CONTAMINATION IDENTIFICATION

Arjun Chandra (CS ‘25) Tia Hannah (DS ‘25)
Heng Chang (DS ‘26) Devon Solheim (DS ‘24)

CLIENTS & MENTORS
CLIENT
Laurent Meunier
CLIENT ASSISTANT
Jason Oh
PROJECT MANAGER
Tanisq Mallpani
SPARK TECH ADVISOR
Michelle Voong
PROGRAM LEAD
SPARK! SUPPORT
Daniel Oh

PROJECT SUMMARY

Spare-It is a company that aims to improve sustainability in workplaces using data. One example of this is the monitoring and analysis of waste disposal for contaminants. These contaminants could be improperly recycled or composted articles which could reduce an organization’s efficiency and increase their carbon footprint.
Spare-It requires a solution that can accurately and efficiently detect contaminants within waste disposal bins, so this project aimed to implement machine learning to accomplish the client’s goal.

DATASET

The dataset provided by Spare-It had approximately 18,000 images and labels for waste bins from overhead. They were formatted in the data for segmentation.

PROCESS

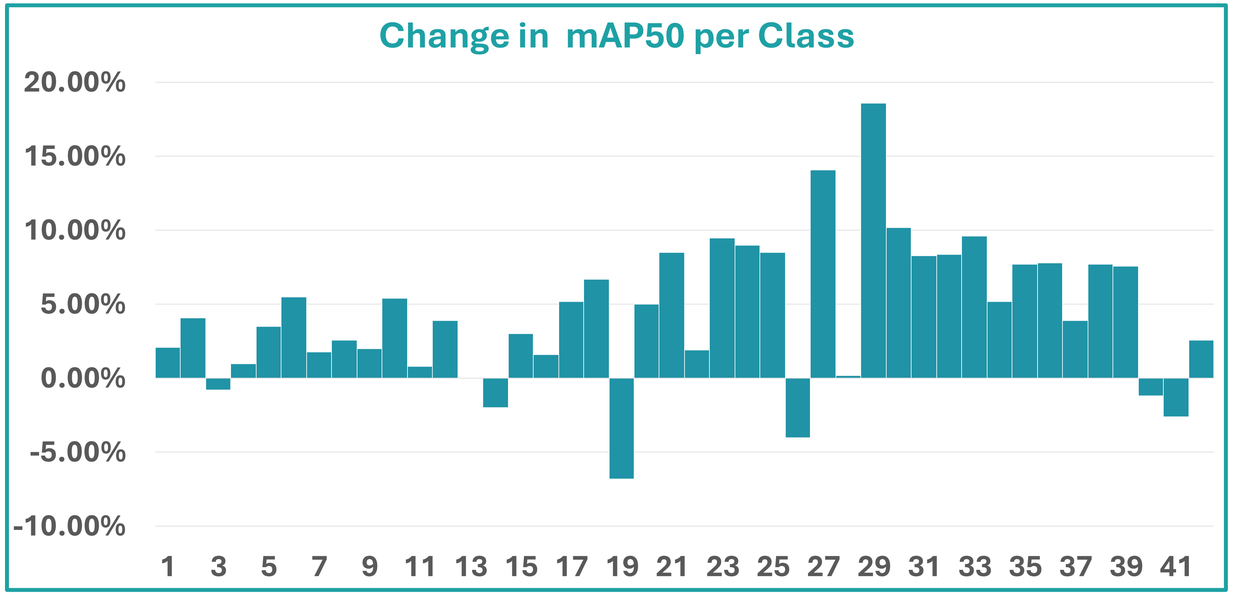
This project’s solution was to train a YOLOv8 segmentation model to classify waste, as well as improve the robustness of the dataset to improve on models generated in past semesters.

Dataset Improvement:

- External Dataset Integration: Images from the TACO dataset were taken and converted into Spare-It’s dataset format
- Copy Paste Augmentation: Using the segmented data, components of the images were extracted and pasted onto other images to augment the dataset further
- AI Image Generation: A stable diffusion model was trained on top of the component images from the dataset in order to generate further data

RESULTS

Significant improvements were made to the dataset after implementing the TACO images and the Copy Paste augmentation. This reduced the imbalance between the classes in the original dataset, as the augmentations made prioritized increasing the prevalence of the underrepresented classes. The integration of TACO provided images with different backgrounds and noise, which helped to alleviate the overfitting from the model. The synthetic images created from stable diffusion were promising but were not fully implemented into the model due to time constraints. A pipeline was created to generate these images from any given class though, allowing for the viable production of synthetic data that is authentic to the original dataset.



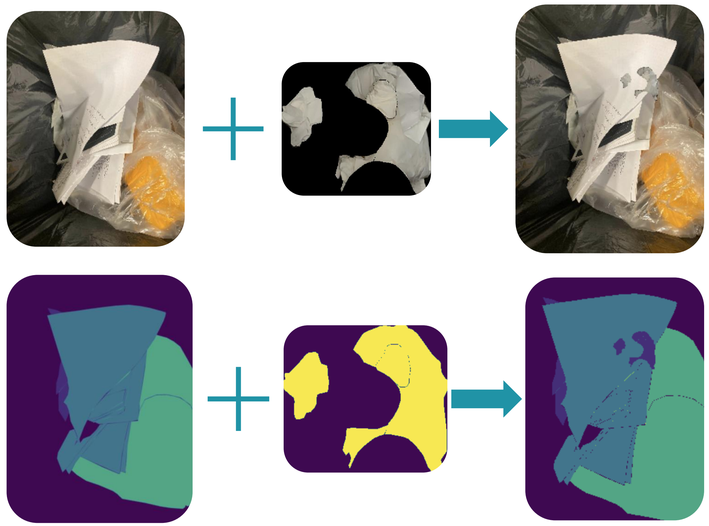
CONTAMINATION IDENTIFICATION

Arjun Chandra (CS ‘25) Tia Hannah (DS ‘25)
Heng Chang (DS ‘26) Devon Solheim (DS ‘24)

TECH STACK

Language: Python
Libraries: YOLO, COCO
Tools: Hugging Face, Stable Diffusion

COPY PASTE SAMPLE OPERATION

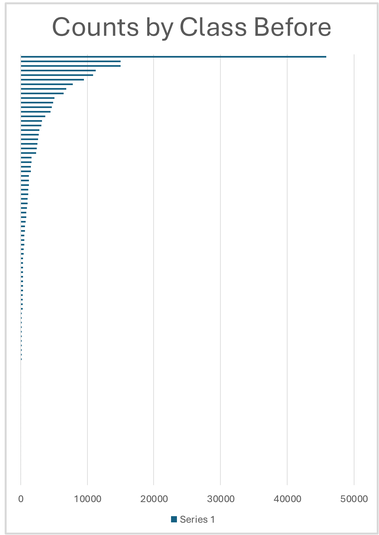


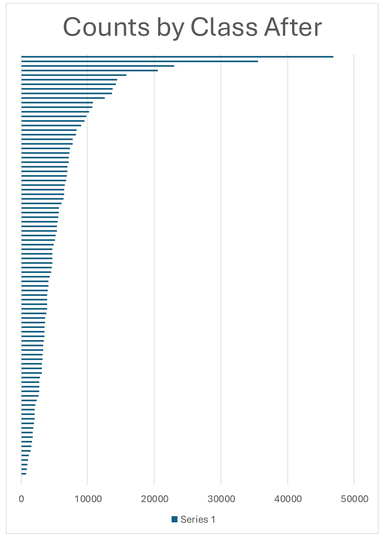
AI IMAGE GENERATION

Base Model Output

Fine-tuned Model Output

CLASS IMBALANCE





DEMO

This demo allows for any image to be classified by the best performing model
NOTE: The demo works best on a desktop or laptop

REFLECTIONS

One key takeaway was how a plethora of ambitious ideas were scrapped during the project. The lesson here was that it was more important to focus on the core mechanics of the project first and doing this allowed us to create a strong and focused solution.
It was also interesting to experience what a real project will entail, with deadlines and clients, as opposed to the typical university project.