Garbage Grouping:
Fine-tuning Convolutional Neural Networks for Waste Classification

Friday Mishra, Junior Undergraduate
Major in Information Systems, Data Science
imishra4@wisc.edu

Joel Myers, Senior Undergraduate
Major in Data Science
jcmyers3@wisc.edu

Lucas Steplyk, Senior Undergraduate
Major in Computer Science, Data Science
Certificate in Mathematics
lsteplyk@wisc.edu

Project Report

1. Abstract

The waste management industry is substantially large, with market size in 2024 being over USD 1.2 Trillion.[1] Waste management across the world is done manually, and with the growing pressure of sustainability and technology, the industry is one that would largely benefit from deep learning and AI. This project aims to create an efficient waste classification method, eliminating the need for manual labor. In this project, several pre-trained Convolutional Neural Networks (CNN) - ResNet50, MobileNetV3, RegNetY, and MobileNetV3 - were fine-tuned in addition to a CNN trained from scratch (simple CNN), to achieve an 88.91% accuracy on the RealWaste dataset, containing 4,752 labeled images across nine categories. The best-performing model, ResNet50, achieved an accuracy of 88.91% while being a lightweight model, demonstrating its potential for scalable, real-world waste classification to optimize the waste management industry.

2. Introduction

Waste management is a pressing challenge facing environments worldwide. As waste production increases, especially in developed countries, traditional methods of waste sorting such as manual classification and visual inspection are becoming inefficient and expensive. With the growing demand for sustainability, the waste management industry must adopt new technologies to improve sorting processes, reduce operational costs, and increase recycling efficiency. One proposed solution is through deep learning, which offers the potential to automate and optimize waste classification. In particular, Convolutional Neural Networks (CNNs) have shown the ability to process visual data and automate the sorting of waste materials as we discuss later. This project aims to develop an efficient CNN model capable of classifying landfill waste accurately and efficiently. Through our work, we aim to demonstrate the applicability of CNNs to waste management and explore the trade-offs between model complexity and computational efficiency.

1.1 Waste Sorting

Waste sorting is the process of categorizing waste materials such as plastics, metals, and glass into separate streams for recycling or disposal. Traditional waste sorting methods typically rely on manual labor and human visual inspection to identify different materials such as plastic, metal, glass, and paper. However, these methods are inefficient, time-consuming, and error-prone, leading to suboptimal recycling rates and increased labor costs. As urbanization continues to rise, the volume of waste along with the cost of labor also grows, further showing a need for an efficient solution to this problem.

Automation in waste sorting has the potential to improve efficiency, reduce human error, and overall contribute to a more sustainable world. The application of image recognition and machine learning, into waste sorting systems could automate the classification of materials, ensuring that they are correctly sorted and processed for recycling.

[1] https://www.marketsandmarkets.com/Market-Reports/waste-management-market-72285482.html

1.2 Image Classification & Convolutional Neural Networks (CNNs)

Image classification is the process of assigning labels to images based on their content, and in this case, classifying waste materials into predefined categories. CNNs are particularly suited for this task due to their ability to learn complex patterns in image data. Unlike traditional machine learning algorithms that require manual feature extraction, CNNs automatically extract relevant features from raw image data, such as edges, textures, and patterns, through layers of convolutions.

CNNs work by applying filters (or kernels) to input images to detect local features in the first few layers. As the network deepens, the filters learn more complex, global features, allowing the model to detect patterns like shapes and textures. CNNs have been highly successful in tasks like facial recognition and medical image analysis, and recently there have been developments applying them to waste classification. Their ability to recognize small differences that may not come into consideration for humans makes CNNs a powerful tool for classifying waste items, even those that are visually similar, such as glass and plastics.

However, despite their potential, CNNs face challenges in waste classification. For example, the visual similarities between materials like metals and plastics can lead to misclassification. In addition, many real-world datasets have an uneven number of data in each class, which can affect model performance. This project addresses these challenges by fine-tuning several pre-trained CNN models, such as ResNet50 and MobileNetV3, to classify waste materials more effectively and efficiently, focusing on both accuracy and computational feasibility.

1.3 Challenges and Opportunities in CNNs for Waste Classification

A key challenge is the computational complexity of CNN models. While deep models like InceptionV3 can achieve high accuracy, they require significant computational power and memory, making them less suitable for real-time deployment in resource-limited environments like landfills or waste sorting facilities. On the other hand, lightweight models like MobileNetV3 or ResNet50 are more computationally efficient but may sacrifice some accuracy. This trade-off between model complexity and deployment feasibility is our main area of investigation for this project, offering us the opportunity to find the right balance for real-world deployment.

In summary, this project explores how CNNs can be optimized for waste classification, comparing the performance of different models to determine which provides the best combination of accuracy and computational efficiency. Through this work, we aim to demonstrate the scalability and practicality of CNN-based waste sorting in improving recycling rates and contributing to a more sustainable future.

3. Existing Literature

Multiple studies have been conducted regarding the application of CNNs to waste classification. Of the studies, the RealWaste dataset paper (Single et al., 2023) achieved an 89.19% accuracy using InceptionV3 but noted limitations due to computational load, poor generalization, and misclassification of metals and plastics. Altikat et al. (2022) built a custom CNN that achieved 83% accuracy on binary classification tasks but struggled with deeper architectures due to overfitting. Other works using lightweight models such as ShuffleNet and MobileNet reported reduced accuracy, often due to a lack of fine-tuning or transformations. Our project addresses these limitations by applying strong augmentations and leveraging our knowledge of fine-tuning pre-trained CNNs.

4. Dataset

The *RealWaste* dataset consists of 4,752 images of individual waste items from an Australian landfill. The data was collected as part of the primary study that was reviewed.[2] Each of the jpg images had a label from one of the nine classes, and there was no extreme class imbalance. Some limitations of this dataset are that the data comes from a single landfill, which poses itself as a challenge for generalization. Including the lack of subdivisions in the dataset such as the 'Plastic' class could have been subdivided into clear and opaque plastics.

Image Label	Count
Cardboard	461
Food Organics	411
Glass	420
Metal	790
Miscellaneous	495
Paper	500
Plastic	921
Textile Trash	318
Vegetation	436

Classes and instance counts in the RealWaste Dataset

5. Methodology

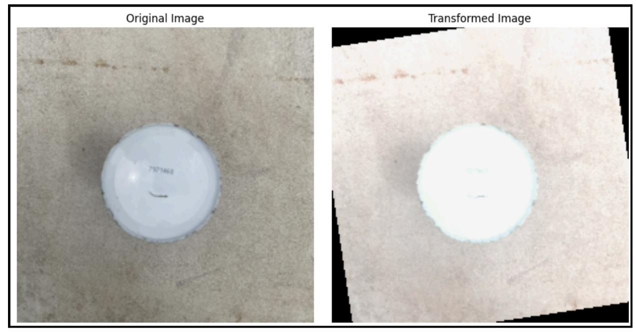
5.1 Setup Details

Firstly, we split our data into train, validate, and test subgroups. 80% of the data was used for training, while 10% were used for validation and testing. We found this to be the perfect ratio since our models were required to inherently learn the structure and patterns of each class, in order to address misclassification and underfitting. Moreover, our hyperparameters were set by initially testing the models to see how quickly the model was learning, with which we were able to set an optimizer, a learning rate, and a number of epochs. Other hyperparameters were a result of trial and error across all models, comparing learning rate schedulers, batch sizes, etc. to yield hyperparameters that worked across all the models. This ensured the integrity of the experiment by setting up hyperparameters that would remain constant throughout. For model evaluation, we focused largely on accuracy since other papers did the same. Confusion matrices were also created to address common misclassifications, such as with plastics and metals. These metrics allowed us to grasp the strengths and weaknesses of each model such as misclassification, overfitting, and dealing with class imbalances.

[2] Single, S., Iranmanesh, S., & Raad, R. (2023). RealWaste: A Novel Real-Life Data Set for Landfill Waste Classification Using Deep Learning. *Information*, *14*(12), 633. https://doi.org/10.3390/info14120633

5.2 Preprocessing & Augmentation

All images were resized to 224x224 pixels to avoid any possible issues with variation in size. To attempt to get a highly accurate and generalizable model, we applied brightness variations, blurring, greyscaling, rotation, mirroring, and random erasing through PyTorch transformation functions.[3] This would simulate real-world variability and improve performance on different datasets. To save memory, training data was transformed and saved to a ‘.pt’ file to be accessed by the DataLoaders. This ensured that our data was not being transformed every training iteration.



Images of a data point before and after transformations

Pre-Trained	
Hyperparameter	**Value**
Loss Function	Cross Entropy Loss
Optimizer	AdamW
Learning Rate	0.01
Weight Decay	0.001
Learning Rate Scheduler	CosineAnnealingLR
Batch Size	64
Epochs	100

5.2 Model Architectures

There were 2 main pre-trained models that were fine tuned: ResNet50 and MobileNetV3Large. For each, the convolutional backbone was frozen and only the final classification head had trainable parameters. 2 other models, RegNetY-3.2GF and ResNet18, were also fine-tuned to get a sense of how intermediate model sizes affected accuracy. Each of the fine-tuned models’ hyperparameters were kept constant to control the learning process. Additionally, these models were pre-trained on the ImageNetV2 dataset.

Scratch Trained	
Hyperparameter	**Value**
Loss Function	Cross Entropy Loss
Optimizer	SGD
Learning Rate	0.02
Weight Decay	0.005
Momentum	0.005
Batch Size	128
Epochs	150

The scratch trained CNN was based loosely on LeNet5. To modernize it, we replaced the tanh activation with ReLU, dedicated max pooling layers replaced with strided convolution, added two additional convolutional layers and batch normalization layers. Hyperparameters were kept constant to control the learning process.

[3] See Appendix

6. Findings and Discussion

Model	Test Accuracy	Parameters
Fine-tuning ResNet50	88.91%	25.6M
Fine-tuning RegNet	80.46%	19.4M
Fine-tuning MobileNetV3	75.42%	5.5M
Training simple CNN	72.06%	1.7M
Fine-tuning ResNet18	65.34%	11.7M

ResNet-50 achieved the highest accuracy (88.91%) and balanced performance across classes. MobileNetV3-Large performed similarly but with significantly fewer parameters, making it more suitable for deployment. Our scratch trained simple CNN had a moderate accuracy (72%) while having only 1.7 million parameters. RegNet also showed strong results, though less accurate than ResNet-50. ResNet-18 had the lowest accuracy, proving that some level of depth and a deep network is required to build a robust model.

6.1 Overfitting and Model Complexity

Fine-tuned models like ResNet50 and MobileNetV3 showed signs of overfitting, a common issue in deep learning models that occurs when a model learns to memorize the training data but fails to generalize to unseen data. While the training loss continued to decrease throughout the training process, the validation accuracy stagnated after a certain point. This suggests that the model had started to memorize the training data instead of learning to generalize, which can lead to poor performance on new data. However, due to the fact that this was a comparison of multiple models, adding an early stopping would not be beneficial as this would mean the number of epochs would differ with each model.

The ResNet50 model, despite achieving high accuracy on the test set, showed a marked difference between its training and validation performance. This discrepancy between training loss and validation accuracy indicates that while the model was becoming increasingly specialized to the training data, it was not improving its ability to recognize patterns in the validation data. In real-world applications, overfitting can be a serious problem, as it affects the model's robustness and ability to handle new or unseen waste types in a waste sorting system.

On the other hand, the simple CNN trained from scratch did overfit to the same extent, likely due to its lower complexity. The simple model, with only 1.7 million parameters, lacked the ability to learn intricate patterns as deeply as ResNet50. However, the model struggled with accuracy because it was not complex enough to learn the training data in the first place. This highlights a critical trade-off: while simpler models may be less accurate, they are less susceptible to overfitting, making them potentially more robust when dealing with new data.

6.2 Confusion Between Plastics and Metals

Another key finding from our experiments was the relatively low confusion between plastics and metals with the ResNet50 and simple CNN models. This was a crucial aspect of the project, as distinguishing between plastics and metals is one of the more challenging tasks in waste classification as we investigated from previous work. The RealWaste dataset contained images of materials that had very similar textures and colors, which often made it difficult for models to classify them accurately.

Interestingly, while ResNet50 had the potential to overfit, its ability to differentiate between these two categories was surprisingly strong. This is due to the model's depth and capacity to capture more complex features in the images. Moreover, the simple CNN, despite its limitations in accuracy, was also able to classify plastics and metals with minimal confusion. This could be due to the fact that the simpler model, while less complex, was less prone to overfitting, allowing it to generalize more effectively to the plastics vs. metals distinction. This finding hints at the fact that while a simple model might not outperform complex ones, it could still hold up surprisingly well for specific tasks where less complexity is needed.

6.3 Impact of Model Complexity on Performance

These findings highlight the trade-off between model complexity and generalization. While ResNet50 and other fine-tuned models demonstrated strong performance in terms of accuracy, they also faced challenges with overfitting, developing the need for careful tuning and regularization practices to prevent performance degradation when deploying models in real-world environments. On the other hand, the simple CNN, despite its lower accuracy, proved to be more stable, suggesting that simpler models may offer better performance in environments with constrained computational resources.

Ultimately, ResNet50 demonstrated the best performance on the training, validation and test sets, but it is clear that the benefits of a powerful model like ResNet50 need to be weighed against the risks of overfitting and computational cost. For real-world deployment, particularly in waste management where computational resources may be limited, simpler models such as the simple CNN may offer an efficient alternative, especially if fine-tuned for specific categories like plastics and metals.

7. Conclusion

In this project, we explored the effectiveness of fine-tuning pre-trained Convolutional Neural Networks (CNNs) for classifying waste items, with the aim of providing an efficient and scalable solution for the waste management industry. Our study leveraged the RealWaste dataset, consisting of 4,752 images across nine categories, to evaluate several models, including ResNet50, MobileNetV3, RegNetY, and ResNet18, as well as a simple CNN trained from scratch. Among the models tested, ResNet50 achieved the highest accuracy of 88.91%, offering a strong balance of performance and computational efficiency.

Despite the challenges faced, such as overfitting, class imbalances, and the difficulty in distinguishing between similar materials like metal and plastic, the results of this study provide valuable insights into the trade-off between accuracy and computational efficiency. Fine-tuning pre-trained models proved to be an effective approach for models to quickly learn patterns in our dataset. On the other hand, lightweight models like MobileNetV3 offered a more computationally efficient solution, achieving reasonable accuracy despite having fewer parameters.

This study is particularly insightful for real-world applications where resources may be limited, such as in landfills or waste management facilities that do not have access to high-end GPUs. Lightweight models, despite slightly lower accuracy, are better suited for deployment in these environments where computational efficiency is crucial.

7.1 Limitations

Through all of the training, we faced computational limits. Due to the limited amount of free Google Colab credits, we were not able to experiment with a more extensive list of hyperparameters. Additionally, even with the assistance of Google Colab, the length of training made it not feasible to explore a wide variety of hyperparameters for out fine-tuning and scratch trained models.

While finetuning only the last layer for ResNet50 and MobileNetV3 helped lessen computational complexity, it may have limited the model's ability to make generalizations for new images. This technique also raises concerns about possible overfitting, as the model depends on just the last layer of the model.

The scratch-trained CNN seemed too simple for the purposes of this task. This is evident in its low accuracy. Furthermore, the closeness of the validation and training loss (appendix 2) suggests that the model was underfitting. This implies that the model lacked the ability to make generalizations on the images due to the model's lack of complexity.

7.2 Future Work

There are several improvements that future work can handle:
Dataset Expansion: A more comprehensive dataset with greater diversity, such as including more granular classifications (e.g., opaque vs. transparent plastics), would enhance model robustness. Expanding the dataset beyond a single landfill would help improve the model's ability to generalize across various environments.

Hyperparameter Tuning: Further experimentation with hyperparameters and model architectures could help achieve better generalization. Additionally, exploring techniques like data augmentation could mitigate issues like overfitting.

Advanced Augmentation Techniques: Implementing more advanced image augmentation strategies, such as introducing artificial noise or adjusting lighting conditions, or even adding dirt overlays, would further improve model robustness for real-world applications.

References

S. Single, S. Iranmanesh, and R. Raad, “RealWaste: A Novel Real-Life Data Set for Landfill Waste Classification Using Deep Learning,” Information (Switzerland), vol. 14, no. 12, 2023, doi: 10.3390/info14120633.

A. Altikat, A. Gulbe, and S. Altikat, “Intelligent solid waste classification using deep convolutional neural networks,” International Journal of Environmental Science and Technology, vol. 19, no. 3, 2022, doi: 10.1007/s13762-021-03179-4.

P. Nowakowski and T. Pamuła, “Application of deep learning object classifier to improve e-waste collection planning,” Waste Management, vol. 109, 2020, doi: 10.1016/j.was man.2020.04.041.

R. Faria, F. Ahmed, A. Das, and A. Dey, “Classification of Organic and Solid Waste Using Deep Convolutional Neural Networks,” in IEEE Region 10 Humanitarian Technology Conference, R10-HTC, 2021. doi: 10.1109/R10-HTC53172.2021.9641560.

A. Jahanbakhshi, M. Momeny, M. Mahmoudi, and P. Radeva, “Waste management using an automatic sorting system for carrot fruit based on image processing technique and improved deep neural networks,” Energy Reports, vol. 7, 2021, doi: 10.1016/j.egyr.2021.08.028.

N. Nnamoko, J. Barrowclough, and J. Procter, “Solid Waste Image Classification Using Deep Convolutional Neural Network,” Infrastructures (Basel), vol. 7, no. 4, 2022, doi: 10.3390/infrastructures7040047.

S. Raschka, cnn-lenet5-cifar10, GitHub repository, 2020
https://github.com/rasbt/stat453-deep-learning-ss20/blob/master/L12-cnns/code/lenet-5/cnn-lenet5-cifar10.ipynb

Appendix

1. Data Augmentations and Transformations

Transformation	Probability	Description
ColorJitter - Brightness	0.5	Dims or brightens image
ColorJitter - Contrast	0.4	Changes brightness between parts of an image
ColorJitter - Saturation	0.4	Changes intensity and vividness of colors in an image
ColorJitter - Hue	0.05	Changes colors in an image
GaussianBlur	0.3	Blurs image with a Gaussian function
RandomGrayscale	0.1	Changes an image to black and white
RandomHorizontalFlip	0.5	Mirrors an image over the Y-axis
RandomErasing	0.4	Removes pixels from a rectangular section of the image

2. Model performance for Simple CNN