Automated Waste Classification Using Deep Learning

Dhruval Bhuva
dhruvalbhuva2000@gmail.com

November 4, 2024

Abstract

This report presents a deep learning approach for automating waste classification at the Grine-Vachine Waste Treatment Center. By employing convolutional neural networks (CNN), residual networks (ResNet), and Vision Transformers (ViT), we aim to classify waste into distinct categories, improving sorting efficiency and minimizing the need for manual labor. Key results demonstrate that our approach significantly enhances sorting accuracy, facilitating operational efficiency at the center.

1 Introduction

The Grine-Vachine Waste Treatment Center currently relies on manual sorting to classify waste before processing, an approach that is both labor-intensive and inefficient. Due to new regulations and staff shortages, automating this process has become a priority. Machine 2, which is responsible for waste classification, lacks an automated sorting algorithm and currently directs waste randomly. Our objective is to develop a deep learning-based solution that enables Machine 2 to classify waste accurately, thereby improving sorting accuracy and reducing manual intervention.

We explore three deep learning architectures: a custom Convolutional Neural Network (CNN), a Residual Network (ResNet), and a Vision Transformer (ViT). Each model is evaluated on its ability to classify waste images into six categories, corresponding to different waste types. This paper discusses the methodologies used, model performances, and the overall impact on sorting efficiency.

2 Methodology

2.1 Dataset Preparation

The dataset consists of 2,527 labeled images of waste items, divided into six categories: cardboard (403 samples), glass (501 samples), metal (410 samples), paper (594 samples), plastic (482 samples), and trash (137 samples). This class imbalance reflects the real-world distribution of waste types encountered at the Grine-Vachine Waste Treatment Center.

To train and evaluate the models effectively, the dataset was split into 80% training and 20% testing sets. This split ensures that the models have sufficient data for learning while maintaining an adequate amount for evaluating generalization performance. Due to the limited sample size, no separate validation set was used.

2.2 Model Architectures

Three distinct architectures were implemented to classify waste images, each designed to explore different deep-learning approaches for image classification:

(a) **CNN:** The custom Convolutional Neural Network (CNN) consists of several convolutional and pooling layers. This architecture includes:

- **Convolutional Layers:** Four convolutional layers with filter sizes of 32, 64, 128, and 256, respectively. Each convolutional layer is followed by ReLU activation to introduce non-linearity, enabling the model to capture complex features.

- **Pooling Layers:** Max pooling layers follow each convolutional layer, reducing spatial dimensions and preserving important features.
- **Fully Connected Layers:** After the convolutional layers, the features are flattened and passed through two fully connected layers, with 128 and 256 neurons, respectively, and ReLU activation.
- **Output Layer:** A final fully connected layer with six neurons (one for each waste category) uses softmax activation for multi-class classification.

CNNs are effective for image classification due to their ability to capture hierarchical features, making this architecture well-suited to identifying visual patterns in waste images.

(b) **ResNet-50 (Basic Implementation):** For the ResNet model, a pre-trained ResNet-50 architecture was fine-tuned on the waste classification dataset.

- **Layers Used:** Only the core convolutional and residual blocks from ResNet-50 were used, with minimal modifications.
- **Final Layers:** The original fully connected layer was replaced with a new classification layer with six output neurons, applying softmax activation.

The use of a pre-trained ResNet-50 allowed for leveraging existing feature representations, which can improve accuracy with limited data by transferring learned knowledge from a larger dataset.

(c) **Vision Transformer (ViT, Basic Implementation):** A pre-trained Vision Transformer model was also fine-tuned on the dataset for waste classification.

- **Patch Embedding and Attention Layers:** Only the standard patch embedding and transformer layers of the pre-trained ViT model were utilized, without additional modifications.
- **Classification Head:** The original classification layer was replaced by a new output layer with six neurons, using softmax activation for multi-class classification.

Fine-tuning the pre-trained Vision Transformer on the waste dataset allows the model to learn task-specific features with limited data, taking advantage of the ViT's global attention mechanism to differentiate waste types.

2.3 Training and Optimization

Each model was trained using the categorical cross-entropy loss function, with accuracy as the primary performance metric. The Adam optimizer was used with a learning rate of 0.0002. Data was fed into the models in batches of 32 to balance memory usage and training efficiency. Each model was trained for 10 epochs.

3 Results

Table 1 presents the accuracy metrics for each model on the test dataset.

Table 1: Model Accuracy on Test Data

Model	Test Accuracy	Precision	Recall
CNN	72%	72%	72%
Vision Transformer (ViT)	85%	86%	85%
ResNet-50	93%	94%	93%

The ResNet-50 model achieved the highest accuracy at 92%, indicating its suitability for waste classification tasks. The CNN model, while less accurate, is computationally efficient, making it a viable option for environments with limited resources. The Vision Transformer model demonstrated strong performance, but with slightly lower accuracy than ResNet, suggesting it may require further fine-tuning or a larger dataset.

4 Conclusion

This study demonstrates the potential of deep learning for automating waste classification at the Grine-Vachine Waste Treatment Center. Among the models tested, ResNet-50 provided the best accuracy, suggesting it is well-suited for implementation on Machine 2 to facilitate automated sorting. The CNN model offers a lighter alternative for scenarios with hardware limitations, while the Vision Transformer could be further explored for scalability and performance improvements.

Future work could involve hyperparameter tuning, model ensemble techniques, and additional data collection to enhance model robustness. Data augmentation techniques can be applied to increase model robustness to variations in waste images. These techniques include random rotations, scaling, and horizontal flipping, which help the model generalize better by simulating real-world variability in waste appearances. By deploying this solution, the Grine-Vachine Waste Treatment Center can achieve greater efficiency and reassign workers to more critical tasks.