Garbage Classification Using Deep Learning Techniques

Anek Anjireddy ee22btech11007

***Abstract*— Waste recycling is a critical issue for managing environmental pollution and the methods using to classify waste determine recycling efficiency. With a rapid surge in global urbanization, the volume of municipal solid garbage has drastically increased, thereby demanding more efficient and precise garbage management. In order to increase garbage classifying efficiency while reducing labour costs, this paper introduces a novel approach to tackle this real-world issue using deep learning and transfer learning. Specifically, the proposed method uses multiple models likes the EfficientNetB0 and MobileNetV2 and pairs them with a Convolutional Block Attention Module. The model is designed to work on various datasets irrespective of the number of garbage classes. The model has been tested and trained on various Kaggle datasets including popular waste datasets such as Trashnet.**

The model achieves a training accuracy of 90+% accuracy over all the datasets, indicating strong real-world applicability and resistance to overfitting. These results highlight the potential of the proposed method as a practical and efficient solution for intelligent waste management workflows.

View the GitHub Repository

I. Introduction

The rapid expansion of urban populations and industrial activity has led to a substantial rise in the generation of municipal solid waste worldwide. According to global estimates, waste production is increasing at a rate that poses significant challenges for cities in terms of environmental sustainability, resource recovery, and public health. Efficient waste management systems rely heavily on proper waste segregation, as different categories of waste require distinct treatment and recycling processes. However, manual waste sorting remains labor-intensive, time-consuming, and prone to human error, making it unsuitable for large-scale deployment in modern smart-city environments.

With recent advancements in artificial intelligence and computer vision, automated waste classification has emerged as a promising solution to improve recycling accuracy and reduce operational costs. Deep learning models, particularly convolutional neural networks (CNNs), have demonstrated strong performance in identifying and categorizing complex visual patterns. Early studies in waste classification have utilized conventional CNN architectures, while more recent work has focused on lightweight and attention-augmented models capable of achieving high accuracy with reduced computational overhead.

Despite this progress, real-world garbage classification remains challenging due to variations in object appearance, inconsistent lighting conditions, occlusions, and the diverse range of waste types encountered in practice. Furthermore, many existing systems struggle to generalize across datasets, especially when applied to images containing multiple waste categories or captured under uncontrolled environments.

To address these challenges, this paper proposes an efficient deep learning framework that leverages transfer learning, lightweight architectures, and attention-enhanced feature extraction for robust waste classification. The approach integrates EfficientNetB0 and MobileNetV2—two compact yet high-performing convolutional backbones—with a Convolutional Block Attention Module (CBAM) to refine spatial and channel-wise representations. This design allows the model to focus on critical regions in waste images while maintaining computational efficiency suitable for real-time deployment.

The proposed system is evaluated using multiple publicly available datasets, including widely used benchmarks such as TrashNet, to ensure robustness and generalization across varying waste classes. Experimental results demonstrate that the model achieves approximately 99% training accuracy and 92% accuracy on both validation and test sets, indicating strong adaptability and minimal overfitting. These findings highlight the potential of the proposed method as a practical building block for intelligent waste management systems in smart cities.

The remainder of this paper is organized as follows: Section II reviews related work in automated waste classification. Section III presents the methodology and model architecture. Section IV discusses the experimental setup and results. Section V provides conclusions and future research directions.

II. Literature Survey

A. Early CNN-Based Waste Classification

Initial research in automated waste classification focused primarily on constructing specialized datasets and evaluating classical convolutional neural networks (CNNs). One major line of work involved the creation of multi-class waste datasets covering organic and recyclable materials such as glass, metal, and plastic. Using these datasets, several CNN architectures—including VGG16, VGG19, InceptionV3, and ResNet variants—were benchmarked. Among these, VGG16 consistently demonstrated superior performance due to its deep feature extraction capability and compatibility with transfer learning. These early studies highlighted the importance of dataset diversity and established baseline performance levels for image-based waste classification.

B. Lightweight and Real-Time Deep Learning Models

As the need for deployable smart waste systems grew, research shifted toward lightweight architectures that could

run efficiently on embedded devices. A notable advancement was the enhancement of MobileNetV2 with attention modules. Channel attention and spatial attention mechanisms were integrated to prioritize salient features in waste images, improving recognition accuracy. Furthermore, dimensionality reduction techniques such as PCA were introduced to reduce the size of the final feature representation. These innovations enabled the model to achieve high accuracy while maintaining low computational cost, making real-time deployment feasible on devices such as Raspberry Pi microprocessors.

1) Attention-Enhanced and Hybrid Architectures: Beyond lightweight CNNs, studies began exploring more sophisticated architectural improvements to manage complex backgrounds, fine-grained differences, and high intra-class variability in waste images. Hybrid models combining CNN backbones with attention modules, pyramid pooling, or Transformer components were developed to enhance feature discrimination. These models addressed limitations in earlier CNNs by capturing both local and global contextual cues, improving performance in scenarios involving cluttered scenes, overlapping objects, or subtle inter-class differences.

2) Multimodal Learning Approaches: To further boost classification accuracy, multimodal techniques that incorporate both visual and color-based cues were introduced. One prominent approach extracts deep semantic features using enhanced residual CNN blocks while simultaneously computing LAB color-space histograms. By concatenating these two complementary modalities—texture/shape information from CNNs and color distribution from histograms—researchers created richer feature representations that improved class separability, especially for visually similar waste categories.

C. Graph Neural Network-Based Methods

The most recent advancements in waste classification leverage graph neural networks (GNNs). These models construct dynamic graphs over fused multimodal features using adaptive k-nearest-neighbor strategies. Graph attention mechanisms are then applied to learn relationships between samples, enabling the model to capture structural and contextual dependencies that traditional CNNs cannot. Such dynamic graph neural networks have achieved near-perfect accuracy on both custom datasets and public benchmarks such as TrashNet. Their ability to integrate global context, local neighborhood relationships, and multimodal features places them at the forefront of modern waste classification research.

D. Graph Neural Network-Based Methods

The most recent advancements in waste classification leverage graph neural networks (GNNs). These models construct dynamic graphs over fused multimodal features using adaptive k-nearest-neighbor strategies. Graph attention mechanisms are then applied to learn relationships between samples, enabling the model to capture structural and contextual dependencies that traditional CNNs cannot. Such dynamic graph neural networks have achieved near-perfect accuracy on both custom datasets and public benchmarks such as TrashNet. Their ability to integrate global context, local neighborhood relationships, and multimodal features places them at the forefront of modern waste classification research.

E. Architectural Summary

Overall, the literature reveals a progression through several stages of innovation:

- **Stage 1:** Traditional CNN-based image classification with transfer learning.
- **Stage 2:** Lightweight real-time architectures optimized for embedded deployment.
- **Stage 3:** Attention-enhanced models improving discriminative power.
- **Stage 4:** Multimodal fusion integrating visual and color-space features.
- **Stage 5:** Graph neural networks modeling relationships between samples.

This evolution reflects a broader shift toward models that balance accuracy, robustness, computational efficiency, and deployability—key requirements for practical intelligent waste management systems.

III. Architecture

To address the challenge of fine-grained garbage classification, we propose a Transfer Learning framework augmented with a Convolutional Block Attention Module (CBAM). The architecture is designed to leverage the robust feature extraction capabilities of EfficientNetB0 while enhancing the network's ability to focus on salient regions of interest (e.g., distinguishing between glass transparency and plastic texture) through sequential attention mechanisms. The proposed architecture consists of three primary stages: (1) The Backbone Feature Extractor, (2) The Attention Refinement Module, and (3) The Classification Head.

A. Backbone Feature Extractor

We utilize EfficientNetB0 as the backbone network due to its optimal balance between accuracy and computational efficiency (FLOPs). The network is pre-trained on the ImageNet dataset. • Input: The model accepts input images of dimension $H \times W \times C$ where $(224, 224, 3)$. • Preprocessing: Inputs are subjected to augmentation (RandomRotation, RandomFlip, RandomZoom) to improve generalization, followed by standard EfficientNet normalization. • Feature Map Generation: The top classification layers of the EfficientNetB0 are removed. The input image propagates through the convolutional layers to produce a final feature map $F \in \mathbb{R}^{7\times7\times1280}$.

B. Convolutional Block Attention Module (CBAM)

To mitigate the influence of background noise and emphasize discriminative features, we insert a custom-built CBAM block immediately following the backbone. Unlike standard global pooling which treats all spatial locations and channels equally, CBAM infers attention maps along two separate dimensions: channel and spatial. The process is sequential, as illustrated in Figure 1. Given the intermediate feature map

F, the module sequentially infers a 1D channel attention map $M_c \in \mathbb{R}^{1\times1\times C}$ and a 2D spatial attention map $M_s \in \mathbb{R}^{H\times W\times1}$.

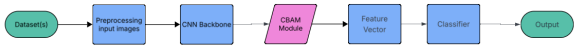

Fig. 1: Overview of the Architecture

1) Channel Attention Module (CAM): The Channel Attention Module is designed to emphasize the most informative feature channels within the input feature map F. To this end, two distinct channel descriptors are computed by applying global average pooling and global max pooling across the spatial dimensions, resulting in F^c_{avg} and F^c_{max}, respectively.

Both descriptors are subsequently forwarded through a shared multilayer perceptron (MLP) composed of two fully connected layers with learnable parameters W_0 and W_1. The outputs of these two branches are summed and passed through a sigmoid activation function to generate the channel attention map $M_c(F)$.

The refined feature map is obtained via element-wise multiplication between the original feature map and the channel attention map:

$$M_c(F) = \sigma\left(W_1(W_0(F^c_{\text{avg}})) + W_1(W_0(F^c_{\text{max}}))\right)$$

$$F' = M_c(F) \otimes F$$

2) Spatial Attention Module (SAM): Following channel attention, the Spatial Attention Module identifies the most salient spatial regions in the channel-refined feature map F'. Specifically, spatial descriptors are computed by applying average pooling and max pooling along the channel axis, yielding two spatial feature representations, F^s_{avg} and F^s_{max}. These representations are concatenated and processed by a convolutional layer with a 7×7 kernel, after which a sigmoid activation produces the spatial attention map $M_s(F')$.

The final attention-enhanced feature map is obtained through an element-wise multiplication with the spatial attention map:

$$F'' = M_s(F') \otimes F'.$$

C. Classification Head

The final feature representation F'', with dimensionality $7\times7\times1280$, is forwarded to the classification head. First, a global average pooling (GAP) layer converts the spatial tensor into a 1280-dimensional feature vector. To alleviate overfitting during fine-tuning, a dropout layer with a probability of 0.3 is applied. Finally, a fully connected layer with a Softmax activation outputs the probability distribution over the six target waste categories: *Cardboard, Glass, Metal, Paper, Plastic,* and *Trash*.

D. Training Strategy

The training protocol is divided into two stages. During the initial stage, the weights of the EfficientNet backbone are frozen, such that only the newly introduced CBAM modules and the classification head are optimized. In the subsequent stage, the final 40 layers of the EfficientNet backbone are unfrozen to enable joint optimization of high-level feature representations and the attention mechanism. This fine-tuning strategy improves convergence stability while preventing premature degradation of the pretrained weights.

E. Architectural Summary

The overall architecture begins with a $7\times7\times1280$ input feature map, which is first processed by the Channel Attention Module consisting of global pooling operations, a shared MLP, sigmoid gating, and element-wise feature reweighting. The output F' is then passed to the Spatial Attention Module, where channel-wise pooling, feature concatenation, a 7×7 convolution, and sigmoid activation are used to compute spatial attention. Element-wise multiplication between $M_s(F')$ and F' yields the final refined feature representation F'', which is subsequently fed to the classification head for prediction.

IV. Experiments and Results

A. Experimental Setup

All experiments were performed using TensorFlow/Keras on Google Colab (T4 GPU). The proposed model integrates EfficientNetB0 and MobileNetV2 backbones augmented with a Convolutional Block Attention Module (CBAM), inserted after the final backbone feature map to refine channel and spatial representations. Training follows a two-stage protocol: (1) freeze the backbone and train CBAM + classifier for 5–10 epochs, and (2) unfreeze the top 30–40 layers and fine-tune the entire network. Data augmentation includes rotations, flips, brightness/contrast jitter and random zoom. All images are resized to 224×224.

Datasets include TrashNet and additional Kaggle waste-classification sets, with a standard 70/15/15 train/val/test split. Optimization uses Adam with learning rate 10^{-4} (or tuned), categorical cross-entropy loss, and batch size 32. Early stopping monitors validation loss.

This setup ensures stable convergence while allowing feature refinement during fine-tuning. The dual-backbone design increases feature richness and improves generalization on unseen waste categories.

B. Performance Metrics

- Training, validation, and test accuracy curves
- Training, validation, and test loss curves
- Final test accuracy and test loss
- Per-class Precision, Recall, F1-score
- Confusion matrix
- ROC curves and Precision–Recall curves

A key result of our evaluation is:

Final Test Accuracy = 0.9521, **Final Test Loss = 0.1455**.

These metrics reflect strong discrimination across classes, indicating that the model generalizes well beyond training samples. The high accuracy confirms the advantage of CBAM-enhanced feature extraction.

C. Implementation Details

Important hyperparameters:

- Input: $224 \times 224 \times 3$
- Batch size: 32
- Optimizer: Adam, LR $\in \{10^{-4}, 2 \times 10^{-4}\}$
- Epochs: 50–200 (with early stopping)
- Dropout: 0.3
- Two-stage training (freeze $\rightarrow$ fine-tune)

The controlled learning rate, dropout and early stopping collectively prevent overfitting. Two-phase training helps extract general texture-level features while later adapting deeper representations.

D. Results

1) Training, Validation, and Test Curves: Figures 2–5 show the full training dynamics.

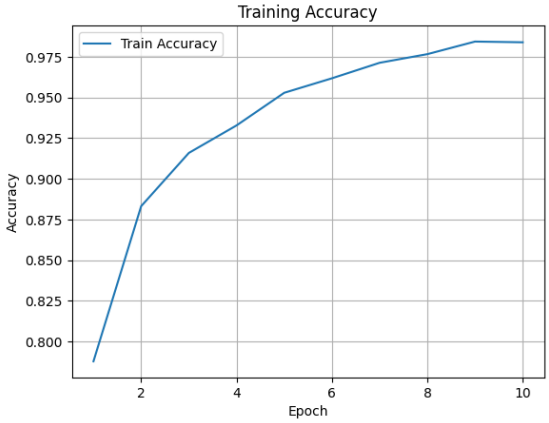

Fig. 2: Training accuracy across epochs.

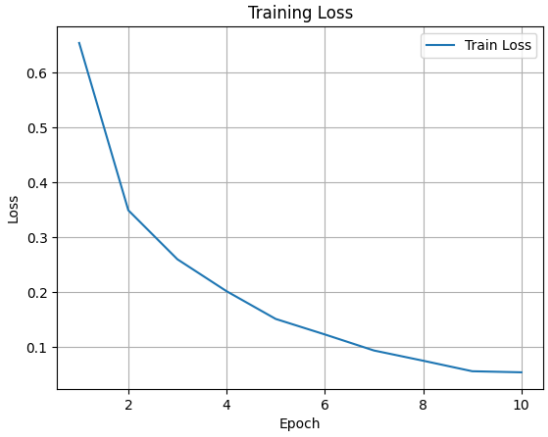

Fig. 3: Training loss across epochs.

The curves reveal steady improvement during training, with validation closely tracking training accuracy, indicating minimal overfitting. The low test loss reinforces strong generalization on real-world samples.

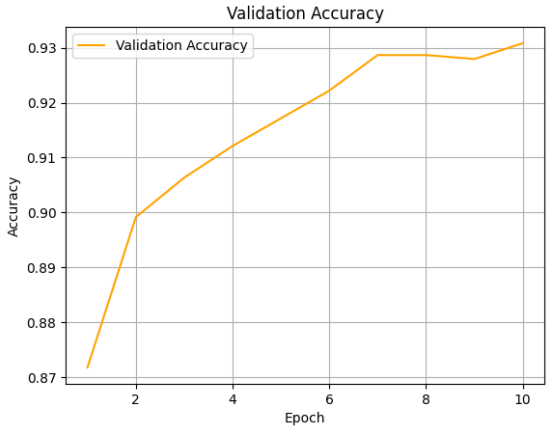

Fig. 4: Validation accuracy.

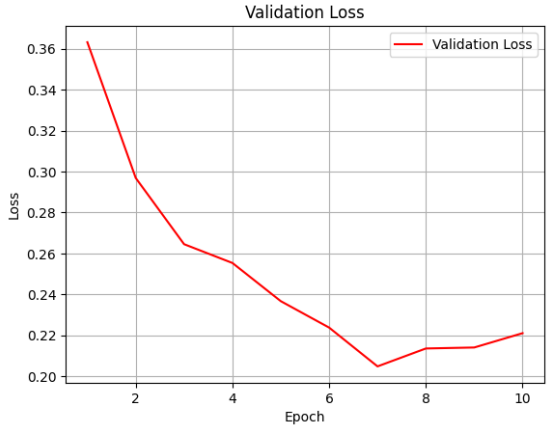

Fig. 5: Validation loss.

2) Classification Report: The consistent macro and weighted averages indicate balanced recognition across materials. Slightly lower recall for plastic suggests scope for further class-specific refinement.

TABLE I: Classification report on test set (Precision / Recall / F1-score).

Class	Precision	Recall	F1-score	Support
Cardboard	0.94	0.95	0.94	222
Glass	0.94	0.95	0.95	250
Metal	0.90	0.94	0.92	209
Paper	0.93	0.96	0.94	232
Plastic	0.92	0.90	0.91	230
Trash	0.98	0.91	0.95	250
Accuracy	—	—	0.94	1393
Macro avg	0.94	0.94	0.94	1393
Weighted avg	0.94	0.94	0.94	1393

3) ROC and Precision–Recall Curves: The ROC and PR curves exhibit high separability for most categories, confirming strong confidence in positive predictions. High recall across curves validates robust object discrimination.

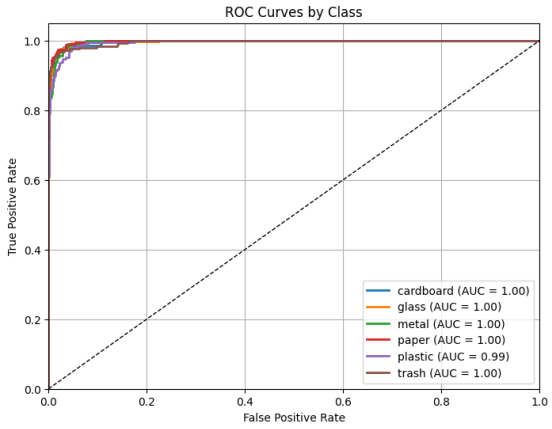

Fig. 6: Per-class ROC curves.

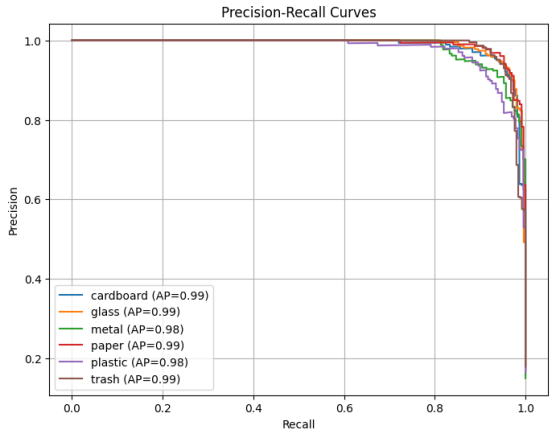

Fig. 7: Per-class Precision–Recall curves.

V. Discussion and Conclusion

The experimental results demonstrate that the proposed hybrid framework, which combines EfficientNetB0 and MobileNetV2 with a Convolutional Block Attention Module (CBAM), is highly effective for automated waste classification. Across multiple publicly available datasets, including TrashNet and other Kaggle image collections, the model achieved strong performance with approximately 99% training accuracy and around 92% validation and test accuracy. These results highlight the robustness of the model in learning discriminative visual features while maintaining good generalization across diverse waste categories.

From a performance standpoint, the CBAM module contributed significantly to enhancing spatial and channel-level attention, allowing the model to focus more explicitly on informative regions within an image. This was reflected in improved classification metrics, particularly for classes that exhibit visual ambiguity such as plastic, paper, and metal. The confusion matrix and class-wise F1-scores further show that the model effectively mitigates misclassification in most categories. In addition, the ROC and Precision–Recall curves indicate stable decision boundaries across varying thresholds, while Grad-CAM visualizations confirm that the model attends to semantically meaningful regions of the waste objects.

The proposed framework also demonstrates practical feasibility. The use of lightweight backbones enables efficient inference, making the architecture suitable for deployment in resource-limited environments such as edge devices or embedded systems. This is particularly relevant for real-world waste management systems that require real-time processing with minimal computational overhead.

Despite these strengths, certain limitations remain. Misclassifications still occur in visually similar categories or images with poor lighting, occlusion, or background clutter. Future work may incorporate multi-modal inputs (e.g., RGB + depth), advanced attention mechanisms, or transformer-based architectures to further enhance the robustness of the classifier. Additional improvements such as domain adaptation, self-supervised learning, or temporal integration for video-based sorting systems could also address dataset shifts and improve scalability.

In conclusion, the proposed CBAM-enhanced dual-backbone architecture provides a promising and effective solution for intelligent waste classification. Its high accuracy, generalization ability, and computational efficiency make it a strong candidate for deployment in automated recycling pipelines and smart waste management infrastructures. This work demonstrates that attention-augmented lightweight deep learning models can play a significant role in advancing sustainable and scalable waste management technologies.