

Garbage Object Detection Based on Improved YOLOv5

Zhiren Xiao^{1, a}, Shangzhou Li^{1, b}, Zhenxian Guan^{1, c}, and Duojiarenqian^{1, d, *}

¹College of Information Science and Technology, Tibet University

^a xiaozhiren@utibet.edu.cn, ^b lishangzhou123456@foxmail.com, ^c 2522729417@qq.com,

^d 20754261@qq.com

Abstract. Traditional waste classification methods often rely on manual feature extraction and manual rule design, which cannot meet the needs of practical application scenarios. This thesis proposes a YOLOv5^[1] target detection method for garbage classification, aiming to solve the problem of automation and efficiency in garbage classification. First, we introduce semi-supervision to enhance the model's detection robustness as well as its extensiveness. Second, we incorporate Mixup, which fuses multiple different datasets to obtain new images with occlusions, which are trained in order to improve the robustness of the model for targets with complex backgrounds. In addition, we introduce a mosaic data enhancement algorithm to defocus the target and improve the accuracy of target detection. The proposed garbage target detection method based on the improved YOLOv5 has made significant progress in improving the detection performance, providing more effective tools and technical support for environmental protection and urban management. This method is expected to be widely popularized in practical applications.

1. Introduction

As the global population grows and urbanization accelerates, waste management has become an increasingly pressing issue. Traditionally, waste classification and disposal relies on manual labor, which is not only inefficient but also prone to errors and delays. Therefore, combining modern computer vision techniques with waste classification to improve classification accuracy and efficiency has become a very promising research direction.

Target detection techniques play a key role in the field of computer vision by recognizing different objects in an image or video and accurately marking their locations. Combining target detection with garbage sorting can realize automatic detection and sorting of garbage, thus improving the efficiency of garbage disposal and environmental protection awareness. By using deep learning models such as Convolutional Neural Networks^[2] (CNNs) and target detection algorithms such as YOLO^[3] (You Only Look Once) and Faster R-CNN^[4], we can achieve a fast and accurate waste classification system.

In this paper, in order to solve the problem of sparse and low accuracy dataset for garbage classification target detection, the study proposes a garbage target detection method based on improved YOLOv5 as well as a homemade dataset.

1. Introducing a semi-supervised algorithm to improve the robustness of the model to targets in complex backgrounds, so that it can be better adapted to the detection needs of waste classification in non-stop scenarios

2. Using Mixup and Mosaic data enhancement algorithms to increase the dataset to make the model more extensive.

2. YOLOv5

YOLOv5 is an advanced target detection model, the latest version developed from the You Only Look Once (YOLO) series. The model is based on a PyTorch implementation with a simple and efficient design designed to balance detection accuracy and inference speed. By combining CSP and Darknet architectures built with CSPDarknet53 as the backbone network, YOLOv5 is able to efficiently extract features and achieve multi-scale target detection with multiple detection heads. This enables the model to simultaneously detect targets of different sizes and demonstrate excellent

performance in different application scenarios. YOLOv5 has low model complexity and parameter count, which makes it suitable for efficient deployment on resource-limited devices.

A key feature is that YOLOv5 employs data enhancement techniques during the training and inference process, which increases the diversity of the training data and improves the generalization ability of the model through random scaling, random cropping and color dithering. In addition, during training, the model employs an adaptive strategy that automatically adjusts the learning rate and other hyperparameters to improve training efficiency and accelerate convergence. This design makes YOLOv5 a flexible and easy-to-use target detection model for a variety of real-world scenarios.

YOLOv5 has a wide range of applications in several fields. It has shown good performance in areas such as pedestrian detection, traffic sign recognition, industrial quality inspection and agricultural image analysis. Due to its efficient detection speed and accurate results, YOLOv5 has become one of the preferred models for many applications that require real-time target detection. In practical use, YOLOv5 demonstrates high stability and reliability, and is able to effectively cope with a variety of complex situations.

3. Semi-supervisory

The application of semi-supervised learning in the field of target detection has received increasing attention, which aims to improve the performance of target detection models by utilizing a small amount of labeled data and a large amount of unlabeled data. Compared to traditional fully supervised learning, semi-supervised learning can reduce the labor and time costs and achieve better detection results with scarce data. Following are some application aspects of semi-supervised learning in target detection.

3.1 Pseudo-labeling approach

A base model is trained by utilizing a supervised dataset and then the model is used to predict and generate pseudo-labels for unlabeled data. The pseudo-labeled data is then used along with the labeled data to train a target detector to improve the generalization ability and detection performance of the model.

3.2 Self-Training Method

This method gradually improves the performance of the target detector by repeated iterations of self-training. The initial model is first trained on labeled data, then the model is used to generate pseudo-labels by predicting the unlabeled dataset, then the pseudo-labeled data is merged with the labeled data for the next round of training, and so on until the model converges.

3.3 Migration Learning

Using a model pre-trained on large-scale data, a small amount of labeled data for the target task is fine-tuned for the target detection task. At the same time, the model is tuned in conjunction with unlabeled data to improve the generalization ability and detection of the type.

3.4 Joint training method

Combine the target detection task with other related tasks, such as image segmentation, image generation, etc., to take advantage of the correlation between different tasks and improve the performance of target detection by sharing features. This approach can effectively reduce the amount of data to be labeled and improve the generalization ability of the model.

4. Mixup[5]

Mixup is a data augmentation technique designed to improve the generalization ability and

robustness of deep learning models. The basic idea of Mixup is to mix different samples in a certain proportion in the input data to generate new "mixed samples" and train them. This method can effectively reduce the overfitting of the model on the training data and improve its generalization ability to unseen data.

4.1 Generate hybrid samples

Linear interpolation of two input samples, i.e., adding the features and labels of the two samples by a certain ratio to obtain new hybrid features and hybrid labels. For example, if the features of sample A and sample B are x_1 and x_2 , and the labels are y_1 and y_2 , then the mixture sample feature is $\lambda x_1 + (1-\lambda)x_2$, and the mixture sample label is $\lambda y_1 + (1-\lambda)y_2$, where λ is the mixing ratio.

4.2 Model Training

The model is trained using mixed samples. During the training process, the optimizer will calculate the loss and update the weights of the model based on the features and labels of the mixture samples. The model is optimized by backpropagation algorithm so that it can better fit the mixed samples.



Figure 1:Original Image



Figure 2:Mixup randomly generates images twice

5. Mosaic

Data Enhancement Technique A mosaic is an image processing method that stitches multiple images together to form a new image. Mosaic technology is mainly used for data enhancement, generating new images by changing the position, size, and rotation angle of the image, increasing the diversity and number of datasets, and improving the model's generalization ability and robustness. Mosaic technology, which is suitable for tasks such as target detection, semantic segmentation, image classification, and others, can effectively improve the performance of the model.



Figure 3: Mosaic represents a new method of data augmentation.

6. Experiments

6.1 Introduction of Dataset and Experimental Environment

In this paper, we establish a waste classification dataset consisting of 15,000 pictures with 44 categories, including three major categories of recyclable waste, food waste, and hazardous waste, each of which consists of 5,000 photographs respectively

This experiment was conducted on the Ubuntu 16.04 LTS operating system environment, with an Intel(R) Core(TM) i5-11500F@2.50GHz CPU, 16.0GB RAM, CUDA version 10.1. Training was performed using a GeForce NVIDIA GTX 3090 GPU, and the program was written and executed in Python 3.8 language under the PyTorch deep learning framework.

6.2 Comparative Experiments

For the selection of the YOLOv5 object detection model, this paper adopts a comparative experimental approach to contrast various object detection models. Multiple experiments were conducted on the dataset of this paper, and the experimental results are shown in the table below.

Table 1: Comparison of experimental results

model	P	R	F1	mAP0.5	FPS
Faster R-CNN ^[1]	61.35%	55.26%	52.38%	55.51%	25.3
SSD ^[6]	82.34%	56.01%	59.05%	54.45%	119.8
YOLOv3 ^[7]	64.62%	62.55%	53.05%	62.56%	63.5
YOLOv5 ^[4]	66.13%	64.24%	60.62%	65.90%	60.3

From the above table, it is evident that YOLOv5 has a significant advantage in most metrics. For garbage detection, where higher accuracy and FPS are desirable, although YOLOv5's FPS is slightly lower than YOLOv3, it compensates for this disadvantage in terms of accuracy. In conclusion, YOLOv5 is the most suitable model for this experiment.

6.3 Ablation Studies

Table 2: Results of ablation experiments

Mosaic	Mixup	Semi-supervisory	P	R	F1	mAP0.5
--------	-------	------------------	---	---	----	--------

			0.66	0.64	0.61	0.66
√			0.65	0.65	0.60	0.67
	√		0.69	0.67	0.71	0.69
		√	0.68	0.66	0.70	0.69
√	√	√	0.71	0.69	0.73	0.72

In this paper, after introducing semi-supervision, a series of experiments are conducted on the improved YOLOv5 to evaluate its performance. The experimental results show that the introduction of semi-supervision significantly improves the detection accuracy on different datasets, especially when dealing with complex scenarios and multi-class objects.

Experiments were conducted on the improved YOLOv5 model with the addition of Mixup and Mosaic. The results show that both improve the model to different degrees in terms of precision, accuracy, and check completeness..

7. Conclusion

In this study, through in-depth exploration and experimental validation of the improved YOLOv5-based target detection garbage categorization method, we have achieved remarkable results. First, the improved strategy of introducing semi-supervision significantly improves the model's ability to detect small targets, making it perform very well in dealing with small litter objects in urban environments. Second, the introduction of Mixup as well as Mosaic data augmentation makes the model more robust in complex contexts, enhances its adaptability to various scenarios, and provides more reliable support for urban management in real-world application scenarios. This paper achieves a significant improvement in detection accuracy and provides a more reliable method to meet the accuracy requirements in the field of target detection waste classification.

Funding

Semi-supervised Learning YOLOV5 Tibetan-Chinese Bilingual System for Recyclable Waste Recognition in Plateau Tourism 2023XCX056.

References

- [1] Z. Rongyun, Z. Kunming, S. Peicheng, M. Ye, L. Haoran, Q. Tian, Traffic Sign Detection Based on the Improved YOLOv5, Applied Sciences, DOI 10.3390/app13179748(2023).
- [2] Chua L O. CNN: A paradigm for complexity[M]. World Scientific, 1998.
- [3] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 779-788.
- [4] Girshick R. Fast r-cnn[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1440-1448.
- [5] L. Guangping, G. Xiang, L. Chenghui, L. Ding kai, L. Bingo Wing-Kuen, Cosine Mixup: A Mixup Strategy for Point Cloud Contrastive Representation Learning, IEEE Transactions on Circuits and Systems II: Express Briefs, DOI 10.1109/tcsii.2023.3310192(2023).
- [6] Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector[C]//Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14. Springer International Publishing, 2016: 21-37.

- [7] Redmon J, Farhadi A. YOLOv3: An incremental improvement[J]. arXiv preprint arXiv:1804.02767, 2018.