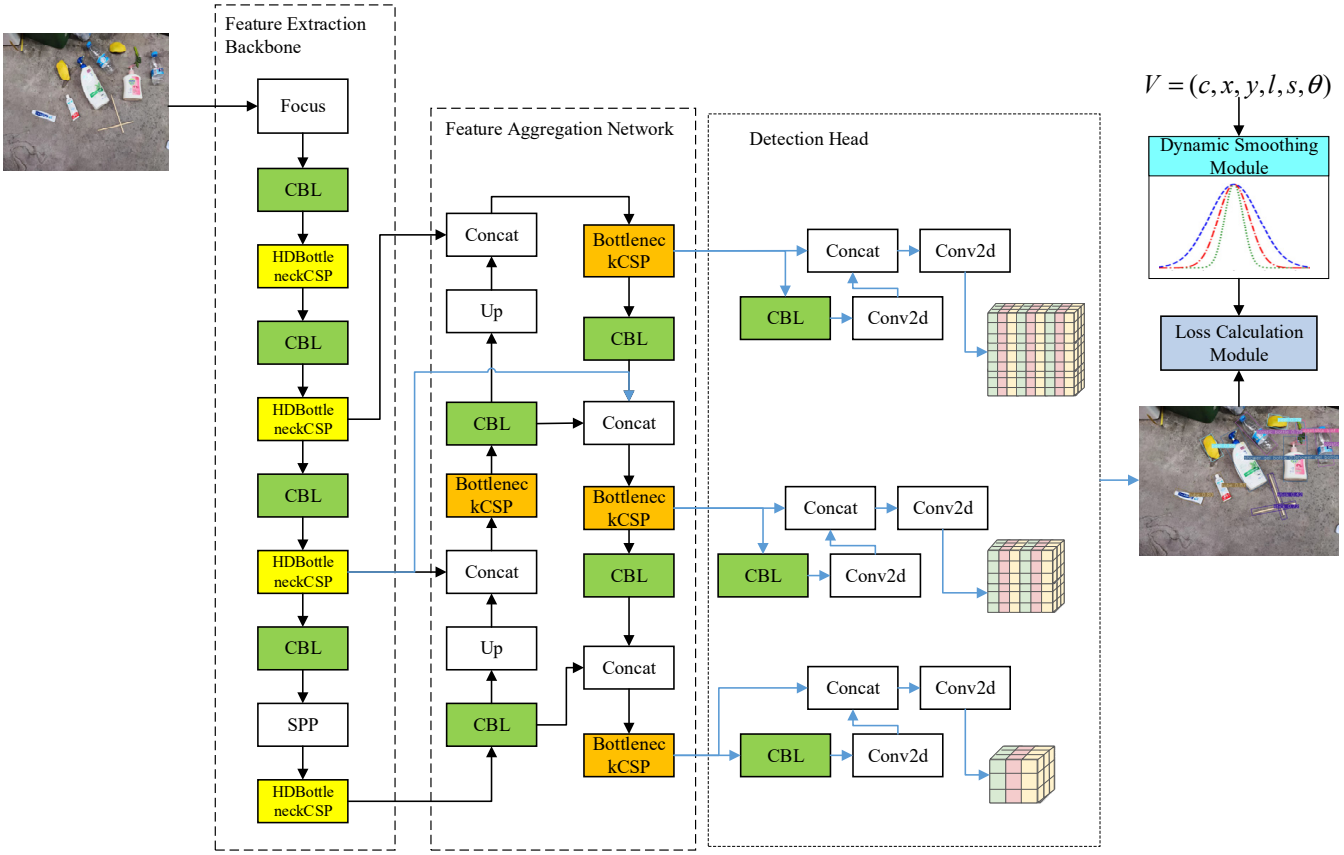



**Figure 9.** Method diagram for YOLOv5m-DSM.

## 3. Experimental Results and Analysis

### *3.1. Datasets*

The dataset for the experiment contained eleven kinds of domestic waste, including a cotton swab, a stick, paper, a plastic bottle, a tube, vegetables, peels, a shower gel bottle, a coat hanger, clothes pegs and an eggshell. The vector of the label contained the category, the center $x$ coordinate of the target box, the center $y$ coordinate of the target box, the long-side value, the short-side value and the angle value. The angle was the angle between the long side of the target frame and the horizontal axis in the clockwise direction, with a range of $(0°, 180°)$.

### *3.2. Evaluation Index*

In order to evaluate the performance of YOLOv5m-DSM, it was compared and analyzed using the recall ($R$), mean average precision ($mAP$) and other indicators. The recall is as follows:

$$R = \frac{TP}{TP + FN} \times 100\% \tag{19}$$

$TP$ represents a "true positive" sample, and $FN$ represents a "false negative" sample. The mean average precision formula is shown in Equation (20).

$$mAP = \frac{1}{m}\sum_{i=1}^{m} AP_i \tag{20}$$

The mean average precision refers to the average precision ($AP$) for each category of samples, which is calculated from the recall rate and precision ($P$) as follows:

$$R = \frac{TP}{TP + FN} \times 100\% \tag{21}$$

$$AP = \int_0^1 P(R)dR \tag{22}$$

### 3.3. Experimental Results and Analysis

In order to better show the advantages of the method described in this paper, YOLOv5-DSM was compared with mainstream horizontal box detection methods and rotating box detection methods in the experiments.

Table 1 shows a comparison of the detection effects for YOLOv5m-DSM and horizontal rectangular box detection methods, such as SSD-OBB, YOLOv3-OBB, YOLOv5s-OBB and YOLOv5m-OBB, which are angle classification network structures commonly added in detection heads based on their original models [32,33].

**Table 1.** Comparison between YOLOv5m-DSM and horizontal frame detection methods.

| Method | Recall/% | mAP/% | AP of "Large Aspect Ratio Category"/% | | | | |
|---|---|---|---|---|---|---|---|
| | | | Cotton Swab | Stick | Plastic Bottle | Shower Gel Bottle | Tube |
| SSD-OBB | 82.1 | 74.9 | 42.6 | 41.9 | 76.2 | 71.6 | 67.8 |
| YOLOv3-OBB | 82.6 | 75.8 | 41.6 | 40.5 | 79.8 | 75.5 | 68.1 |
| YOLOv5s-OBB | 84.7 | 77.5 | 43.4 | 40.7 | 85.8 | 76.5 | 71.5 |
| YOLOv5m-OBB | 87.2 | 82.3 | 52.1 | 57.1 | 83.8 | 79.6 | 72.1 |
| YOLOv5m-DSM (Cos) | 94.5 | 93.3 | 70.1 | 80.7 | 100 | 100 | 99.9 |
| YOLOv5m-DSM (Linear) | 94.8 | 93.9 | 78.7 | 81.0 | 100 | 100 | 100 |

Table 1 shows that, compared with SSD-OBB, YOLOv3-OBB and YOLOv5s-OBB, the recall rate and average precision of YOLOv5m were better. Compared with the original network, YOLOv5m-DSM showed improvements of 7.6% and 11.6% in the recall rate and the average precision, respectively. This proves that the modified waste detection algorithm has obvious improvements. Furthermore, YOLOv5m-DSM showed a good detection effect for oriented waste with a large aspect ratio, demonstrating an obvious improvement over the original model. The good performance of DSM (Cos) and DSM (Linear) proves that the dynamic smoothing label was effective and strong.

The detection effects of YOLOv5m, YOLOv5m-OBB and YOLOv5m-DSM are shown in Figure 10. It can be seen from Figure 10a that the YOLOv5m network only generated a horizontal detection box. It did not provide the grasping angle information for the waste object. Therefore, the robotic arm could not set the optimal grasp mode according to the inherent shape and placement angle of the target object, which could easily lead to the object falling and to grabbing failure, especially in cases involving a large aspect ratio.

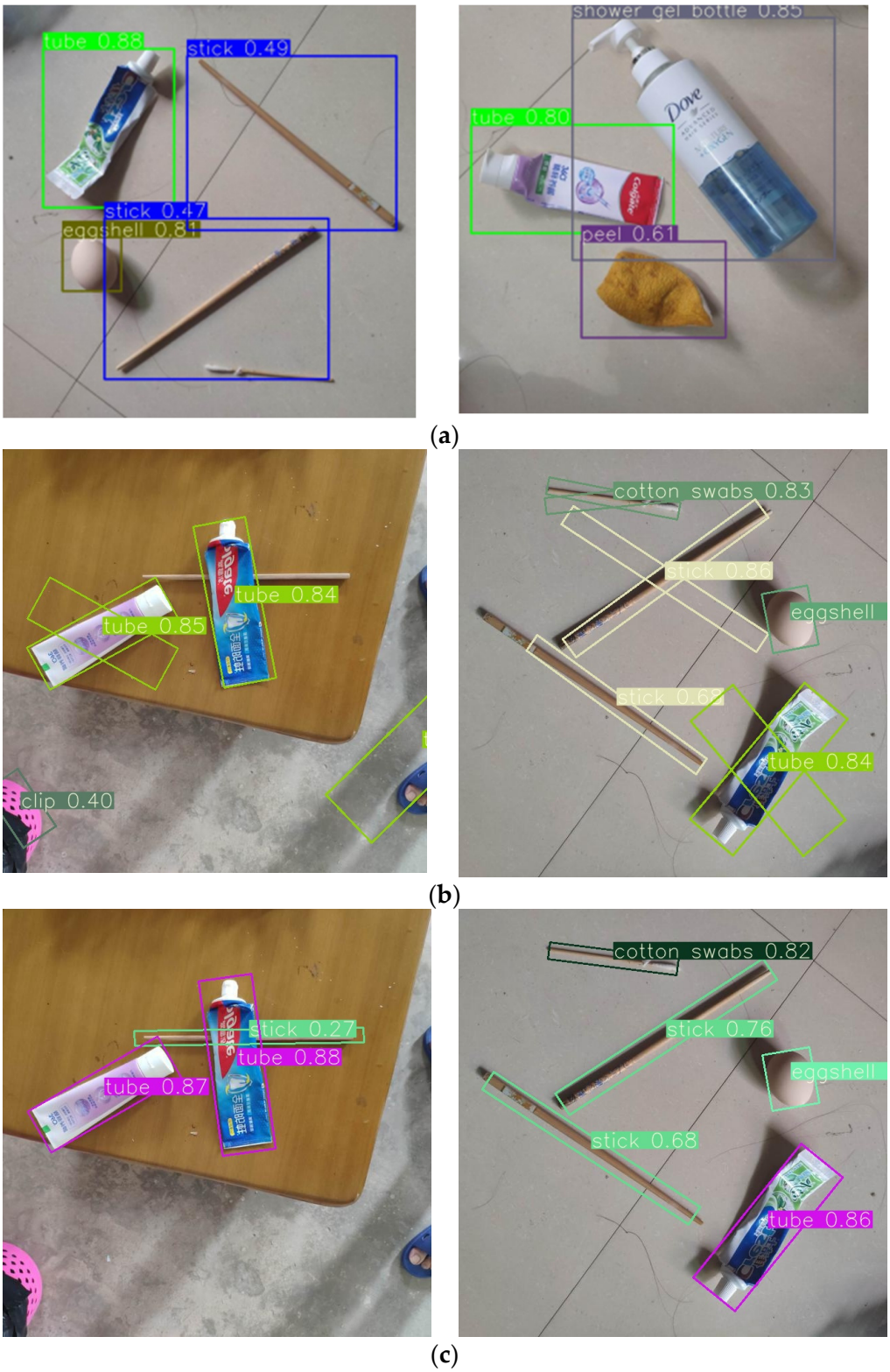



**Figure 10.** Comparison of detection effects of the methods. (**a**) Detection using YOLOv5m. (**b**) Detection using YOLOv5m-OBB. (**c**) Detection using YOLOv5m-DSM.

Figure 10b shows that, when the angular classification network was added to the detection head, YOLOv5m-OBB could generate a waste object detection box at any angle, but the angle of the generated detection box was not accurate enough, especially in cases involving a large aspect ratio. Due to the large aspect ratio, a slight deviation in the prediction box resulted in a smaller IoU for the prediction box and the true box, which

resulted in difficulties in the model training. Therefore, a large aspect ratio makes effective learning difficult.

Figure 10c shows the detection results for YOLOv5m-DSM. It can be seen that YOLOv5m-DSM could generate a waste object detection box at any angle and could detect objects involving a large aspect ratio accurately. It can be seen that, with the optimization of the feature extraction backbone network and feature aggregation network, and after the optimization of the loss function through the dynamic smoothing algorithm, YOLOv5m-DSM had better precision and performance in the detection of oriented waste.

Table 2 shows a comparison of the detection effects of YOLOv5m-DSM and the mainstream rotating rectangular box detection methods.

**Table 2.** Comparison of YOLOv5m-DSM and rotating frame detection methods.

| Method | Recall/% | mAP/% | Params (M) | GFLOPs | FPS |
|---|---|---|---|---|---|
| RoI Trans | 88.6 | 87.3 | 55.4 | 265.4 | 5.8 |
| Gliding-Vertex | 92.4 | 89.6 | 41.4 | 224.8 | 7.6 |
| R3Det | 91.4 | 90.1 | 48.0 | 250.5 | 7.0 |
| S2A-Net | 94.3 | 93.1 | 38.9 | 153.8 | 8.3 |
| YOLOv5m-DSM (Cos) | 94.5 | 93.3 | 23.7 | 76.5 | 15.5 |
| YOLOv5m-DSM (Linear) | 94.8 | 93.9 | 23.7 | 76.5 | 15.5 |

It can be seen that, when compared to RoI Trans [34], the average recall rate and average precision of detection increased by 6.2% and 6.6%, respectively. Compared to Gliding-Vertex [35], they increased by 2.4% and 4.3% respectively. Compared to R3Det [36], they increased by 3.4% and 3.8%, respectively. The recall rate and average precision of the YOLOv5m-DSM model were also better than those of S2A-Net [37], and our method had fewer parameters and a detection rate twice as high. In addition, we extended the flops counter tool to calculate the floating point operations (FLOPs) in the methods, and the computation load of YOLOv5m-DSM was lower than the comparison algorithm, making the model more suitable for deployment and application in embedded devices.

### *3.4. Network Model Ablation Experiment*

The network model ablation comparison experiment was used to evaluate the optimization effects of each improvement scheme. The experimental comparison results are shown in Table 3. Optimization 1 involved using the dynamic smoothing algorithm to densify the angle label conversion and calculate the loss function (1a is linear annealing and 1b is the cosine annealing angle). Optimization 2 involved the improvement of the feature extraction backbone network based on the proposed HDBottleneckCSP module. Optimization 3 involved an improvement of the feature aggregation network of the YOLOv5 based P2P-PANet structure.

**Table 3.** Ablation experiment.

| Method | Angle | op 1a | op 1b | op 2 | op 3 | Recall/% | mAP/% |
|---|---|---|---|---|---|---|---|
| YOLOV5m | × | × | × | × | × | - | - |
| YOLOV5m-OBB | √ | × | × | × | × | 87.2 | 82.3 |
| Optimization model 1 | √ | √ | × | × | × | 92.5 | 90.5 |
| Optimization model 2 | √ | √ | × | √ | × | 93.6 | 91.7 |
| Optimization model 3 | √ | √ | × | × | √ | 93.4 | 91.5 |
| YOLOv5m-DSM (Cos) | √ | × | √ | √ | √ | 94.5 | 93.3 |
| YOLOv5m-DSM (Linear) | √ | √ | × | √ | √ | 94.8 | 93.9 |

It can be seen from Table 3 that, after adding the linear dynamic smoothing algorithm and the corresponding loss function, the recall rate and mean average precision of optimization model 1 increased by 5.3% and 8.2%, respectively. After adding the linear dynamic

smoothing algorithm and HDBottleneckCSP module, these values increased by 6.4% and 9.4% in optimization model 2, respectively. After adding the linear dynamic smoothing algorithm and P2P-PANet module, they increased by 6.2% and 9.2% in optimization model 3, respectively. For the YOLOV5m-DSM (Linear) model, the detection recall rate and average precision of the model increased by 7.6% and 11.6%, respectively, with the above optimization methods.

In order to analyze the effects of replacing the original module structure with the HDBottleneckCSP structure and P2P-PANet network on the image waste detection algorithm more clearly, as well as the reasons for these effects, the intermediate characteristic graphs of YOLOv5 and YOLOv5m-DSM were extracted for comparison, as shown in Figure 10.

Figure 11a,b show an input image and label image, and Figure 11c–f and Figure 11g–j show the 1/8, 1/16 and 1/32 down sampling feature maps of YOLOv5 and YOLOv5m-DSM in the backbone network. It can be seen from Figure 11c,d,g,h that shallower feature information was extracted from the model after using the HDBottleneckCSP network, and the edge information and feature details of the waste were obtained more clearly. As can be seen from Figure 11e,i, two network structures obtained high-level semantic features through multi-layer convolution operations. Finally, from the comparison of Figure 11f,j, we can see that the YOLOv5m-DSM network generated a clearer edge for the target object, which led to an improvement in the recall and accuracy of the waste detection.

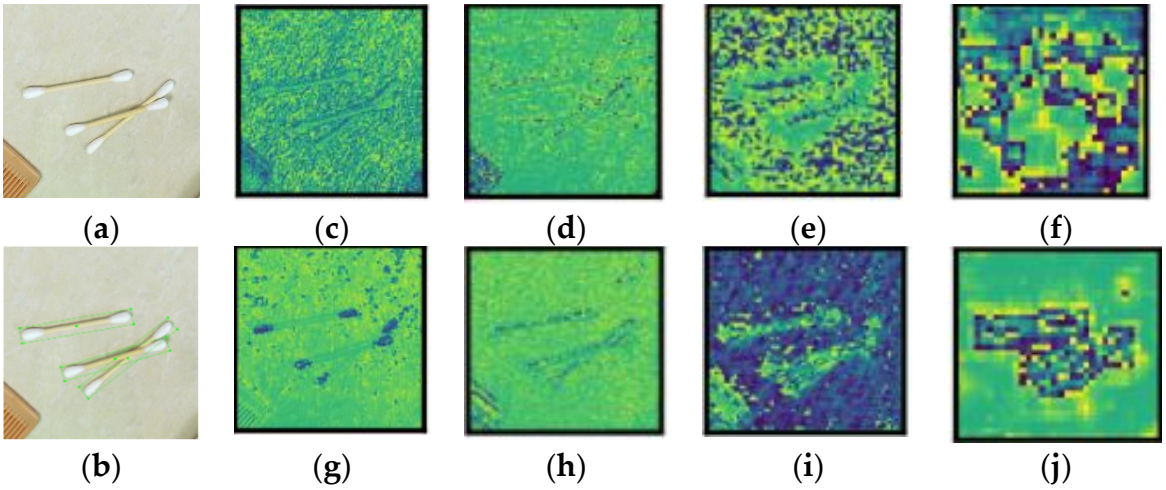



**Figure 11.** Intermediate characteristic diagrams of YOLOv5 and YOLOv5m-DSM. (**a**) Input image. (**b**) Label image. (**c**–**f**) Down sampling feature maps of YOLOv5. (**g**–**j**) Down sampling feature maps of YOLOv5m-DSM.

Table 4 shows a comparison of the detection effects of "interlayer merging" and "layer by layer merging" on the characteristic layer of the HDBottleneckCSP network.

**Table 4.** Effect comparison of "interlayer merging" and "layer by layer merging".

| Model | mAP/% | Recall/% | Parameters |
|---|---|---|---|
| Layer by layer merging | 90.9 | 92.8 | 894,842 |
| Interlayer merging | 91.7 | 93.6 | 717,645 |

It can be seen from Table 4 that, compared with "layer by layer merging", the "interlayer merging" used for feature map aggregation had fewer training parameters and a better detection effect. This was mainly because the "layer by layer merging" led to excessive duplication of the use of feature maps, which can easily cause feature redundancy and increase the difficulty of learning for the model. In addition, overly dense feature map aggregation increases the number of channels in the feature map, thus increasing the number of parameters and consuming more computing resources.

Table 5 shows the effects of the proposed method with different backbones. It can be seen that, compared with other backbones, such as VGG19, Resnet50, and CSPDarknet, the backbone proposed in this paper achieved a better detection effect.

**Table 5.** Effect comparison with different backbones.

| Backbone | Recall/% | mAP/% | AP of "Large Aspect Ratio Category"/% | | | | |
|---|---|---|---|---|---|---|---|
| | | | Cotton Swab | Stick | Plastic Bottle | Shower Gel Bottle | Tube |
| VGG19 | 92.4 | 90.5 | 69.5 | 65.1 | 88.9 | 98.2 | 96.8 |
| Resnet50 | 93.8 | 92.1 | 68.9 | 77.9 | 90.9 | 100 | 100 |
| CSPDarknet | 93.4 | 91.5 | 67.5 | 75.2 | 90.0 | 100 | 98.2 |
| Ours | 94.8 | 93.9 | 78.7 | 81.0 | 100 | 100 | 100 |

In order to analyze the effects of replacing the original module structure with the HDBottleneckCSP and P2P-PANet network on image waste detection clearly, as well as the reasons for these effects, Figure 11 shows maps of the feature aggregation network and the detection results for YOLOv5 and YOLOv5m-DSM.

Figure 12a–h show the multi-scale feature maps and detection results for the network obtained from the convergence of the original YOLOv5 and YOLOv5m-DSM features. It can be seen from the graph analysis that the YOLOv5 model converged the feature map but, in the generated multi-scale feature map, the contour of the detected object was not clear enough, the feature differentiation from the background map was not obvious and a situation occurred involving mixing with the background feature. The YOLOv5m-DSM algorithm uses the P2P-PANet structure and the smoothing labels of the angle, which makes the model's learning of image features more obvious and the feature contour of the detection object clearer and more differentiated from the background features, thus making the final detection effect more accurate.

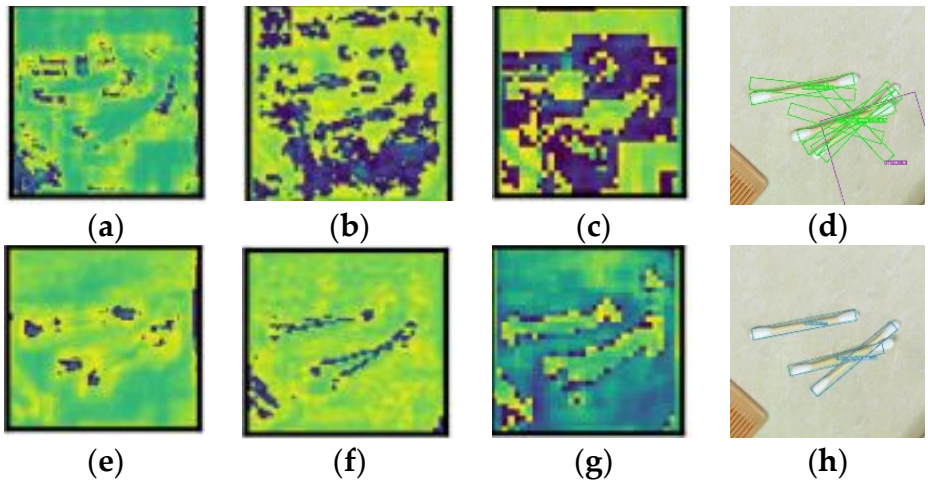



**Figure 12.** Feature aggregation network maps of YOLOv5 and YOLOv5m-DSM. (**a**–**d**) Multi-scale feature maps and detection result of YOLOv5m. (**e**–**h**) Multi-scale feature maps and detection result of YOLOv5m-DSM.

Table 6 shows a comparison of the effects of dynamic smoothing and the circular smooth label with different hyper-parameters.

**Table 6.** Comparison of dynamic smoothing and the circular smooth label.

| Method | Recall/% | mAP/% |
|---|---|---|
| CSL (r = 7) | 92.8 | 91.0 |
| CSL (r = 6) | 93.2 | 91.8 |
| CSL (r = 5) | 93.2 | 91.3 |
| DSM-Cos (c = 5, e = 4) | 94.3 | 93.2 |
| DSM-Cos (c = 4, e = 4) | 94.1 | 93.1 |
| DSM-Cos (c = 4, e = 3) | 94.5 | 93.3 |
| DSM-Linear (c = 8, e = 4) | 94.2 | 93.5 |
| DSM-Linear (c = 7, e = 4) | 94.6 | 93.8 |
| DSM-Linear (c = 7, e = 3) | 94.8 | 93.9 |

It can be seen from the table that different detection effects were obtained by adjusting the range of the circular smoothing algorithm. However, the performances of the two kinds of dynamic smoothing were better than the best result with the circular smooth label. This proves that the dynamic smoothing was strong. Dynamic smoothing controls angle learning by shrinking the range of smoothness gradually. In the initial stage of model training, a larger smoothing range was set to reduce the difficulty of model learning and improve the recall rate for waste detection. With the iteration of the model learning, the angle smoothing range was gradually reduced through the attenuation function to reduce the angle deviation in target detection, thus improving the detection accuracy. Higher accuracy for angle prediction can improve the recall rate for oriented waste, especially in cases involving a large aspect ratio.

### *3.5. Detection Application Results*

In order to demonstrate the waste detection performance of the improved method proposed in this paper, the method was used for actual testing in different scenarios with different levels of illumination, such as a waste station, garage, corridor, lawn, and so on. The results are shown in Figure 13. It can be seen that the method detected the waste objects effectively in a series of scenarios. It was proven that the method described in this paper is able to carry out the detection of oriented waste effectively.


**Figure 13.** *Cont.*


**Figure 13.** Waste detection with YOLOv5m-DSM in different scenarios.

## 4. Conclusions

This paper focused on waste detection for a robotic arm based on YOLOv5. In addition to object identification in general image analysis, a waste-sorting robotic arm not only needs to identify a target object but also needs to accurately judge its placement angle, so that the robotic arm can set the appropriate grasping angle. In order to address this need, we added an angular prediction network to the detection head to provide the grasping angle information for the waste object and proposed a dynamic smoothing algorithm for angle loss to enhance the model's angular prediction ability. In addition, we improved the method's feature extraction and aggregation abilities by optimizing the backbone and feature aggregation network of the model. The experimental results showed that the performance of the improved method in oriented waste detection was better than that of comparison methods; the average precision and recall rate were 93.9% and 94.8%, respectively, which were 11.6% and 7.6% higher than those of the original network, respectively.

**Author Contributions:** Conceptualization, W.Y., Y.X. and P.G.; methodology, W.Y. and Y.X.; software: W.Y. and P.G.; validation: W.Y.; writing—original draft preparation, W.Y.; writing—review and editing, W.Y., Y.X. and P.G. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was partially supported by Guangdong Key Discipline Scientific Research Capability Improvement Project: no. 2021ZDJS144; Projects of Young Innovative Talents in Universities of Guangdong Province: no. 2020KQNCX126; Key Subject Projects of Guangzhou Xinhua University: no. 2020XZD02; Scientific Projects of Guangzhou Xinhua University: no. 2020KYQN04; and the Plan of Doctoral Guidance: no. 2020 and no. 2021. Xie's work was supported by the Natural Science Foundation of China (no. 61972431).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

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

## References

1. Adedeji, O.; Wang, Z. Intelligent Waste Classification System Using Deep Learning Convolutional Neural Network. In Proceedings of the 2nd International Conference on Sustainable Materials Processing and Manufacturing (SMPM), Sun City, South Africa, 8–10 March 2019; pp. 607–612.
2. Alsubaei, F.S.; Al-Wesabi, F.N.; Hilal, A.M. Deep Learning-Based Small Object Detection and Classification Model for Garbage Waste Management in Smart Cities and IoT Environment. *Appl. Sci.* **2022**, *12*, 2281. [CrossRef]
3. Song, Q.; Li, S.; Bai, Q.; Yang, J.; Zhang, X.; Li, Z.; Duan, Z. Object Detection Method for Grasping Robot Based on Improved YOLOv5. *Micromachines* **2021**, *12*, 1273. [CrossRef]
4. Wu, Y.; Zhang, Y.; Li, M. Fine classification model for plastic waste based on FV-DCNN. *Transducer Microsyst. Technol.* **2021**, *40*, 118–120.
5. Chen, Z.-C.; Jiao, H.-N.; Yang, J.; Zeng, H.-F. Garbage image classification algorithm based on improved MobileNet v2. *J. Zhejiang Univ. (Eng. Sci.)* **2021**, *55*, 1490–1499.
6. Liu, Y.; Ge, Z.; Lv, G.; Wang, S. Research on Automatic Garbage Detection System Based on Deep Learning and Narrowband Internet of Things. In Proceedings of the 3rd Annual International Conference on Information System and Artificial Intelligence (ISAI), Suzhou, China, 22–24 June 2018.
7. Kang, Z.; Yang, J.; Li, G.; Zhang, Z. An Automatic Garbage Classification System Based on Deep Learning. *IEEE Access* **2020**, *8*, 140019–140029. [CrossRef]
8. Cui, W.; Zhang, W.; Green, J.; Zhang, X.; Yao, X. YOLOv3-DarkNet with adaptive clustering anchor box for garbage detection in intelligent sanitation. In Proceedings of the 2019 3rd International Conference on Electronic Information Technology and Computer Engineering (EITCE), Xiamen, China, 18–20 October 2019; pp. 220–225.
9. Xu, W.; Xiong, W.; Yao, J.; Sheng, Y. Application of garbage detection based on improved YOLOv3 algorithm. *J. Optoelectron. Laser* **2020**, *31*, 928–938.
10. Chen, W.; Zhao, Y.; You, T.; Wang, H.; Yang, Y.; Yang, K. Automatic Detection of Scattered Garbage Regions Using Small Unmanned Aerial Vehicle Low-Altitude Remote Sensing Images for High-Altitude Natural Reserve Environmental Protection. *Environ. Sci. Technol.* **2021**, *55*, 3604–3611. [CrossRef] [PubMed]
11. Majchrowska, S.; Mikolajczyk, A.; Ferlin, M.; Klawikowska, Z.; Plantykow, M.A.; Kwasigroch, A.; Majek, K. Deep learning-based waste detection in natural and urban environments. *Waste Manag.* **2022**, *138*, 274–284. [CrossRef] [PubMed]
12. Meng, J.; Jiang, P.; Wang, J.; Wang, K. A MobileNet-SSD Model with FPN for Waste Detection. *J. Electr. Eng. Technol.* **2022**, *17*, 1425–1431. [CrossRef]
13. Li, J.; Qiao, Y.; Liu, S.; Zhang, J.; Yang, Z.; Wang, M. An improved YOLOv5-based vegetable disease detection method. *Comput. Electron. Agric.* **2022**, *202*, 107345. [CrossRef]
14. Chen, Z.; Wu, R.; Lin, Y.; Li, C.; Chen, S.; Yuan, Z.; Chen, S.; Zou, X. Plant Disease Recognition Model Based on Improved YOLOv5. *Agronomy* **2022**, *12*, 365. [CrossRef]
15. Ling, L.; Tao, J.; Wu, G. Research on Gesture Recognition Based on YOLOv5. In Proceedings of the 33rd Chinese Control and Decision Conference (CCDC), Kunming, China, 22–24 May 2021; pp. 801–806.
16. Wang, Z.; Wu, L.; Li, T.; Shi, P. A Smoke Detection Model Based on Improved YOLOv5. *Mathematics* **2022**, *10*, 1190. [CrossRef]
17. Gao, P.; Lee, K.; Kuswidiyanto, L.W.; Yu, S.-H.; Hu, K.; Liang, G.; Chen, Y.; Wang, W.; Liao, F.; Jeong, Y.S.; et al. Dynamic Beehive Detection and Tracking System Based on YOLO V5 and Unmanned Aerial Vehicle. *J. Biosyst. Eng.* **2022**, *47*, 510–520. [CrossRef]
18. Yang, X.; Yan, J. Arbitrary-oriented object detection with circular smooth label. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; pp. 677–694.

19. Rezatofighi, H.; Tsoi, N.; Gwak, J.; Sadeghian, A.; Reid, I.; Savarese, S.; Soc, I.C. Generalized Intersection over Union: A Metric and A Loss for Bounding Box Regression. In Proceedings of the 32nd IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 16–20 June 2019; pp. 658–666.
20. Chen, K.; Zhu, Z.; Deng, X.; Ma, C.; Wang, H. Deep Learning for Multi-scale Object Detection: A Survey. *J. Softw.* **2021**, *32*, 1201–1227.
21. Hariharan, B.; Arbelaez, P.; Girshick, R.; Malik, J. Hyper columns for Object Segmentation and Fine-grained Localization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 447–456.
22. Shelhamer, E.; Long, J.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 640–651. [CrossRef] [PubMed]
23. Yang, S.; Ramanan, D. Multi-scale recognition with DAG-CNNs. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 11–18 December 2015; pp. 1215–1223.
24. Kaiming, H.; Xiangyu, Z.; Shaoqing, R.; Jian, S. Deep residual learning for image recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
25. Huang, G.; Liu, Z.; van der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the 30th IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 2261–2269.
26. Hu, J.; Shen, L.; Sun, G. Squeeze-and-Excitation Networks. In Proceedings of the 31st IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
27. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. CBAM: Convolutional Block Attention Module. In Proceedings of the 15th European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
28. Hou, Q.; Zhou, D.; Feng, J.; Ieee Comp, S.O.C. Coordinate Attention for Efficient Mobile Network Design. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 19–25 June 2021; pp. 13708–13717.
29. Lin, T.-Y.; Dollar, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature Pyramid Networks for Object Detection. In Proceedings of the 30th IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 936–944.
30. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path Aggregation Network for Instance Segmentation. In Proceedings of the 31st IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 8759–8768.
31. Mingxing, T.; Ruoming, P.; Le, Q.V. EfficientDet: Scalable and Efficient Object Detection. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA, 14–19 June 2020; pp. 10778–10787.
32. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. Ssd: Single shot multibox detector. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; pp. 21–37.
33. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
34. Ding, J.; Xue, N.; Long, Y.; Xia, G.-S.; Lu, Q.; Soc, I.C. Learning RoI Transformer for Oriented Object Detection in Aerial Images. In Proceedings of the 32nd IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 16–20 June 2019; pp. 2844–2853.
35. Xu, Y.; Fu, M.; Wang, Q.; Wang, Y.; Chen, K.; Xia, G.S.; Bai, X. Gliding vertex on the horizontal bounding box for multi-oriented object detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *43*, 1452–1459. [CrossRef] [PubMed]
36. Yang, X.; Yan, J.C.; Feng, Z.M.; He, T.; Assoc Advancement Artificial, I. R(3)Det: Refined Single-Stage Detector with Feature Refinement for Rotating Object. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtual, 2–9 February 2021; pp. 3163–3171.
37. Han, J.; Ding, J.; Li, J.; Xia, G.S. Align deep features for oriented object detection. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–11. [CrossRef]