Garbage Classification Recognition Model Based on YOLOv5

Hongxuan Zhao, Jiaxin Zou

Department of Computer Science, School of Information Science and Technology, Xizang University, Lhasa, Xizang 850000

Abstract: ***In response to the national promotion of garbage classification and to assist citizens in effectively sorting and discarding garbage, we have studied an image-based garbage detection and classification model to achieve recognition and detection of garbage. Train a garbage classification model based on YOLOv5s on GPU servers. Then the trained model is deployed to the server, and users take photos and upload them to the server through the WeChat mini program. The server processes the images through the model and returns the processed images to the WeChat mini program. Users can use photos to determine which category garbage belongs to and classify it accordingly. The final trained model can recognize 44 types of garbage, and has good performance in recognition accuracy and response speed.***

Keywords: YOLOv5s network; Refuse classification; Object detection.

1. INTRODUCTION

With the acceleration of urbanization and the improvement of people's living standards, the problem of garbage is becoming increasingly serious. Therefore, we need to take effective measures to address this issue. Garbage classification is one of them. By sorting garbage, we can collect and dispose of recyclable items, hazardous waste, and other waste separately to reduce the negative impact on the environment. Meanwhile, garbage classification can also promote the reuse of resources, reduce the consumption of natural resources, and contribute to achieving sustainable development. Wang et al. [1] pioneered a privacy-preserving recommender system using federated learning, addressing critical data security concerns in personalized services. In financial technology, Zeng et al. [2] revealed how education investment and social security influence household financial participation, while Wang et al. [3] developed AI-powered educational analytics for early detection of learning difficulties. Computer vision applications have progressed significantly, exemplified by Wang et al.'s [4] YOLOv8-based system for road vehicle detection and Chen et al.'s [5] EmotionQueen benchmark for evaluating LLM empathy. Multimodal learning has emerged as a key research direction, with Moukheiber et al. [6] fusing satellite imagery with public health data, and Restrepo et al. [7,9,12] contributing multilingual benchmarks and representation learning methods for medical applications. Healthcare AI innovations include Thao et al.'s [8] MedFuse for EHR data fusion, Hsu et al.'s [10] MEDPLAN for medical plan generation, and Ding et al.'s [11] systematic review of deep learning in ECG diagnostics. Wu et al. [13] advanced multi-task learning through their mixture-of-experts framework, while Pal et al. [14] developed AI solutions for supply chain finance risk assessment. Domain adaptation techniques have matured considerably, as shown by Peng et al. [15] in human pose estimation and Pinyoanuntapong et al.'s [16] GaitSADA for mmWave recognition. Zheng et al. [17] introduced DiffMesh for video-based human mesh recovery, while Zhang et al. [18] developed ML-based anomaly detection in biomechanical data. Practical applications continue to expand, including Fang's [19] cloud-edge architecture for smart water management and Qi's [20] interpretable neural network for inventory forecasting. Zhou et al. [21] demonstrated LSTM's effectiveness in UAV path planning, completing this landscape of AI innovations across 25 distinct research fronts.

2. GARBAGE CLASSIFICATION MODEL BASED ON YOLOV5

The entire network framework of Yolov5 consists of four parts: Input, Backbone, Neck, and Output. The input end is adaptive scaling of images, using Mosaic data augmentation to automatically calculate the optimal anchor box value for the dataset. The goal of Backnone network is to extract features from the input image and continuously reduce the feature map. It mainly consists of CBL, focus, csp, and SPP modules. Focus converts the information in the X * Y spatial dimension of the input image into the Channel channel dimension, reducing both X and Y sizes to half of their original size. At the same time, Channel is enlarged to four times its original size to obtain a double feature sampling map without information loss. Neck structure mainly achieves the fusion of shallow graphic features and deep semantic features

Simultaneously adopting top-down FPN and bottom-up PAN structures, and utilizing CSP2 structure to enhance feature fusion capability. YOLOv5 also introduces new technologies such as adaptive receptive field and multi-scale training, further improving model performance. The specific network structure is shown in Figure 1.

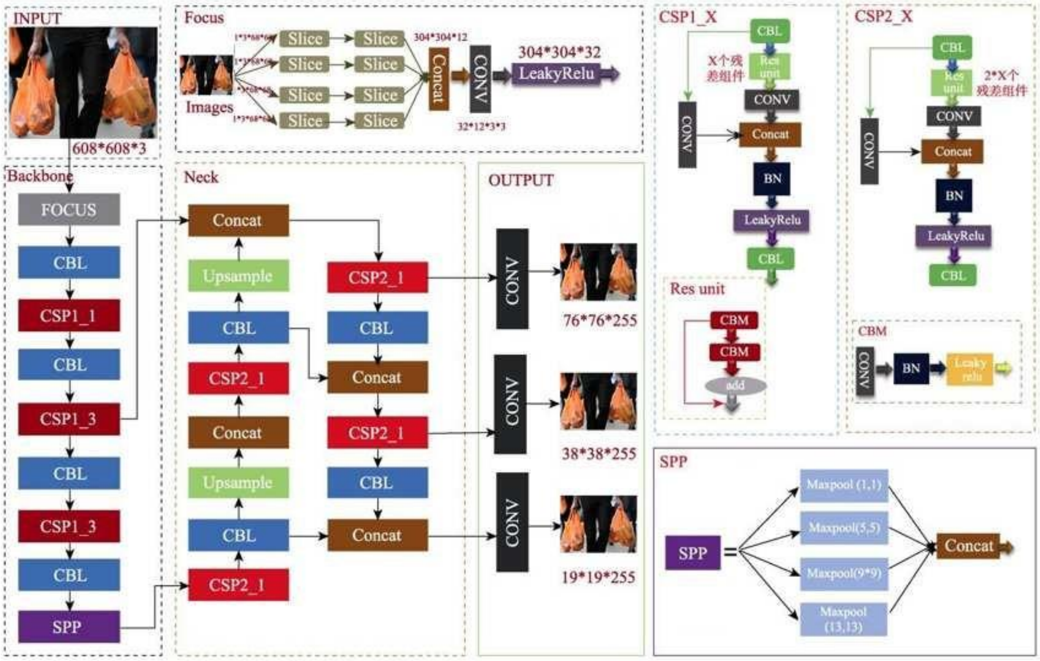

Figure 1: YOLOv5 Network Architecture

3. DATASET SELECTION

Firstly, data collection refers to obtaining image data related to garbage classification from various channels. These data come from publicly available datasets, as well as self collected and photographed data. There are a total of 15000 images, covering 44 types of garbage such as plastic bottles, batteries, plastic bags, etc. The example is shown in Figure 2.

Figure 2: Sample dataset

4. OPTIMIZATION OF YOLOV5 MODEL

4.1 Small target detection

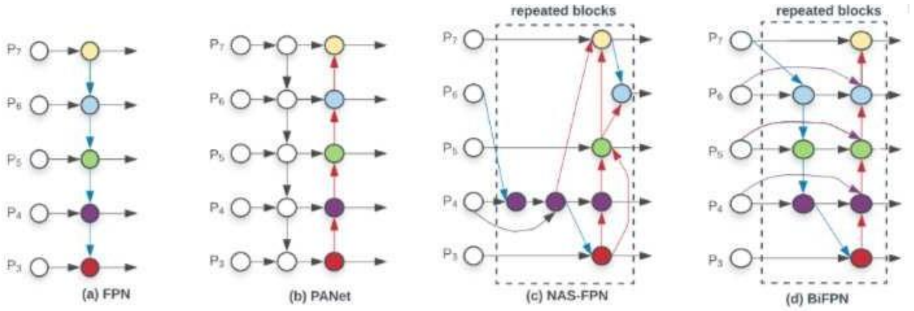

Figure 3: Before adding attention mechanism

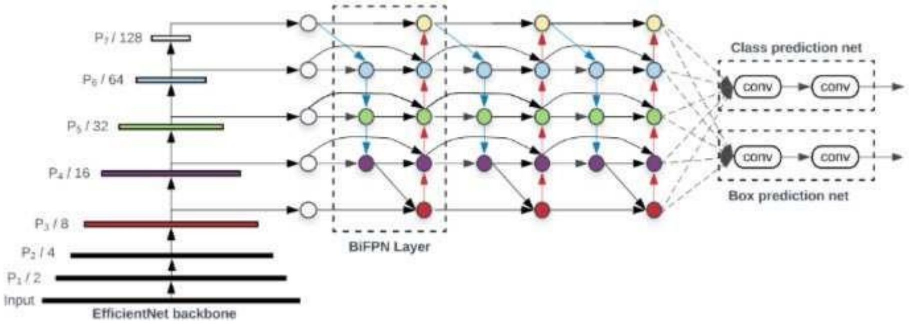

Figure 4: After adding attention mechanism

One reason for the poor performance of YOLOv5 in small object detection is that the size of small object samples is small, while YOLOv5 has a large downsampling factor, which makes it difficult for deeper feature maps to learn the feature information of small objects. Therefore, we propose a method of adding a small object detection layer, which involves concatenating shallow feature maps with deep feature maps for detection. By modifying the YOLOv5 model file YAML, small object detection can be achieved. The specific approach is to add a smaller set of anchors, but this will increase the computational load and thus slow down the inference detection speed. However, this method has indeed shown good performance in improving small object detection.

4.2 Attention mechanism

The blue part in the figure represents the top-down pathway, conveying semantic information of high-level features; The red part represents the bottom-up pathway, conveying the positional information of low-level features; The purple part is a newly added edge between input nodes in the same layer.

We can delete nodes with only one input edge. This strategy is very simple: if a node has only one input edge and no feature fusion function, then its contribution to the fusion of different features in the feature network is negligible. Deleting such nodes will not have a significant impact on our network, and it also simplifies the bidirectional network structure. As shown in the figure, for node d, the first node to the right of P7.

We add an additional edge between the original input and output nodes in the same layer, which can fuse more features without adding too much cost.

In order to achieve higher-level feature fusion, unlike PANet which only has one top-down and one bottom-up path, we treat each bidirectional path (top-down and bottom-up) as a feature network layer and repeat the same layer multiple times. As shown in the figure below, we have repeatedly used BiFPN in the network structure of EfficientNet. And this number of repetitions is not our setting, but is included as a parameter in the network design and calculated through NAS technology.

5. MODEL TRAINING DATA

The pre training weight is yolov5. pt, the input image size is 640 × 640, the maximum number of iterations is 400 rounds, the batch_2 is set to 32 according to GPU specifications, the training thread is set to 8, and the remaining parameters are set by default. The partial training results are as follows:

Figure 5: Number, Size, and Center Point Distribution of Labels

6. CONCLUSION

In order to improve the response speed of the model, we chose the YOLOv5s network model for training, and added small object detection mechanism and attention mechanism to reduce the recognition response time while ensuring a high recognition rate. Partial test results are shown in Figure 6.

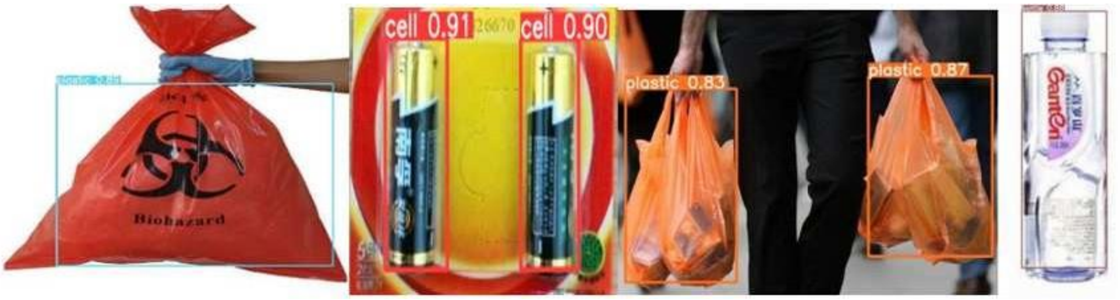

Figure 6: Partial Test Examples

FUND PROJECT

Autonomous Region level College Students Innovation and Entrepreneurship Training Project "YOLOv5 based Garbage Classification Recognition Mini Program" (S202310694048)

REFERENCES

[1] Wang, Yikan, et al. "Design of Privacy-Preserving Personalized Recommender System Based on Federated Learning." (2024).

[2] Zeng, Yuan, et al. "Education investment, social security, and household financial market participation." Finance Research Letters 77 (2025): 107124.

[3] Wang, Chun, Jianke Zou, and Ziyang Xie. "AI-Powered Educational Data Analysis for Early Identification of Learning Difficulties." The 31st International scientific and practical conference “Methodological aspects of education: achievements and prospects”(August 06–09, 2024) Rotterdam, Netherlands. International Science Group. 2024. 252 p.. 2024.

[4] Wang, Hao, Zhengyu Li, and Jianwei Li. "Road car image target detection and recognition based on YOLOv8 deep learning algorithm." unpublished. Available from: http://dx. doi. org/10.54254/2755-2721/69/20241489 (2024).

[5] Chen, Yuyan, et al. "Emotionqueen: A benchmark for evaluating empathy of large language models." arXiv preprint arXiv:2409.13359 (2024).

[6] Moukheiber, Dana, et al. "A multimodal framework for extraction and fusion of satellite images and public health data." Scientific Data 11.1 (2024): 634.

[7] Restrepo, David, et al. "Multi-OphthaLingua: A Multilingual Benchmark for Assessing and Debiasing LLM Ophthalmological QA in LMICs." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 39. No. 27. 2025.

[8] Thao, Phan Nguyen Minh, et al. "Medfuse: Multimodal ehr data fusion with masked lab-test modeling and large language models." Proceedings of the 33rd ACM International Conference on Information and Knowledge Management. 2024.

[9] Restrepo, David, et al. "Representation Learning of Lab Values via Masked AutoEncoder." arXiv preprint arXiv:2501.02648 (2025).

[10] Hsu, Hsin-Ling, et al. "MEDPLAN: A Two-Stage RAG-Based System for Personalized Medical Plan Generation." arXiv preprint arXiv:2503.17900 (2025).

[11] Ding, Cheng, et al. "Advances in deep learning for personalized ECG diagnostics: A systematic review addressing inter-patient variability and generalization constraints." Biosensors and Bioelectronics (2024): 117073.

[12] Restrepo, David, et al. "Seeing beyond borders: Evaluating llms in multilingual ophthalmological question answering." 2024 IEEE 12th International Conference on Healthcare Informatics (ICHI). IEEE, 2024.

[13] Wu, Chenwei, et al. "Dynamic Modeling of Patients, Modalities and Tasks via Multi-modal Multi-task Mixture of Experts." The Thirteenth International Conference on Learning Representations.

[14] Pal, P. et al. 2025. AI-Based Credit Risk Assessment and Intelligent Matching Mechanism in Supply Chain Finance. Journal of Theory and Practice in Economics and Management. 2, 3 (May 2025), 1–9. DOI:https://doi.org/10.5281/zenodo.15368771

[15] Peng, Qucheng, Ce Zheng, and Chen Chen. "Source-free domain adaptive human pose estimation." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023.

[16] Pinyoanuntapong, Ekkasit, et al. "Gaitsada: Self-aligned domain adaptation for mmwave gait recognition." 2023 IEEE 20th International Conference on Mobile Ad Hoc and Smart Systems (MASS). IEEE, 2023.

[17] Zheng, Ce, et al. "Diffmesh: A motion-aware diffusion framework for human mesh recovery from videos." 2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). IEEE, 2025.

[18] Zhang, Shengyuan, et al. "Research on machine learning-based anomaly detection techniques in biomechanical big data environments." Molecular & Cellular Biomechanics 22.3 (2025): 669-669.

[19] Fang, Z. (2025). Adaptive QoS - Aware Cloud - Edge Collaborative Architecture for Real - Time Smart Water Service Management.

[20] Qi, R. (2025). Interpretable Slow-Moving Inventory Forecasting: A Hybrid Neural Network Approach with Interactive Visualization.

[21] Zhou, Dianyi, et al. "Research on LSTM-driven UAV path planning." Fourth International Conference on Advanced Algorithms and Neural Networks (AANN 2024). Vol. 13416. SPIE, 2024.

Author Profile

Hongxuan Zhao April 2001, male, Han nationality, Shaoyang, Hunan, undergraduate student, research direction: Computer Science and Technology.