# Object Detection and Classification for Waste Disposal

Ethan Ko
Stanford University
Department of Computer Scence
ethanko@stanford.edu

Peng Hao Lu
Stanford University
Department of Computer Scence
penglu@stanford.edu

## Abstract

*Human waste is a monumental issue for our planet. Proper waste management has great potential for cutting down on pollution and ensuring accurate, efficient sorting for recycling. We propose an image recognition application to tackle the detection and classification problem of properly and efficiently sorting different types of waste. We propose 2 novel approaches which combines existing models and approaches(CNN with Softmax/SVM). In addition, we will compare our our model with 3 baseline classifiers.*

## 1. Introduction

### 1.1. Inputs and Outputs

The input to our algorithm is an image. We then use either SVM, KNN, or softmax model as our baselines and novel combined CNN+SVM and CNN+Softmax models to output a predicted ”organic” or ”recyclable” label for each of the detected objects in the image.

### 1.2. Problem Statement

Detecting, classifying, and sorting trash objects is a highly motivating and important task for the health of our planet and for renewability practices. However, due to high throughput nature of this task and oftentimes muddled test images, an efficient and accurate solution is needed. Currently, human workers are responsible for this task. Potentially eliminating the need for human workers in dirty working conditions is motivating. Existing strategies are time costly and are trained on datasets of pristine images of trash objects, so they are inaccurate as well. Here, we build upon this approach by utilizing data augmentation techniques for model robustness, considering the dirty nature of test images while testing out a novel SVM+CNN approach to hopefully improve accuracy.

## 2. Related Work

Convolutional Neural Networks (CNNs) have been proven to be applicable to image processing through studies such as Guo et al. [1]. In applying CNNs to trash classification, Yang et al. [10] compares CNNs and Support Vector Machines (SVM) on their performance in the classification of trash into glass, paper, metal, plastic, cardboard, and unrecyclable trash. They achieved an accuracy of 22% with CNN and 63% with SVM, with the hypothesis that the CNN performed poorly due to suboptimal hyperparameters. They utilized a 70/30 training/test split and publicly released their image dataset for use. Inspired by this, we propose a model which combines both an SVM and a CNN.

Recent attempts at categorization within recyclable waste have shown that CNNs with the correct architecture are able to attain high accuracy, with Ruiz et al. [8] reaching 88.6% accuracy through using a hybrid Inception-ResNet architecture on the TrashNet baseline. Masand et al. [6] further build on this with their ScrapNet which is able to achieve a 98.4% accuracy on the TrashNet baseline utilizing a ResNet architecture. Gyawali et al. [2] is another example of a successful Resnet with an overall accuracy of 87.8%. Mao et al. [5] uses DenseNet121 to achieve an accuracy of 99.6% on the TrashNet baseline, with improvements on DenseNet being made through a genetic algorithm.

Although there are many successful methods of classifying subcategories of recyclable trash, it is difficult to imagine a realistic implementation of them. In the real world, trash is often an amalgamation of various items and rarely as cleanly divided into ”pure” recyclable trash as in the datasets used by the previous studies. To achieve the final goal of identifying recyclable trash types, the process would likely be to first identify overall trash amongst an assortment of items, then to identify recyclable trash within that trash, and finally to classify the recyclable trash into the various types (paper, metal, etc.). Sultana et al. [9] use two CNNs to first to identify trash and then identify recyclable trash within the identified trash, achieving an overall accuracy of 92%. Their approach proves the concept of a multi-step solution that utilizes more than one model.

The approach to the trash classification used by Huang et al. [3] involved implementing a self-attention-based transformer. CNN-based methods are costly and impractical for high throughput applications such as trash classification. As such, a transformer-based method with higher efficiency is greatly desirable, not to mention an accuracy that reportedly beats CNN-based methods. Pairing this architecture with an object detection step could prove extremely impactful.

Kulkarni [4] approaches the issue of identifying multiple and overlapping trash items through data augmentation and the generation of composite images. By combining a Knapsack problem-solving approach to learn object placement within the collage and a pretrained GAN to blend different masks, Kulkarni was able to generate rich and diverse collages and train a model to identify categories of recyclable trash despite occlusion. Limitations they mention include that they were unable to test on images of real trash piles and difficulties in hybrid training. A general study of image quality affecting CNN classification was conducted by Pei et al. [7] and found that image degradation strongly decreases the classification accuracy of CNNs and that existing algorithms for correcting certain types of degradation were not able to improve accuracy significantly. Settings in which our trash classification CNN could be applied would likely be dirty and encounter these occlusion and degradation effects due to factors such as limited space, poor hardware, and even temperature and humidity, thus we approach the problem with these limitations in mind.

# 3. Methods

## 3.1. Baseline Model

We will compare our novel methodology to 3 common baseline classifiers:

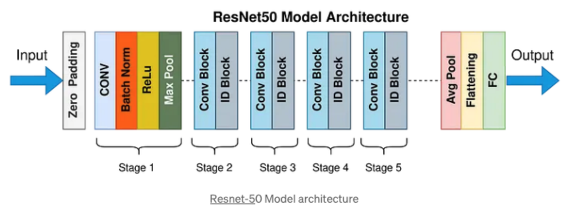
1. **K-nearest neighbors(kNN) classifier**, which uses $L_2$ distances to compute image similarities between test images and labelled train images. Within the k(a hyperparameter) most similar train images(ie neighbors), choose the majority to assign the test image. This method has fast training but slow predictions, hence the need for improvement considering the high throughput nature of our problem.
   L2 Distance: $d_2(I_1, I_2) = \sqrt{\sum_p (I_1^p - I_2^p)^2}$

2. **Support Vector Machine(SVM)** linear classifier, which uses a loss function of
   $L_i = \sum_{j \neq y_i} max(0, s_j - s_{y_i} + 1)$, with the scores being the product between an image and a learned weight matrix.

3. **Softmax Linear Classifier**, which uses a loss function of
   $L_i = -logP(Y = y_i | X = x_i)$
   where $P(Y = k | X = x_i) = \frac{e^{s_k}}{\sum_j e^{s_j}}$, with the scores being the product between an image and a learned weight matrix.

## 3.2. Novel Model

CNNs are effective for this classification task because of the small nature of the differentiating features between classes. To improve upon previous methods, we propose combining a CNN and baseline SVM/Softmax approach. For the CNN, we implement a model based on ResNet-50. Using the learned features and matrices, we then fix the weights of the network, and replace the final FC layer with a SVM or Softmax layer.




# 4. Dataset and Features

We used a waste classification dataset from Kaggle available here: `https://www.kaggle.com/datasets/techsash/waste-classification-data/data`. It contains a total of 25,077 labeled .jpg images pre-split between a training(85%) and testing set(15%). Within these, images are labeled as either 1 of 2 classes: organic or recyclable. We standardized image size to 256 x 256 via padding and cropping. In addition, to improve model robustness, we implemented a preprocessing step where images were randomly flipped, rotated, scaled, and sheared. We also altered image brightness and contrast.

Table 1. Dataset Distribution

| | O | R |
|---|---|---|
| Train(85%) | 12.6k | 9999 |
| Test(15%) | 1401 | 1112 |

Figure 1. Sample Train Image

Figure 2. Sample Test Image

## 5. Experimental Details

Early iterations of a shallow CNN resnet were extremely limited in their abilities to extract and learn complex features from images, which are likely crucial in this classification task. As such, we increased our depth of our network by adding CONV layers, batch-norm, pooling, and Dropout to prevent overfitting, akin to the ResNet-50 network. We used Stochastic Gradient Descent with Momentum for training the weights and biases. In addition, we used Mini-batches of 256.

The experiment was run on Google Colab notebooks. We followed the suggested dataset train/test ratios. For the kNN baseline, we followed the recommended $k = \sqrt{n} \approx 116$ where $n$ was the number of training samples. We elected to not do hyperparameter cross-validation because the number of candidate $k$ values are so high and we were limited by compute power. We ran with 10 epochs.

## 6. Results and Discussion

We evaluated model performance by comparing the accuracy of the binary classification task with the baseline models.

| | Accuracy(%) | Precision(%) | Recall(%) |
|---|---|---|---|
| kNN | 53.4 | 49.3 | 74.0 |
| SVM | 61.9 | 54.8 | 79.3 |
| Softmax | 62.2 | 55.0 | 80.9 |
| CNN+SVM | 86.3 | 79.2 | 74.5 |
| CNN+Softmax | 86.9 | 83.9 | 74.5 |

We believe that the combined CNN and SVM/Softmax models performed well because it combined the best of two worlds. By applying SVM and Softmax to learned complex features(via CNN), we were able to attain results significantly better than using the baselines alone. In all of our baseline and novel models, the following image was incorrectly labeled as Recyclable, and not Organic(as required).

We believe this is due to the fact that the vast majority of training images were of a singular object, unlike this conglomerate of multiple objects, despite all of them being organic. Because we did not implement an object detection algorithm and operated under the assumption that all train and test images were singular, there were inaccuracies such as this one.

Figure 3. Incorrectly Labelled Image

## 7. Future Work

1. **Dataset:** Our dataset is oversimplified, as it lacks negative samples and diverse classes. Introducing images that fall under these categories would make our model more realistic.

2. **Object Detection:** Real images of trash are not of only one image. As such, an edge detection step is needed prior to object classification. Therefore, we would need to repeat our experiments on a dataset with images containing multiple, labeled objects. Another approach would be to manually create mosaics of trash objects from the existing dataset.

3. **Efficiency vs Accuracy:** Due to the high throughput nature of the task, the efficiency of the classification model is extremely important. An efficient model with an imperfect or even less than desirable accuracy could be acceptable if paired with a human to verify results. In the future, we would like to experiment with less complex CNN architectures. Although they would be able to learn less complex features, they might provide decreased runtimes, which is more suited for this task.

## 8. Contributions

Ethan implemented data augmentation and adapted kNN, SVM, and Softmax algorithms for the task. He also helped Peng with the novel models.

Peng spearheaded the implemention of the CNN+SVM and CNN+Softmax models. He also researched existing methods and approaches.

Both members contributed to the milestone, final report, and poster writeups/design equally.

## References


[1] T. Guo, J. Dong, H. Li, and Y. Gao. Simple convolutional neural network on image classification. In *2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA)*, pages 721–724, 2017.

[2] D. Gyawali, A. Regmi, A. Shakya, A. Gautam, and S. Shrestha. Comparative analysis of multiple deep cnn models for waste classification. *arXiv preprint arXiv:2004.02168*, 2020.

[3] K. Huang, H. Lei, Z. Jiao, and Z. Zhong. Recycling waste classification using vision transformer on portable device. *Sustainability*, 13(21), 2021.

[4] H. N. Kulkarni and N. K. S. Raman. Waste object detection and classification. *Stanford CS230 Reports*.

[5] W.-L. Mao, W.-C. Chen, C.-T. Wang, and Y.-H. Lin. Recycling waste classification using optimized convolutional neural network. *Resources, Conservation and Recycling*, 164:105132, 2021.

[6] A. Masand, S. Chauhan, M. Jangid, R. Kumar, and S. Roy. Scrapnet: an efficient approach to trash classification. *IEEE access*, 9:130947–130958, 2021.

[7] Y. Pei, Y. Huang, Q. Zou, X. Zhang, and S. Wang. Effects of image degradation and degradation removal to cnn-based image classification. *IEEE transactions on pattern analysis and machine intelligence*, 43(4):1239–1253, 2019.

[8] V. Ruiz, Á. Sánchez, J. F. Vélez, and B. Raducanu. Automatic image-based waste classification. In *From Bioinspired Systems and Biomedical Applications to Machine Learning: 8th International Work-Conference on the Interplay Between Natural and Artificial Computation, IWINAC 2019, Almería, Spain, June 3–7, 2019, Proceedings, Part II 8*, pages 422–431. Springer, 2019.

[9] R. Sultana, R. D. Adams, Y. Yan, P. M. Yanik, and M. L. Tanaka. Trash and recycled material identification using convolutional neural networks (cnn). In *2020 SoutheastCon*, pages 1–8. IEEE, 2020.

[10] M. Yang and G. Thung. Classification of trash for recyclability status. *CS229 project report*, 2016(1):3, 2016.