

A Campus Waste Sorting Terminal Based on Lightweight YOLOv5n - With Offline Voice Interaction and Edge Computing

Jing Zhao^{1*}

¹ Chuzhou Polytechnic, Chuzhou, China

*Correspondence: Jing Zhao

Received: 20 March 2026; **Accepted:** 25 April 2026; **Published:** 16 May 2026

Abstract:

To address the problem of domestic waste disposal in university campus scenarios, this study proposes an intelligent terminal featuring a closed-loop workflow of infrared wake-up, offline voice guidance, visual recognition, and category-specific lid opening. The visual recognition module is built on YOLOv5n. GhostConv is introduced into the backbone and neck networks to reduce model complexity, CBAM is embedded into the feature fusion layers to enhance the representation of small targets, and the bounding-box regression loss is replaced with SIoU to improve localization accuracy. Based on a self-constructed four-category campus waste detection dataset containing approximately 12,000 images, the improved model achieves an mAP@0.5 of 96.8%, which is 2.4 percentage points higher than that of the baseline model, while reducing the number of parameters by 28%. After TensorRT optimization on Jetson Nano, the inference speed reaches 26 FPS, and the average end-to-end response time is approximately 1.2 s. The success rate of offline voice interaction exceeds 95%. The proposed terminal enables real-time edge inference and interactive disposal guidance in campus environments.

Keywords:

Campus Waste Sorting, Lightweight Object Detection, YOLOv5n, CBAM, Offline Voice Interaction, Edge Computing

1. Introduction

With the advancement of carbon peaking and carbon neutrality initiatives and the continuous refinement of urban governance, domestic waste sorting has become an important component of public management and ecological civilization construction. University campuses are characterized by high population density, large volumes of waste generation, and diverse waste categories, making misclassification and improper disposal particularly prominent. Existing campus waste sorting practices still rely heavily on manual guidance and students' self-discipline, while traditional sorting bins remain inadequate in terms of interactive prompting, sanitation management, and real-time supervision. In recent years, deep-learning-based object detection has been widely applied to waste recognition, and studies on waste detection and classification based on the YOLO series have continued to increase [1,2,3]. However, the practical deployment of intelligent terminals in real campus disposal

scenarios still faces two major constraints. First, high-accuracy detection models often have large parameter sizes and high computational requirements, making real-time deployment on low-cost edge devices difficult. Second, most existing terminals merely perform the sequence of recognition and lid opening, lacking interactive guidance mechanisms such as voice prompts, which limits their effectiveness in reducing sorting errors and reinforcing correct disposal behavior.

To address these issues, this study designs and implements a campus waste disposal guidance terminal that integrates lightweight visual recognition and offline voice interaction. The terminal uses Jetson Nano as the edge-computing core for real-time on-device inference with an improved YOLOv5n model, while STM32 is employed for control execution in combination with an offline voice module, thereby forming a closed-loop process of perception, recognition, guidance, and execution. The main contributions of this study are as follows. First, a four-category waste detection dataset for campus scenarios is constructed. Second, GhostConv, CBAM, and SIOU are introduced into YOLOv5n to balance detection accuracy and lightweight deployment. Third, a prototype terminal is developed, and its edge-side inference speed, end-to-end response time, and interaction reliability are experimentally evaluated.

2. Campus Domestic Waste Detection Dataset

2.1. Category System and Data Collection

Considering typical campus environments such as teaching buildings, dormitories, and canteens, the detection targets are divided into four categories: recyclables (plastic bottles, cans, parcel cartons, etc.), food waste (fruit peels, leftovers, etc.), hazardous waste (used batteries, expired medicine boxes, etc.), and residual waste (tissues, snack packaging bags, etc.). The data sources include both filtered samples from public datasets and on-site photographs captured on campus. To improve model adaptability to different backgrounds and lighting conditions, the collected images cover multiple scenarios, including indoor and outdoor environments, strong illumination and shadows, floor textures, and reflective bin interiors. In total, approximately 4,500 valid images were collected.

2.2. Annotation Protocol and Data Augmentation

The images were annotated using the Labellmg tool. Each waste object was labeled using its minimum enclosing bounding rectangle, and multiple objects in the same image were annotated separately. To reduce overfitting and enrich samples in complex scenarios, data augmentation strategies including Mosaic augmentation, random cropping, horizontal flipping, and color jittering based on HSV transformation were applied to the training set. Ultimately, a dataset of approximately 12,000 images was constructed and divided into training, validation, and test sets at a ratio of 8:1:1. (Table 1)

Table 1. Overview of the campus waste detection dataset.

Item	Quantity (approx.)	Description	Remarks
Raw collected images	4,500	Filtered public datasets + on-campus photography	Covers diverse lighting and background conditions
Augmented images	12,000	Mosaic / cropping / flipping / HSV, etc	Used for training and evaluation
Data split	8:1:1	Training / validation / test	Proportional split

3. Improved YOLOv5n Object Detection Algorithm

YOLOv5n is the smallest model in the YOLOv5 family [8] in terms of depth and width, offering an inherent lightweight advantage suitable for embedded deployment. To further reduce computational cost while improving the detection performance of small waste objects in complex backgrounds, three modifications are introduced to the original YOLOv5n. GhostConv is used to replace part of the standard convolutions to reduce redundant feature computation. CBAM is embedded into the feature fusion layers to enhance the representation of key information. SIoU loss is adopted to optimize bounding-box regression, thereby improving localization accuracy and convergence efficiency.

3.1. *Lightweight GhostConv Module*

Traditional convolution operations tend to generate a large number of redundant feature maps during feature extraction, increasing both the number of parameters and floating-point operations. GhostNet proposes the idea of generating more features from cheap operations [5]. It first obtains intrinsic features through a small number of conventional convolutions and then uses linear transformations to generate additional ghost features, thereby reducing the computational burden. In the improved model, GhostConv replaces some standard convolution layers in the backbone and neck of YOLOv5n, which reduces the number of parameters without significantly sacrificing representational capability and reserves sufficient computational resources for real-time edge inference.

3.2. *Embedding of the CBAM Attention Mechanism*

In campus disposal scenarios, waste objects vary greatly in shape and are often accompanied by complex background interference, such as floor textures, reflections from bin interiors, and object occlusion or stacking. The Convolutional Block Attention Module (CBAM) performs sequential modeling of channel attention and spatial attention [6], enabling the network to adaptively focus on target-related feature regions while suppressing background noise. By introducing CBAM into the feature fusion stage, the discriminability of small targets, such as batteries and bottle caps, under complex backgrounds can be enhanced, thereby reducing the risk of missed detections.

3.3. *Optimization with the SIoU Loss Function*

YOLOv5 uses CIoU as the default bounding-box regression loss. However, its penalty term does not explicitly consider the directional consistency between the predicted box and the ground-truth box, which may lead to relatively slow convergence during the early stage of training. SIoU introduces additional constraints, such as angle cost, into the loss formulation [7], enabling the predicted box to align with the ground-truth box more rapidly along a more reasonable direction, thereby improving regression stability and localization accuracy. In the improved model, the bounding-box regression loss is replaced with SIoU to enhance localization performance under occlusion and overlap conditions.

4. Design of the Intelligent Sorting Terminal

4.1. *Overall Architecture*

The terminal consists of a perception layer, an edge-computing layer, a control execution layer, and an interaction layer, forming a closed-loop process from target perception and recognition decision-making to interactive prompting and mechanical actuation, as shown in Figure 1. The perception layer includes a USB camera, an infrared distance sensor, and an ultrasonic fill-level sensor. The edge-computing layer uses Jetson Nano as the core platform to run the improved YOLOv5n model. The control execution layer employs an STM32F103C8T6 microcontroller to drive servos for lid opening and closing. The interaction layer adopts the SU-03T offline speech recognition and voice broadcast module to realize voice wake-up and disposal guidance.

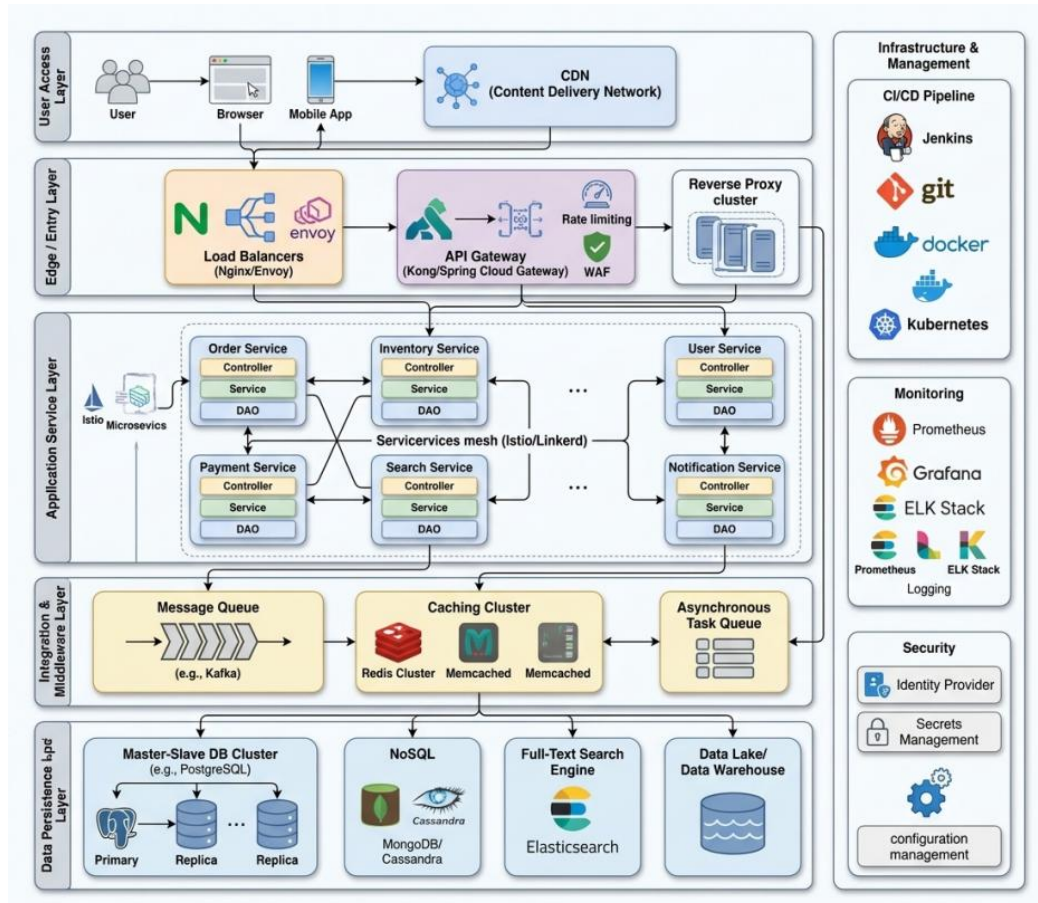


Figure 1. Overall architecture of the system.

3.2. Hardware Workflow and Interaction Strategy

The system operates as follows. When the infrared sensor detects that a user is approaching within a preset threshold distance, the terminal is awakened and plays the voice prompt, “Please place the waste in the recognition area.” The camera then captures an image and sends it to the Jetson Nano for inference. The recognized category is transmitted via serial communication to the STM32 controller and the voice module, and the voice module simultaneously broadcasts the recognition result. The STM32 then controls the lid of the corresponding waste bin to open and automatically close after a delay. When the ultrasonic sensor detects that a certain category bin is full, the system triggers a red-light alarm and plays the message, “This bin is full. Please do not dispose of waste here,” thereby preventing overflow. Offline speech recognition and voice broadcast are adopted so that the system can operate

stably without network connectivity while reducing privacy concerns and network dependence.

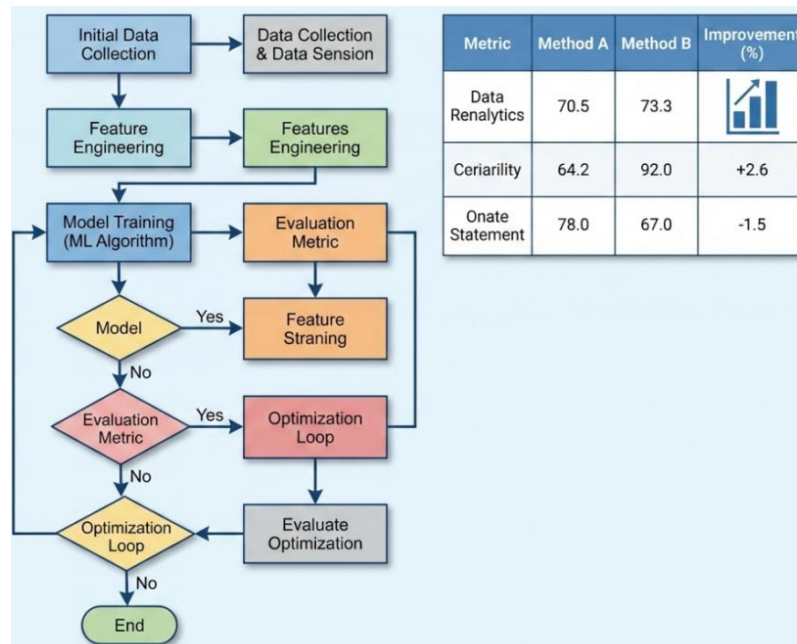


Figure 2. Software control flow.

5. Experimental Results and Analysis

5.1. Experimental Setup and Evaluation Metrics

Model training was conducted on a computer equipped with an Intel i7-12700H CPU and an NVIDIA RTX 3060 GPU, using the PyTorch deep-learning framework. Deployment and terminal-side testing were performed on a Jetson Nano 4 GB development board, and TensorRT was used for inference acceleration. The evaluation metrics include mean Average Precision at an IoU threshold of 0.5 (mAP@0.5), parameter count (Params), and edge-side inference speed in frames per second (FPS). Here, mAP@0.5 refers to the mean of the average precision values for all categories under an IoU threshold of 0.5.

5.2. Validation of the Improved Model

On the self-constructed campus waste dataset, the improved YOLOv5n model achieved an mAP@0.5 of 96.8%, which is 2.4 percentage points higher than that of the baseline YOLOv5n. Meanwhile, the number of parameters was reduced by 28%. The inference speed on Jetson Nano reached 26 FPS. Table 2 presents a comparison of the key performance indicators between the baseline model and the improved model.

Table 2. Comparison of key performance metrics between the baseline and improved models.

Model	mAP@0.5/%	Params (relative)	FPS on Jetson Nano	End-to-end response time / s
YOLOv5n (baseline)	94.4	1.00	—	—
Improved YOLOv5n (GhostConv + CBAM + SIoU)	96.8	0.72	26	1.2(average)

5.3. Comparison with Lightweight Models

To verify the performance of the improved model in lightweight detection tasks, it was compared with lightweight detectors such as YOLOv4-tiny and MobileNet-SSD. The results show that the improved model outperforms YOLOv4-tiny in terms of mAP@0.5 by approximately 4.1 percentage points. In addition, its inference speed is significantly higher than that of MobileNet-SSD, while maintaining real-time performance on Jetson Nano.

5.4. Prototype Testing of the Terminal

A prototype of the intelligent sorting terminal was fabricated and fully integrated for joint debugging. A total of 100 common campus waste items were randomly selected for real disposal tests, and the overall sorting success rate reached 95%. The average end-to-end response time was approximately 1.2 s, satisfying the requirements of interactive disposal scenarios. Under an ambient noise level of approximately 65 dB, the offline voice wake-up and broadcast instructions remained clear, and the interaction success rate exceeded 95%.

5.5. Sources of Error and Future Improvements

During testing, several typical sources of error were observed. First, reflective objects such as transparent plastic bottles caused confidence fluctuations because local textures were easily affected by specular reflection and background interference. Second, small objects, such as batteries and bottle caps, were prone to missed detection in scenarios involving object stacking and occlusion. Third, residual waste with packaging visually similar to recyclables could be confused under low-resolution input. Future work will focus on expanding fine-grained categories and hard samples, introducing more refined occlusion-aware data augmentation strategies, and incorporating multi-frame temporal information to further improve robustness. From an engineering perspective, the system can also be integrated with Internet of Things technologies for remote fill-level monitoring and sanitation scheduling, thereby improving the efficiency of campus logistics management.

6. Conclusion and Future Work

This study develops an intelligent waste-sorting terminal integrating lightweight visual recognition and offline voice interaction for campus waste disposal scenarios. By incorporating GhostConv, CBAM, and SIoU into YOLOv5n, the proposed model achieved an mAP@0.5 of 96.8% on a self-constructed dataset while reducing the number of parameters by 28% and maintaining a real-time inference speed of 26 FPS on Jetson Nano. Prototype experiments showed an overall sorting success rate of 95% and an average end-to-end response time of approximately 1.2 s, demonstrating its feasibility for practical campus deployment. Future work will expand the dataset with finer-grained categories, strengthen cross-scenario generalization evaluation, and integrate IoT and cloud platforms for fill-level monitoring, data statistics, and sanitation scheduling, thereby advancing the digital management of campus waste sorting.

Conflicts of Interest

The author declares that there is no conflict of interest regarding the publication of this article.

Funding

This article is Chuzhou Polytechnic ‘Research on Small Object Detection Method Based on Faster R-CNN and Temporal Multi-Feature Fusion Dominant Mechanism(Project No.ZKY-2025-14)’; ‘Teaching Practice Research on an Intelligent Attention Evaluation System for Programming Courses (Project No.2025jyxm06)’; ‘Research and Application of High-Quality Green Development for Prefabricated Rural Houses (Project No.QDJZ-2025-5)’ research findings.

References

- [1] Ning, K.; Zhang, D.; Yi, F, et al. Waste detection and classification for intelligent sweeping robots based on visual perception. *Journal of Image and Graphics*, 2019, 24(8), 1358-1368. DOI: 10.11834/jig.180475.
- [2] Ye, H.; Guo, L.; Li, Y, et al. Design of an intelligent domestic waste classification system based on an improved YOLOv5 algorithm. *Environmental Engineering*, 2025, 43(4), 232-241. DOI: 10.13205/j.hjgc.202504023.
- [3] Yuan, B.; Zhang, C.; Li, C. Design of an automatic waste classification system based on a lightweight MobileViT vision model. *Packaging Engineering*, 2023, 44(23), 208-215.
- [4] Duan, C.; Li, G. Simulation modeling of intelligent classification of construction waste in green building projects. *Computer Simulation*, 2022, 39(3), 234-237.
- [5] Han, K.; Wang, Y, Tian, Q, et al. GhostNet: More Features from Cheap Operations[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020, 1580-1589.
- [6] Woo, S.; Park, J.; Lee, J Y, et al. CBAM: Convolutional Block Attention Module[C]//Proceedings of the European Conference on Computer Vision. 2018, 3-19.
- [7] Gevorgyan, Z. SIOU Loss: More Powerful Learning for Bounding Box Regression. arXiv:2205.12740, 2022.
- [8] Jocher, G.; Chaurasia, A.; Qiu, J. YOLO by Ultralytics. Zenodo, 2020.



© 2026 by the author(s); licensee International Technology and Science Publications (ITS), this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (<http://creativecommons.org/licenses/by/4.0/>)