International Core Journal of Engineering Volume 12 Issue 11, 2026
ISSN: 2414-1895 DOI: 10.6919/ICJE.202611_12(11).0013

Detection of Waste Plastic Bottles based on an Improved YOLOv8s Model

Xiangchi Li, Liying Deng

Shenyang University of Technology, Liaoyang 111003, China

Abstract

Waste plastic bottles are frequently compressed, deformed, and stacked during recycling. To improve their detection, an enhanced model is developed from YOLOv8s. BiFPN is introduced into the Neck to strengthen feature fusion across scales; Coordinate Attention is added to the P3/8 and P4/16 branches to emphasize bottle mouths, body contours, and edge regions; and SIoU is adopted to improve bounding-box regression. Experimental results show that the proposed model achieves a Precision of 0.927, a Recall of 0.871, and an mAP@50 of 0.913, representing improvements of 5.5, 9.9, and 6.2 percentage points over the baseline, respectively. The model reduces missed detections and localization errors and improves the detection of waste plastic bottles.

Keywords

Waste Plastic Bottles; YOLOv8s; Object Detection.

1. Introduction

With socioeconomic development and rising living standards, the use of plastic products in daily life has continued to increase [1]. Waste beverage and mineral-water bottles, which are frequently used and have relatively high recycling value, have become an important component of recyclable waste [2][3]. If these bottles are not sorted and recycled promptly and effectively, they not only cause resource waste but may also produce long-term adverse effects on the environment [4]. Improving the recognition, classification, and sorting efficiency of waste plastic bottles is therefore important for promoting resource recycling and the intelligent development of waste classification[5,6].Xie Shilong et al. [7] introduced BiFPN and WIoU v3 into YOLOv8n to modify feature fusion and box regression. Luo Qianqian et al. [8] applied YOLOv8 to recyclable-waste recognition. For waste plastic bottles, detection errors are more likely to occur when targets overlap, vary in scale, or become deformed and partially occluded. These cases place greater demands on feature aggregation and box localization. Based on this observation, YOLOv8s is used as the baseline, and the subsequent improvements are made to the Neck, attention module, and regression loss. Existing studies indicate that the application of YOLO-based detectors to recyclable-waste recognition has gradually shifted from direct model deployment toward task-oriented structural modification. Lightweight modules have been introduced to reduce computational cost, while multi-scale feature-fusion structures and improved bounding-box regression losses have been used to enhance detection accuracy under cluttered backgrounds and target occlusion [6,7]. In addition, YOLOv8 has been applied to recyclable-waste recognition, showing that the detector can provide a practical basis for subsequent sorting and localization tasks [8]. These studies also suggest that feature representation, multi-scale information exchange, and bounding-box regression remain important factors affecting detection performance in waste-sorting scenarios.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202611_12(11).0013

2. YOLOv8 Algorithm

YOLO formulates object recognition and localization as a single-stage prediction task [9]. YOLOv8 retains this detection scheme while making further adjustments to the network structure and prediction strategy [10,11]. Figure 1 presents the YOLOv8s architecture adopted in this study. Image features are first extracted by the Backbone, passed to the Neck for cross-scale aggregation, and finally delivered to a decoupled Head, where category prediction and bounding-box regression are performed separately.The overall architecture of YOLOv8 is shown in Figure 1.

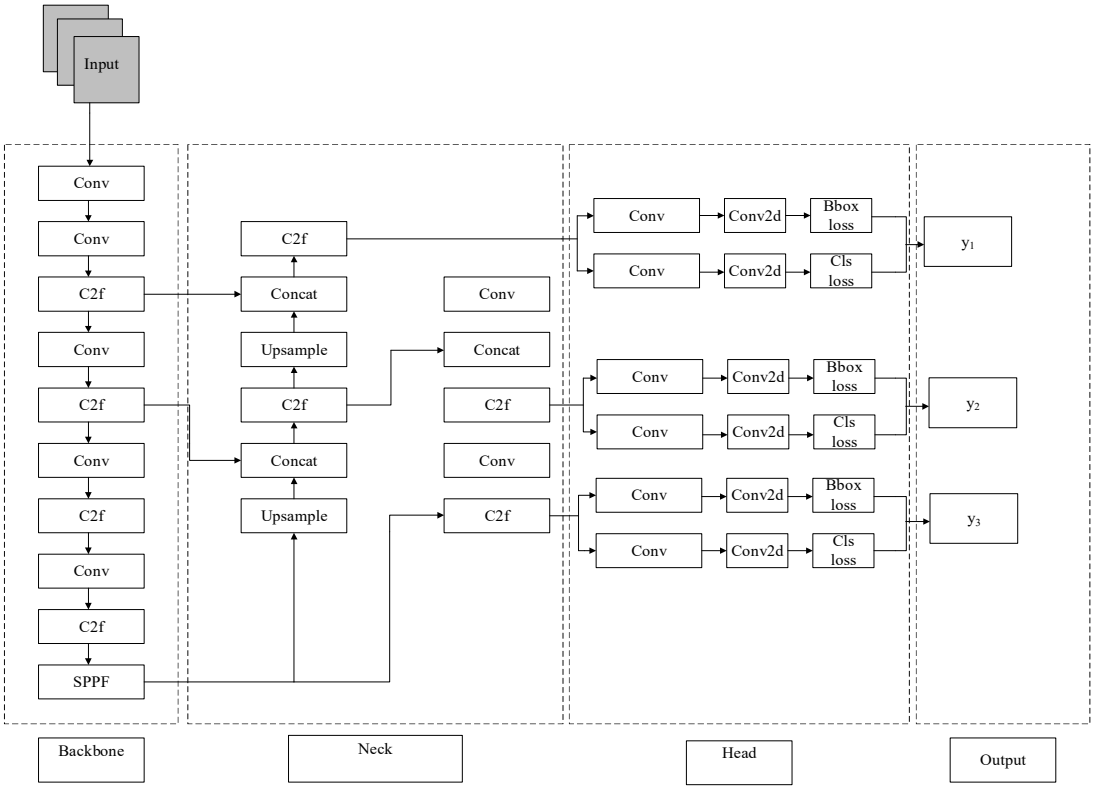


Figure 1. Network architecture of YOLOv8

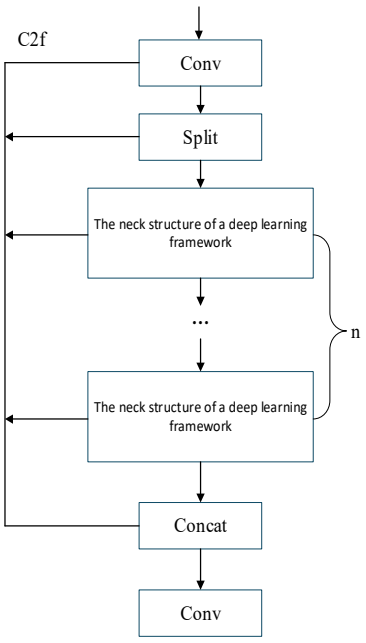


Figure 2. Structure of the C2f module

ISSN: 2414-1895 DOI: 10.6919/ICJE.202611_12(11).0013

In the Backbone, YOLOv8s uses C2f as one of its principal feature-extraction modules [10,11]. Instead of passing all features through a single sequential path, C2f divides the intermediate features into different branches and reconnects them after repeated Bottleneck operations. This arrangement preserves information from different processing depths and improves gradient propagation without introducing an excessively large computational burden. The internal structure of the C2f module is illustrated in Figure 2.

For waste plastic bottle images, such hierarchical feature transmission is useful because bottle contours, mouths, labels, and partially occluded regions appear at different spatial scales. Nevertheless, the baseline feature-fusion process still has limitations when multiple bottles overlap or when local contours are incomplete. The subsequent modifications therefore retain the Backbone structure and concentrate mainly on multi-scale feature fusion, feature attention, and bounding-box regression.

The Neck receives feature maps produced at different stages of the Backbone and combines spatial details from shallow layers with higher-level semantic information. This multi-scale representation is important for waste plastic bottle detection because the apparent size of a bottle changes with its position, orientation, and degree of overlap. The detection Head then converts the fused features into classification and localization outputs. YOLOv8 adopts a decoupled Head and an anchor-free prediction strategy, allowing classification and box regression to be optimized through separate branches [10]. Although this design provides a suitable baseline, errors are still observed when bottle contours are partially occluded or when several targets overlap closely.

3. Improvements to YOLOv8

3.1 Improved YOLOv8 Network

The modifications were determined from the failure cases observed in the baseline YOLOv8s results. Missed detections were more likely to occur when adjacent bottles overlapped or when only part of a bottle contour remained visible, while inaccurate boxes were mainly associated with deformation and inclined placement. On this basis, three parts of the baseline model were adjusted. BiFPN was introduced into the Neck to improve information exchange across feature scales; Coordinate Attention was inserted into the P3/8 and P4/16 branches to strengthen responses to bottle mouths, body contours, and other local structures; and SIoU was used in the regression branch to provide additional constraints on box localization. The modified network is shown in Figure 3.

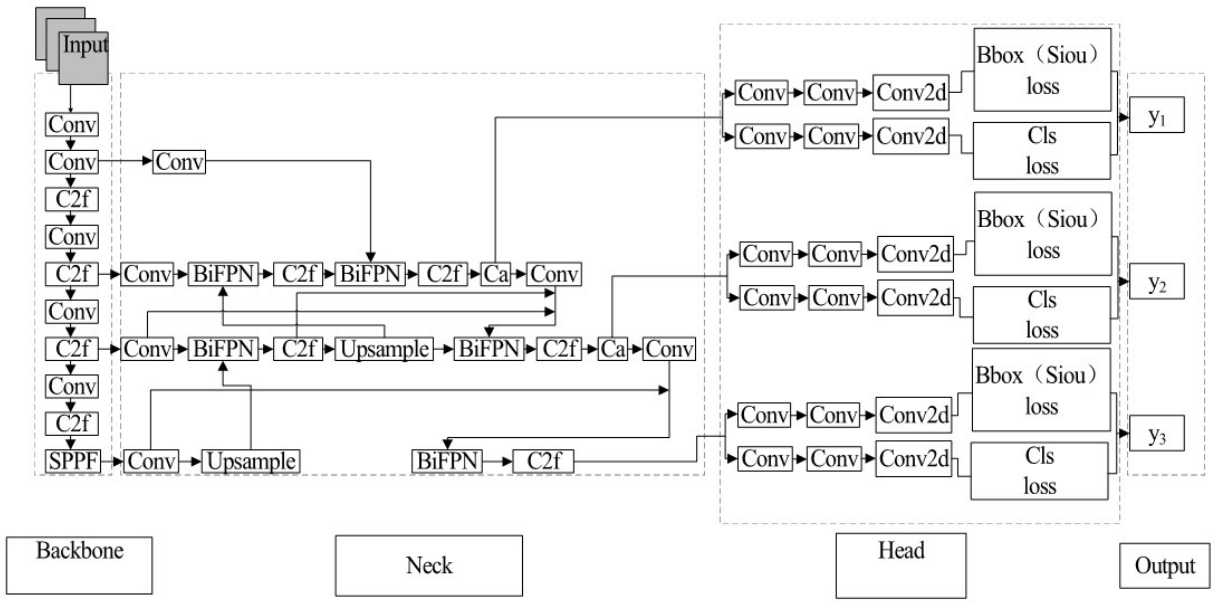


Figure 3. Architecture of the improved YOLOv8s model

ISSN: 2414-1895 DOI: 10.6919/ICJE.202611_12(11).0013

3.2 Improvement of the Feature-Fusion Network

The Neck in the original YOLOv8s adopts a PANet-based multi-scale fusion structure. Feature maps from neighboring levels are transferred and combined through upsampling, downsampling, and concatenation operations. Although this design enables information exchange across different resolutions, the features entering the same fusion stage are still combined in a relatively fixed manner. As a result, differences in the contribution of individual feature levels to waste plastic bottle detection are not fully reflected during fusion.

The Neck of the baseline YOLOv8s uses a PAN-style structure to combine features at different resolutions. In this study, the original fusion nodes are replaced with BiFPN nodes [12], while the Backbone and detection Head are kept unchanged. The modified fusion structure is shown in Figure 4.

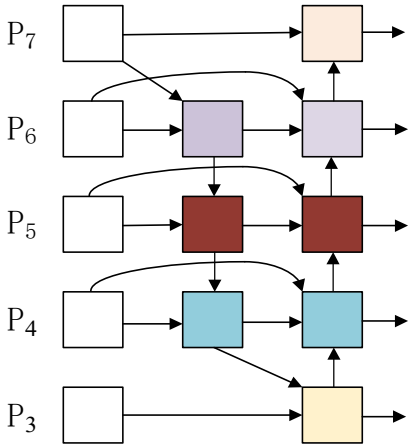


Figure 4. Structure of BiFPN

BiFPN assigns learnable weights to features entering the same fusion node. For the input features F_i, the fusion result is calculated according to Eq. (1).

$$F_{out} = \sum_i \frac{\omega_i}{\varepsilon + \sum_j \omega_j} \cdot F_i \tag{1}$$

In Eq. (1), $\boldsymbol{F_{out}}$ denotes the fused feature, $\boldsymbol{F_i}$ represents the input feature at the $\boldsymbol{i}$-th scale, $\boldsymbol{\omega_i}$ is the corresponding trainable weight, and ε is a small positive constant introduced to avoid numerical instability. The fusion weights are updated together with the network parameters during training.

3.3 Introduction of Coordinate Attention

Coordinate Attention (CA) introduces coordinate information into channel attention, allowing the network to model channel dependencies while retaining target-location information [13]. The process mainly consists of coordinate-information embedding and attention-weight generation. For an input feature map, one-dimensional global average pooling is performed separately along the width and height directions:

$$z_c^h(h) = \frac{1}{W} \sum_{i=0}^{W-1} x_c(h,i) \tag{2}$$

ISSN: 2414-1895 DOI: 10.6919/ICJE.202611_12(11).0013

$$z_c^w(w) = \frac{1}{H}\sum_{j=0}^{H-1} x_c(j,w) \tag{3}$$

In Eqs. (2) and (3), $z_c^{h(h)}$ and $z_c^{w(w)}$ denote feature vectors obtained by pooling along the width and height directions, respectively. The two features are concatenated and processed by a 1 × 1 convolution and a nonlinear activation function to obtain the intermediate feature . *The calculation is expressed as follows:*

$$f = \delta\left(F_1\left([z^h, z^w]\right)\right) \tag{4}$$

The intermediate feature is then split into two directional branches. Convolution and the Sigmoid function are used to generate the attention weights g^h and g^w. The final output is:

$$y_c(i,j) = x_c(i,j) \times g_c^h(i) \times g_c^w(j) \tag{5}$$

Through the above operations, CA highlights target regions and suppresses irrelevant information. Waste plastic bottles often have small mouths, elongated body contours, and local features that are easily lost under occlusion. Accordingly, CA is embedded in the small- and medium-target detection branches of YOLOv8s, corresponding to the P3/8 and P4/16 feature layers. These layers retain abundant edge, texture, and positional information; after CA weighting, the responses to bottle mouths, body boundaries, and occluded regions are further strengthened. The P5/32 branch is mainly responsible for large targets whose overall structures are relatively complete, so CA is not added to this branch. This targeted arrangement better matches the characteristics of waste plastic bottle detection and improves the recognition of small, medium, and overlapping targets.

3.4 Improvement of the Loss Function

SIoU is adopted in the regression branch of YOLOv8s to improve the localization of waste plastic bottles. Traditional bounding-box regression losses usually evaluate the predicted box according to overlap, center distance, and differences in width and height. For the bottle samples considered in this study, target geometry is often affected by inclination, stacking, occlusion, and compression because of the elongated shape of the bottle body. Under these conditions, the discrepancy between a predicted box and its ground-truth counterpart may involve both positional deviation and changes in orientation or shape. SIoU introduces angle, distance, shape, and overlap terms into the regression process [14], allowing these geometric differences to be considered simultaneously during box optimization. This modification provides additional constraints for irregular bottle poses and helps alleviate box displacement or incomplete coverage of the bottle body in complex recycling scenes.

SIoU considers overlap area, center distance, width-height differences, and regression direction in bounding-box regression. Its total loss is expressed as follows:

$$L_{\mathrm{SIoU}} = 1 - \mathrm{IoU} + \frac{\Delta + \Omega}{2} \tag{6}$$

In Eq. (6), L_{SIoU} denotes the bounding-box regression loss, Δ is the distance cost, and Ω is the shape cost. The angle cost contributes to box regression by modulating the distance cost.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202611_12(11).0013

4. Experiments and Comparative Analysis

4.1 Experimental Dataset

Since few public datasets are tailored specifically to waste plastic bottle detection, the images used in this study were collected and organized as a dedicated dataset. During image collection, scenes containing bottle overlap and background clutter were deliberately retained to reflect practical recycling conditions. The original samples were then processed with translation, brightness variation, and noise perturbation to broaden the range of training conditions. These augmentation operations increased the diversity of the available samples and helped reduce excessive dependence on a limited set of image patterns during training. Representative augmented images are presented in Figure 5.



Figure 5. Examples of images after data augmentation

The dataset was divided into training, validation, and test sets at a ratio of 7:2:1, containing 2,275, 650, and 325 images, respectively.

4.2 Experimental Environment and Training Settings

The hardware and software configuration adopted in the experiments is summarized in Table 1. During training, the input images were resized to 640 × 640, with a batch size of 16, and the model was optimized for 150 epochs. SGD was selected as the optimizer, with the initial learning rate and weight decay set to 0.01 and 0.0005, respectively. The learning rate was updated during training according to a cosine annealing schedule.

Table 1. Experimental environment configuration

Parameter	Configuration
Operating system	Windows 11
CPU	AMD Ryzen 7 7840H
GPU	NVIDIA GeForce RTX 4060 Laptop GPU
Python version	Python 3.10.20
Acceleration environment	CUDA 12.4

4.3 Evaluation Metrics

The evaluation metrics used in this study include Precision, Recall, and mean Average Precision (mAP).

Recall represents the proportion of actual targets that are correctly detected and is calculated as follows:

ISSN: 2414-1895 DOI: 10.6919/ICJE.202611_12(11).0013

$$R = \frac{TP}{TP + FN} \tag{7}$$

In Eq. (7), TP denotes the number of targets correctly detected by the model, and FN denotes the number of actual targets that are not detected.

As shown by the Recall curves in Figure 6, the improved model consistently outperforms the original YOLOv8s. At the end of training, Recall increases from 0.772 to 0.871, an improvement of 9.9 percentage points, indicating that the proposed model reduces missed detections of waste plastic bottles.

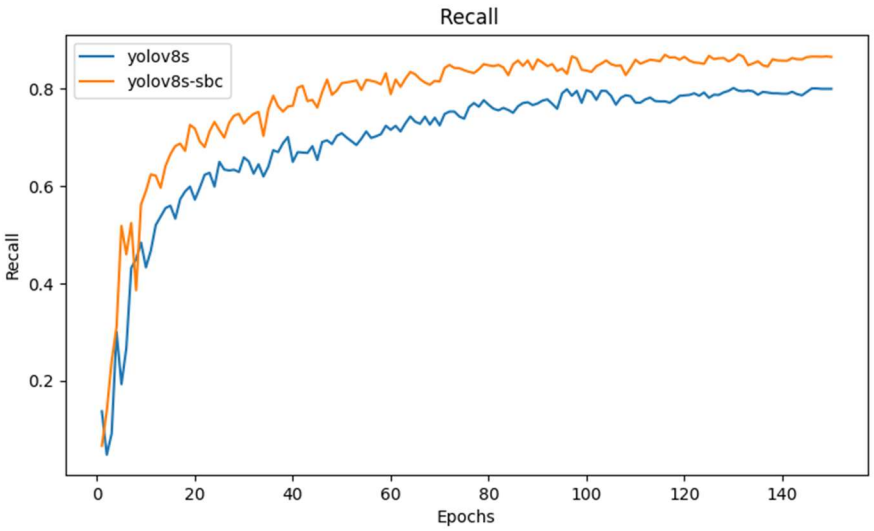


Figure 6. Comparison of Recall curves

As shown in Figure 7, the mAP@50 curve of the improved model remains above that of the original YOLOv8s and rises more rapidly during the early training stage. At the end of training, mAP@50 increases from 0.851 to 0.913, an improvement of 6.2 percentage points. This result shows that the improved model provides better recognition of waste plastic bottles and higher overall detection accuracy.

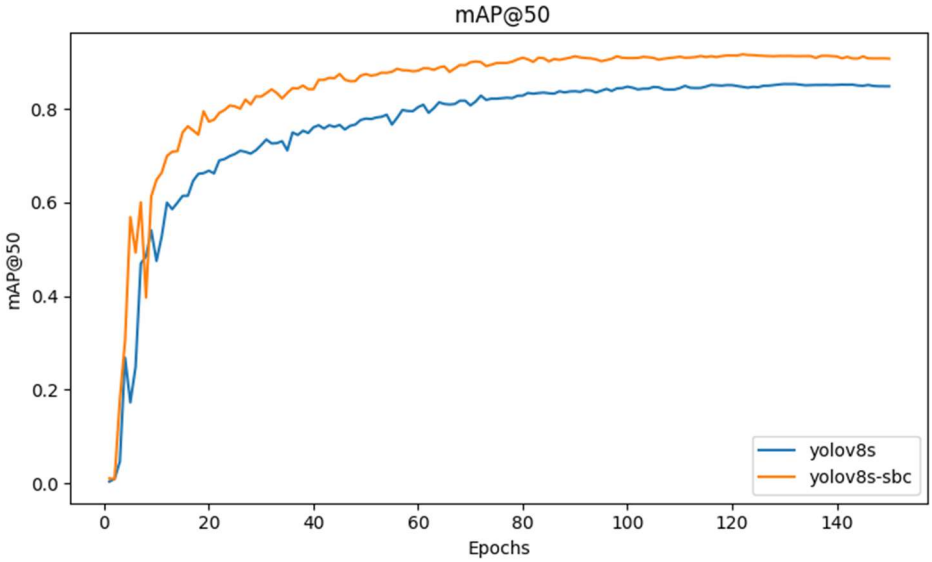


Figure 7. Comparison of mAP@50 curves

ISSN: 2414-1895 DOI: 10.6919/ICJE.202611_12(11).0013

4.4 Ablation Study

To assess the contribution of each modification, ablation experiments were performed using YOLOv8s as the baseline, as shown in Table 2. Each module improved performance: SIoU mainly increased Recall, whereas CA produced the largest gain in Precision. Two-module combinations provided further improvements. When SIoU, BiFPN, and CA were used together, Precision, Recall, and mAP@50 reached 0.927, 0.871, and 0.913, respectively, which were 5.5, 9.9, and 6.2 percentage points above the baseline. These results indicate that the three modules have complementary effects.

Table 2. Comparison of ablation-study results

Model	SIoU	BiFPN	CA	Precision	Recall	mAP@50
YOLOv8s				0.872	0.772	0.851
YOLOv8s-S	√			0.889	0.867	0.875
YOLOv8s-B		√		0.894	0.854	0.868
YOLOv8s-C			√	0.917	0.848	0.867
YOLOv8s-SB	√	√		0.924	0.868	0.894
YOLOv8s-SC	√		√	0.919	0.863	0.900
YOLOv8s-BC		√	√	0.923	0.856	0.895
Proposed model	√	√	√	0.927	0.871	0.913

5. Conclusion

This study improves YOLOv8s for waste plastic bottle detection. To address occlusion, deformation, scale variation, and indistinct boundaries, BiFPN is introduced into the Neck to strengthen feature fusion across levels. CA is added to the small- and medium-target detection branches to emphasize bottle mouths, body contours, and edge regions, while SIoU is adopted to improve bounding-box regression and localization. Experimental results show that the improved model achieves a Precision of 0.927, a Recall of 0.871, and an mAP@50 of 0.913, which are 5.5, 9.9, and 6.2 percentage points higher than those of the original YOLOv8s, respectively. The improved model therefore provides better recognition accuracy, target recall, and localization performance for waste plastic bottles and is more suitable for complex recycling scenes.

References

[1] Wang, J. S. (2022). What does China ranking first in the world in waste plastic recycling volume mean? Resource Recycling, (8), 46–47.

[2] Xue, Z. H., Liu, P., & Gao, Y. L. (2021). Current status of waste plastic recycling and reuse. Plastics Science and Technology, 49(4), 107–110.

[3] Xue, X. C., Wang, S. L., & Ju, K. L. (2022). Review of plastic waste resource utilization. Qinghai Science and Technology, 29(1), 113–118.

[4] Xing, J. K., Liu, T. T., & Wang, B. (2023). Research progress on closed-loop recyclable plastics. Sino-Global Energy, 28(9), 92–100.

[5] Wang, Y. Y. (2023). Design of a waste plastic bottle sorting system based on vision and industrial robots [Master's thesis]. Yancheng Institute of Technology.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202611_12(11).0013

[6] Li, X. Y., Gong, X., Fu, B., et al. (2026). Binocular-vision recognition and localization method for waste plastic bottles based on YOLOv5. Journal of Packaging, 18(1), 98–106.

[7] Xie, S. L., Wu, H., Mao, W. J., et al. (2024). Efficient color recognition of waste plastic bottles based on deep learning. Plastics Science and Technology, 52(11), 140–146.

[8] Luo, Q. Q., Peng, W., & Yang, G. H. (2024). Recyclable-waste recognition method and engineering application based on YOLOv8. Computer Knowledge and Technology, 20(17), 29–32.

[9] Redmon, J., Divvala, S., Girshick, R., et al. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 779–788).

[10] Terven, J., Córdova-Esparza, D. M., & Romero-González, J. A. (2023). A comprehensive review of YOLO architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-NAS. Machine Learning and Knowledge Extraction, 5(4), 1680–1716.

[11] Yaseen, M. (2024). What is YOLOv8: An in-depth exploration of the internal features of the next-generation object detector. arXiv. https://arxiv.org/abs/2408.15857

[12] Tan, M., Pang, R., & Le, Q. V. (2020). EfficientDet: Scalable and efficient object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 10781–10790). IEEE.

[13] Wu, C. Y., Zhang, C. H., Kong, D. J., et al. (2025). Infrared image detection algorithm for power transmission and transformation equipment based on CA-YOLOv8. Infrared Technology, 47(9), 1128–1134.

[14] Shen, X. L., & Guo, X. (2026). Attention-based small-object detection in remote-sensing images. Opto-Electronic Engineering, 53(3), 116–132.