Large-scale High-resolution Trash Dataset for trash detection

Min-Seok Jo[1], Seong-Soo Han[2], Chang-Sung Jeong[3],

[1] Dept. of Electrical and Engineering Korea University, Seoul, Republic of Korea
[2] Dept. of Division of Liberal Studies, Kangwon National University, Samcheok-Si, Republic of Korea
[3] Dept. of Electrical and Engineering Korea University, Seoul, Republic of Korea
{jms0923}@korea.ac.kr
{sshan1}@kangwon.ac.kr
{csjeong}@korea.ac.kr

Abstract. We produce large-scale high-resolution trash datasets for solving trash-detection problems. It was intended for general household goods. And dataset was classified based on guidelines for separating and discharging recyclables from the Ministry of Environment. There are 10 categories, and 33,434 objects were annotated for 4977 images with 1280 x 720 resolution.

Keywords: Dataset, Deep learning, Trash detection, Trash

1 Introduction

In the aftermath of Corona 19, the use of online shopping and delivery food is increasing rapidly, and disposable waste emissions are increasing rapidly. If the separation is not properly collected, it cannot be collected or recycled again at the staging area and is incinerated or reclaimed. This results in large incidental costs, accompanied by environmental degradation.

Therefore, research is emerging to automate separation collection using artificial intelligence or to determine separation collection items instead. In order to automate, the problem of trash recognition, which determines the location, item and material is essential and an important issue underlying automation. All studies using artificial intelligence are possible only with sufficient datasets. However, when it comes to the problem of trash recognition, there is still a lack of quality datasets.

So, for the purpose of creating a trash dataset that can be used for object detection model training using deep learning, a large-capacity, high-resolution trash dataset was produced to solve the object detection problem. It was intended for general household goods. And dataset was classified based on guidelines for separating and discharging recyclables from the Ministry of Environment. There are 10 categories, and 33,434 objects were annotated for 4977 images with 1280 x 720 resolution.

2 Related Work

Datasets are a very important factor in deep learning techniques. It is difficult to determine the optimal category, resolution, and size for the purpose. The results depend on how the dataset is constructed through these values. In this section, we will look at the dataset studies related to waste detection.

2.1 Trash Datasets

Existing Trash Dataset-related studies mostly focus on the Classification problem and consist of specifying one category on one image. Furthermore, its size is small and limited, which is not sufficient for learning deep learning models. There are also no unified standards in determining categories, so each is different. Therefore, sufficient-scale datasets are needed for Object detection models that can detect and classify recyclables. We show detailed information about existing datasets in Table 1.

Table 1. A summary of trash datasets.

Paper	Categories	Images	Annotations	Format	Resolution
[1]	3	2,000		Classification	256 x 256
[2]	4	5,000		Classification	240 x 240
[3]	240	17,690		Classification	320 x 240
[4]	9	681		Bounding Box	420 x 400
[5]	2	450		Bounding Box	256 x 256
[6]	25	469	4,338	Bounding Box	640 x 480
Ours	10	4,977	33,434	Bounding Box	1280 x 720

In the case of [3], the dataset for classification was collected and classified through crawling based on Shanghai's policy. In [4] and [6], a dataset for object detection was constructed and experimented with garbage on city streets. In [5], garbage images in the city were collected through crawling and only the coordinates were annotated. In addition, it is possible to share geo-tags as well as detect trash through a smartphone.

3 Dataset Description

In this section, we will discuss about datasets we have constructed.

3.1 Categories

The purpose is to detect and classify recyclable household items, so the material was set as an important criterion. So, based on the material, the categories of Paper, Can, Bottle, Pet, Plastic, and Vinyl were set. In the case of Paper, paper corresponding to general books such as books and magazines, boxes for packaging, and paper cups have very different appearances, so it is said that classification of categories can learn more quality features due to the nature of CNN based on computer vision. It was judged and divided into Paper, Paper Pack, and Paper Cup. In addition, it was determined that it is a different material, but it is separate waste collection was the least, so labels and caps were added to the category.

In the case of label like Nutrition Facts, the material is mostly made of paper and vinyl. Paper labels are generally used for glass bottles and cans, and vinyl labels are often used for PET bottles. However, it was considered that two categories could not be accurately judged by only visual judgment used in deep learning without classification mark information, and thus, it was unified into one label category.

In the case of caps, there are lids made of cans, which are often used for glass bottles, and lids made of plastic, which are often used for PET bottles. However, the cap occupied less area in the entire image compared to other categories, so it was considered that it could not be visually judged like the label, so it was composed of one cap. Finally, 10 categories were composed of Paper, Paper Pack, Paper Cup, Can, Bottle, Pet, Plastic, Vinyl, Cap, and Label.

3.2 Setup

On average, there are three or more images in a single image, with a single object or a large number of objects randomly. An object can be made up of several substances. Objects placed on the floor were photographed at a height of about 160cm. The target object was kept in its original shape or arbitrarily transformed, such as crumpling or tearing. The angle between the camera and the ground was maintained at around 90 degrees, and the direction and position of the object looking at the camera were randomly changed. The brightness was controlled in two stages using lighting.

3.3 Annotations

Like other object-detection related studies, our datasets must also exist in a form that can be used in deep learning models. For a single image, there should be data on the bounding box, the coordinates of the target object you want to learn, and what category the object is located in that area. An example of annotated data is shown in Fig. 2.

3.4 Preprocessing

Most of the images were composed of various resolutions of 2560 x 1440 (QHD) or higher, such as 2024 x 2024 and 3024 x 4032. In order to create a dataset optimized for deep learning, the image was resized collectively to a resolution of 1280 x 720 (HD), which satisfies the large input size of the latest deep learning models and does not break the image significantly. Secondly, it was unified in a 3-channel JPG format to be used as an input for a deep learning model. Third, by randomly performing Augmentation during training, it was possible to learn about various data while maintaining the total amount of images.

3.4 Statistics

A total of 4977 images were composed of 33,434 Annotated data. As can be seen in Table 1, it is the largest among the datasets that can be used for object detection. The resolution is also the largest and has enough size to be applied to deep learning. The number of Annotated data for each category is shown in Fig. 1. The largest number of caps is 6,921, and the lowest number of paper packs is 1,745. There is an average of 6.71 annotated data per image.

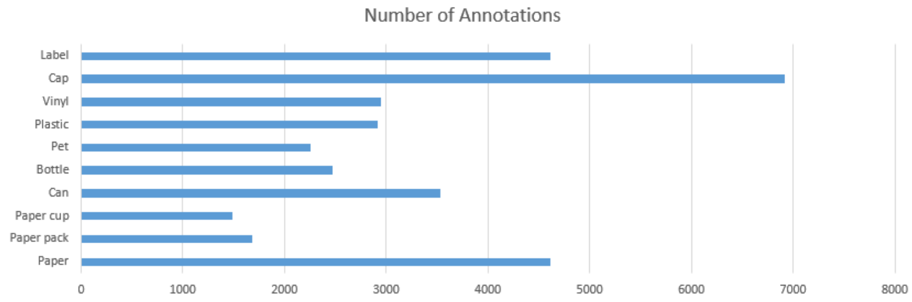

Fig. 1. The number of Annotations per category in our dataset. The number of annotated data in paper cups is the lowest at 1491, and the number of annotated data in caps is the highest at 6,921.

4 Experiments

We conducted an experiment to find out the performance of trash detection using our dataset. Three models of SSD, YOLO v3, and Cascade R-CNN with different backbones were used. The SSD used VGG-16 as the backbone, and YOLO v3 used Darknet-53 and the Cascade R-CNN used Resnet-50 as the backbone.

4.1 Implementation

Of a total of 4,977 images, 4,577 were randomly divided into training sets and 400 into validation sets. Each model was trained for 300 epochs using three v100s, and the epoch with the best performance for the validation set was used in Fig. 2. Only one v100 was used for inference. Average were used as evaluation metrics.

The dataset was resized to the appropriate input size for each model and padded appropriately. Each model used transfer learning, which trained in advance on the COCO[7] dataset and learned on our dataset. During learning, several augmentation techniques were applied. Horizontal flip, vertical flip, Hue, Saturation, Contrast, Brightness, and center crop were applied. In addition, hyper parameters such as running rate and optimizer used the settings of the original paper as they were.

4.2 Results

Table 2 allows us to compare AP performance of models for each category. By category, Can had the highest average of 0.578 and the lowest cap at 0.316 on average. Can was composed of good data to extract features from the model's perspective, even though the number of annotated data was average, and Cap had the largest number of annotated data, but had a smaller footprint and ambiguous criteria for extracting features from the model's perspective.

Table 2. AP performance by Category of each model.

Model	SSD	YOLO v3	Cascade RCNN	Average
Paper	0.314	0.536	0.33	0.393
Paper pack	0.335	0.502	0.387	0.408
Paper cup	0.282	0.569	0.327	0.393
Can	0.44	0.595	0.698	0.578
Bottle	0.392	0.488	0.604	0.495
Pet	0.337	0.366	0.327	0.343
Plastic	0.294	0.314	0.436	0.348
Vinyl	0.289	0.405	0.404	0.366
Cap	0.26	0.285	0.404	0.316
Label	0.164	0.395	0.406	0.322

5 Conclusions

We collected images of general household waste. A dataset was created by annotating the images so that they can be used for deep learning. For the purpose of separate waste collection, categories were created and classified based on materials. We implemented trash detection using three different deep learning models, SSD, YOLO v3, and Cascade RCNN through our dataset. By category, Can has the best detection, and Cap has the most difficult to detect. We plan to release our dataset for further study and further study.

Acknowledgments. This work was supported by Institute for Information & communications Technology Promotion(IITP) grant funded by the Korea government(MSIP) (No.2020-0-02219,Development of technology for irregularly shaped waste classification based on deep learning)

References

1. Sakr, G. E., Mokbel, M., Darwich, A., Khneisser, M. N., & Hadi, A. (2016). Comparing deep learning and support vector machines for autonomous waste sorting. 2016 IEEE

International Multidisciplinary Conference on Engineering Technology, IMCET 2016, 207–212.
2. Chu, Y., Huang, C., Xie, X., Tan, B., Kamal, S., & Xiong, X. (2018). Multilayer hybrid deep-learning method for waste classification and recycling. Computational Intelligence and Neuroscience, 2018.
3. Liao, Y. (2020). A Web-Based Dataset for Garbage Classification Based on Shanghai's Rule. International Journal of Machine Learning and Computing, 10(4), 599–604.
4. Zhang, P., Zhao, Q., Gao, J., Li, W., & Lu, J. (2019). Urban Street Cleanliness Assessment Using Mobile Edge Computing and Deep Learning. IEEE Access, 7, 63550–63563.
5. Mittal, G., Yagnik, K. B., Garg, M., & Krishnan, N. C. (2016). SpotGarbage: Smartphone app to detect garbage using deep learning. UbiComp 2016 - Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, 940–945.
6. Rad, M. S., von Kaenel, A., Droux, A., Tieche, F., Ouerhani, N., Ekenel, H. K., & Thiran, J. P. (2017). A computer vision system to localize and classify wastes on the streets. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 10528 LNCS, 195–204.
7. Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C. L., & Dollár, P. (2014). Microsoft COCO: Common Objects in Context.

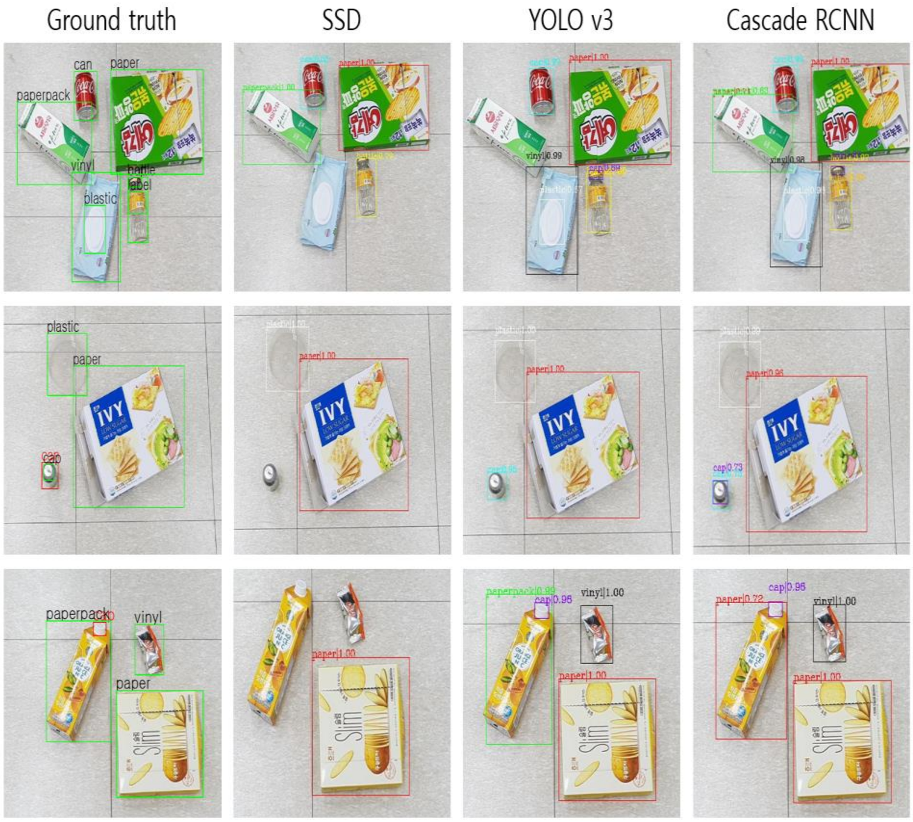

Fig. 2. They are cropped images of the correct dataset and the inference results for each model.