Final Report: Smart Trash Net: Waste Localization and Classification

Oluwasanya Awe oawe@stanford.edu

Robel Mengistu robel@stanford.edu

Vikram Sreedhar vsreed@stanford.edu

December 15, 2017

Abstract

Given an image of a jumbled waste, we seek to categorize the different pieces of the waste into three categories: landfill, recycling, paper. This project utilizes Faster R-CNN to get region proposals and classify objects [6]. In this report, we will give an overview of our project and what we have done so far in terms of solving this problem. First, we define our waste sorting problem and present current research and solutions to similar object detection problems. We then give an outline of our proposed architecture and model for approaching our specific task, which entails using a fine-tuned Faster R-CNN. We will also describe the nature and generation of our dataset, as well as the results we achieved from our experiments. Lastly, we outline our next steps in terms of optimizing and improving upon our solution.

1 Introduction

Americans produce more than 250 million tons of waste every year. According to Environmental Protection Agency, 75% of this waste is recyclable. However, currently, only 30% of it is recycled [1]. We want to increase this recycling rate by automating waste sorting. Given an image of jumbled trash/waste which contains two or more different pieces of waste of different types, we want to localize the image and classify the different forms of waste into three categories: recyclable, paper, and landfill.

1.1 Definition

Our automated waste classifier takes in a 768×1024 image containing 2 or more pieces (objects) of waste on a white background. These pieces can be overlapping or non-overlaping and of different size on the white background.

We will fine-tune a Faster R-CNN model pre-trained on PASCAL VOC dataset [2] [6]. Our fine-tuned model will produce anchors (region proposals) and classify objects into three classes: landfill, recycling, and paper.

Our image dataset is generated by composing (stitching together) images in TrashNet dataset [5]. We will discuss this in detail later in this paper.

2 Related Work

Waste Classification. Prior research on waste sorting was done by a previous CS 229 project group, Mindy Yang and Gary Thung [5], where they used a support vector machine (SVM), with scale-invariant feature transform (SIFT) features, and a convolutional neural network (CNN) to classify images of a single object of waste and classify them into six different categories: metal, paper, glass, plastic, trash, and cardboard. They achieved a 63% classification accuracy for the trained SVM and 22% classification accuracy for the CNN. However, their implementation involved classifying a single object image (single piece of trash) as opposed to a jumbled mix of waste.

Deep Networks for region proposals and object detection. Faster R-CNN is a state-of-the-art object detection network that include nearly cost-free Region Proposal Networks (RPNs) that share convolutional layers with state-of-the-art object detection networks (Fast R-CNN) [3] [6]. There is also Mask R-CNN that extends Faster R-CNN by adding a branch for predicting segmentation masks on each Region of Interest (RoI), in parallel with the existing branch for classification and bounding box regression [4]. Since our problem does not require segmentation, we will use Faster R-CNN as a base model.

3 Methods and Architecture

3.1 Faster R-CNN

Faster R-CNN has a Region Proposal Network (RPN) that shares full-image convolutional features with the detection network, thereby allowing nearly cost-free region proposals. An RPN is a fully convolutional network that simultaneously predicts object bounds and objectness scores at each position. The RPN is trained end-to-end to generate high-quality region proposals, which are used by Fast R-CNN for detection.

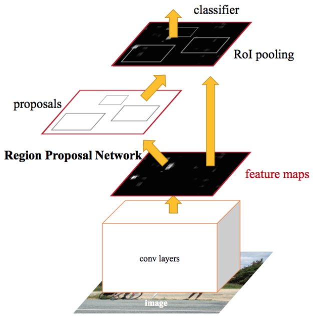

Figure 1: Faster RCNN Architecture based off [6]

4 Image Dataset

Faster R-CNN requires a large image dataset of 10,000+ images for effective training. We could not find a jumbled waste images dataset of this size anywhere. Hence, we decided to augment Mindy Yang and Gary Thung's dataset [5] 2,500 images of single pieces of waste which they used for their trash classification project. We were able to obtain their approval and dataset upon request.

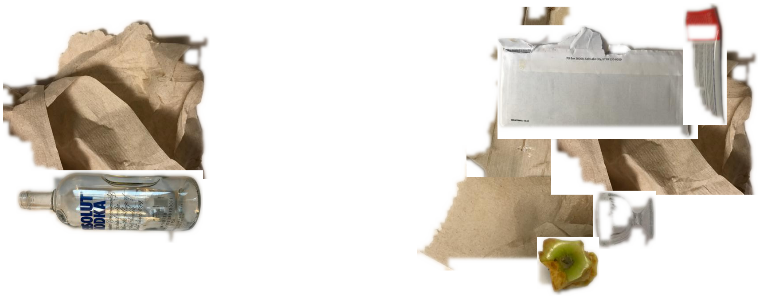

Figure 2: Sample from the dataset showing a combination of waste items.

4.1 Data Preprocessing

We wrote a python script that removed the background of each image in order to isolate a bounding box around each piece of waste. Next, we sorted the edited images based on the classes we were looking at, which were landfill, recycling, and paper. We then would use a 768 × 1024 plain white background and merge 2-6 pieces of trash onto this background at random locations on the white space. Thus, we created new images with corresponding labels and the locations of each trash's bounding box within the white background.

We were able to generate our 10,000 images of 'piles' of trash and could always modify our code in order to merge even more pieces of trash onto one image. We split the (train, validation, test) data into roughly $(6000, 2000, 2000)$ images respectively.

5 Experiments/Results/Discussion

5.1 Hyperparameters

We worked on the following hyperparameters:

1. RPN Batch Size
2. Batch Size (Number of Regions of Interest)
3. Learning Rate
4. Model: RPN/SS + Fast R-CNN

5.2 Experiments

For our baseline, we fine-tuned a pre-trained Faster R-CNN model by modifying the last fully connected layers – 'clsscore' and 'bboxpred'. Hence, we utilized pre-trained lower level features. We trained on 2000 images containing only two objects each with a fairly even representation of every class across the different examples.

Similar to the standard Faster R-CNN [6], we used mean average precision (mAP) to evaluate the performance of our model. The mAP equals to the integral over the precision-recall curve p(r)

$$\int_0^1 p(r)dr.$$

Where the precision-recall curve is determined by calculating the IoU (intersection of union) between the predicted and ground truth bounding boxes. We also use the same multi-task loss function described in Faster R-CNN paper [6]:

$$L(\{p_i\}, \{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \frac{\lambda}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*).$$

This is a sum of two normalized losses over anchors with indices, i, such that p_i is the model's probability of an anchor representing an object, $p_i^* \in \{0, 1\}$ is the ground truth on whether or not the anchor is actually an object. t_i is the coordinates of the model's proposed bounding box while t_i^* is the ground truth of the coordinates. [6] give more detail on the loss functions used.

5.3 Results

We tuned hyperparameter values for RPN Batch Size, Batch Size, and Learning Rate.

Class	Learning Rate (0.0012)	Batch RPN Size (128)	Batch Size (128)
Landfill	0.514	0.725	0.722
Paper	0.433	0.596	0.596
Recycling	0.585	0.767	0.767

Table 1: Validation AP scores from best hyperparameters tested

After running our Faster R-CNN on our dataset using our proposed data split, we achieved a mean Average Precision (mAP) of 0.683 overall on classification of the trash images, in which we got the following AP values for the different classes:

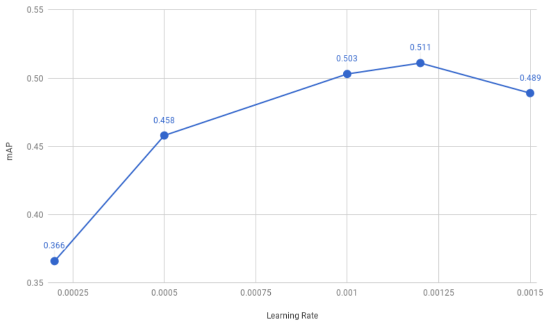

Figure 3: Average mAP scores over the 3 classes for different learning rates

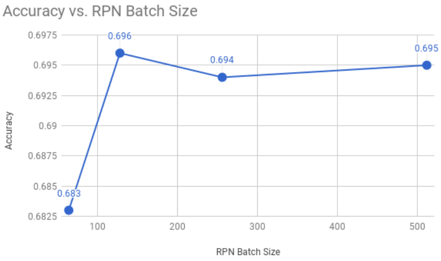

Figure 4: Accuracy over the 3 classes for different RPN batch sizes

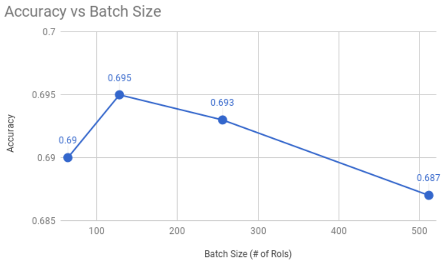

Figure 5: Accuracy over the 3 classes for different batch sizes

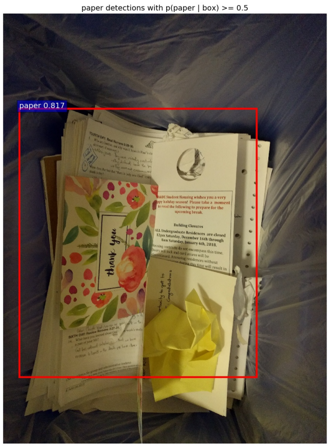

Figure 6: Sample model output without preprocessing

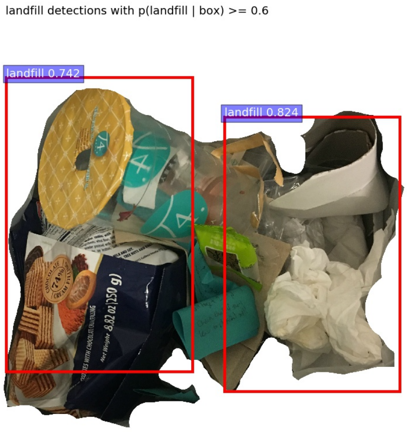

Figure 7: Sample output after preprocessing

5.3.1 Categorizing Photos of Trash

Given that we fine-tuned our model on an artificial dataset, we wanted to see if our model would still be able to identify waste items from arbitrary photos pooled from Stanford students.

5.4 Discussion

We had hoped to get a higher accuracy for each category, however this is due to our baseline using a larger dataset from PASCAL VOC rather than our custom dataset. We also realized training our model took a lot of time and if we had more CPUs working in parallel, we would have been able to try out more hyperparameter optimizations.

In general, we noticed that the 'paper' category had the worst performance from our model.

Class	AP of Optimal Hyperparameters
Landfill	0.699
Paper	0.607
Recycling	0.744
mAP	**0.683**

We thought this could have been a bias issue with the number of examples trained on but there was a roughly even representation in our models. The only other plausible cause for this would be the fact that we used a white background which tends to have a similar color with members of the paper class.

In terms of processing raw photos, in most cases it was only after preprocessing the image to look like the trained dataset were we able to get our model to return reasonable values.

5.5 Future Work

Some of the key areas we hope to work on are:

1. **Training from Scratch:** Given that we can theoretically generate an infinite number of images with our original dataset, we have sufficient data to train the model from scratch. Time and resource permitting, we will train the model from scratch and use that instead of the pre-trained model.

2. **Architecture:** We have only used ZF Net [8] that has 5 convolutional layers and 3 fully-connected layers, but will look into using the alternative pre-trained model based on VGG-16 model [7] that has 13 convolutional layers and 3 fully-connected layers.

3. **Test on Real Images:** We hope to test our model on real images of piles of trash in order to determine how our model does on the actual problem.

6 Contributions

1. **Oluwasanya Awe**: Worked on model set-up, fine-tuning, and milestone/final report writeup. Researched and refined the problem.

2. **Robel Mengistu**: Worked on model and gpu set-up, and milestone/final report writeup. Researched and refined the problem.

3. **Vikram Sreedhar**: Worked on dataset generation/pre-processing and milestone/final report writeup. Researched and refined the problem.

References

[1] Texas Natural Resource Conservation Commission et al. *Municipal solid waste plan for Texas*. Texas Natural Resource Conservation Commission, 1995.

[2] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88(2):303–338, 2010.

[3] Ross B. Girshick. Fast r-cnn. *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 1440–1448, 2015.

[4] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross B. Girshick. Mask R-CNN. *CoRR*, abs/1703.06870, 2017.

[5] Gary Thung Mindy Yang. Classification of trash for recyclability status. *CS229 Project Report 2016*, 2016.

[6] Shaoqing Ren, Kaiming He, Ross B. Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39:1137–1149, 2015.

[7] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[8] Matthew D Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. In *European conference on computer vision*, pages 818–833. Springer, 2014.