Sifting Through Trash (CSE512 Final Report)

Reilly Browne, Sai Tanmay Reddy Chakkera, Dylan Scott

November 2023

1 Introduction and Related Work

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. ML/Vision models that can classify different types of waste can increase the amount of waste which is recycled. Existing research shows that these vision classification models can achieve impressive levels of accuracy, however, not without misclassifications. Certain misclassfications may be worse than others, such as classfiying paper as cardboard as opposed to classfiying metal as cardboard (the metal may damage cardboard recycling machines if this model is directly deployed). This project aims to analyse these misclassifications in state of the art vision models and improve them using cost-sensitive losses. Further, these cost-sensitive losses are also analysed in the class-imbalance setting.

Many prior works have tackled the primary dataset of this project, i.e., trashnet. [5], [3], [6], [4] use CNNs and deep-learning to classify trash. These methods analyse the misclassification performance of their algorithms through confusion matrices, F1-scores, accuracies, etc. In contrast, we analyse more metrics such as macro and micro F1-score as well as using different losses such as cost-sensitive cross entropy [2] and OVA regression based modeling [1].

2 Experiments & 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.

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. Later we fine-tune ViT and ResNet-50 on trashnet using our custom loss functions. Further, to analyse the effect of our cost-sensitive losses we train these models on a synthetically imbalanced fashion-mnist dataset. The imbalance is described in table 1.

Fashion-MNIST Class-Wise Population										
	0	1	2	3	4	5	6	7	8	9
Original	6000	6000	6000	6000	6000	6000	6000	6000	6000	6000
Imbalanced	2000	2000	5600	6000	600	6000	6000	6000	6000	600

Table 1: Synthetically imbalanced Fashion-MNIST

The following sections describe the various cost sensitive losses that were used to fine-tune ViT and ResNet.

2.1 Cost Sensitive Cross Entropy

The vanilla cross entropy loss is well known as a standard loss for multi-class classification tasks. Given $d_{i,n}$ to be the ground truth probability of example n being in class i and $p_{i,n}$ to be the prediction of a model to the probability of nth example being in class i, the vanilla cross entropy is computed as follows over all samples

$$L_{CE} = \sum_{n \in \{1,\ldots,N\}} \left(- \sum_{i} d_{i,n} \log p_{i,n} \right)$$

where n is the iterator over samples in the dataset, i is the iterator over classes and N is the number of samples. The cost sensitive cross entropy loss is described in [2]. Given a cost matrix c defined such that $c_{p,q}$ represents the cost of classifying a sample of class p as class q. They define a change to the softmax function to incorporate the cost matrix c.

$$p_{i,k,n} = \frac{c_{k,i} e^{o_i}}{\sum_j c_{k,j} e^{o_j}}$$

where i is an iterator over classes, k is the ground truth class of sample n, n is the iterator over samples and o is the output of the final layer of the classifier. Given this changed definition of of softmax, the cross entropy loss is computed similarly:

$$L_{CE} = \sum_{n \in \{1,\ldots,N\}} \left(- \sum_{i} d_{i,n} \log p_{i,k,n} \right)$$

Since $c_{p,q}$ is just a constant over the problem, this change doesn't impact the gradient computation, and thus can be simply implemented using standard pytorch functions.

2.2 One-Vs-All Regression

In One-Vs-All Regression, the softmax layer at the end of DNN based classifiers is removed and make them regressors of "risk" or "cost" of classifying the particular input into a particular class as opposed to the other classes. This is achieved by replacing the logistic function with the identity, since the cost sums risks of misclassification rather than just misclassifications. The model is trained to minimize this risk over the training set. They come up with a smooth approximation of this loss (the one that minimizes the risk/cost) so that DNNs can be trained over it using back-propagation. This method is described below:

Each sample in the dataset is assumed to also have an associated vector c_n where $c_{n,k}$ gives the cost of classifying the nth sample as class k. Note that this is a generalization of the notion that the whole problem has class wise classification costs, in that, here each sample may have a different misclassification cost. Further, instead of predicting a probability, the model is made to regress on the cost. That is, let $g(x)$ be the classifier that gives a class label as an output for input x. Let the final layer of $g(x)$ be $r_k(x)$ which is then fed to a softmax function to predict probabilities. The paper suggests that we model $r_k(x) \approx ck$. Then, a prediction can be made by

$$g_r(x) = \underset{k \in 1,\ldots,K}{\operatorname{argmin}} r_k(x)$$

where the $\{1...K\}$ is the set of K classes. Further they define $z_{n,k} = 1$ if $k = y_n$ and $= -1$ if $k \neq y_n$ where k is an iterator over classes y_n is the correct class for the nth sample. In that case the loss function they define is:

$$L = \sum_{n \in 1,\ldots,N} \sum_{k} \max(z_{n,k}(r_k(x_n) - c_{n,k}), 0)$$

However, this function is not smooth, so they derive an upper bound to this loss:

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

For the two methods, we used the cost matrices in Tables 2 and 3.

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

Table 2: Cost matrix for CSCE

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

Table 3: Cost matrix for OVA Regression

3 Results

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

Table 4: Results on selected metrics before cost sensitive learning

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. Note that these losses can also be applied to fine-tune CLIP. However, due to resource and time constraints we were not able to experiment on it. However, since CLIP also uses variants of ViT as a backbone, we believe that its performance can be interpolated on these cost-sensitive losses.

	ViT (CSCE)	ViT (OVAReg)	ResNet (CSCE)	ResNet (OVAReg)
Mac.-F1	**94.21%** **99.73%**	**95.00%** **99.73%**	**88.53%** **94.43%**	**85.54%** **91.75%**
Mic.-F1 / Acc.	**95.26%** **99.75%**	**96.25%** **99.70%**	**90.71%** **96.43%**	**87.55%** **94.11%**
Bal. Acc.	**93.47%** **99.78%**	**94.98%** **99.73%**	**88.72%** **94.08%**	**85.58%** **90.57%**

Table 5: Results on selected metrics after cost sensitive learning

True → Predicted Class	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%**

Table 6: Selected entries from the confusion matrices (All confusion matrices in Github)

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:

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.

4 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.

True → Predicted Class	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%**

Table 7: Selected entries from the confusion matrices for CSCE and OVAReg

	ViT (Default)	ViT (CSCE)	ViT (OVAReg)	ResNet (OVAReg)
Mac.-F1	**92.96%** **94.80%**	**93.21%** **95.30%**	**92.59%** **93.88%**	**85.54%** **91.75%**
Mic.-F1 / Acc.	**93.01%** **94.87%**	**93.27%** **95.34%**	**92.63%** **93.92%**	**87.55%** **94.11%**
Bal. Acc.	**93.01%** **94.87%**	**93.27%** **95.34%**	**92.63%** **93.92%**	**85.58%** **90.57%**

Table 8: Results after applying to the Fashion-MNIST data set.

5 Teamwork

- Reilly Browne - Model evaluation; figures and organization of poster and report
- Sai Tanmay Reddy Chakkera - Implementation of loss functions (CSCE and OVAReg), fine-tuning models, and extensions to FashionMNIST; writing and organization of report
- Dylan Scott - Fine tuning of baseline models (ViT, ResNet, CLIP); writing and organization of poster.

References

[1] Yu-An Chung, Hsuan-Tien Lin, and Shao-Wen Yang. Cost-aware pre-training for multiclass cost-sensitive deep learning, 2016.

[2] Salman H. Khan, Munawar Hayat, Mohammed Bennamoun, Ferdous A. Sohel, and Roberto Togneri. Cost-sensitive learning of deep feature representations from imbalanced data. *IEEE Transactions on Neural Networks and Learning Systems*, 29(8):3573–3587, 2018.

[3] Wei-Lung Mao, Wei-Chun Chen, Haris Imam Karim Fathurrahman, and Yu-Hao Lin. Deep learning networks for real-time regional domestic waste detection. *Journal of Cleaner Production*, 344:131096, 2022.

[4] Cuiping Shi, Ruiyang Xia, and Liguo Wang. A novel multi-branch channel expansion network for garbage image classification. *IEEE Access*, 8:154436–154452, 2020.

[5] Yuheng Wang, Wen Jie Zhao, Jiahui Xu, and Raymond Hong. Recyclable waste identification using cnn image recognition and gaussian clustering, 2020.

[6] Zhihu Yang and Dan Li. Wasnet: A neural network-based garbage collection management system. *IEEE Access*, 8:103984–103993, 2020.