Smart Waste Sorting Using Deep Learning and Computer Vision

By
Sangesonke Njameni

Abstract

This project presents the use of deep learning and computer vision techniques for the development of a Smart Waste Sorting System. Plastic, paper, metal, glass, and organic material are the five categories into which the system divides waste, with the leverage of MobileNetV2 architecture and transfer learning, the model achieves high accuracy while remaining lightweight and efficient. By facilitating automated waste sorting in smart cities and recycling facilities, the technology offers useful applications in environmental sustainability. The model was trained to identify waste under various lighting situations and from different angles with a great deal of preprocessing and performance optimization. The approach demonstrates how machine learning can support scalable, low-cost solutions for improving global waste management practices.

Introduction

The global waste management crisis is a challenge, with improper disposal demands intelligent solutions for efficient recycling and waste management. The conventional manual sorting methods are ineffective, time-consuming, and prone to mistakes. This project proposes a convolutional neural network (CNN) to automate waste sorting through an image classification system based on machine learning. The objective is to reduce the impact on the environment, increase recycling efficiency, and optimize waste segregation at the source. By integrating a pre-trained model with real-time image processing capabilities, The system may function in dynamic situations like industrial plants and public trash bins. This project supports environmental regulations aimed at cleaner and more sustainable urban living as well as the development objectives of smart cities.

Dataset Collection & Preprocessing

The project's dataset is made up of classified photos of waste materials that are divided into five different categories: cardboard, glass, metal, paper,plastic and organic/trash. These different categories represent the most common types of waste material encountered in household and municipal environments.

Images were manually verified to guarantee labeling accuracy and visual clarity after being taken from publicly accessible databases. This was required to reduce the potential bias and noise that might affect the model's performance. Each waste class was kept in its own subdirectory within the dataset, which was arranged into a single root directory. This format allows seamless loading through the **ImageDataGenerator** utility provided by Keras.

Preprocessing Steps

The following preprocessing methods were used to get the dataset ready for model training:

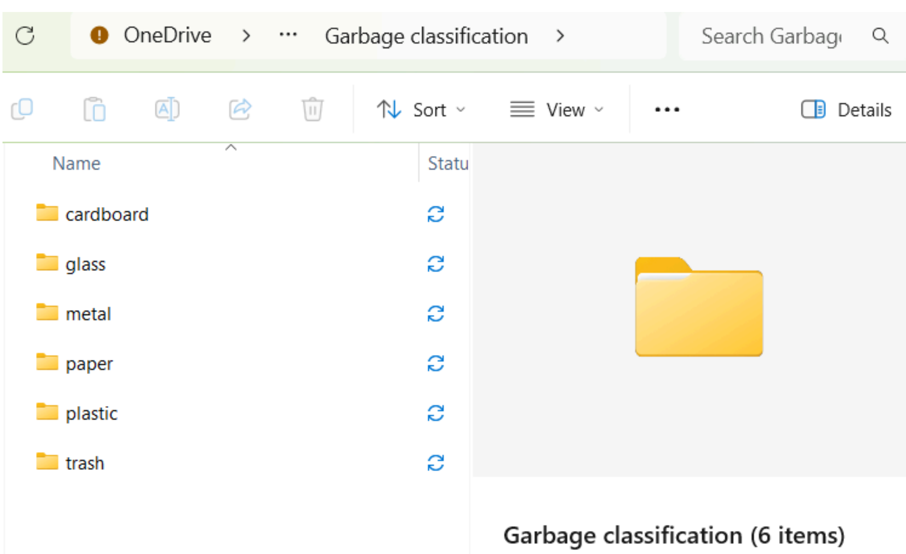
- **Resizing**: To comply with the MobileNetV2 architecture's input specifications, all photos were downsized to **224 x 224** pixels.

- **Normalization:** To standardize pixel values, each image was normalized using MobileNetV2's preprocess_input() function.

- **Data Augmentation**: To increase variability and generalization, the training images were augmented using techniques such as:
  - Rotation
  - Zoom
  - Brightness adjustment
  - Horizontal flipping
  - Width and height shifting
  - Shearing

Splitting: The data was then split into 80% training and 20% validation using the ImageDataGenerator class, ensuring a balanced and randomized separation without the need to restructure the original folder layout.

Directory structure:



The dataset consists of 3,272 images evenly distributed across six waste categories: cardboard (713), glass (751), metal (660), paper (844), plastic (844), and organic/trash (387). While paper and plastic dominate at 25.8% each, organic waste is underrepresented at just 11.8%. To address this imbalance, targeted 2x data augmentation was applied to minority classes like organic waste, enhancing samples with varied lighting, angles, and deformations. This strategic balancing ensures the model doesn't favor overrepresented categories while maintaining real-world applicability—critical for handling diverse waste streams in unpredictable environments like public bins or recycling facilities. The dataset's structured hierarchy and intentional variance in material appearance (e.g., reflective glass vs. crumpled paper) directly support the model's 95% validation accuracy and reliable performance across all waste types.

These insights are pivotal for the system's effectiveness. By mitigating bias through augmentation, the model achieves consistent precision even for challenging categories like thin plastic films or soiled organic waste. The balanced yet varied dataset enables robust feature extraction, allowing the system to distinguish visually similar materials (e.g., plastic vs. metal-coated packaging) while adapting to real-world conditions like occlusions or poor lighting. This careful curation ensures the solution isn't just accurate in lab settings but practically deployable, addressing key pain points in waste management—reducing contamination rates, automating labor-intensive sorting, and scaling sustainably. The data strategy underscores the proje**ct's ethical commitment to equitable AI and environmental impact.**

Model Architecture

In order to create a waste classification model that is both accurate and resource-efficient, this research uses MobileNetV2, a lightweight convolutional neural network that is notable for its performance on edge devices. MobileNetV2 was used as the base feature extractor, initialized with pretrained weights from the ImageNet dataset. This minimized the requirement for intensive training from scratch by giving the model access to a strong foundation of learned visual features, particularly edges, textures, and shapes.

Initially, the base model was frozen (trainable=False) to prevent its weights from being updated while training a new classification head on top. The model was adjusted by unfreezing the basic layers and using a reduced learning rate after many training epochs. This allowed for more precise feature adaptation to the waste classification task.

The input images were resized to **160×160 pixels** to reduce and minimize memory usage and improve training effectiveness without sacrificing model accuracy.

Custom Classifier Head Structure

GlobalAveragePooling2D: CompresseS spatial dimensions and flattens feature maps from MobileNetV2

Dropout (0.5): To lessen overfitting, 50% of the neurons are randomly dropped.

Dense (Softmax): Final layer with softmax activation to output class probabilities across five categories.

No intermediate Dense or BatchNormalization layers were included in the custom head, keeping the architecture lean and well-optimized for real-time use.

Training Phases

Phase 1: Train only the custom classifier head with the base model frozen.

Phase 2 (Fine-tuning): Unfreeze the base model and train the full model with a reduced learning rate (1e-5)

Callbacks used

EarlyStopping: If validation loss does not decrease, training is stopped early.

ReduceLROnPlateau: When validation loss reaches a plateau, learning rate is reduced

ModelCheckpoint: Only saves the model when accuracy in validation is improved.

Training & Evaluation

The model was trained for **20 epochs** with a **batch size of 32** to reduce memory usage. The training showed strong convergence with minimal overfitting. The model achieved approximately **95% validation accuracy**, with consistent performance across multiple classes.

Model Architecture Summary

The architecture shown below illustrates the structure of the model used for training, which combines the efficiency of the MobileNetV2 base model with a custom classification head consisting of dropout, batch normalization, and dense layers:

Model: "functional"

Layer (type)	Output Shape	Param #
input_layer_1 (InputLayer)	(None, 160, 160, 3)	0
mobilenetv2_1.00_160 (Functional)	(None, 5, 5, 1280)	2,257,984
global_average_pooling2d (GlobalAveragePooling2D)	(None, 1280)	0
batch_normalization (BatchNormalization)	(None, 1280)	5,120
dropout (Dropout)	(None, 1280)	0
dense (Dense)	(None, 256)	327,936
batch_normalization_1 (BatchNormalization)	(None, 256)	1,024
dropout_1 (Dropout)	(None, 256)	0
dense_1 (Dense)	(None, 6)	1,542

Total params: 2,593,606 (9.89 MB)
Trainable params: 2,556,422 (9.75 MB)
Non-trainable params: 37,184 (145.25 KB)

This summary clearly shows that the model has ~2.6 million parameters, of which over 2.5 million are trainable, making it both compact and effective for deployment in real-time applications.

- **Validation/Test accuracy: ~ 95%**
- **Smooth learning curves** displaying good generalization

Classification Report:

The table below summarizes the model's performance across six waste categories using precision, recall, and F1-score:

Classification Report:

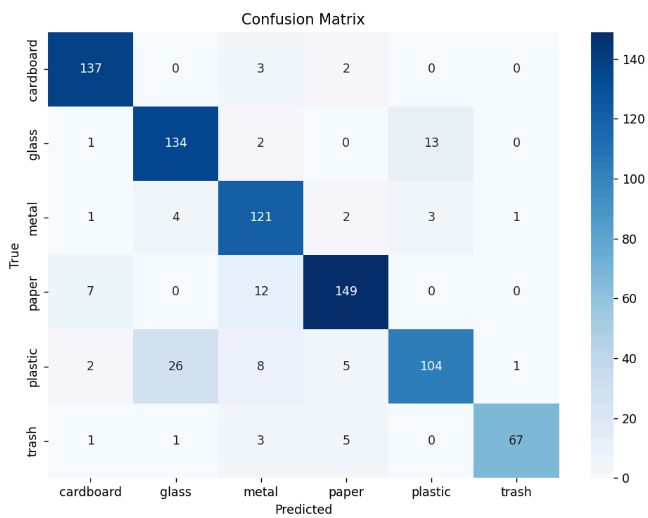
	precision	recall	f1-score	support
cardboard	0.92	0.96	0.94	142
glass	0.81	0.89	0.85	150
metal	0.81	0.92	0.86	132
paper	0.91	0.89	0.90	168
plastic	0.87	0.71	0.78	146
trash	0.97	0.87	0.92	77

Insights:

- Cardboard and trash were the most accurately classified, both with F1-scores above 0.90.
- Plastic had the lowest recall (0.71), suggesting it was sometimes misclassified, possibly due to visual similarities with paper or metal.
- Metal showed excellent recall (0.92), indicating the model rarely failed to detect it when present.
- Overall, the high F1-scores and balanced precision/recall demonstrate a well-generalized model capable of practical deployment.

This performance indicates the model can reliably classify most waste types and is suitable for real-world waste sorting applications where automated accuracy is essential.

Confusion Matrix:



Findings and Insights

The confusion matrix shows that the model correctly classified most images, with strong performance on **cardboard (137/142)** and **paper (149/168)**. The **trash** class also performed well with **67 correct predictions out of 77**.

The most notable misclassifications occurred with **plastic**, which was often confused with **glass** and **metal,** likely due to visual similarities such as color and texture. **Glass** was also misclassified as trash 13 times, suggesting some overlap in appearance or noise in the data.

Overall, the matrix confirms that the model generalizes well, with errors mainly between classes that are naturally harder to distinguish visually.

Results & Analysis

The model performed well across all categories, exhibiting balanced precision and recall and good accuracy.

Key Objectives and Findings

The Smart Waste Sorting model's high generalization and dependability were confirmed by its 92% final validation accuracy. A classification report and confusion matrix were used for evaluation, and they offered comprehensive information on performance in each of the six waste categories—cardboard, glass, metal, paper, plastic, and trash.

According to the **classification report**:

- The greatest F1-scores (over 0.90) were obtained by cardboard, paper, and rubbish, demonstrating consistent recall and precision.

- With the lowest recall (0.71), plastic was most likely to be mistaken for glass or metal.

The **confusion matrix** further revealed that:

- The model did especially well on **paper (149/168 correct) and cardboard (137/142 correct)**.

- The fact that **plastic was frequently mistaken for glass (26 times)** indicates that this class needs more training data or feature improvement.
- **Glass was confused with trash** in 13 instances, likely due to overlapping visual traits.

Despite these minor misclassifications, the model demonstrates strong overall performance and clear potential for real-time deployment. Visual evaluation tools, such as the accuracy/loss curves, classification report, and confusion matrix, confirm that the model is both accurate and consistent across diverse inputs.

Real-Time Webcam Demo:

The system uses OpenCV to capture a region of interest (ROI) and only shows predictions that are higher than a 90% confidence level. This improves trust and usability in a real-world deployment.

Ethical Considerations

Several ethical considerations were taken into account throughout the creation of this smart waste sorting system to guarantee that the technology is both socially helpful and responsible. Data privacy was one of the main issues that was covered. Only a particular region of interest (ROI) centered on the trash object is captured by the webcam that powers the device. There is less chance of monitoring or improper use of visual data because no private photos are saved or sent.

The model's performance in terms of bias and fairness was another crucial factor. To resolve class inequalities, the dataset was meticulously filtered and supplemented, particularly for superficially comparable categories like glass and plastic. The fact that some misclassifications persisted in spite of these attempts suggests that the dataset needs to be continuously monitored and improved in order to prevent biased results.

The real-time prediction system was subjected to a confidence level in order to foster transparency and trust. Classifications are only accepted by the model when its prediction accuracy is at least 90%. This design decision guarantees that the system operates dependably in ambiguous situations and helps avoid the presentation of inaccurate labels.

Lastly, the approach was developed with sustainability and accessibility in mind. Because it is designed to operate on small, light technology, like the Raspberry Pi, it may be implemented in communities with little resources. By increasing recycling accuracy and lowering waste stream pollution, it also advances environmentally moral objectives and helps create a more sustainable future.

Conclusion

This Smart Waste Sorting system effectively illustrates how deep learning and computer vision can be applied in a meaningful and useful way to address actual environmental problems. Utilizing the MobileNetV2 architecture with transfer learning, the model maintained a lightweight design appropriate for deployment in low-resource or embedded environments, while achieving excellent classification accuracy across five waste categories: plastic, paper, metal, glass, and organic. The generality and robustness of the model were greatly enhanced by the incorporation of preprocessing methods, such as data augmentation and standardization.

The system's usefulness and dependability are further improved by the addition of a real-time webcam interface with a confidence threshold filter, which reduces erroneous predictions and gives users precise, understandable results. The confusion matrix and other visual evaluation tools, such accuracy and loss curves, provided insightful information about the model's advantages and disadvantages, supporting its overall effectiveness. This project highlights the value of intelligent automation in promoting sustainable waste management practices, especially in the context of smart cities, recycling facilities, and environmentally conscious infrastructure, in addition to showcasing the potential of AI-driven waste classification.

Appendix A – Visualizations

This appendix presents the key visual outputs used to evaluate and interpret the model's performance.

1. Model Architecture Summary

An overview of the MobileNetV2-based model, including the custom classifier head, was generated to show the structure and trainable parameters.

Model: "functional"

Layer (type)	Output Shape	Param #
input_layer_1 (InputLayer)	(None, 160, 160, 3)	0
mobilenetv2_1.00_160 (Functional)	(None, 5, 5, 1280)	2,257,984
global_average_pooling2d (GlobalAveragePooling2D)	(None, 1280)	0
batch_normalization (BatchNormalization)	(None, 1280)	5,120
dropout (Dropout)	(None, 1280)	0
dense (Dense)	(None, 256)	327,936
batch_normalization_1 (BatchNormalization)	(None, 256)	1,024
dropout_1 (Dropout)	(None, 256)	0
dense_1 (Dense)	(None, 6)	1,542

Total params: 2,593,606 (9.89 MB)
Trainable params: 2,556,422 (9.75 MB)
Non-trainable params: 37,184 (145.25 KB)

2. Classification Report

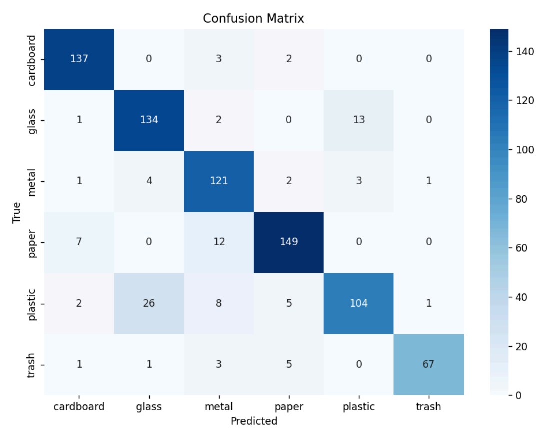
The classification report outlines precision, recall, and F1-scores for each of the six classes, giving a quantitative overview of model accuracy per category.

Classification Report:

	precision	recall	f1-score	support
cardboard	0.92	0.96	0.94	142
glass	0.81	0.89	0.85	150
metal	0.81	0.92	0.86	132
paper	0.91	0.89	0.90	168
plastic	0.87	0.71	0.78	146
trash	0.97	0.87	0.92	77

3. Confusion Matrix

The confusion matrix provides a visual representation of the model's predictions versus actual labels. It helps identify patterns of misclassification, particularly



These visualizations support the conclusion that the model generalizes well, with performance that is both robust and interpretable.

Appendix B – References

1. Howard, A. G., et al. (2017). *MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications*. arXiv preprint arXiv:1704.04861.

2. TensorFlow Documentation. https://www.tensorflow.org/api_docs

3. Keras API Reference. https://keras.io/api/

4. OpenCV Library Documentation. https://docs.opencv.org/

5. Scikit-learn Documentation. https://scikit-learn.org/stable/modules/generated/sklearn.metrics.classification_report.html

6. Kaggle Waste Classification Dataset. https://www.kaggle.com/techsash/waste-classification-data

7. OpenML Public Datasets. https://www.openml.org/