
COSE474-2021F: Final Project Proposal

Trash Classification for Recycling with Object Detection

Yusung Min

Abstract

The goal of this project is to create a garbage classification model for recycling separation using object detection technology, and eventually to detect garbage in photos taken through smartphone cameras and tell them how the garbage should be classified. I got the dataset which has six classes from roboflow. For object detection model, and I used YOLOv5 Model for object detection.

1. Introduction

Currently, garbage emissions are increasing in our society. The government is trying several methods to encourage recycling and improving the recycling system. However, it is not easy to solve because people have to classify garbage themselves. The biggest reason for not properly separating trash is that they do not know how to do well. Therefore, We are going to try to develop a trash classification model. For developing model, object detection is necessarily required to distinguish garbage in image. After that, image classification should be learned to distribute which material these detected objects are.

2. Problem definition & challenges

For sustainable society, recycling trash is inevitable for human being. According to Ministry of Environment of Republic of Korea, each Korean citizen produces 55 tons of garbage over his lifetime. Although there are only 5 years left until current landfill is available, in capital area, it is hard to find new site for creating landfill. Also garbage made of materials such as plastic do not rot easily and generate environmental hormones during incineration. For solving those problems, separating trash for recycling is necessary.

3. Related Works

Object Detection. Object detection is still one of the fields actively studied. Since it is a difficult task to satisfy two

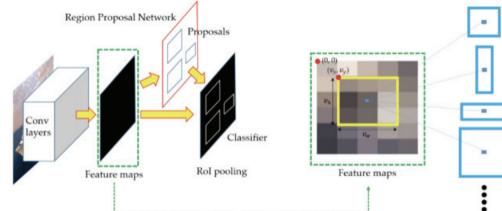


Figure 1. Structure of Faster-RCNN

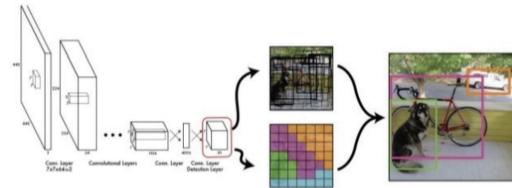


Figure 2. Structure of YOLO

tasks, object localization and classification, various models have been proposed so far. For example, various models have been developed, ranging from Faster-RCNN(Ren et al., 2015), which built a two-stage detection pipeline using Region Proposal Network (RPN) and pre-trained convolution network in Fig. 1, to YOLO(Zhu et al., 2021), which is based on single-stage model. Among these detection models, we plan to construct a garbage object detection model by fine-tuning YOLOv5(Zhu et al., 2021), a one-stage model using CNN backbone in Fig. 2.

4. Methods

4.1. Main challenges

In order to create a model that can help recycle, it is necessary to match the class label of what kind it is, beyond simply being able to determine whether it is trash or not. In addition, for using this model in practice, even if there are multiple objects in one picture, each class must be matched, also test speed must be fast. In addition, AI Model API Server is necessary to interact with the model and users. To solve the above problem, the YOLO model was selected.

Although I considered between YOLO and Faster-Rcnn, YOLO operates faster as a single stage and does not fall far behind in terms of performance, compared to Faster-Rcnn which is a multi-stage-based model.([Kim et al., 2020](#)) Then I used Flask uses python to create API Server because model is created by python. Finally, I used React Native to create simple application to show the model works.

4.2. Architecture

$$GIoU(A, B) = IoU(A, B) - \frac{|C| - |A \cup B|}{|C|} \quad (1)$$

$$Loss_{GIoU} = 1 - GIoU(A, B) \quad (2)$$

In this project, YOLOv5s Model was selected where C is the smallest box which is containing A and B. GIoU overcomes the distributed situation of A and B in IoU. Furthermore it has the same scale-invariance as IoU. GIoU is less than or equal to IoU, so it can be considered the lower bound of IoU. GIoU not only pays attention to the overlapping area but also to other non-overlapping areas, which can be better to reflect the degree of overlap between of them. When the two boxes have no overlapping area, GIoU is a number that varies in the range of [1, 0], and there is a gradient, so optimization can be performed. The optimization direction gradually narrows the distance between the two boxes.([Wang et al., 2022](#))

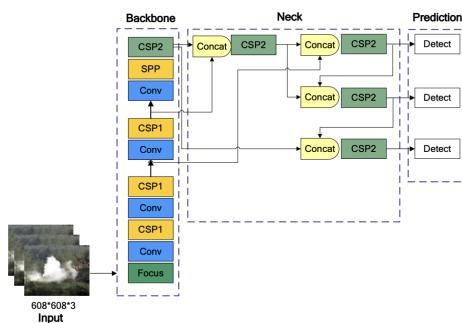


Figure 3. YOLOv5 Network Structure

At Fig. 3, it describes Network Structure of YOLOv5 model. The network consists of three main parts: backbone, neck, and output. Backbone part focuses on extracting feature information from input images, neck part fuses the extracted feature information and generates three scales of feature maps, and the output part detects the objects from these generated feature maps.([Zhu et al., 2021](#))

5. Experiments

5.1. Project Environment

For the smooth progress of the project, the project was carried out using the cloud-based collaboration service provided by Google.

5.2. Datasets

The data acquisition process is simple. We found a dataset includes garbage images from Roboflow.([Identification](#)) The dataset contains images of 7 types of trash, types are cardboard, glass, metal, paper, plastic, cloth and biodegradable. It contains total 10,464 images. However, we felt the need to increase the number of images to improve the performance of the model. Therefore, about 20% of images were added through data augmentation. For proper work for data augmentation, as shown in Fig. 4, we reversed the random training images from left to right. In addition, it is also necessary to do operation to reverse boundary box's coordinate values. Furthermore, we tried to improve the performance of the model by resizing the image to 320x320. Finally, 70% of images are used for training, 20% for validation and 10% are used for testing.

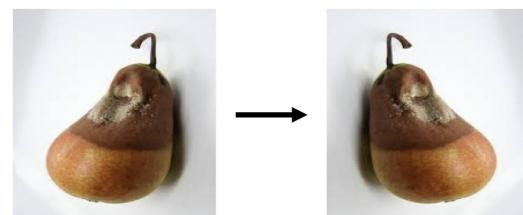


Figure 4. Data Augmentation

5.3. Implementations

The first step in the process of this project is preprocessing data. For this, we resized images 320 x 320 size and augmented data, and also divided data for training, validation, and test. Next, we constructed a model. We used YOLOv5s Model and trained model with above train dataset. After that, we got best fit model and validated with validation dataset. At Fig. 5, it shows about the model's performance of YOLO. Mean average precision called MAP is used to evaluate model. If it is low, it means that the performance is not good, and if it is high, it means that it is good. When confidence is adjusted to 0.5, MAP is measured high. And the value was improved by 0.5 compared to when the data was not augmented.

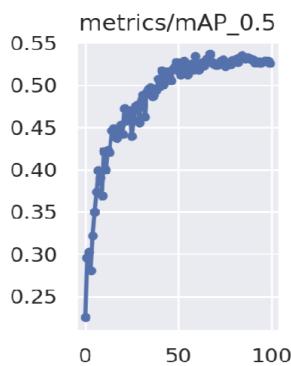


Figure 5. MAP of Model

5.4. Compared with SOTA

DETReg which is about Class-agnostic Object Detection with Multi-modal Transformer has 84.16% of MAP.(Maaz et al., 2022) By contrast, my model has only about 53.19% of MAP. But this model is faster than DETReg to object detect in real time, so our model is more appropriate for our project.

5.5. Application

Our project's final goal is to use this model in practice, so we developed simple client with React Native. And then, we construct API server for our model by Flask. Fig. 6 is result of our project. We actually took a photo a bottle of yogurt made of plastic and a torn box of snack. This model also detected real objects well. The torn box was recognized as two objects, but in fact, it is almost half torn, so it is not judged that there is a performance problem.



Figure 6. Result of applying to device

6. Discussion & Future Direction

This project was quite successful, but not perfect. The advantage of YOLO model is that object detection can be performed with real time, but it could not be performed with real time by loading the model through the API server, not putting it directly into the device. Therefore, in order to take advantage of real time, it seems necessary to put it directly into the device using a smaller model. Another problem is that it does not have a certain probability for all classes and detects only specific classes well. However, since there is not much difference, this will need to be addressed through data supplementation.

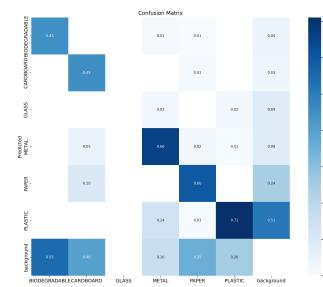


Figure 7. Confusion Matrix

References

- Identification, M. Garbage classification 3 dataset.
- Kim, J.-a., Sung, J.-Y., and Park, S.-h. Comparison of faster-rcnn, yolo, and ssd for real-time vehicle type recognition. In *2020 IEEE international conference on consumer electronics-Asia (ICCE-Asia)*, pp. 1–4. IEEE, 2020.
- Maaz, M., Rasheed, H., Khan, S., Khan, F. S., Anwer, R. M., and Yang, M.-H. Class-agnostic object detection with multi-modal transformer. In *European Conference on Computer Vision*, pp. 512–531. Springer, 2022.
- Ren, S., He, K., Girshick, R. B., and Sun, J. Faster R-CNN: towards real-time object detection with region proposal networks. *CoRR*, abs/1506.01497, 2015.
- Wang, Z., Wu, L., Li, T., and Shi, P. A Smoke Detection Model Based on Improved YOLOv5. *Mathematics*, 10 (7):1–13, April 2022.
- Zhu, L., Geng, X., Li, Z., and Liu, C. Improving yolov5 with attention mechanism for detecting boulders from planetary images. *Remote Sensing*, 13(18):3776, 2021.

Contents

Github Address

https://github.com/alsdbtjd0103/COSE474_Final_Project.git