

EcoVision: Multimodal Waste Material Recognition using Vision Language Models (VLMs)

Manoah Edwin Paul

Department of Computer Intelligence, SRM Institute of Science and Technology, Chennai, Tamil Nadu, India

B. Amutha

CTECH, SRM Institute of Science and Technology, Chennai, Tamil Nadu, India

Abstract

Effective waste segregation is important for increasing recycling efficiency and reducing environmental impact. Because conventional management systems depend heavily on mechanical techniques that categorize waste by size rather than material composition, they often give poor recovery rates for recyclable resources. To resolve these issues, we present Ecovision, a multimodal waste recognition framework combining open-vocabulary object detection with vision language classification.

The system combines Grounding DINO for object detection with a CLIP based model for material classification. We achieve zero-shot classification via prompt based inference, enabling the system to identify material categories without task specific training. To improve performance in complex, real world conditions, we apply Low-Rank Adaptation (LoRA) to the visual encoder using a small dataset of cluttered field images. This architecture is specifically designed to combine with existing mechanical segregation pipelines at landfill and recycling facilities.

We tested the model using the TrashNet and TACO datasets and real-world images. By giving few fewshot adaptation improves classification accuracy while maintaining zero shot generalization. Ultimately, EcoVision is a system that can easily grow and be used to identify wastes automatically, making recycling and waste management smarter.

Keywords

CLIP, Grounding DINO, Open vocabulary object detection, Vision language models, Waste material recognition, Waste segregation, Zero shot learning.