

Article

Object Detection for Construction Waste Based on an Improved YOLOv5 Model

Qinghui Zhou *, Haoshi Liu, Yuhang Qiu and Wuchao Zheng

School of Mechanical-Electronic and Vehicle Engineering, Beijing University of Civil Engineering and Architecture, Beijing 100044, China

* Correspondence: zhouqinghui@bucea.edu.cn

Abstract: An object detection method based on an improved YOLOv5 model was proposed to enhance the accuracy of sorting construction waste. A construction waste image sample set was established by collecting construction waste images on site. These construction waste images were preprocessed using the random brightness method. A YOLOv5 object detection model was improved in terms of the convolutional block attention module (CBAM), simplified SPPF (SimSPPF) and multi-scale detection. Then, the improved YOLOv5 model was trained, validated and tested using the established construction waste image dataset and compared with other conventional models such as Faster-RCNN, YOLOv3, YOLOv4, and YOLOv7. The results show that: based on the improved YOLOv5 model, the mean average precision (mAP) on the test dataset can reach 0.9480. The overall performance of this model is better than that of other conventional models in object detection, which verifies the accuracy and availability of the proposed method.

Keywords: construction waste; computer vision; deep learning; YOLOv5; waste sorting



check for updates

Citation: Zhou, Q.; Liu, H.; Qiu, Y.; Zheng, W. Object Detection for Construction Waste Based on an Improved YOLOv5 Model. *Sustainability* **2023**, *15*, 681. <https://doi.org/10.3390/su15010681>

Academic Editors: Carlos Morón Fernández and Daniel Ferrández Vega

Received: 1 December 2022

Revised: 23 December 2022

Accepted: 28 December 2022

Published: 30 December 2022



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

1. Introduction

The rapid rise in construction activities has produced a large amount of construction waste with the global increase in population and urbanization [1]. According to a literature review and survey, construction waste accounts for more than 25% of the world's waste [2]. In China, the average recovery rate of construction waste is approximately 5%. Additionally, the annual construction waste level is about 1.55 billion tons to 2.4 billion tons [3], accounting for nearly 30–40% of urban waste, causing many environmental issues [4].

Due to the lack of proper recovery schemes and effective disposal technologies, construction waste without any treatment will be transported to the suburban landfill, causing land-use threats [5]. However, some materials are potentially valuable in construction and easily reused/recycled, including concrete, stone masonry, bricks, etc. These sustainable materials should be sorted out and turned into recycled aggregates that can be used in new building projects after crushing and separation, thus reducing the need to mine and process virgin materials. Therefore, reducing, reusing, and recycling construction waste has become an important and essential issue.

Currently, the traditional method of sorting construction waste is mixing, crushing and screening by means of mechanical operation, while preselecting, rejecting and diverting by manual work. However, there are problems of low recycling purity and low efficiency of manual work and especially serious harm to health in dusty and noisy environments. Increasingly, computer vision (CV), robotics, and other-artificial intelligent technologies are being used for construction waste sorting [6]. Usually, a robot for sorting construction waste is used to finely sort a large number of objects before mixing and crushing. Smart technologies can improve the reuse and recycling of construction waste. For example, the company ZenRobotics began to manufacture robots that used artificial intelligence and other recognition technologies to identify and sort household, industrial, and construction and demolition waste in 2007 [7]. Other well-known commercial robots have also been

tried for use in the waste management industry, such as Sadako [8], SamurAI [9], and AMP Robotics [10]. The majority of the existing systems capitalize on the agility of robots to rapidly transfer recyclables from a conveyor belt to a bin [11]. However, many factors affect the accuracy and efficiency of sorting construction waste. In a real work environment, the stacking of construction waste on the conveyor belt, the irregular shapes, and the small-sized objects lead to errors in detection. Measures should be taken to improve the accuracy of object detection.

Machine learning can improve the efficiency and accuracy of sorting construction waste. With sufficient data, a CV model can identify different waste materials by machine learning. Previous research has found that CV performed well in construction waste recycling. Several algorithms for CV have been used to identify and classify waste, but inter occlusion and small object detection were not fully considered. Convolutional neural network (CNN) is an algorithm that has become the standard in image classification and object recognition. Therefore, several model developments based on CNN have emerged. For example, Adedeji and Wang (2019) [12] employed such a technique, which extracted features learned by the ResNet-50, and performed waste classification with SVM. Chen et al. (2021) [13] developed a hybrid model that integrated visual features extracted by a DenseNet-169 network and physical features such as weight and depth collected by other sensors. These methods improve the accuracy of sorting waste, but still do not consider the inter occlusion and small object detection. Yang et al. (2021) [14] adopted a “ResNeXt + k-NN” structure. Lau Hiu Hoong et al. (2020) [15] improved the performance of sorting construction waste through the residual network. Chen et al. (2017) [16] employed Fast R-CNN to detect and locate waste objects on conveyor belts, which demonstrated a false negative rate (FNR) of 3% and a false positive rate (FPR) of 9%. Awe et al. (2017) [17], Wang et al. (2019) [18], and Nowakowski and Pamula (2020) [19] applied Faster R-CNN for the detection of residential and municipal waste, construction and demolition waste, and electronic waste, respectively. Ku et al. (2021) [6] proposed a deep learning method for grab detection based on R-CNN. Zhou et al. [20] selected the RepVGG residual network as the basic feature network based on the Faster-RCNN algorithm to retain more information of small-sized objects. Li et al. [21] built an RGB detection platform and used color cameras and laser line scanning sensors to collect RGB images to detect construction waste. Lin et al. [22] proposed a CVGGNet model based on knowledge transfer together with data enhancement and periodic-learning-rate technology to classify construction waste. From the above-mentioned studies, the applications of CV in waste sorting had been specifically focused on. The existing object detection algorithms were mainly improved from different perspectives: multi-scale feature fusion, data augmentation, training algorithm, and context-based detection.

In order to enhance the accuracy rate of object detection, the YOLO model has been used to identify and classify waste, such as YOLOv3 [23], YOLOv4 [24] and YOLOv5. The YOLO model uses multiple lower sampling layers, and the target features learned from the network are not exhaustive so the detection effect will be improved [25]. Liu et al. [26] improved the network structure and multi-scale detection based on the YOLOv3 algorithm. The mAP value could reach 91.96%. Chen et al. [27] designed a waste robot with a YOLOv4 model that can identify beverage bottles, cans, wastepaper, and banana peels in an unobstructed environment. Yuan et al. [28] proposed an improved algorithm based on YOLOv5 for underwater waste detection. Gamma transform was added in the preprocessing stage to improve the gray and contrast of underwater images, and the CBAM attention mechanism was embedded in the YOLOv5 detection part to highlight object features and suppress secondary information, thus improving the accuracy of detection. Therefore, YOLO algorithms have also been used for waste sorting. Similarly, YOLO algorithms were also improved from the following different perspectives: CBAM attention mechanism, multi-scale feature learning, data augmentation, and training strategy.

It is well known that small object detection and inter-occlusion are still challenging problems in computer vision. Currently, with the increasing need for recycling onsite

construction waste, a higher velocity of the conveyor belt and more stacking of waste will lead to difficulty in finely sorting, which is probably one of the most critical problems in construction waste management. Under the conditions of inter occlusion and small object detection, the accuracy of the CV may decrease [29].

Therefore, there are two problems in sorting construction waste: one is inter-object occlusion, i.e., multiple objects are overlaid on each other, occluding one another. The other problem is small-object detection. Hence, in order to solve these problems and further improve the accuracy of classification and detection of construction waste, an improved YOLOv5 model was proposed by applying the CBAM attention mechanism and SimSPPF module, adding a shallow detection layer to detect small construction waste objects and inter-occlusion and increasing the fourth scale feature fusion to the feature fusion part correspondingly. Additionally, a dataset was established, and a data enhancement method was used to expand the diversity of training samples.

2. Materials and Methods

2.1. YOLOv5 Architecture

Any robotic sorting system needs to accurately categorize recyclables from various waste material type. It is important for the system to develop an effective model. With the recent developments in deep learning, the YOLO family of models provides a scalable means for categorizing recyclables into various classes. Construction waste can be roughly split into 4 classes: brick, wood, stone and plastic, according to the material. Considering the accuracy and efficiency in object detection, the YOLOv5 model was chosen, which consists of three main architectural blocks: Backbone, Neck and Head, as shown in Figure 1.

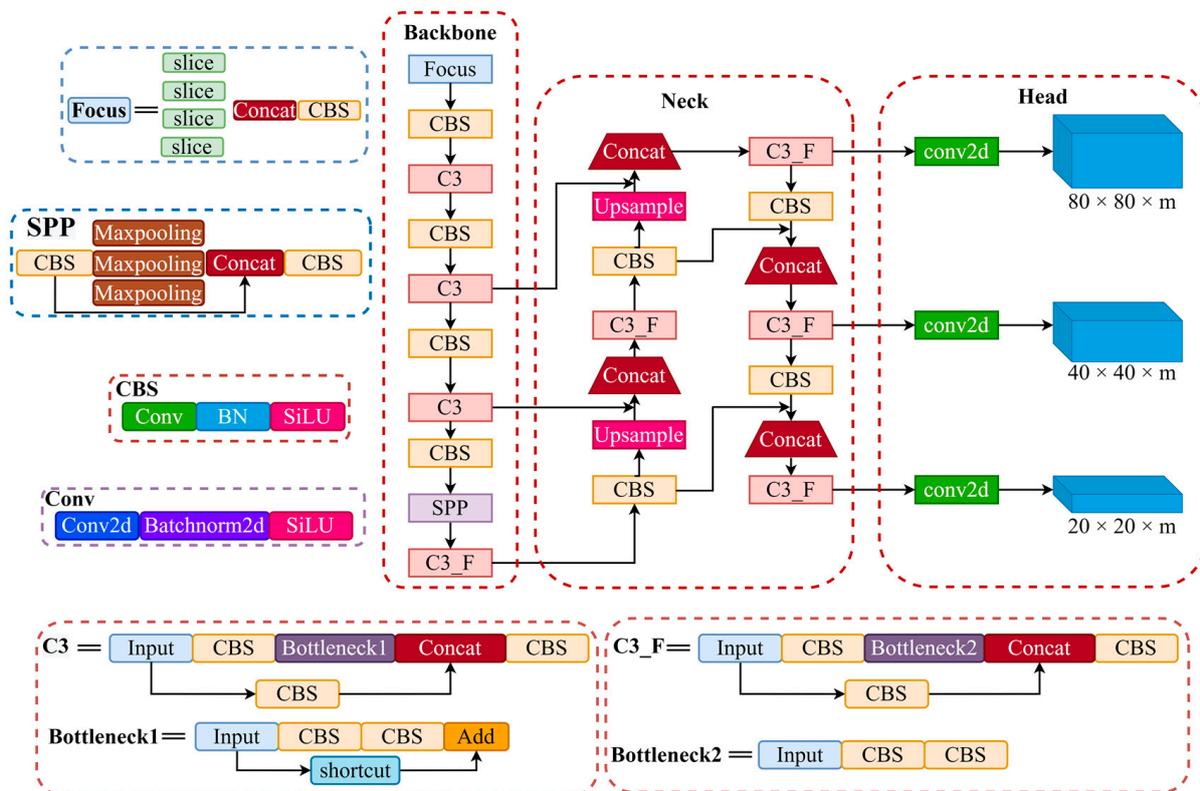


Figure 1. Architecture of YOLOv5 Model.

2.1.1. YOLOv5 Backbone

YOLOv5 Backbone employs CSPDarknet as the backbone for feature extraction from images consisting of cross-stage partial networks. The focus module, which rapidly down-samples the images of the dataset, can pass the image information into the channel, while

ensuring that the image information is not missing, i.e., more fully extracting the image information features. The backbone layer uses C3, C3_F and Spatial Pyramid Pooling (SPP) modules. C3 and C3_F modules can improve the ability of feature extraction from images, simplify the YOLOv5 model and make the detection speed faster [30]. The SPP module can improve the scale invariance of the dataset image, effectively increase the receiving range of the backbone features, make it easier to converge the network, and enhance the accuracy [31].

2.1.2. YOLOv5 Neck

YOLOv5 Neck uses PANet to generate a feature pyramid network to perform aggregation on the features and pass it to Head for prediction. The bottleneck layer of YOLOv5 combines feature pyramid network (FPN) and path aggregation network (PAN) structures. Deep feature images have stronger semantic information and weaker location information, while shallow feature images have stronger location information and weaker semantic information. FPN can transfer semantic information from the deep feature image to the shallow feature image [32]. Conversely, PAN can transfer location information from the shallow feature layer to the deep feature layer [33]. The combination of FPN and PAN can aggregate parameters of different detection layers from different trunk layers, which greatly strengthens the feature fusion ability of the network [24].

2.1.3. YOLOv5 Head

YOLOv5 Head consists of layers that generate predictions from the anchor boxes for object detection. The head includes two parts: loss function and non-maximum suppression (NMS). In YOLOv5, the binary cross entropy loss function is used to calculate classification loss and confidence loss, while the complete Iou (CIoU) loss function is applied to calculate location loss (bounding box regression loss). All the losses add up to the total loss. The CIoU loss function fully considers three key geometric parameters: the overlap area, the distance from the center point, and the aspect ratio, thus improving the speed and accuracy of the regression of the prediction box [34]. The NMS is mainly used to remove redundant detection boxes and reserve the candidate box with the highest prediction probability as the final prediction box [35].

YOLOv5 is a family of compound-scaled object detection models trained on the COCO dataset [36]. YOLOv5 is the latest object detection model developed by Ultralytics, which offers open-source research into future object detection methods. An open-source network, such as the COCO dataset, was employed since it has been particularly successful at similar tasks: object segmentation, object detection and classification. However, any open-source dataset cannot be used in all circumstances due to the various classes of object. Once a class is varied, the model should be retrained according to a new dataset. Meanwhile, the model needs to be further optimized in the case of object occlusion and small-object detection.

2.2. Improved YOLOv5

The convolutional block attention module (CBAM) is added to the Backbone of the original YOLOv5 model, the feature fusion of the fourth scale is embedded in the Neck, and the shallow detection layer is employed in the Head, which will be of benefit in detecting the small objects and inter occlusion. The improved model architecture is shown in Figure 2.

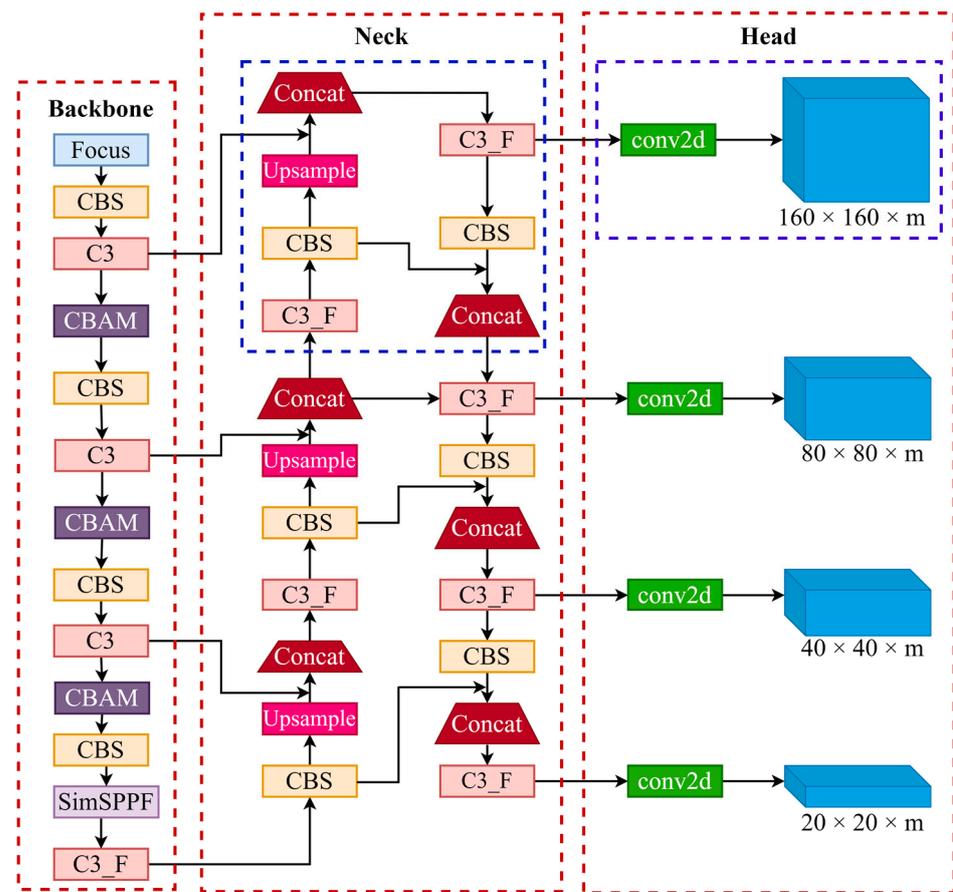


Figure 2. Architecture of the improved YOLOv5 Model.

2.2.1. CBAM-CSPDarknet53

Due to the problem of small objection with low pixels in the picture, it is easy for there to be missing information in construction waste sorting. The CBAM attention mechanism was added in the YOLOv5 backbone, named CBAM-CSPDarknet53. The CBAM attention mechanism module is mainly divided into a channel attention module and a spatial attention module [37]. The channel attention module pays more attention to the core information in the image of the dataset and can squeeze the spatial size, while the channel size remains uniform. The spatial attention module focuses on the position information of the object and can squeeze the channel size without modifying the spatial dimension. The structure of the CBAM attention mechanism module is shown in Figure 3.

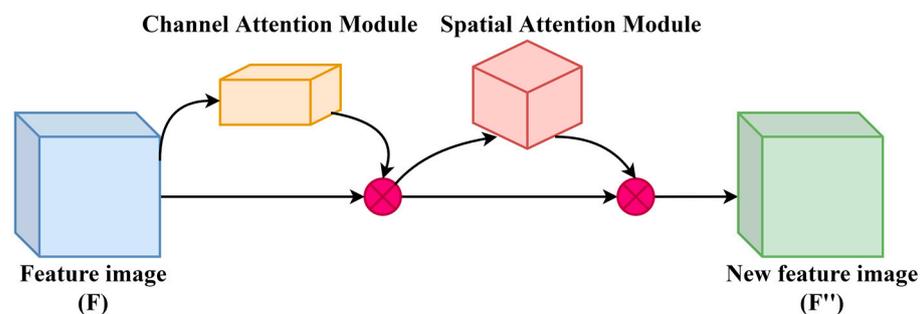


Figure 3. Structure of CBAM attention mechanism.

The structure of the channel attention module is described in Figure 4. The feature image F ($F \in R^{C \times H \times W}$) is processed by average pooling and maximum pooling, and then, the feature image size changes from $C \times H \times W$ to $C \times 1 \times 1$. Next, the new feature image

is sent to the Multi-Layer Perception (MLP) and the number of neurons in the first layer of the MLP is C/r , where r is the decline rate, and C is the number of neurons in the second layer of the MLP. Then, after processing by the Sigmoid function, the weight coefficient M_c will be obtained and calculated. The equation is shown in Equation (1).

$$M_c(F) = \sigma(W_1(W_0(F_{avg}^C)) + W_1(W_0(F_{max}^C))) \tag{1}$$

where: σ is the Sigmoid function; avg is global average pooling; max is the maximum pooling; $W_0 \in R^{C \times \frac{C}{r}}$; $W_1 \in R^{C \times \frac{C}{r}}$; F_{avg}^C is the global average pooled feature image of size $1 \times 1 \times C$; F_{max}^C is the largest pooled feature image with size $1 \times 1 \times C$.

Finally, the weight coefficient M_c is multiplied by the feature image F ($F \in R^{C \times H \times W}$). Therefore, the new feature image can be obtained.

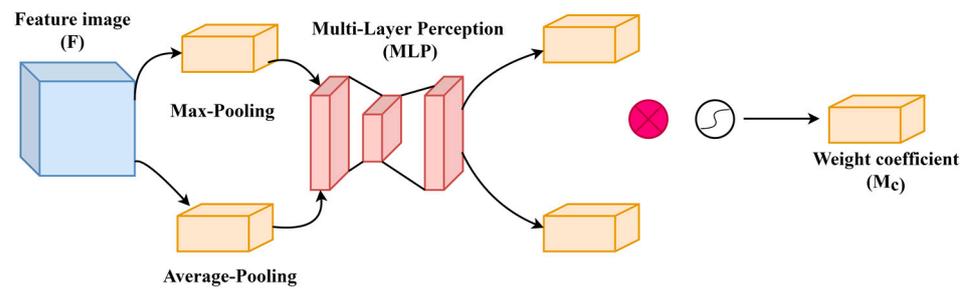


Figure 4. Structure of Channel Attention Module.

The structure of the spatial attention module is shown in Figure 5. The feature images of the new dataset obtained in the previous step are again processed by maximum pooling and average pooling, and then divided into two channels of size is $1 \times H \times W$. Then, the obtained tensors are stacked together by joining operations, and the weight coefficient M_s is obtained after convolution and Sigmoid function operations. The equation is shown in Equation (2).

$$M_s(F) = \sigma(f^{7 \times 7}([F_{avg}^S; F_{max}^S])) \tag{2}$$

where: σ is the Sigmoid function; avg is global average pooling; max is the maximum pooling; $f^{7 \times 7}$ is the convolution of 7×7 ; F_{avg}^S is the feature after the average pooling operation, and the size is $1 \times H \times W$; F_{max}^S is the feature after the maximum pooling operation, and the size is $1 \times H \times W$.

Finally, the calculated weight coefficient M_s is multiplied by the feature image F' which can obtain a new feature image F'' .

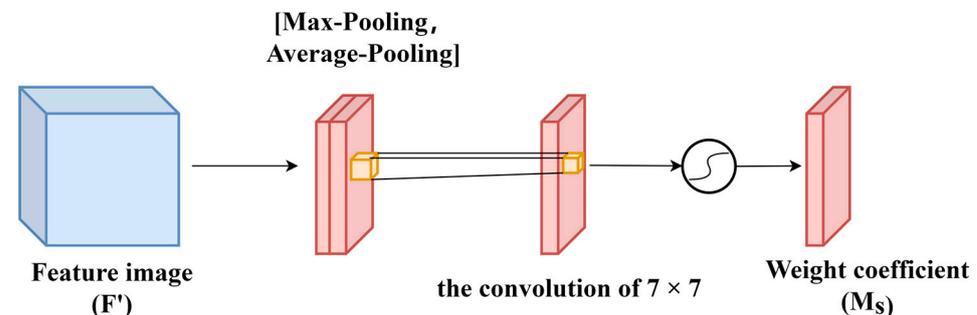


Figure 5. Structure of Spatial Attention Module.

2.2.2. SimSPPF (Simplified SPPF)

The SimSPPF module based on the maximum pooling layer at the same size is proposed to replace the SPP module in the original YOLOv5s model. The structure of the SimSPPF module is shown in Figure 6.

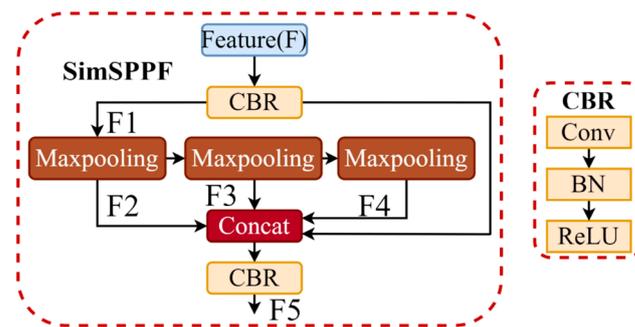


Figure 6. Structure of SimSPPF Module.

The SimSPPF module has a scale of 5×5 for the input feature images in the construction waste dataset. Because of the different maximum poolings in the three stages, channels should be connected for the output of the pooling layer. The equations are shown in Equations (3)–(7).

$$F1 = CBR(F) \quad (3)$$

$$F2 = Maxpooling(F1) \quad (4)$$

$$F3 = Maxpooling(F2) \quad (5)$$

$$F4 = Maxpooling(F3) \quad (6)$$

$$F5 = CBR([F; F2; F3; F4]) \quad (7)$$

The SimSPPF module can avoid the local feature loss of the construction waste dataset images, effectively reduce the residual parameter information, and retain the core texture features of construction waste dataset images. The calculation speed of the forward propagation of the SimSPPF module is faster than that of the SSP module. Meanwhile, after the SimSPPF module is embedded, the ability of the YOLOv5 model to extract image features can be improved greatly.

2.2.3. Multi-Scale Detection

In the process of sorting construction waste, small objects need to be detected. However, the original YOLOv5 model can only detect an 8×8 receptive field for an input image with a size of 640×640 . If the height or width of the detected object is less than 8 pixels in the dataset images, the image information features will be missing after convolution processing. In order to solve this problem, a special detection layer for small objects is added. This special detection layer is a 160×160 size output feature image, which can identify a receptive field size of more than 4×4 objects, basically meeting the detection requirements of sorting construction waste. Meanwhile, the three-scale feature fusion of the original YOLOv5 model was correspondingly increased to four-scale feature fusion, and a 160×160 feature detection layer was added. Therefore, the 80×80 feature detection layer was up-sampled twice. The twice up-sampled layer fused with the newly added 160×160 feature detection layer, which was used to detect small objects. The overall network structure of the improved YOLOv5 model is shown in Figure 7, where the dotted line in the Head denotes the increased detection of the fourth scale, and the dotted line in the Neck describes the corresponding increased part of the four-scale feature fusion.

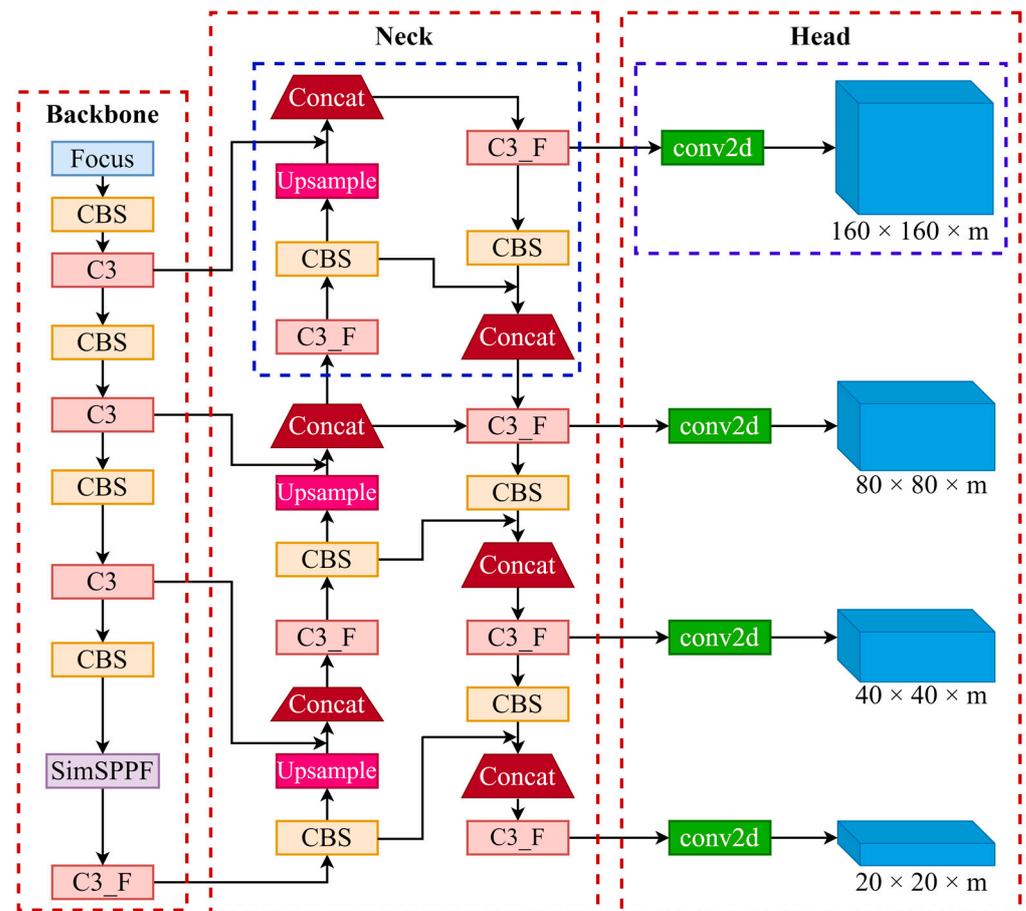


Figure 7. Structure of the Improved Multi-Scale Detection.

2.3. Dataset Construction and Evaluation Index

2.3.1. Dataset

To evaluate the improved YOLOv5 model, it was tested in the two datasets. One is the public dataset, PASCAL VOC [38], and the other is the self-built construction waste dataset. The PASCAL VOC dataset is a common object-detection dataset, which includes 4 categories and 20 subcategories. Here, the train + val parts of VOC2007 and VOC2012 are applied for the training set, which consists of 5011 training samples of VOC 2007 and 11,540 training samples of VOC2012. Additionally, the test part of the VOC2007 dataset is given as the test set, which consists of 4952 test samples.

A rich dataset including many construction waste images of different types should be implemented in sorting waste. Only two open-source datasets—TrashNet [39] and Taco [40]—are available, but they are not suitable for a robotic sorting system because the detected objects transferred on a conveyor belt would be irregular, dirty and piled up on one another. Developing a new dataset is the first important step. Thus, sample images were collected in the construction site, as shown in Figure 8. The dataset consisted of 3046 construction waste images divided into 4 classes: bricks, wood, stones, and plastics. To create a more effective dataset, further data enhancement processing should be carried out, such as image flipping, translation, rotation, cropping, scaling, adding noise and random occlusion operations [41], so as to effectively avoid the overfitting problem in the training and improve the robustness of the model. A graphical image annotation tool, Labeling, was used to label the image in the construction waste dataset. Finally, the dataset was divided into 3 subsets: the training set accounted for 80%, the verification set 10% and the test set 10%.



Figure 8. Construction waste image on site.

2.3.2. Model Performance Evaluation Index

Average precision (AP), $F1$ score ($F1$ -score) and mean average precision (mAP) were used as evaluation indicators to test the performance of the model. Average precision is a measure that combines recall and precision for ranked retrieval results. The recall rate reflects the ability to find positive samples, the precision expresses the ability to classify samples, and the AP shows the overall performance for object detection. The precision (P)–recall (R) curve can be plotted with the calculated P and R as the ordinate and abscissa, while the area under the curve is AP , and the mean value of AP is mAP . In addition, $F1$ -score is commonly used for multiple classification problems, which is considered the harmonic average of precision and recall. The equations are shown in Equations (8)–(12).

$$P = \frac{TP}{TP + FP} \quad (8)$$

$$R = \frac{TP}{TP + FN} \quad (9)$$

$$AP = \int_0^1 P \cdot RdR \quad (10)$$

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (11)$$

$$F1 = 2 \frac{P \cdot R}{P + R} \quad (12)$$

where: TP (true positive) is the number of positive samples correctly predicted as positive samples; FP is the number of negative samples wrongly predicted as positive; FN is the number of positive samples wrongly predicted as negative samples; N is the number of sample classes of dataset. Judgment of positive and negative samples was based on the threshold of the Intersection over Union (IoU) which is the area of overlap between the predicted segmentation and the ground truth divided by the area of union between the predicted segmentation and the ground truth. If the IoU is greater than the threshold, it is classified as a positive sample, and if the IoU is less than the threshold, it is a negative sample. When the IoU is 0.5, the average accuracy of the YOLOv5 model is expressed as $AP_{0.5}$, and the mean average accuracy is described as $mAP_{0.5}$.

2.3.3. Experimental Platform and Parameter Setting

The experimental configuration consists of the software environment of Pytorch1.8.0 and the hardware equipment of the CPU of Intel (R) Core (TM) i5 and GPU of NVIDIA GeForce RTX 3060 Ti with 16 GB memory and the Windows10 operation system, as shown in Table 1. During the training process of the YOLOv5 model, Adaptive Moment Estimation (Adam), an algorithm for stochastic optimization, is used - which only requires little memory. The momentum factor was set at 0.937 and the initial learning rate was 0.001.

The learning rate is adjusted by the cosine annealing method [42]. The weight attenuation coefficient, the Batch size, and the epoch times of training was set to 0, 16, and 300, respectively. The method of label smoothing was used to smooth the classification label of the image, which could help to avoid overfitting. The value of label smoothing was 0.01.

Table 1. Experimental platform configuration.

Configuration Name	Parameter
Operation System	Win10
GPU	NVIDIA GeForce RTX 3060 Ti
CPU	Intel (R) Core (TM) i5-10400F CPU @2.90 GHz
Memory	16 G
Deep Learning Framework	Pytorch1.8.0

3. Results

3.1. Comparison of Experimental Results on a Public Dataset

In order to compare the performance of the improved model with that of the original YOLOv5 model, an experimental comparison was made on the PASCAL VOC dataset. The experimental results are shown in Table 2.

Table 2. Experimental results on PASCAL VOC dataset.

Method	Training Dataset	Test Dataset	mAP
YOLOv5_Y	VOC07 + 12	VOC-Test07	0.8620
Ours	VOC07 + 12	VOC-Test07	0.8731

The improved YOLOv5 model can increase the mAP by 1.11% and enhance the effect of objection detection.

3.2. Ablation Study

In order to compare the performance of the improved model with other different models, ablation studies were carried out. The label smoothing method was used to process the images in the training experiment. Evaluation was performed every 10 epochs of training. A total of 300 epochs were trained in the experiment. The experimental results are shown in Table 3.

Table 3. Ablation study results for model on self-built construction waste dataset.

Method	AP				mAP	F1			
	Brick	Wood	Stone	Plastic		Brick	Wood	Stone	Plastic
YOLOv5_Y	0.8711	0.9138	0.9158	0.8959	0.8991	0.84	0.82	0.89	0.85
YOLOv5_C	0.9141	0.9572	0.9511	0.9581	0.9451	0.89	0.90	0.93	0.87
YOLOv5_S	0.9075	0.9565	0.9430	0.9447	0.9379	0.88	0.92	0.92	0.90
YOLOv5_D	0.9119	0.9581	0.9534	0.9605	0.9460	0.89	0.91	0.92	0.91
YOLOv5_CS	0.9132	0.9551	0.9506	0.9680	0.9467	0.89	0.92	0.94	0.90
YOLOv5_CD	0.9215	0.9485	0.9422	0.9545	0.9417	0.89	0.91	0.93	0.92
YOLOv5_SD	0.9095	0.9555	0.9412	0.9718	0.9445	0.88	0.93	0.93	0.92
Ours	0.9222	0.9659	0.9555	0.9485	0.9480	0.89	0.91	0.93	0.92

YOLOv5_Y, YOLOv5_C, YOLOv5_S, YOLOv5_D, YOLOv5_CS, YOLOv5_CD, and YOLOv5_SD represent the original YOLOv5 model, the model with CBAM, the model with the SimSPPF module, the model with improved multi-scale detection, the model with the CBAM and SimSPPF module, the model with CBAM and improved multi-scale detection, and the model with the SimSPPF module and improved multi-scale detection, respectively. Our study denotes the proposed YOLOv5 model, i.e., the model with all of the CBAM,

the SimSPPF module and improved multi-scale detection. Additionally, AP values of the four classes for the YOLOv5 original model are described in Figure 9 and for the improved YOLOv5 model in Figure 10.

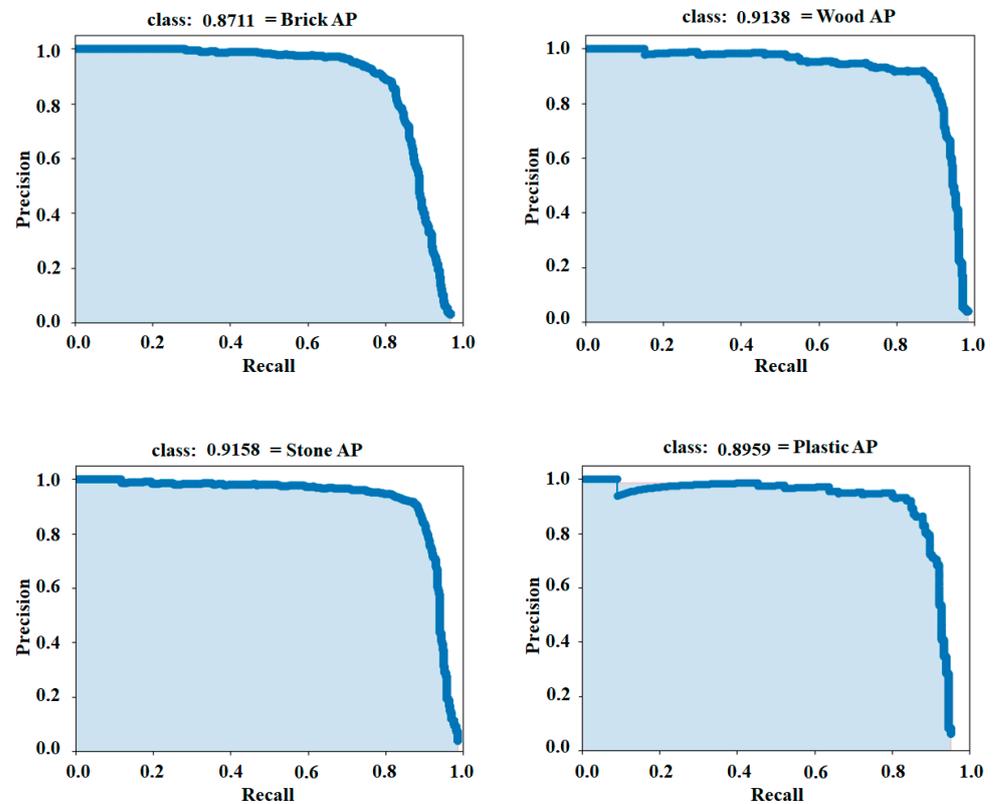


Figure 9. The AP Value of the Four Classes for YOLOv5 Original Model.

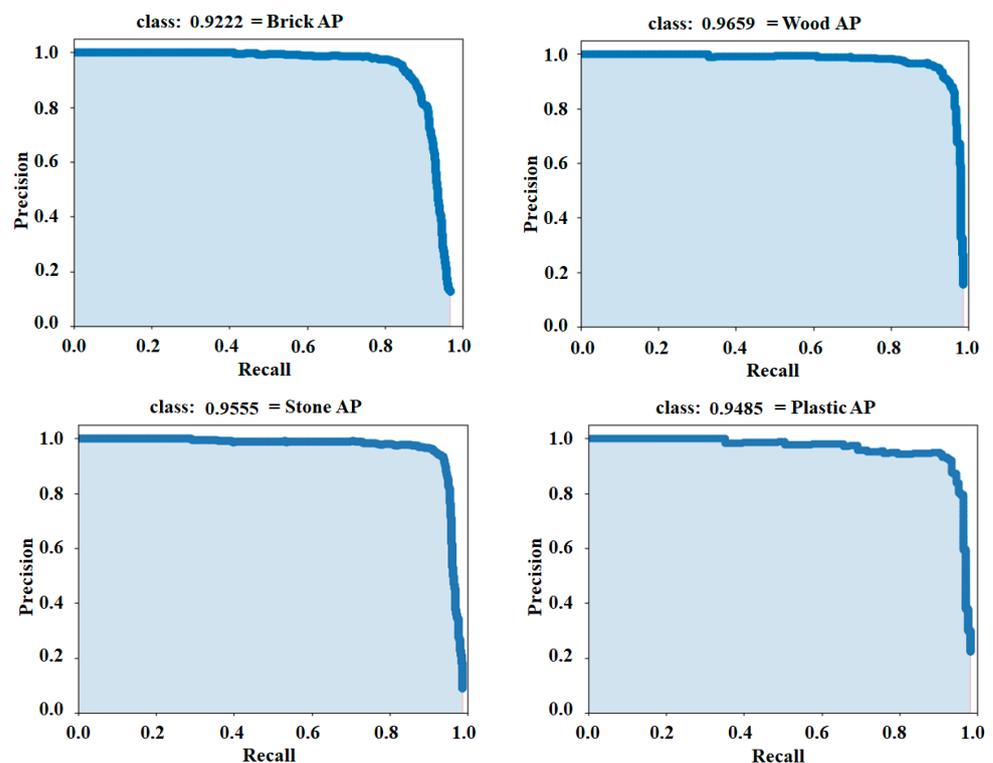


Figure 10. The AP Value of Four Types of objects for Our Improved Method.

It can be seen from Table 2 that the mAP of the original model of YOLOv5 on the construction waste dataset is the lowest, at only 0.8991. When CBAM was added, mAP can be improved by 4.6%, up to 0.9451, but with the replacement of the SPP module with the SimSPPF module, mAP could only be increased by 3.88%. Even so, the method of improvement of multi-scale detection can increase mAP by 4.69%. When both CBAM and the SimSPPF module were added, mAP increased by 4.76%. When increasing CBAM and improving multi-scale detection, mAP could be increased by 4.26%. With the addition of the SimSPPF module and the improvement of multi-scale detection, mAP was increased by 4.54%. However, in the experiment with our proposed method, i.e., when adding the CBAM, the SimSPPF module and improving multi-scale detection, mAP can be improved by 4.89%, the highest of all the different models.

Similarly, the F1-scores of brick, wood, stone and plastic of the original YOLOv5 model are 0.84, 0.82, 0.89 and 0.85, respectively, as shown in Table 2. However, in the proposed method, which adds all the CBAM and the SimSPPF module and improves multi-scale detection, the F1-scores of brick, wood, stone and plastic are 0.89, 0.91, 0.93 and 0.92, increasing by 5%, 9%, 4% and 7%, respectively. Therefore, the improved YOLOv5 model has a higher accuracy and better availability in objection detection, thereby improving the efficiency of sorting construction waste.

3.3. Contrast Experiment

To further verify the advantages and effectiveness of the improved YOLOv5 model on the construction waste dataset, a contrast experiment was also carried out. Comparing the improved model with other conventional models, such as YOLOv7, YOLOv5, YOLOv4, YOLOv3 and Faster-RCNN models, the loss indicator and mAP of every model in the training experiment and testing experiment are shown in Figure 11.

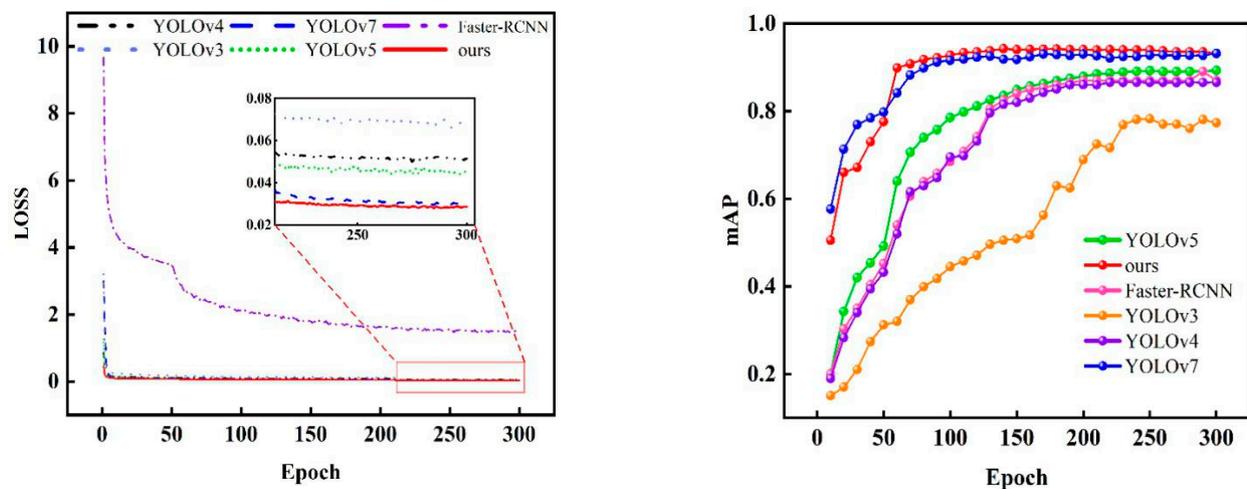


Figure 11. The Value of Loss and mAP of Different Models.

The loss of every model decreases rapidly in the first 20 epochs, which shows the training does not reach a stable state. If the training became stable, the loss in the curve would be flat rather than steep. When the training reaches a relatively stable state, the loss in our model is lower than that of other models. Meanwhile, the mAP value of every model increases rapidly at the first 60 epochs. Similarly, ours has the most obvious improvement among all the models. After 200 epochs of training, all models tend to be steadier, and the mAP of ours is significantly higher than that of other models.

The values of the evaluation indicators are also described in Table 4. Compared with YOLOv7, YOLOv5, YOLOv4, YOLOv3 and the Faster-RCNN model, the mAP value of our model increased by 2.15%, 4.89%, 8.24%, 16.12% and 7.78%, respectively. It shows that the improved YOLOv5 model can improve the accuracy to classify and detect construction waste.

Table 4. The Performance Contrast for Different Models.

Method	AP				mAP	F1			
	Brick	Wood	Stone	Plastic		Brick	Wood	Stone	Plastic
Ours	0.9222	0.9659	0.9555	0.9485	0.9480	0.89	0.91	0.93	0.92
YOLOv7	0.9006	0.9416	0.9250	0.9388	0.9265	0.88	0.93	0.91	0.92
YOLOv5_Y	0.8711	0.9138	0.9158	0.8959	0.8991	0.84	0.82	0.89	0.85
YOLOv4	0.8063	0.8972	0.8809	0.8782	0.8656	0.74	0.85	0.85	0.84
YOLOv3	0.7016	0.8157	0.8239	0.8058	0.7868	0.69	0.77	0.77	0.79
Faster-RCNN	0.7933	0.8967	0.8886	0.9021	0.8702	0.75	0.85	0.83	0.88

4. Conclusions

In construction waste sorting, inter-object occlusions and small-object detection are the two most important problems, affecting the effective performance of the construction waste detection system. In order to increase the accuracy of object detection, an improved YOLOv5 model is proposed for intelligent construction waste sorting and is trained by a dataset consisting of 3046 construction waste images of, for example, bricks, wood, stones, and plastics. The following conclusions can be drawn:

- (1) An improved YOLOv5 can be obtained through the fourth-scale feature fusion, shallow detection layer, CBAM and SimSPPF. The increase in CBAM and SimSPPF in the backbone layer of YOLOv5 can strengthen the characteristics of small objects and mutual occlusion and enhance the effect and accuracy of detection, thus improving the generalization ability and robustness of the model. The fourth-scale feature fusion is added to the feature fusion part in the Neck and a shallow detection layer is added to the Head, which can aid in the detection of small objects and inter-occlusion.
- (2) Compared with the conventional models of Faster-RCNN, YOLOv3, YOLOv4 and YOLOv7, the detection accuracy of the proposed model is higher, and its mAP can reach up to 0.9480, which verifies the accuracy and the availability of the improved YOLOv5 model.

Author Contributions: Project administration, Q.Z.; methodology, H.L.; Data collection, Y.Q. and W.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the “Research Project of the Ministry of Housing and urban-rural Development of the People’s Republic of China, grant number 2022-K-079”.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data used in this research can be provided upon request.

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

References

1. Huang, B.; Gao, X.; Xu, X.; Song, J.; Geng, Y.; Sarkis, J.; Fishman, T.; Kua, H.; Nakatani, J. A life cycle thinking framework to mitigate the environmental impact of building materials. *One Earth* **2020**, *3*, 564–573. [CrossRef]
2. Teh, S.H.; Wiedmann, T.; Moore, S. Mixed-unit hybrid life cycle assessment applied to the recycling of construction materials. *J. Econ. Struct.* **2018**, *7*, 13. [CrossRef]
3. Duan, H.; Miller, T.R.; Liu, G.; Tam, V.W.Y. Construction debris becomes growing concern of growing cities. *Waste Manag.* **2019**, *83*, 1–5. [CrossRef] [PubMed]
4. Lei, J.; Huang, B.; Huang, Y. Life cycle thinking for sustainable development in the building industry. In *Life Cycle Sustainability Assessment for Decision-Making*; Elsevier: Amsterdam, The Netherlands, 2020; pp. 125–138.
5. Yu, B.; Wang, J.; Li, J.; Lu, W.; Li, C.Z.; Xu, X. Quantifying the potential of recycling demolition waste generated from urban renewal: A case study in Shenzhen, China. *J. Clean. Prod.* **2020**, *247*, 119127. [CrossRef]
6. Ku, Y.; Yang, J.; Fang, H.; Xiao, W.; Zhuang, J. Deep learning of grasping detection for a robot used in sorting construction and demolition waste. *J. Mater. Cycles Waste Manag.* **2021**, *23*, 84–95. [CrossRef]
7. Zen Robotics. Available online: <https://zenrobotics.com/> (accessed on 1 October 2022).

8. Sadako Technologies. Applications/Max-AI. Available online: <https://sadako.es/max-ai/> (accessed on 1 October 2022).
9. Machinex. SAMURAI-Recycling Sorting Robots. Available online: <https://www.machinexrecycling.com/products/samurai-sorting-robot/> (accessed on 1 September 2022).
10. AMP Robotics. Available online: <https://www.amprobotics.com/> (accessed on 1 October 2022).
11. Koskinopoulou, M.; Raptopoulos, F.; Papadopoulos, G.; Mavrakis, N.; Maniadakis, M. Robotic waste sorting technology: Toward a vision-based categorization system for the industrial robotic separation of recyclable waste. *IEEE Robot. Autom. Mag.* **2021**, *28*, 50–60. [[CrossRef](#)]
12. Adedeji, O.; Wang, Z. Intelligent waste classification system using deep learning convolutional neural network. *Procedia Manuf.* **2019**, *35*, 607–612. [[CrossRef](#)]
13. Chen, J.; Lu, W.; Xue, F. “Looking beneath the surface”: A visual-physical feature hybrid approach for unattended gauging of construction waste composition. *J. Environ. Manag.* **2021**, *286*, 112233. [[CrossRef](#)] [[PubMed](#)]
14. Yang, J.; Zeng, Z.; Wang, K.; Zou, H.; Xie, L. GarbageNet: A unified learning framework for robust garbage classification. *IEEE Trans. Artif. Intell.* **2021**, *2*, 372–380. [[CrossRef](#)]
15. Hoong, J.D.L.H.; Lux, J.; Mahieux, P.-Y.; Turcry, P.; Ait-Mokhtar, A. Determination of the composition of recycled aggregates using a deep learning-based image analysis. *Automat. Constr.* **2020**, *116*, 103204. [[CrossRef](#)]
16. Zhihong, C.; Hebin, Z.; Yanbo, W.; Binyan, L.; Yu, L. A vision-based robotic grasping system using deep learning for garbage sorting. In Proceedings of the 2017 36th Chinese Control Conference (CCC), Dalian, China, 26–28 July 2017; pp. 11223–11226.
17. Awe, O.; Mengistu, R.; Sreedhar, V. Smart trash net: Waste localization and classification. *arXiv* **2017**. preprint. Available online: <http://cs229.stanford.edu/proj2017/final-reports/5226723.pdf> (accessed on 1 January 2022).
18. Wang, Z.; Li, H.; Zhang, X. Construction waste recycling robot for nails and screws: Computer vision technology and neural network approach. *Automat. Constr.* **2019**, *97*, 220–228. [[CrossRef](#)]
19. Nowakowski, P.; Pamuła, T. Application of deep learning object classifier to improve e-waste collection planning. *Waste Manag.* **2020**, *109*, 1–9. [[CrossRef](#)] [[PubMed](#)]
20. Zhou, H.; Zhao, L. Intelligent detection and classification of domestic waste based on improved faster-RCNN. *J. Fuyang Norm. Univ. Nat. Sci.* **2022**, *39*, 49–55.
21. Li, J.; Fang, H.; Fan, L.; Yang, J.; Ji, T.; Chen, Q. RGB-D fusion models for construction and demolition waste detection. *Waste Manag.* **2022**, *139*, 96–104. [[CrossRef](#)] [[PubMed](#)]
22. Lin, K.; Zhou, T.; Gao, X.; Li, Z.; Duan, H.; Wu, H.; Lu, G.; Zhao, Y. Deep convolutional neural networks for construction and demolition waste classification: VGGNet structures, cyclical learning rate, and knowledge transfer. *J. Environ. Manag.* **2022**, *318*, 115501. [[CrossRef](#)]
23. Redmon, J.; Farhadi, A. Yolo3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
24. Bochkovskiy, A.; Wang, C.-Y.; Liao, H.-Y.M. Yolo4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
25. Jiang, P.; Ergu, D.; Liu, F.; Cai, Y.; Ma, B. A Review of Yolo algorithm developments. *Procedia Comput. Sci.* **2022**, *199*, 1066–1073. [[CrossRef](#)]
26. Liu, W.; Peng, J.; Wu, B.; You, T. Improved YOLOv3 life article detection method for sorting. *Transducer Microsyst. Technol.* **2022**, *41*, 134–137.
27. Chen, Y.; Li, L.; Xie, H.; Lu, S.; Dong, J. Garbage sorting robot based on machine vision. *Instrum. Anal. Monit.* **2022**, *1*, 30–34.
28. Yuan, H.; Zang, T. Underwater garbage target detection based on the attention mechanism Ghosty-YOLOV5. *Environ. Eng.* **2022**, *9*, 1–14. Available online: <https://kns.cnki.net/kcms/detail/11.2097.X.20220913.1006.004.html> (accessed on 1 January 2022).
29. Wang, Z.; Li, H.; Yang, X. Vision-based robotic system for on-site construction and demolition waste sorting and recycling. *J. Build. Eng.* **2020**, *32*, 101769. [[CrossRef](#)]
30. Chen, L.; Cao, Y.; Huang, M.; Xie, X. Flame detection method based on improved YOLOv5. *Comput. Eng.* **2022**, *10*, 1–17. [[CrossRef](#)]
31. 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](#)]
32. Lin, T.-Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.
33. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path aggregation network for instance segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 8759–8768.
34. Jiang, D.; Jiang, Z.; Huang, Z.; Guo, C.; Li, B. Uav vehicle target detection algorithm based on Efficientnet. *Comput. Eng. Appl.* **2022**, *10*, 1–11. Available online: <https://kns.cnki.net/kcms/detail/11.2127.TP.20221027.0859.002.html> (accessed on 1 January 2022).
35. Su, S.; Chen, R.; Zhu, Y.; Jiang, B. Relocation non-maximum suppression algorithm. *Opt. Precis. Eng.* **2022**, *30*, 1620–1630.
36. Lin, T.-Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft coco: Common objects in context. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 8–11 September 2014; pp. 740–755.
37. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
38. Everingham, M.; van Gool, L.; Williams, C.K.I.; Winn, J.; Zisserman, A. The pascal visual object classes (voc) challenge. *Int. J. Comput. Vis.* **2010**, *88*, 303–338. [[CrossRef](#)]

39. Thung, G.; Yang, M. Classification of Trash for Recyclability Status. CS229 Project Report. 2016, pp. 1–6. Available online: <http://cs229.stanford.edu/proj2016/report/ThungYang-ClassificationOfTrashForRecyclabilityStatus-report.pdf> (accessed on 1 January 2022).
40. Proença, P.F.; Simões, P. Taco: Trash annotations in context for litter detection. *arXiv* **2020**, arXiv:2003.06975.
41. Zhao, X.; Zhang, Q.; Wang, W.; Xu, Z. Image detection method of combustible dust cloud. *China Saf. Sci. J.* **2020**, *30*, 8–13.
42. Qiu, Z.; Zhu, X.; Liao, C.; Shi, D.; Kuang, Y.; Li, Y.; Zhang, Y. Detection of bird species related to transmission line faults based on lightweight convolutional neural network. *IET Gener. Transm. Dis.* **2022**, *16*, 869–881. [[CrossRef](#)]

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