Garbage Segmentation and Classification using Mask R-CNN and EfficientNet

EcoSort

By: Harsh Patel, Adem Chylah, Sharat Krishnan

harshd.patel@mail.utoronto.ca

adem.chlyah@mail.utoronto.ca

sharat.krishnan@mail.utoronto.ca

CSC490

https://github.com/Sharat21/EcoSort

Anthony Bonner

Abstract

Garbage classification and segmentation are crucial for efficient waste management. This paper presents a dual-stage pipeline approach to tackle identifying garbage on a conveyor belt and classifying it as one of 6 classes. In addition to that, the classification also identifies if the garbage item is biodegradable or not and incorporates subclassification on different kinds of plastic. This study addresses challenges in classification accuracy due to reflective surfaces and class imbalances while focusing on classifying garbage items relative to conveyor belts to focus on application at recycling plants. Automating this process at recycling plants would reduce the need for human labour, which is more practical at recycling plants that have more hazardous waste. By utilizing a segmentation and classification pipeline, this information can be passed to a robot or automated sorter to sort garbage. By creating a synthetic dataset of garbage items on a conveyor belt, the segmentation and classification models can be trained on that dataset to yield a higher pipeline accuracy to prepare for real-world scenarios.

1. Introduction

1.1 Problem Statement

Sorting garbage manually is an inefficient and labour-intensive process that poses several challenges. Manual sorting is prone to errors due to worker fatigue and inconsistency, leading to misclassification and contamination of recyclable materials. The process is also time-consuming, limiting the throughput of waste processing facilities. There is also an inherent health risk when working with trash due to exposure to hazardous materials. To overcome these limitations, AI-based automation provides a viable alternative by enhancing sorting speed and accuracy. Automated systems can operate continuously without human intervention, reducing labour costs and improving overall efficiency. By integrating machine learning into waste-sorting facilities, we can develop a more reliable and scalable solution to waste classification and recycling.

1.2 Motivation

Key motivations for this research include:

- **Enhancing Recycling Efficiency**: AI-driven sorting systems improve the accuracy of waste classification, leading to higher recycling rates and better resource utilization.
- **Reducing Landfill Waste**: Proper waste categorization prevents nonrecyclable materials from contaminating recyclable waste streams, minimizing landfill accumulation.
- **Improving Workplace Safety**: Automation reduces direct human exposure to hazardous waste, lowering health risks for workers in waste sorting facilities.
- **Increasing Sorting Throughput**: AI-based models operate at higher speeds than manual sorting, enabling large-scale waste processing with minimal human intervention.

1.3 Input and Output

- **Input**: High-resolution images of waste items on a conveyor belt, captured from a bird's-eye view perspective.
- **Output**:

    - **Segmented Waste Items**: Individual trash objects detected and isolated from the conveyor belt background.
    - **Classified Labels**: Each detected item is categorized into one of six waste classes.
    - **Sub-Classified Labels**: Each detected plastic item is categorized into a type of plastic
    - **Biodegradability Prediction**: Each classified item is labelled as biodegradable or non-biodegradable.

1.4 Learning Problem

The problem is formulated as a supervised learning task with distinct segmentation and classification components:

- **Segmentation Problem**: The model is trained to detect and isolate individual waste items using annotated bounding boxes.
- **Classification Problem**: Each segmented object is assigned a predefined category, and its biodegradability is determined.
- **Optimization Objective**: The system must balance classification accuracy and computational efficiency to enable real-time deployment in waste-sorting facilities.

1.5 Challenges

Despite the benefits of AI-based waste classification, several challenges must be addressed to develop a strong system:

1. **Overlapping Objects**: Waste items often appear in clusters on conveyor belts, making object segmentation difficult.
2. **Dataset Limitations**: Existing waste classification datasets do not fully capture the nature of a conveyor belt environment
3. **High Intra-Class Variation**: Garbage items have variations in shape, texture, and lighting conditions, making it harder to classify
4. **Computational Constraints**: Real-time sorting requires a model that is both computationally efficient and highly accurate.
5. **Reflection Issues**: Reflective surfaces, such as metal and glass, can cause misclassification due to inconsistent lighting conditions. The model must effectively distinguish between these materials while minimizing classification errors.

2. Related Work

Deep learning-based garbage classification has been explored using CNNs, ResNet, and MobileNet. Object detection models like Mask R-CNN and YOLO have been applied for waste localization. Recent advancements in EfficientNet have shown promise in classification tasks, and robotic waste sorting systems.

- **Recycleye Robotics (2024)**: Developed an AI-powered robotic arm for waste sorting that integrates with existing conveyor belt systems for seamless operation. This system employs unsupervised learning to improve classification without manual labelling.

- **Chen & Lu (2022)**: Explored computer vision-based solid waste sorting using CNNs. However, their approach focused on classification and did not include segmentation models, thus limiting its applicability to conveyor belt environments.
- **Zhang et al. (2023):** An AI robotic sorting system, incorporating machine learning for waste classification. However, this study did not include conveyor-specific datasets, making its applicability to industrial sorting systems limited.

Our approach expands upon these studies by integrating Mask R-CNN for segmentation with EfficientNet for classification in a conveyor belt setting, addressing real-world challenges such as overlapping objects, reflection issues, and dataset limitations.

3. Methodology

3.1 Datasets

We utilized multiple datasets to build a pipeline. The datasets used include:

- **BD WASTE (Mendeley Data)**: This dataset focuses on determining the biodegradability of waste items
- **Garbage Classification 6 Classes (Kaggle)**: This dataset provides a multi-class categorization of waste items and serves as the foundation for our classification model.
- **Homemade Conveyor Dataset (Synthetic Data)**: To better capture real-world conveyor belt waste sorting conditions, we generated a synthetic dataset containing real conveyor belt images with trash items. This dataset enhances model performance in scenarios closely resembling industrial waste processing settings.
- **Plastic Recycling Codes Dataset (Kaggle)**: This dataset was used for subclassification, allowing the model to distinguish between different types of plastic waste, which is crucial for improving recycling efficiency. One issue was we could not find suitable datasets for further subclassification of the other five categories. As a future task, we aim to create an augmented dataset to enhance subclass-level classification across all waste types.

3.1.1 Data Preprocessing: Create Dataset (create_dataset.py)

To ensure robust training, preprocessing steps include:

- Resizing images to match the input dimensions of an EfficientNet.
- Create Synthetic Dataset(appendix)
- Normalization and standardization to reduce variations caused by lighting conditions.
- Data augmentation (random rotation, flipping, and synthetic glare effects) to improve generalization.

3.2 Model Architecture

This project aims to develop a two-stage AI-driven pipeline for automated garbage classification. The proposed system consists of the following components:

1. **Segmentation Model**: Utilizes Mask R-CNN to detect and isolate individual waste items from a conveyor belt background.
2. **Classification Model**: Employs EfficientNet to categorize each segmented item into one of six predefined waste categories and determine its biodegradability status.

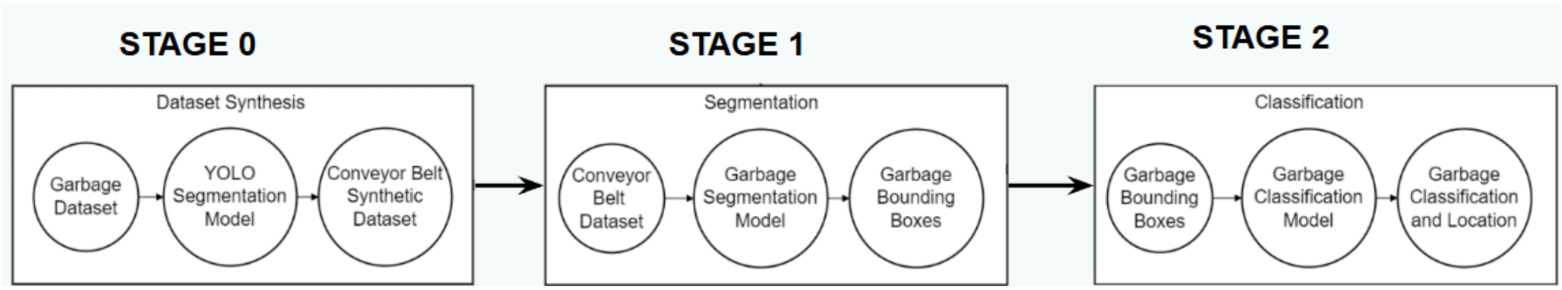

Figure 1: Figure 1 showcases the pipeline stages, with stage 0 handling creating the synthetic dataset stage 1 having a segmentation train on the conveyor belt dataset, then stage 2 using the segmented bounding box to classify the garbage.

3.2.1 Mask R-CNN (ResNet-50 FPN) for Segmentation

We employ a Mask R-CNN architecture with a ResNet-50 backbone and Feature Pyramid Network (FPN) to perform instance segmentation on synthetic waste data. This model first uses a Region Proposal Network (RPN) to generate candidate object locations and then refines these proposals to generate accurate bounding boxes and pixel-level binary masks. The segmentation model is trained entirely on a custom synthetic dataset stored in an SQLite database, with each image containing 10 or more annotated waste items, each labelled with its precise binary mask and bounding box. To improve mask quality, we introduced a custom loss function that combines IoU and Dice losses with a boundary-aware term, progressively weighted through training. This combination ensures both spatial coverage and edge alignment. The model is evaluated using mean box IoU and mean mask IoU metrics, tracking how well the predicted masks align with ground-truth regions.

3.2.2 EfficientNet for Classification

Our classification model uses image detection to determine how to label a piece of trash. The model leverages its compound scaling and depth optimization to improve classification performance, using the EfficientNet B-3 model with a modified classification sequence that includes additional linear layers applying downscaling. This upgrade to the classification layers saw a significant increase in model accuracy, with numerous combinations and types of layers tested using Bayesian optimization. The model's output layer is a series of 7 nodes, each corresponding to a different class. We use cross-entropy loss when training to find the probability of the model predicting each output and classifying each piece of trash based on the resultant probability distribution. The model uses the pre-trained weights from EfficientNet, which were first fine-tuned on the TrashNet dataset and then further fine-tuned from the synthetic dataset. This combination of fine-tuning pushed the model's accuracy significantly when applying the model to the synthetic dataset, pushing the model from sub-40 % accuracy to above 80%.

3.3 Reflection Mitigation

Reflective surfaces in waste items cause misclassification due to inconsistent lighting. We mitigate this through:

- Adaptive histogram equalization
- Polarized filtering during data collection
- Augmentation with synthetic glare effects
- Background subtraction to reduce the impact of reflections

4. Results

4.1 Segmentation Performance

Before training, the model couldn’t detect objects nor masks. This meant the entire pipeline couldnt function right at the start, since there were no detections to classify. But after training on our synthetic dataset images and the ground truth stored (stored in a SQLite database). The segmentation model reached a Box IoU of 0.77 and a Mask IoU of 0.85, and it successfully covered about 90% of all ground-truth objects. What sets our approach apart is that the segmentation is tied to the classification step, and how well it adapts to synthetic data. If you look at the comparison images, the untrained model didn’t recognize anything, while the trained model outlined and segmented each piece of trash. This really highlights how important domain-specific training and customized loss functions are for making segmentation actually work in practice.

Figure 2: Classification before and after the training (Evaluation)

4.2.1 Classification Accuracy

EfficientNet achieves an accuracy of 93.7% on the publicly sourced, outperforming baseline CNNs (85.3%) and ResNet (88.7%). This model also improved based on fine-tuning, with the initial implementation performing with a 48.33% accuracy on the synthetic dataset post-first-stage fine tuning, improved to 81.58% after second-stage fine-tuning based on our synthetic dataset.

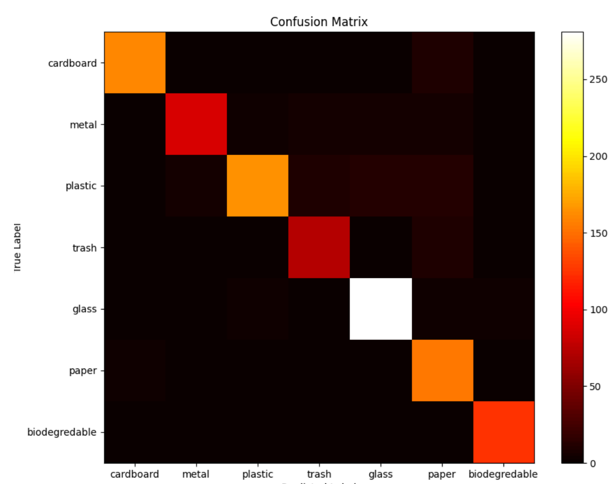

Figure 3

Alongside classification accuracy, another evaluation metric that was essential for the classification model was the classification speed. This pipeline is intended to work in an actual recycling facility, where it can classify and pick up trash in real time. For the model to be able to do this, it needs to be fast enough to classify trash in real time before it gets off of the conveyor belt. To this end, we used the B3 variant of the EfficientNet model, which is an even balance of performance and complexity. Adding complexity to the image detection model added diminishing returns in terms of accuracy, with an exponential gain in model inference time, while there was a significant improvement going from the B2 variant to the B3 variant. Going from the B2 model to the B3 model increased accuracy by 4%-6% on average, depending on chosen hyperparameters, and took on average 0.1s less time to classify. On the other hand, going from B3 to B4 did not have any change in average model accuracy but took 0.3s more time to classify each image.

Compared to other models that already exist, the classification model is unique in its ability to classify more labels, though further stages to the pipeline are used to fully extend the functionality of the model and introduce novel concepts, such as classifying the specific type of plastic as opposed to just grouping all plastics.

To do this, we use the classification model to generate preliminary labels and then pass the classification into another model to specifically determine sub-classes, such as identifying specific types of plastic or metal that a piece of trash is made of.

Based on the confusion matrix, we identified that the model performed well, though its main issues lay with correctly predicting paper and often mispredicting glass and paper as plastic.

4.2.2 Sub-Classification Accuracy

On top of classification, we added to the classification model to introduce a new label specific for plastic due to the limited datasets. Plastic was categorized in one of 6 categories: PET, HDPE, PVC, LDPE, PP, and PS. Through tuning the hyper parameters and having the model train on this plastic dataset, the following Figure was achieved.

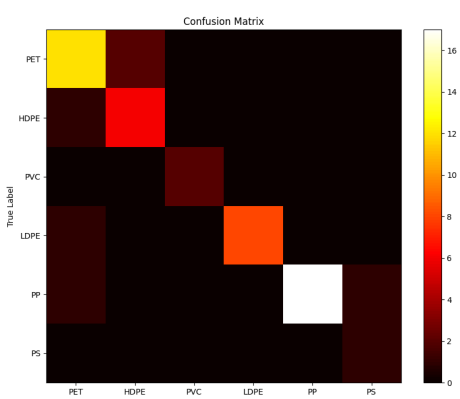

Figure 4: (Confusion Matrix of Plastic Categories)

The following figure showcases that the subclassification model was able to fairly classify plastics well achieving an accuracy of about 88.46% (correct predictions / total predictions) after training the model and tuning the hyper parameters.

4.3 Pipeline Results

The evaluation results of the pipeline, based on 50 synthetic images containing approximately 2,000 trash objects, demonstrate the model's performance and its effectiveness following domain-specific fine-tuning. The segmentation component, based on Mask R-CNN, achieved an average box IoU of 0.7594 and an average mask IoU of 0.8150, indicating strong localization and pixel-level segmentation accuracy. These metrics confirm the model's ability to reliably detect and separate individual waste items in densely packed synthetic scenes, even when objects are overlapping or occluded.

For the classification module, the pipeline reached an overall accuracy of 69.99%, with correct predictions averaging a confidence score of 0.9690, and incorrect ones still showing a relatively high average confidence of 0.9442. The model, though the narrow confidence gap suggests that additional work on uncertainty estimation or post-classification filtering could further reduce misclassifications.Since each image contains a high number of labeled objects, this evaluation covers a substantial sample of object-level predictions, making the results meaningful and statistically significant even at the image subset level. These metrics collectively validate the model's robustness in both segmentation and classification, showcasing the success of the combined architecture and custom training approach on synthetic waste datasets.

```
===== Final Model Metrics =====
Total Images Processed: 50
Segmentation - Average Box IoU: 0.7594, Average Mask IoU: 0.8150
Classification - Overall Accuracy: 69.99%
Highest Classification Confidence: 1.0000, Lowest Classification Confidence: 0.2775
Average Confidence (Correct Predictions): 0.9690
Average Confidence (Incorrect Predictions): 0.9442
Total time elapsed: 69.51 seconds
```

Figure 5: Model final Metrics as mentioned.

5. Limitations Addressed in Our Approach

- **Overlapping Objects**
  • Cause: Hard matching via Hungarian algorithm.
  • Solution: We introduced soft assignment and multi-matching strategies for ambiguous occlusions.

- **Weak Early Mask Supervision**
 • Cause: Noisy predictions in early training mislead supervision.
 • Solution: Linearly schedule $\alpha(t)$ to delay strong loss penalties until predictions stabilize.
- **Edge Loss Sensitivity**
 • Cause: Sobel operator is too reactive to visual artifacts.
 • Solution: Replaced with trainable edge detectors (e.g., HED) for adaptive edge learning.
- **Limited Dataset Size**
 • Cause: Small batches due to SQLite backend and imbalanced representation.
 • Solution: Augmented data with synthetic generation and self-supervised techniques.

Remaining Limitations

- **Real-world Deployment**
- Pending: Needs domain generalization and test-time adaptation for robustness.
- **Dataset Bias**
 • Challenge: Under-represented waste categories reduce model generalizability.
 • Pending: Requires curated, balanced datasets or class-reweighted loss.
- **Computational Cost**
 • Challenge: EfficientNet offers high accuracy but demands significant computing power, limiting its use on edge devices.
 • Pending: Potential improvements via lightweight backbones (e.g., MobileNetV3) or model distillation.

6. Novelty

- Two Stage pipeline: Integrates Mask R-CNN for waste segmentation and utilizes EfficientNet Version 3 for classifying garbage into 6 categories.
- Synthetic Conveyor Belt Dataset: Created a custom synthetic dataset to simulate a real-world conveyor belt for the segmentation and classification models to train and improve on.
- Performance: Optimized EfficientNet B3 for speed and accuracy balance to classify segmented objects more efficiently
- Additional Classifications: Further classified for biodegradability and subclassification for plastic types

7. Conclusion

EcoSort is an AI-driven pipeline utilizing both segmentation and classification designed for real world recycling facilities. By integrating Mask R-CNN for precise segmentation and Efficient for classification (including biodegradability and subclassification), we can address key challenges in recycle sorting, including some of the ongoing issues, such as overlapping objects and dataset limitations. In the end, the pipeline was able to achieve a 70% accuracy rate for 1.39s per image, which is a solid result, while also showcasing room for improvement to further enhance the pipeline for a real-world recycling plant scenario.

8. Appendix

8.1 Team Contributions

Adem: Researched and conducted second training with EfficientNet; applied reflection mitigation techniques on the segmentation layer; . Contributed heavily to the report.

Sharat: Performed initial EfficientNet training, researched reflection mitigation techniques for segmentation and made the final plastic subclassification model. Contributed heavily to the report.

Harsh: Developed initial segmentation model and data creation scripts and made the final classification model. Contributed heavily to the report.

Mark: Created conveyor belt script using dataset images, collaborated on dataset creation.

All Members: Collaborated on dataset creation, model training, and evaluation.

Extra models:
https://drive.google.com/drive/u/0/folders/1ShzTkV4nZuxKNSvBo6p1cbRgATyAbA76

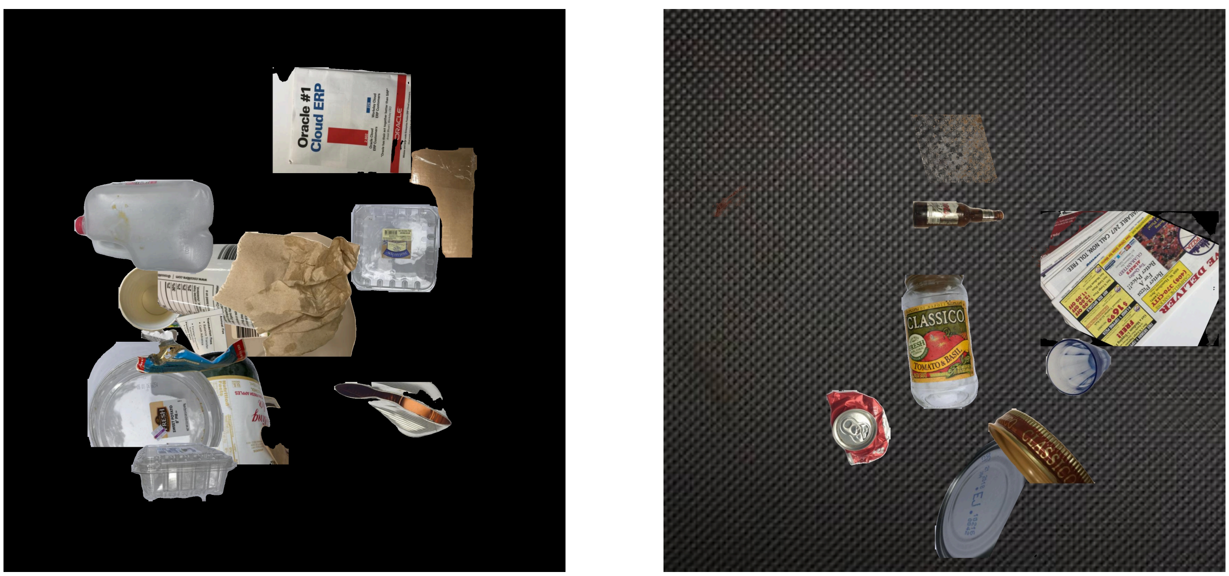

Figure 6 : Dataset pictures

7. References

1. Akter, M. (2023, November 17). BDWASTE: A comprehensive image dataset of digestible and indigestible waste in Bangladesh. *Mendeley Data.* https://data.mendeley.com/datasets/96g5pgfnfw/1
2. CCHANGCS. (2018, November). Garbage classification. *Kaggle.* https://www.kaggle.com/datasets/asdasdasasdas/garbage-classification
3. Chen, J., & Lu, W. (2022, April). *Computer vision for solid waste sorting: A critical review of academic research.* ResearchGate. https://www.researchgate.net/publication/358580288_Computer_vision_for_solid_waste_sorting_A_critical_review_of_academic_research
4. ETM Recycling. (2020, April 27). *Waste management – Everything you need to know about waste management.* https://www.recyclingbristol.com/waste-management-everything-you-need-to-know-about-waste-management
5. Fan, Q., Yang, J., Hua, G., Chen, B., & Wipf, D. (2017). A generic deep architecture for single image reflection removal and image smoothing. *IEEE International Conference on Computer Vision (ICCV),* 3238-3247. https://doi.org/10.1109/ICCV.2017.349
6. Kunwar, S. (2024, December 12). Garbage dataset. *Kaggle.* https://www.kaggle.com/datasets/sumn2u/garbage-classification-v2
7. Mohamed, M. (2021, January 24). Garbage classification (12 classes). *Kaggle.* https://www.kaggle.com/datasets/mostafaabla/garbage-classification
8. Recycleye. (2023, December 11). *Wastenet.* https://recycleye.com/wastenet/
9. Sustainable Waste Management in Indore: A case study. (n.d.). *Earth5R.* https://earth5r.org/sustainable-waste-management-in-indore-a-case-study/
10. Wei, K., Zhao, X., Zhang, K., Zhou, W., & Meng, D. (2019). Single image reflection removal exploiting misaligned training data and network enhancements. *IEEE Transactions on Image Processing, 28*(7), 3665-3679. https://doi.org/10.1109/TIP.2019.2897162
11. WMW. (2021, July 12). *Waste-to-energy to curb US pandemic garbage boom.* https://waste-management-world.com/resource-use/waste-to-energy-to-curb-us-pandemic-garbage-boom/
12. Zamzamy, R. (2021, June 12). Non and biodegradable material dataset. *Kaggle.* https://www.kaggle.com/datasets/rayhanzamzamy/non-and-biodegradable-waste-dataset
13. Zhang, M., Xu, Q., Luo, A., Kong, Y., Chi, W., & Sun, L. (2023, September 18). *An intelligent garbage sorting robot system based on machine vision and knowledge base. IEEE Xplore.* https://ieeexplore.ieee.org/document/10240196