Computer Vision Models for Waste Classification

Thanh Trinh

Abstract—**Automatic methods for waste classification are becoming increasingly important as many countries face an overloaded amount of waste. In this paper, we trained a model to effectively identify different types of waste in images. Specifically, we finetuned a ResNet-50 model by adding some data augmentations and unfreezing some network layers. Our model obtained a classification accuracy of 94.20%, thereby outperforming two strong baselines. We believe that our model can help people sort out waste efficiently.**

I. INTRODUCTION

The current amount of pollution, especially mismanaged waste, is increasing at an alarming rate [1]. This problem has negative consequences on physical habitats and aquatic life [2], but the lack of attention and human resources makes it difficult to solve immediately. Fortunately, modern technology may provide opportunities to effectively address this problem.

In this paper, we studied three machine learning models for classifying waste in images based on ResNet-50 [3]. ResNets address the vanishing gradient problem [4] through residual blocks, which introduce skip connections that the gradient can use to bypass layers. First, we utilized a model pretrained on ImageNet [5] and mapped the labels to our corresponding waste labels. Second, we ran a Kaggle notebook by Mtszkw [6], who retrained the final logit layer of an ImageNet-pretrained ResNet-50 backbone. Lastly, we further improved Mtszkw's model by adding image data augmentations and unfreezing more of the model's last layers.

Based on quantitative results on a waste classification dataset [7], we found that Mtszkw's model achieved 67.43% better accuracy than the pretrained ImageNet model since the latter was not trained for our task. We also found that our model further improves on Mtszkw's model by 3.96% thanks to data augmentation and the finetuning of more layers.

A limitation of our analysis is that the dataset we used simplifies the waste classification problem (it contains ideal images with objects in the middle of the frame with a clear white background). However, we believe that our improvements generalize to more complex images, thereby providing steps to improve automated waste classification systems in the real world.

II. DATA

We used one main Kaggle dataset for waste classification [7]. The dataset contains around 2,500 images from six classes of waste materials: cardboard, glass, metal, plastics, paper, and trash. The dataset comes with three different data splits—1768 training, 328 validation, and 328 testing images—which we used for development and evaluation.

III. METHODOLOGY

To tackle our problem, we used three different approaches. For the first approach (which we denote as `ImageNet`), we used a ResNet-50 image classification model which was pre-trained on the ImageNet dataset [5]. We manually mapped its ImageNet class predictions to the six classes in our Kaggle dataset, as shown in Table 1.

TABLE I. TABLE OF MATCHING MAPPING BETWEEN KAGGLE/IMAGENET LABELS FOR THE IMAGENET ARPPOACH

Kaggle Class	ImageNet Classes
Cardboard	carton
Glass	beer_bottle, beer_glass, pop_bottle, red_wine, syringe, water_jug, wine_bottle, petri_dish, hourglass, perfume
Plastics	water_bottle, plastic_bag, Petri_dish, bottle_cap, nipple, syringe
Paper	paper_towel, menu, envelope, comic_book, book_jacket, packet, paper_towel, toilet_tissue
Metal	bottlecap, combination_lock, Crock_Pot, can_opener, corkscrew, hook, shield
Trash	All ofther labels in the ImageNet data

For the second approach (denoted as `Mtszkw`), we reproduced a notebook on Kaggle provided by the user Mtszkw [6]. They utilized a frozen ResNet-50 backbone pre-trained on ImageNet and trained the final logit layer via ADAM optimization [8] to predict the six Kaggle classes. For the third approach (denoted as `Ours`), we improved on Mtszkw's method by adding some data augmentations. Specifically, during training, we applied random transformations to the images, including panning, zooming, and vertical/horizontal flipping. Additionally, we unfroze the last 55 layers of the ResNet structure. As shown in Section IV, these adjustments helped increase performance substantially.

IV. RESULT

We deployed the three models on the test split in the Kaggle dataset. In order to evaluate the performance of each model, we calculated its accuracy in identifying each object class as well as its overall accuracy. For the `ImageNet` method, due to the fact that there are 1,000 classes in ImageNet, we also included the top 5 accuracy for comparison, which counts an image as correct if any of the top 5 predicted ImageNet classes map to the correct Kaggle class.

Table 2 shows the class-wise and overall accuracy of the three methods. For the `ImageNet` method, we can see that the class-wise accuracy was not high. The highest accuracy was from the cardboard class at 51.71%, while the lowest accuracy is from the plastics class at 2.28%. Moreover, this model was biased toward predicting the trash class because we mapped the majority of ImageNet classes to the Kaggle trash class; as a result, it achieved high accuracy in this class. However, the overall accuracy was only 22.79%.

TABLE II. ACCURACY OF THE THREE METHODS DESCRIBED IN SECTION IV

	ImageNet (top 5)	Mtszkw	Ours
Cardboard	53.71% (81.68%)	93.47%	100.00%
Glass	36.45% (63.74%)	89.23%	93.85%
Plastics	02.28% (10.97%)	88.52%	88.52%
Paper	04.03% (15.13%)	95.18%	96.39%
Metal	12.65% (36.25%)	91.07%	98.21%
Trash	64.49% (97.82%)	76.47%	76.47%
Overall	**22.79% (73.41%)**	**90.24%**	**94.20%**

Moving on to Mtszkw, the overall performance was much higher, at 90.24%. This is because this model's logit layers were explicitly trained on the Kaggle dataset. Finally, our method performed best with an overall accuracy of 94.20%, due to data augmentation and unfreezing more layers. Compared to Mtszkw, our model achieved better class-wise accuracy for all classes except trash. This was likely because in the training set, we have much fewer images of trash than of other classes.

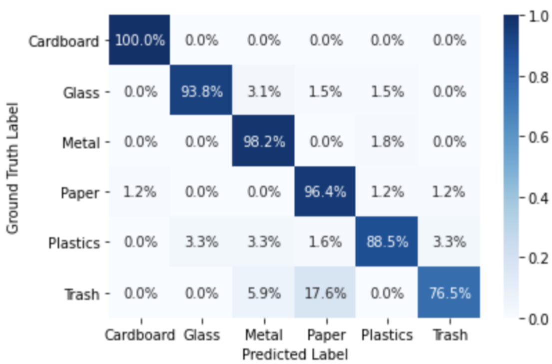

Figure 1. Confusion matrix of our approach.

In Figure 1, we include the classification confusion matrix for our method. The most common mistake that our method made was that it classifies images of trash as paper. Specifically, it tended to misinterpret wrappers as paper even though they should be classified as trash (Figure 2).

Figure 2. Trash images that were misclassified as paper.

Moreover, our model also misinterpreted some plastics images as glass. As shown in Figure 3, the misclassified plastics are transparent; even humans could easily misidentify these examples as glass.

Figure 3. Plastics images that were misclassified as glass.

V. CONCLUSION AND DISCUSSION

The goal of our research was to improve the performance of existing waste classification systems. To this end, we took a ResNet model pre-trained on ImageNet and finetuned it by adding data augmentations and modifying the number of frozen layers. When comparing the results of our model against the baseline models ImageNet and Mtszkw, we observed significantly greater accuracy in our model.

It is important to note the limitations of our work. For example, the Kaggle dataset is clean, meaning objects are placed in the middle of the image with a white background. Therefore, our model may struggle with real-life situations or photos, where objects may not be in the center, or the background may have complex textures from roads, grass, etc. Additionally, some countries may not have large enough recycling programs to utilize our models (e.g., Chile only recycles around 1% of its total waste [9]).

Although our experiments were run on a simplified version of the waste classification problem, our improvements are general and can be applied to more complex real-world data. Therefore, we believe that our work can promote the deployment of more advanced waste classification systems.

ACKNOWLEDGMENT

Thank you to Dr. Ryan Szeto for assisting in the development of this research paper.

REFERENCES

[1] 100+ Ocean Pollution Statistics & Facts (2020-2021). Condor Ferries. (n.d.). Retrieved September 20, 2021, from https://www.condorferries.co.uk/marine-ocean-pollution-statistics-facts.

[2] Environmental Protection Agency. (n.d.). Impacts of Mismanaged Trash. EPA. Retrieved September 20, 2021, from https://www.epa.gov/trash-free-waters/impacts-mismanaged-trash.

[3] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. In IEEE Conference on Computer Vision and Pattern Recognition (pp. 770-778).

[4] Wang, C.-F. (2019, January 8). The Vanishing Gradient Problem. Medium. Retrieved September 21, 2021, from https://towardsdatascience.com/the-vanishing-gradient-problem-69bf08b15484.

[5] Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009, June). ImageNet: A Large-Scale Hierarchical Image Database. In IEEE Conference on Computer Vision and Pattern Recognition (pp. 248-255).

[6] Mtszkw. (2020, August 12). Garbage Keras (90% at first run, no preprocessing). Kaggle. Retrieved September 20, 2021, from https://www.kaggle.com/mtszkw/garbage-keras-90-at-first-run-no-preprocessing.

[7] Cchangcs. (2018, November 24). Garbage Classification. Kaggle. Retrieved September 20, 2021, from https://www.kaggle.com/asdasdasasdas/garbage-classification.

[8] Kingma, D. P., & Ba, J. L. (2015, May). ADAM: A Method for Stochastic Gradient Descent. In International Conference on Learning Representations (pp. 1-15).

[9] The 5 Worst countries for Plastic Recycling. Plastic Expert. (2019, November 27). Retrieved September 20, 2021, from https://www.plasticexpert.co.uk/the-5-worst-countries-for-plastic-recycling.