

# A Study on Garbage Classification Detection Employing the Faster R-CNN Framework

Zhixian Zhang, Dan Li

Jincheng College, Sichuan University, Chengdu 610065, China

**Abstract:** *Since the comprehensive implementation of mandatory garbage classification in Shanghai in 2019, accompanied by the introduction of relevant management regulations, garbage classification has not only repeatedly captured public attention but has also emerged as a "new social norm." Nevertheless, classification accuracy rates remaining below 50% have accelerated the expansion of related markets and catalyzed technological innovation in this domain. In response to the dual challenges of continuously increasing waste volumes and persistently low disposal accuracy, this paper proposes the application of the Faster R-CNN algorithm for garbage classification detection. By obtaining garbage categories through object detection, the proposed approach provides technical support for subsequent waste disposal operations, thereby reducing the workload associated with environmental governance and enhancing the efficiency of garbage recycling processes.*

**Keywords:** Garbage classification; Faster R-CNN; Target detection.

## 1. INTRODUCTION

In recent years, China's economic growth and urbanization development have been very rapid, resulting in an increase in the amount of various types of production and household waste, which means that the difficulty of disposal will also be relatively high. Garbage classification may seem easy, but in reality, it is not easy to promote and implement. As early as December 2016, the central government held a meeting emphasizing garbage classification and regarded it as a major focus of follow-up work. At the same time, the meeting also pointed out the need to vigorously promote and popularize garbage classification knowledge, implement garbage classification system. In the past five years, most cities will implement garbage classification.

In the context of national policies and regulations on garbage classification management, there are still many problems in garbage classification in China. In 14 districts and 2 counties of Beijing, the accuracy rate of residents' garbage classification is only 16.92%~34.56%, indicating that residents' awareness in this area is still insufficient, and there is a situation of perfunctory publicity and education, which has not been implemented properly; There is still a situation of insufficient funding for grassroots communities. In Sancun, Xuhui District, Shanghai, street communities will provide subsidies, but over time, the pressure of high subsidies is not conducive to the long-term development of garbage classification; There is also the problem of inadequate policies in cities, where fines for garbage classification cannot be too high, but too low may not achieve the desired effect and may even lead to a series of problems.

Yi [1] proposed a real-time fair-exposure ad allocation framework for small and medium-sized businesses and underserved creators using contextual bandits-with-knapsacks. In photonic device design, Tang et al. [2] presented the design and optimization of shallow-angle grating couplers for vertical emission from indium phosphide devices. Deng and Yang [3] developed multi-layer defense strategies and privacy-preserving enhancements against membership reasoning attacks in a federated learning framework. Wu, Luo, and Liao [4] addressed small-sample object detection of surface cracks in concrete structures of high-rise buildings via multi-level transfer learning. Zhou [5] introduced a hierarchical needs framework for analyzing US automotive customer feedback linking sentiment to functional requirements, while Zhou [6] further applied gradient boosting trees to diagnose bottlenecks in international automotive sales funnels based on cross-regional team efficiency evaluation. Wensi [7] investigated AI-assisted marketing content generation for non-standard industrial automation solutions, and Li [8] optimized AI-driven bid pricing models for non-standard automation projects by leveraging historical financial data and machine learning algorithms. In natural language processing, Ren [9] proposed a Semantic Transformation Network for dialogue summarization using contrastive learning and attention mechanisms, and Ren [10] developed an enhanced graph convolutional network model (VGCN) for text classification. Ximeng and Yiming [11] applied offline conservative reinforcement learning to transaction authorization balancing fraud risk and customer friction, while Yang and Zhang [12] designed an edge-enabled real-time fraud detection system for network lending terminals under low-latency constraints. Zhao et al. [13] optimized deep learning models for

dynamic market behavior prediction, and Yang, Zheng, and Lu [14] constructed multi-dimensional credit-related transaction risk maps using graph neural networks. Shen et al. [15] researched the whale optimization algorithm for financial payment fraud detection. Yang [16] applied LightGBM in the Chinese stock market, and Tian, Wang, and Cui [17] proposed an improved U-Net for brain tumor image segmentation using GSConv and ECA attention mechanisms. Yang [18] further confirmed the utility of LightGBM in the Chinese stock market, and Deng and Yang [19] additionally developed multi-layer defense strategies and privacy-preserving enhancements for membership reasoning attacks in federated learning frameworks.

## 2. DATASETS

A garbage dataset consisting of 540 garbage images was constructed through crawling, downloading, and custom collection, which includes five types of garbage: banana peels, beverage bottles, watermelon peels, vegetable leaves, and shoes. For the convenience of future use, the tool x2coco is used here to create the COCO dataset. In the selection of images, for the types of fruits and vegetables, multiple angle images of the ripening and rotting stages of fruits and vegetables, as well as the compressed and torn images of fruits and vegetables were collected; For beverage bottles, pictures of different shapes of beverage bottles that have been squeezed were collected. This can eliminate the influence of some objective factors on the results, thus better detecting the target garbage.

## 3. FASTER R-CNN OBJECT DETECTION ALGORITHM

The Faster R-CNN algorithm is an upgraded version of the Fast R-CNN algorithm. As the name suggests, "Faster" means that the algorithm introduced in this section has a faster detection speed and is closer to "real-time" performance than the Fast R-CNN algorithm. The trained model can identify the location and category of garbage in the image, providing support for various industrial deployments.

### 3.1 Faster R-CNN has faster detection speed compared to Fast R-CNN

Fast renn uses SS algorithm for candidate box selection, but it can only be implemented on CPU, while Faster renn algorithm can use GPU, greatly improving computing speed. Although the two algorithms may have slight differences in other aspects, the difference in computational speed is the greatest.

### 3.2 Faster R-CNN Algorithm Process

The first step is to input the image to be detected into the CONV layer and extract the subsequent shared feature maps. RPN has two branches. One branch generates anchors, which are then pruned and filtered using a threshold. Finally, the activation function is used to determine whether the anchors are objects; Another route is to use bounding box regression to adjust the anchor, making it closer to the true position of the object. After RPN, candidate boxes were obtained. Inputting it into the RoI pooling layer along with feature mapping, its function is to standardize RoIs of different sizes, and finally perform a second classification and regression training.

The following section will discuss RPN anchor, Provide a detailed introduction to the loss function.

### 3.3 RPN

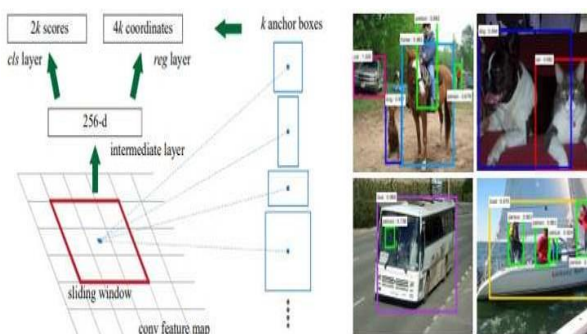
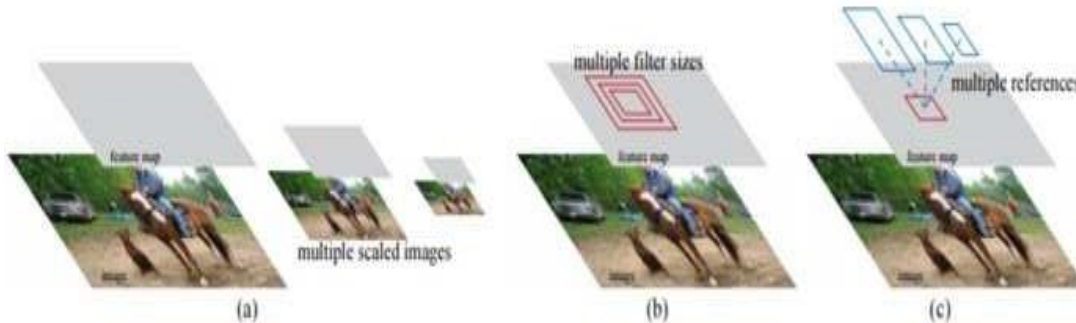


Figure 1: Regional Proposal Network RPN

The input of RPN is feature mapping, and the output contains 4 coordinates representing the offset of the real box and 2k scores (Figure 1). RPN is unique to this algorithm and may appear complex, but it is essentially a convolutional neural network. Simply put, it consists of an  $n * n$  convolutional layer (shown in Figure 1 as  $n=3$ ), as well as two parallel fully connected layers for classification and regression. Finally, the NMS algorithm is used to filter out proposals and complete the task of locating the object to be recognized.

Next, explain the  $3 * 3$  window sliding shown in Figure 2. During the sliding process, the center of this  $3 * 3$  always coincides with the center of the original image, which means that the area and position of the original image can be inferred from the proportion at various positions of the window. And the  $3 * 3$  window is obtained by pooling the original image through SSP, and the pooled part is the anchors mentioned earlier.



**Figure 2:** Multi scale prediction method

### 3.4 Anchors

Based on the center of the window, use 3 sizes and 3 ratios to obtain 9 anchors. So for an image with a width of  $W$  and a length of  $H$ , the final number of anchors is  $W * H * K$ . This setting improves the generalization ability on images of different sizes and aspect ratios, saves computation time on running image pyramids and filter pyramids in multiple dimensions, and also provides a greater experience than basic selection.

The algorithm needs to maintain the feature of invariant translation between the function (anchor and candidate box). When the target object undergoes translation or position movement, the corresponding candidate box should also move, and the same function can still be used to generate predicted candidate boxes at different positions. It is precisely because of this feature that the model is lighter and involves fewer parameters than previous methods [8], which also reduces the risk of overfitting on smaller datasets.

There are two commonly used methods for multi-scale prediction before the emergence of this anchor. As shown in Figure 2 (a), a method based on the "image pyramid". Scale the image and perform feature mapping calculations at different scales. This method is very useful, but every time scaling and calculation is performed, it consumes a lot of time, which leads to a relatively long algorithm time. Another method is the "filter pyramid" [8], as shown in Figure 2 (b). It can be seen that after processing the original image to obtain feature maps, a "filter pyramid" composed of three different sizes of filters was used in the image, achieving the goal of multi-scale prediction. However, these two methods are usually used in combination. Compared to the anchor here, it is both fast and accurate. It uses a feature map generated from a size image, and on this basis, multiple anchors are used to adjust the bounding box. This not only enables prediction at multiple scales, but also shares features, eliminating the need for repeated calculations and improving the speed of the algorithm.

During RPN training, anchors will receive two types of labels. In both cases of anchors, a label of 'being an object' (positive label) will be obtained. The first type is when the intersection to union ratio with the real bounding box is the highest, and the second type is when the intersection to union ratio with the real bounding box is higher than the set threshold (here set to 0.7 IoU) for anchors. When using it, the first allocation method is usually used because the second allocation method may result in the intersection to union ratio being less than the set threshold, that is, there is no positive label allocation, which is not a desirable situation.

### 3.5 Loss Function

The loss function consists of two parts: classification loss and regression loss. The first half is classification loss, and the second half is regression loss. Under these two types of losses, there are still two main network losses, which means that if subdivided, there are four parts of losses.

In RPN, anchors are divided into two categories, namely the binary classification of objects and backgrounds mentioned earlier. Therefore, the Lcls used here is binary cross entropy, and Ncls represents the total number of anchors. Sum up the cross entropy losses of each anchor and then average them. The number of categories in Fast rcnn is usually greater than 2, using multi class cross entropy loss.

In the regression loss of the latter half of the formula,  $t_i$  is a vector composed of four values of the predicted box coordinates.  $t_i^*$  has the same structure as  $t_i$ , except that it is a vector of the real box coordinates. These four quantities are the coordinates  $x$  and  $y$  of the center point, as well as the width  $w$  and height  $h$  of the box. The  $R$  in the regression loss  $L_{reg}$  [11] is the smooth L1 loss function. When a  $p_i^*$  is multiplied before the regression loss, with a value of 1, which is positive, the regression of the loss is performed; When it is 0, it is not considered.

For box regression, the predicted bounding boxes generally deviate from the actual object's box and need to be adjusted by translation and scaling. Here, we use the learning method from R-CNN to learn the parameters of position and size changes. Then apply these parameters to the box regression formula in Faster R-CNN to obtain new predicted bounding boxes. Due to the design of the anchor, adjustments can be made to predict bounding boxes of various sizes.

#### 4. EXPERIMENTAL RESULTS AND ANALYSIS ON GARBAGE CLASSIFICATION

This article uses the target detection library released by PaddlePaddle, which is an end-to-end target detection suite. The modular design and model compression capability provide convenience for subsequent algorithm deployment. The experiment was conducted on GPU computing in the Tesla V100 environment.

The experiment used Resnet101 as the backbone extraction network to build a Faster R-CNN model, and adopted transfer learning, which uses pre trained weights to initialize the network and accelerate the convergence speed of the model. The model underwent 50000 training iterations here, and the results are shown in Figure 5. It is evident from the graph that after 40000 iterations, the trained mAP remains stable at around 0.88.

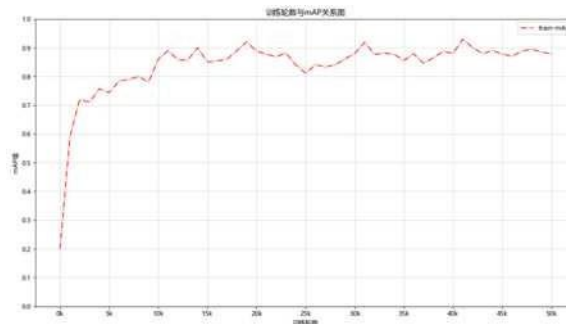
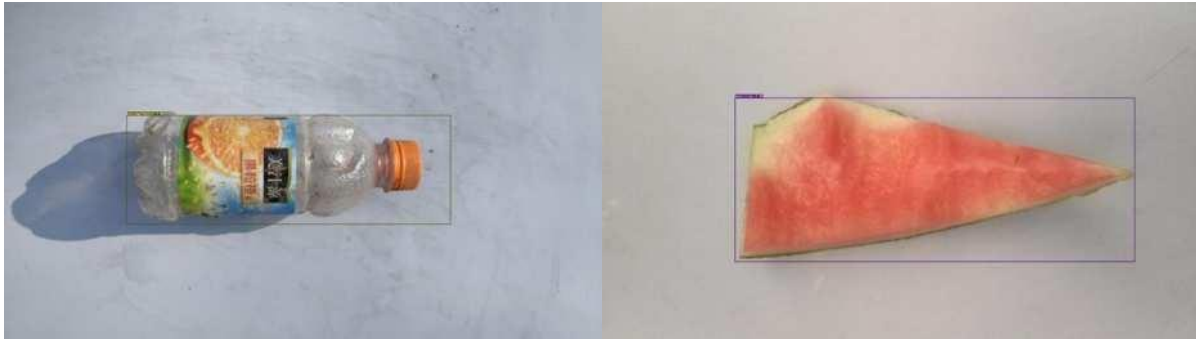


Figure 3: Training rounds and mAP

When detecting single garbage images, it can be seen from the detection results that the success rate is very high. Figure 6 shows the detection results of two types of garbage (watermelon peels and beverage bottles).



**Figure 6:** Detection performance of two types of garbage

## 5. CONCLUSION

This article proposes a method based on Faster R-CNN algorithm to recognize and classify several types of garbage. Although this method can achieve a certain high mAP value, it has also demonstrated certain effectiveness. However, there are still some areas for improvement in analysis and implementation, and further in-depth research will be conducted from multiple aspects. Due to device and time limitations, only a small number of datasets and garbage types were selected here, which also affected the final results. (2) There have been no major changes made to the structure proposed in the original paper for this algorithm. However, in further in-depth research, efforts can be made to select feature extraction networks and improve the loss function to further enhance the model's detection capability and achieve more accurate recognition of garbage, providing support for various industrial deployments.

## REFERENCES

- [1] Yi, X. (2025, October). Real-Time Fair-Exposure Ad Allocation for SMBs and Underserved Creators via Contextual Bandits-with-Knapsacks. In Proceedings of the 2025 2nd International Conference on Digital Economy and Computer Science (pp. 1602-1607).
- [2] Tang, Y., Kojima, K., Gotoda, M., Nishikawa, S., Hayashi, S., Koike-Akino, T., ... & Klamkin, J. (2020). Design and Optimization of Shallow-Angle Grating Coupler for Vertical Emission from Indium Phosphide Devices.
- [3] Deng, X., & Yang, J. (2025, August). Multi-Layer Defense Strategies and Privacy Preserving Enhancements for Membership Reasoning Attacks in a Federated Learning Framework. In 2025 5th International Conference on Computer Science and Blockchain (CCSB) (pp. 278-282). IEEE.
- [4] Wu, J., Luo, L., & Liao, N. (2025). Small-Sample Object Detection of Surface Cracks in Concrete Structures of High-Rise Buildings via Multi-Level Transfer Learning. *Innovation & Technology Advances*, 3(2), 57–72. <https://doi.org/10.61187/ita.v3i2.262>
- [5] Zhou, Z. (2026). Hierarchical Needs in US Automotive Customer Feedback and the Sentiment–Function Nexus. *Journal of Industrial Engineering and Applied Science*, 4(1), 27-33.
- [6] Zhou, Z. (2026). Bottleneck Diagnosis in International Automotive Sales Funnels Using Gradient Boosting Trees: Evidence from Cross-Regional Team Efficiency Evaluation. *Journal of Computer Technology and Applied Mathematics*, 3(1), 11-18.
- [7] Wensi, L. (2026). AI-Assisted Marketing Content Generation for Non-Standard Industrial Automation Solutions. *Journal of Economic Theory and Business Management*, 3(1), 18-25.
- [8] Li, W. (2026). Optimizing AI-Driven Bid Pricing Models for Non-Standard Automation Projects: Leveraging Historical Financial Data and Machine Learning Algorithms.
- [9] Ren, Z. (2024). Semantic Transformation Network: Improving Dialogue Summarization Through Contrastive Learning and Attention. *Journal of Theory and Practice in Engineering and Technology*, 1(3), 1–8. Retrieved from <https://woodyinternational.com/index.php/jtpet/article/view/59>
- [10] Ren, Z. (2024). VGCN: An Enhanced Graph Convolutional Network Model for Text Classification. *Journal of Industrial Engineering and Applied Science*, 2(4), 110–115. <https://doi.org/10.5281/zenodo.13118430>
- [11] Ximeng, Y., & Yiming, Z. (2026). Offline Conservative RL for Transaction Authorization: Smartly Balancing Fraud Risk and Customer Friction. *Journal of Economic Theory and Business Management*, 3(1), 1-9.
- [12] Yang, X., & Zhang, Y. (2026). Edge-Enabled Real-Time Fraud Detection for Network Lending Terminals under Low-Latency Constraints. *Journal of Computer Technology and Applied Mathematics*, 3(1), 55-62.

- [13] Zhao, S., Lin, Y., Yang, X., Lu, Q., Xue, H., & Jiang, G. (2025). Optimization of Deep Learning Models for Dynamic Market Behavior Prediction. arXiv preprint arXiv:2511.19090.
- [14] Yang, X., Zheng, X., & Lu, Q. (2025, October). Construction and early warning of multi-dimensional network credit-related transaction risk maps by integrating graph neural network (GNN). In Proceedings of the 2025 2nd International Conference on Digital Economy and Computer Science (pp. 919-923).
- [15] Shen, Zepeng, et al. "Research on Application of Whale Optimization Algorithm in Financial Payment Fraud Detection." 2025 4th International Conference on Artificial Intelligence, Internet and Digital Economy (ICAID). IEEE, 2025.
- [16] Yang, J. (2025, January). Application of lightgbm in the chinese stock market. In Proceedings of the 2025 4th International Conference on Big Data, Information and Computer Network (pp. 853-858).
- [17] Tian, Q., Wang, Z., & Cui, X. (2024). Improved Unet brain tumor image segmentation based on GSConv module and ECA attention mechanism. arXiv preprint arXiv:2409.13626.
- [18] Yang, J. (2025). Application of LightGBM in the Chinese Stock Market.
- [19] Deng, X., & Yang, J. (2025, August). Multi-Layer Defense Strategies and Privacy Preserving Enhancements for Membership Reasoning Attacks in a Federated Learning Framework. In 2025 5th International Conference on Computer Science and Blockchain (CCSB) (pp. 278-282). IEEE.