Stony Brook University

Computer Science

Sifting Through Trash

Reilly Browne, Sai Tanmay Reddy Chakkera, Dylan Scott

CSE 512 Course Project (Group 6)

BACKGROUND

As global populations continue to grow, an ever-increasing amount of waste is produced. This magnitude of waste has the potential to cause environmental disaster as landfills get overwhelmed and it ends up accumulating in our communities and oceans. One of the frontline defenses to mitigate excessive waste is recycling. We see several ways in which ML/Vision models that can recognize different types of waste can increase the amount of waste which is recycled. In the waste management industry, ML-powered robots can be used to separate recyclables in municipalities which use single-stream recycling and even pull recyclables out from the regular waste stream. In a consumer context, we see that ML-powered apps could better educate the public on what they can recycle and where they should recycle it could improve community participation.

INTRODUCTION

Much of ML's usefulness in recycling depends on the effectiveness of Vision models in classifying waste and recyclables. Existing research has already shown that impressive levels of accuracy can be achieved, and we aim to explore how state of the art models, such as Vision Transformers, could lead to even greater accuracy and flexibility.

We also aim to explore the novel improvements which can be made using Cost-Sensitive Learning (CSL). In the recycling industry, the cost of misclassifying paper as cardboard is much lower than the cost of misclassifying metal as plastic, yet existing research treats both equally. Through CSL, we consider these inherent imbalances in the loss functions of our models.

METHODS

For our training and evaluation, we utilized the TrashNet dataset which contains 2,527 pictures of waste sorted into trash, glass, paper, cardboard, plastic, and metal.

Sample images from TrashNet

We then fine-tuned and evaluated the following three state-of-the-art models — Google's ViT, Microsoft's ResNet, and OpenAI's CLIP — using their default loss functions. We also tested CLIP's baseline capabilities as a Zero-Shot Image Classifier.

We then adapted ViT and ResNet to utilize to CSL approaches – Cost-Sensitive Cross Entropy (CSCE) and OVA Regression (OVAReg)

For CSCE, we implemented the following loss function using PyTorch, which uses a modified SoftMax function which incorporates a cost matrix (c):

$$L_{CE} = \sum_{n\in\{1,\ldots,N\}} \left(-\sum_i d_{i,n} \log p_{i,k,n} \right) \quad p_{i,k,n} = \frac{c_{k,i} e^{-o_i}}{\sum_j c_{k,j} e^{-o_j}}$$

For OVA Regression, we implemented the following loss function (also using PyTorch), which sums up the risks / costs associated with wrong classifications instead of just determining whether a misclassification occurred:

$$L = \sum_{n\in\{1,\ldots,N\}} \sum_k \max(z_{n,k}(r_k(x_n) - c_{n,k}), 0) \quad \hat{L} = \sum_{n\in\{1,\ldots,N\}} \sum_k \ln(1 + e^{z_{n,k}(r_k(x_n) - c_{n,k})})$$

(smooth approximation)

Where $z_{n,k} = 1$ if $k = y_n$ and $z_{n,k} = -1$ if $k \neq y_n$ where k is an iterator over the classes and y_n correct class for the nth sample.

For the two methods, we used the following cost matrices:

CSCE Cost Matrix

	Cardboard	Glass	Metal	Paper	Plastic	Trash
Cardboard	0.1	0.1	0.1	0.1	0.1	0.1
Glass	0.1	0.1	0.4	0.1	0.3	0.1
Metal	0.1	0.4	0.1	0.1	0.1	0.2
Paper	0.1	0.1	0.1	0.1	0.1	0.1
Plastic	0.1	0.4	0.5	0.1	0.1	0.2
Trash	0.4	0.1	0.2	0.4	0.1	0.1

OVA Reg Cost Matrix

	Cardboard	Glass	Metal	Paper	Plastic	Trash
Cardboard	0	5	5	5	5	5
Glass	5	0	20	5	15	5
Metal	5	20	0	5	5	10
Paper	5	5	5	0	5	5
Plastic	5	5	25	5	0	10
Trash	20	5	10	20	5	0

RESULTS

For our baseline tests, without the Cost-Sensitive Loss Functions, we achieved the following results with each model:

Test Data / *Training Data*

	CLIP (Base)	CLIP (Fine-Tuned)	ViT	ResNet
Macro-F1	**65.84%** *64.29%*	**90.93%** *94.92%*	**93.78%** *98.39%*	**89.61%** *98.29%*
Micro-F1 / Accuracy	**71.73%** *71.00%*	**92.09%** *95.89%*	**95.06%** *98.91%*	**91.50%** *98.86%*
Balanced Accuracy	**66.22%** *64.57%*	**92.68%** *96.54%*	**94.91%** *98.96%*	**90.23%** *98.45%*

Upon analyzing the full confusion matrices, we identified three pairings as the primary "pain points" for this dataset:

True → Predicted	CLIP (Base)	CLIP (Fine-Tuned)	ViT	ResNet
Trash → Paper	**39.29%** *28.44%*	**0.00%** *0.00%*	**3.57%** *0.92%*	**7.14%** *0.92%*
Plastic → Glass	**13.59%** *16.22%*	**6.80%** *2.37%*	**1.94%** *0.53%*	**5.83%** *0.53%*
Glass → Plastic	**13.68%** *20.69%*	**3.16%** *3.20%*	**3.16%** *0.00%*	**3.16%** *0.25%*

Then, with the Cost-Sensitive Loss Functions, we achieved the following results:

	ViT (CSCE)	ViT (OVAReg)	ResNet (CSCE)	ResNet (OVAReg)
Macro-F1	**94.21%** *99.73%*	**95.00%** *99.73%*	**88.53%** *94.43%*	**85.54%** *91.75%*
Micro-F1 / Accuracy	**95.26%** *99.75%*	**96.25%** *99.70%*	**90.71%** *96.43%*	**87.55%** *94.11%*
Balanced Accuracy	**93.47%** *99.78%*	**94.98%** *99.73%*	**88.72%** *94.08%*	**85.58%** *90.57%*

Here are the same three pairings identified as "pain points" in our confusion matrix:

True → Predicted	ViT (CSCE)	ViT (OVAReg)	ResNet (CSCE)	ResNet (OVAReg)
Trash → Paper	**0.00%** *0.00%*	**0.00%** *0.00%*	**3.57%** *1.83%*	**7.14%** *0.92%*
Plastic → Glass	**2.91%** *0.26%*	**1.94%** *0.26%*	**3.88%** *1.58%*	**7.77%** *6.07%*
Glass → Plastic	**2.11%** *0.25%*	**2.11%** *0.00%*	**5.26%** *0.31%*	**8.42%** *0.49%*

We see that ViT is the clear winner in all categories. The Cost-Sensitive Cross Entropy loss gives ViT a modest performance improvement while OVA Regression provides us with the best accuracy on TrashNet we know of, across our own experiments and existing research. Unfortunately, the same is not true for ResNet and our custom loss functions cannot be applied to CLIP at all.

To put ViT to the test further, we fine-tuned ViT with the three losses (default, CSCE, and OVAReg) on the Fashion MNIST dataset, which contains 70,000 grayscale images of various articles of clothing. Since TrashNet is quite small, this analysis shows how these techniques could scale up to larger datasets although the dataset is not in the domain of waste identification. We achieved the following results:

	ViT (Default)	ViT (CSCE)	ViT (OVAReg)
Macro-F1	**92.96%** *94.80%*	**93.21%** *95.30%*	**92.59%** *93.88%*
Micro-F1 / Accuracy	**93.01%** *94.87%*	**93.27%** *95.34%*	**92.63%** *93.92%*
Balanced Accuracy	**93.01%** *94.87%*	**93.27%** *95.34%*	**92.63%** *93.92%*
Coat → Pullover	**11.40%** *10.18%*	**8.30%** *7.87%*	**5.80%** *5.72%*

We see that ViT continues to show strong image classification performance on a much larger dataset. We also see that CSCE continues to offer a modest performance gain while, interestingly, OVAReg offered worse overall results. However, analyzing our confusion matrices, we identified Coat / Pullover as a "pain point" for our non-CSL trained model, and OVAReg does provide a significant improvement in worst-case pairwise accuracy.

CONCLUSION

We believe our research demonstrates the potential ML / Vision models have to revolutionize recycling. By combining state-of-the art Vision Transformers with Cost-Sensitive Learning techniques, we were able to achieve an impressive 96.25% accuracy on our test set. Beyond the baseline accuracy, investigating the types of misclassifications, we were able to greatly reduce costly mistakes – even reducing the rate of misclassifying trash as paper to 0% on our test set.

For further research, we would certainly have liked to have access to a larger waste classification dataset with a wider variety of labels (for example, identifying different types of plastic). We also think adding object detection would increase the real-world usefulness of our models.

What our research certainly shows is how far ML has come in the past few years. When the TrashNet dataset was first introduced in 2016, the highest test accuracy achieved was 63% with an SVM (Wang et al.). With humans serving as the greatest barrier to recycling, we hope that in the near future, sufficient accuracy can be achieved to allow recycling to become a fully automated task that humans don't need to think about.