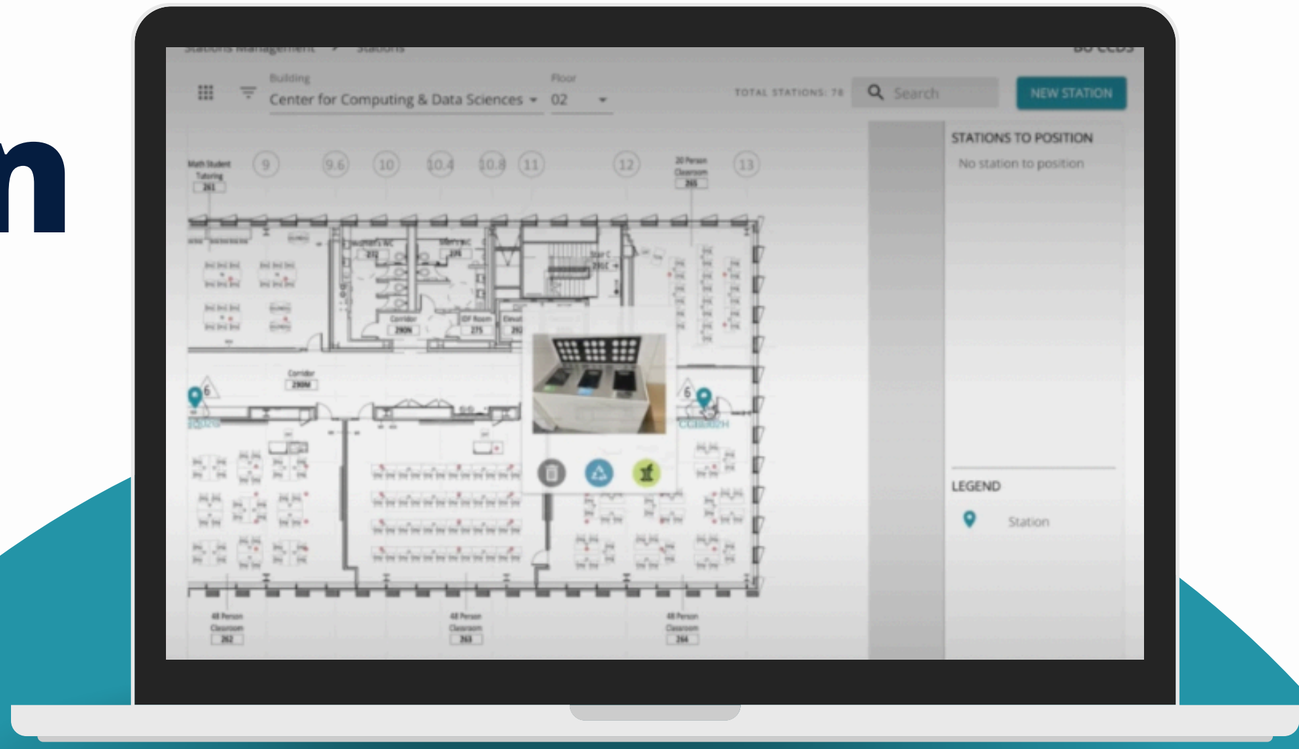
spare-it

Contamination Identifier

Cheolmin Oh, Mustafa Taibah, Tessa Wu

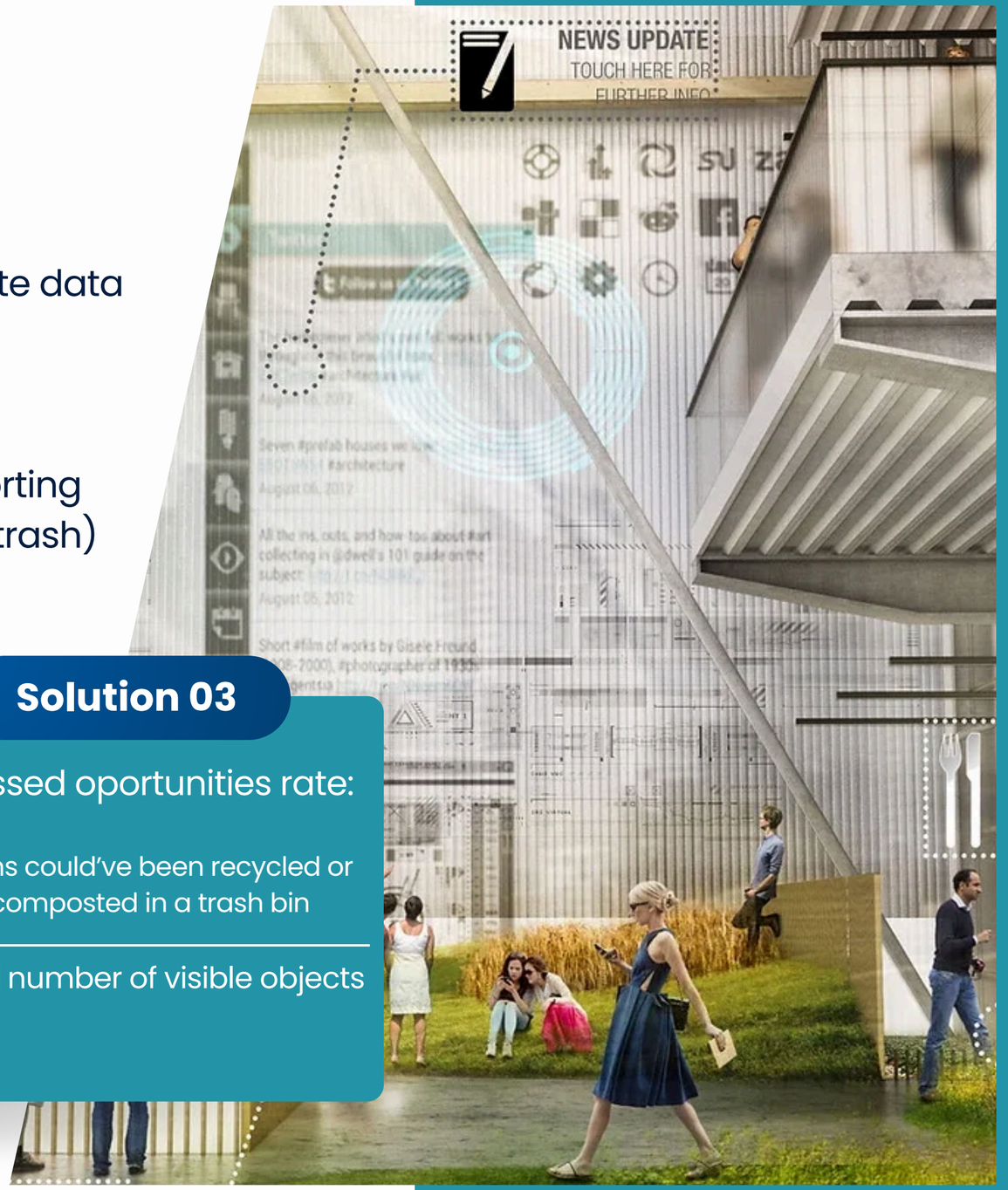


Problem Statement

There is a lack of transparency and awareness in true waste data

Goal: estimate composition of the waste visible in images

Solution count each material type in each image, apply sorting rules based on the type of bin (compost/mixed recycling/trash)



Solution 01

composition of waste:

% of items of each material type

Solution 02

waste contamination rate:

$$\frac{\text{items mistakenly composted}}{\text{total number of visible objects}}$$

Solution 03

missed oportunities rate:

$$\frac{\text{items could've been recycled or composted in a trash bin}}{\text{total number of visible objects}}$$

Use Case

Each bin sits on a scale that reports data in real time where analysts can see which floors and bins accumulate the most waste

The combination of weight data and photos are reveals how each floors sorts their waste and which need more awareness to disposal

Company Overview

Spare-It provides real time waste data to help create sustainable programs, enhance recycling, and reduce waste

Spare-It helps businesses, office owners and universities access to live intelligence for all kinds of workspace waste and encourage employees and students to reduce

Boston University
Center for Computing and Data Science

Process Overview

We are developing an ML model to enhance waste management by identifying, classifying, and estimating contamination rates of waste in trash recycling and compostable bins

Literature Review

Assessed recent research and open-source models

Focused on integrating **YOLO framework** and advanced data augmentation techniques

Explanatory Data Analysis (EDA)

11,666 images with corresponding JSON files

Significant class imbalance among 100 object categories

Model Development

Main Model: **YOLOv8** for real-time object detection of waste types

Class Imbalance: oversampling and undersampling techniques

Data Augmentation: Albumentation library to handle low quality images

FINDINGS

Dataset & Training Issues

100 labels

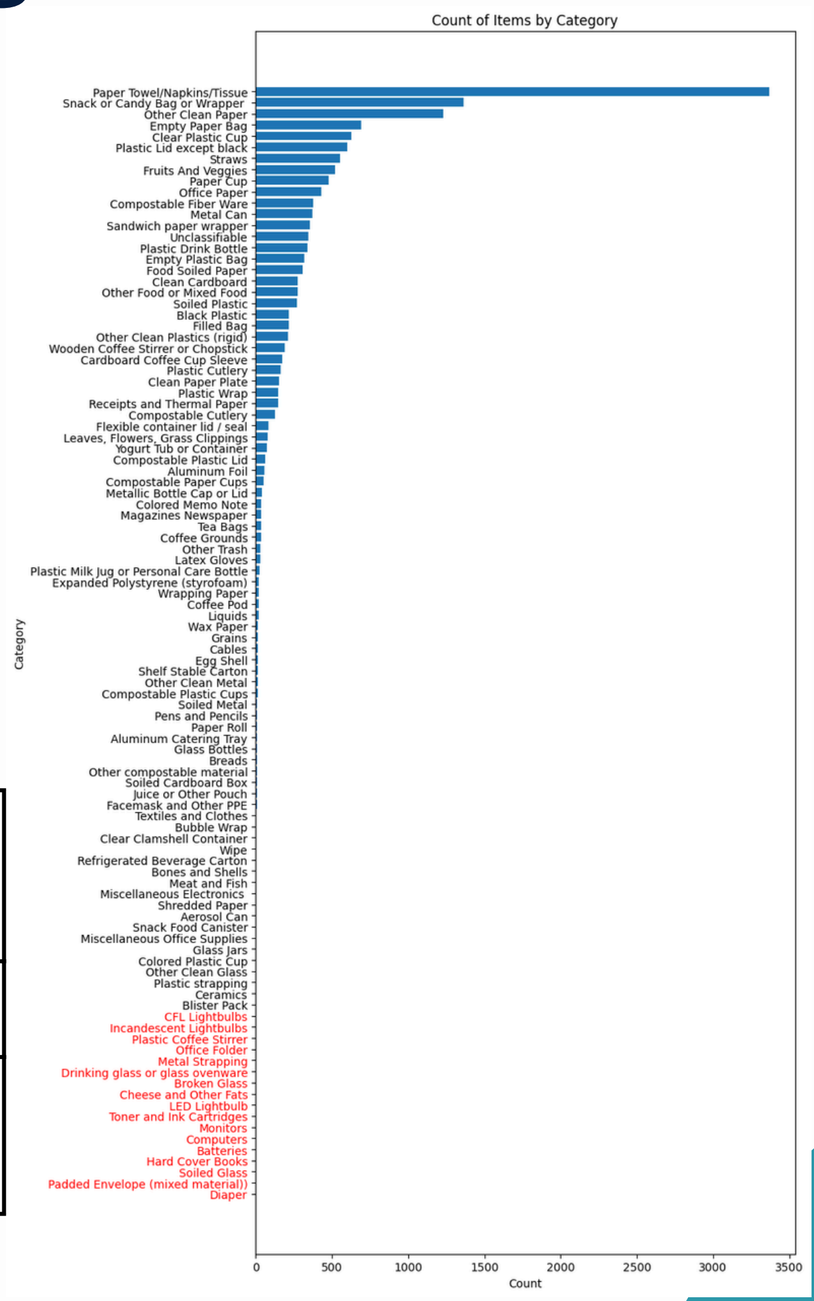
2,024 Images and Labels > 11,666

Issues:

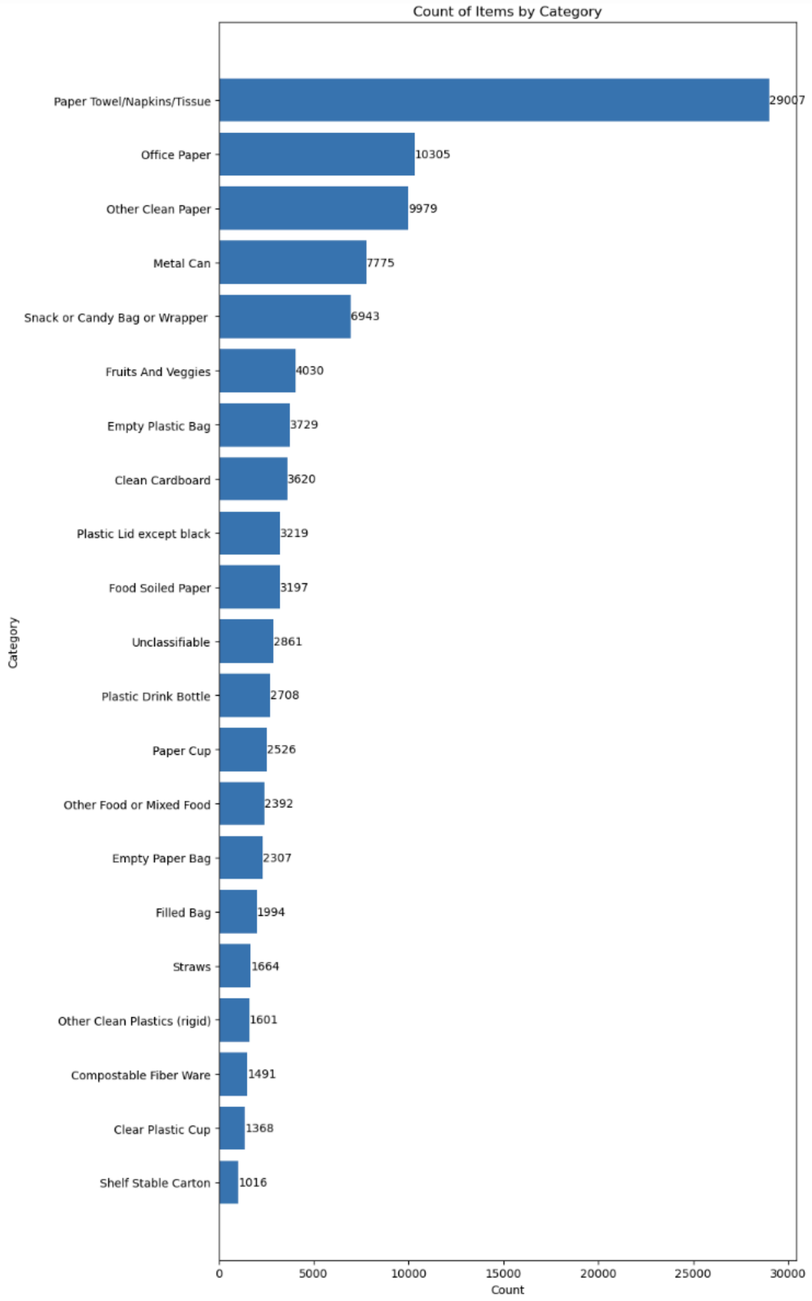
- Class Imbalance
- 1 Precision due to no False Postive

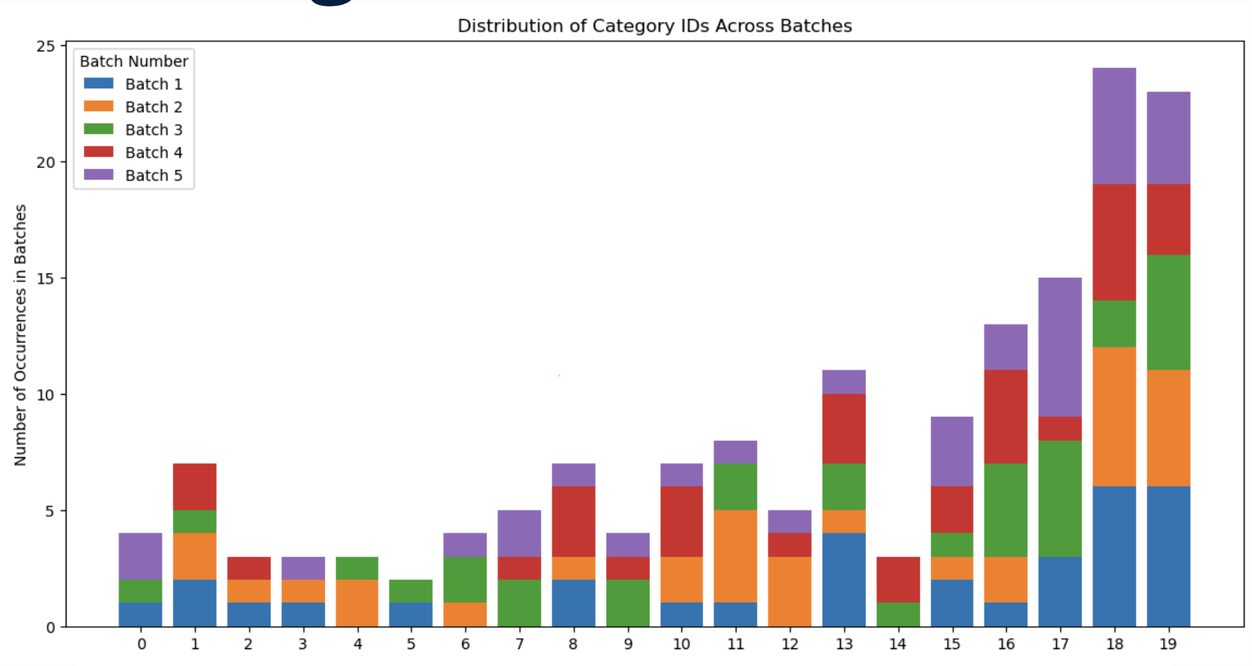
ultralytics YOLOv8

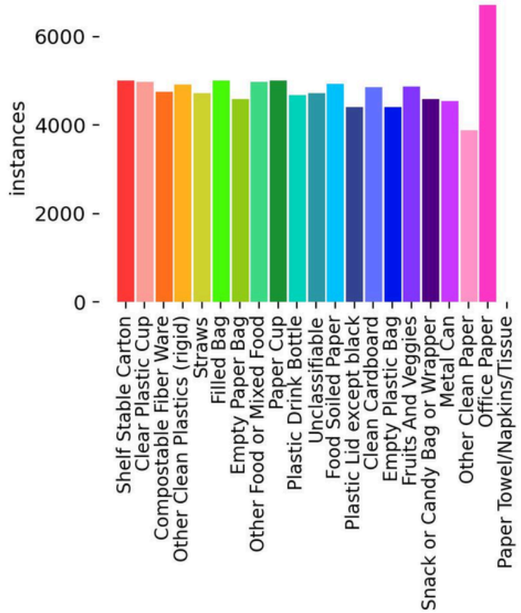
	Total Images	Samples	Precision	Recall	mAP50	mAP50-95
Wipe	**1750**	**6**	**1**	**0**	**0**	**0**
Wax Paper	**1750**	**11**	**1**	**0**	**0**	**0**

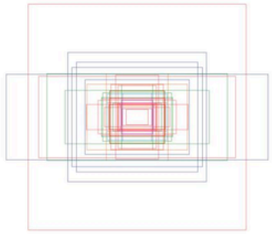


Dataset Changes



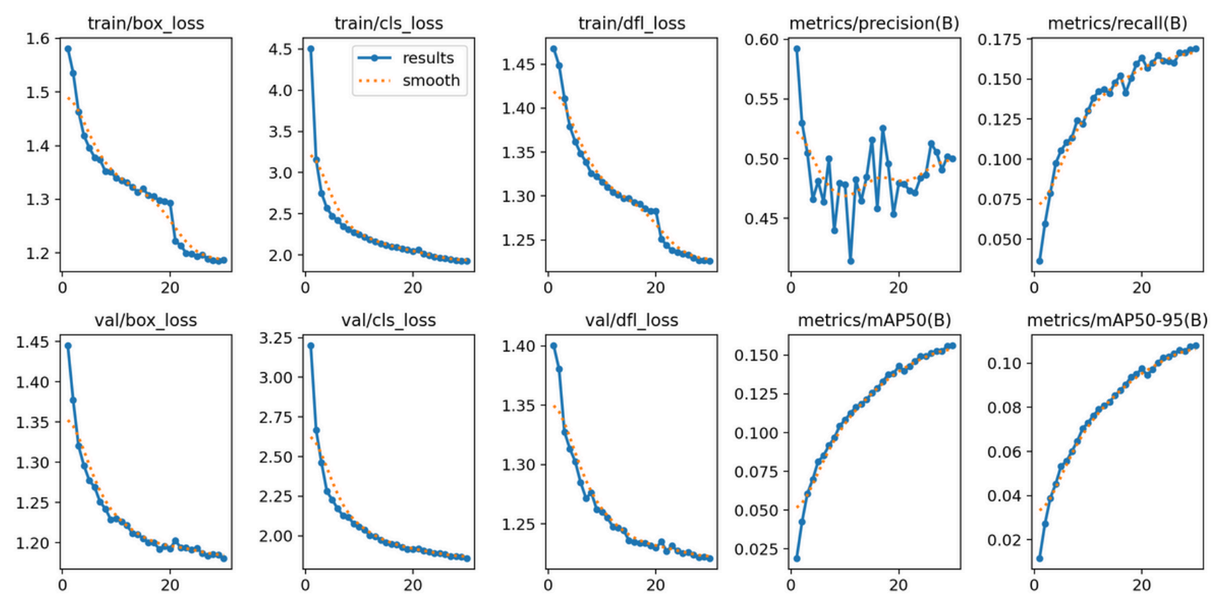




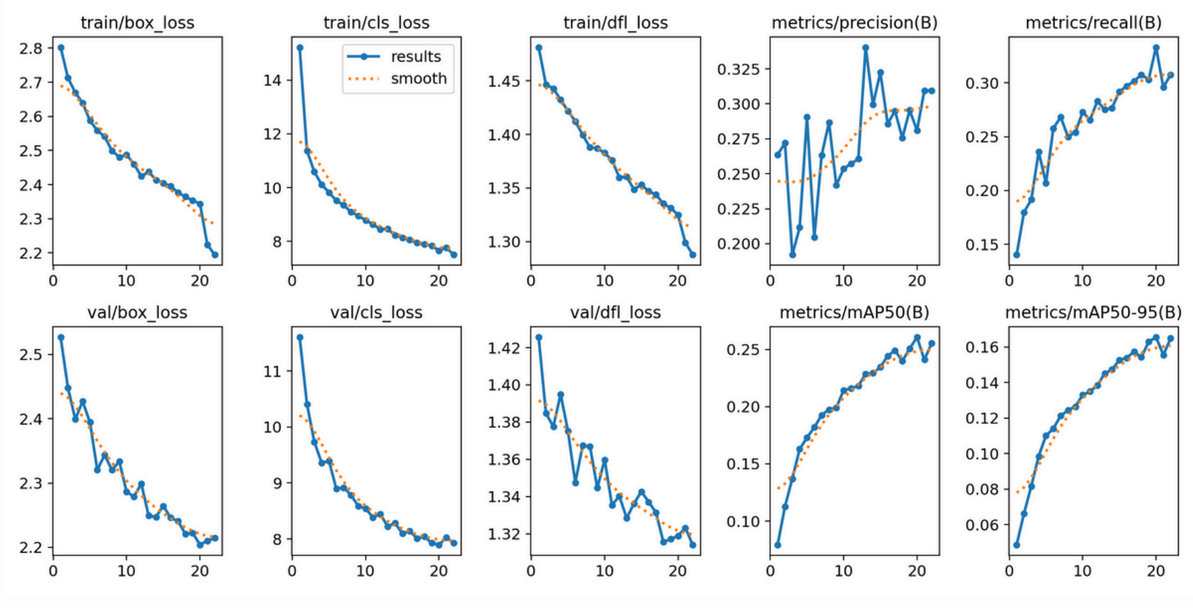


Sampling Results

No Sampling - 100 Class

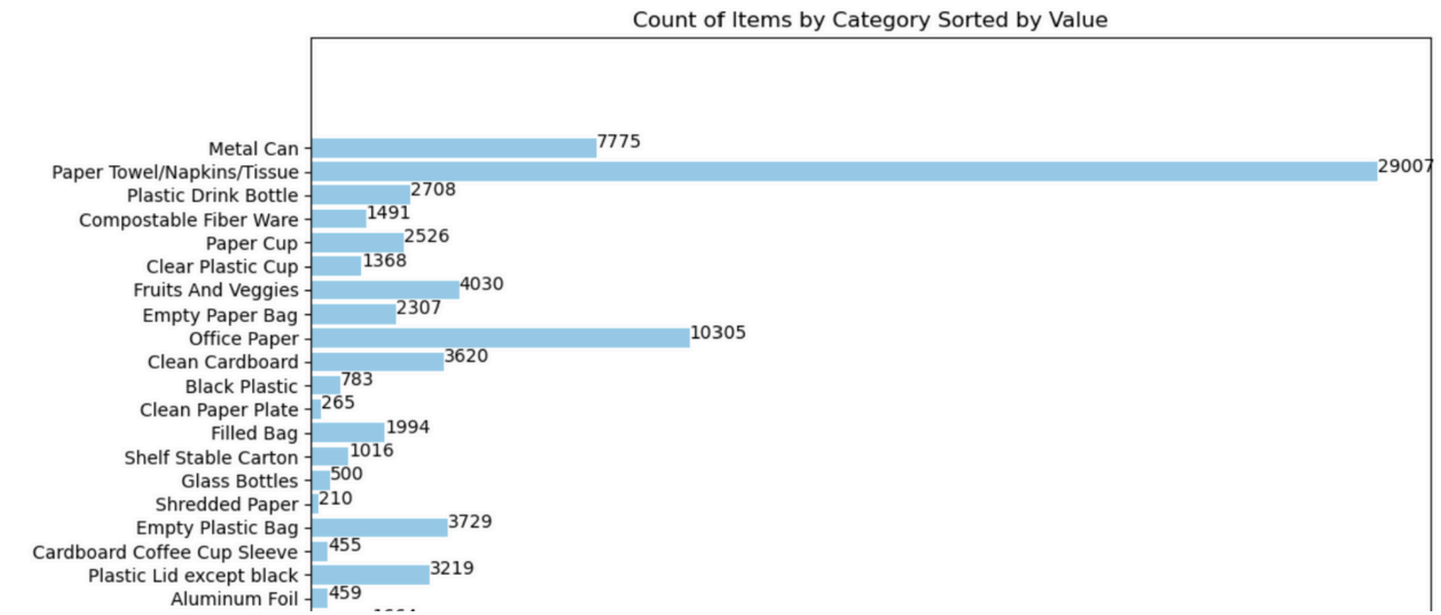


Sampling - 20 Class



New Classes with High Recall Rate

Category	Value
Metal Can	0.84600
Paper Towel/Napkins/Tissue	0.66100
Plastic Drink Bottle	0.63800
Compostable Fiber Ware	0.62200
Paper Cup	0.61300
Clear Plastic Cup	0.57200
Fruits And Veggies	0.50400
Empty Paper Bag	0.50100
Office Paper	0.48200
Snack or Candy Bag or Wrapper	0.48200
Clean Cardboard	0.47700
Black Plastic	0.46500
Clean Paper Plate	0.42300
Filled Bag	0.42100
Shelf Stable Carton	0.41900
Glass Bottles	0.41300
Shredded Paper	0.40000
Empty Plastic Bag	0.38100
Cardboard Coffee Cup Sleeve	0.36400
Plastic Lid except black	0.36100
Aluminum Foil	0.35400
Straws	0.33200
Magazines Newspaper	0.31000
Drinking glass or glass ovenware	0.30100
Soiled Plastic	0.28700
Sandwich paper wrapper	0.27700
Compostable Plastic Lid	0.25000



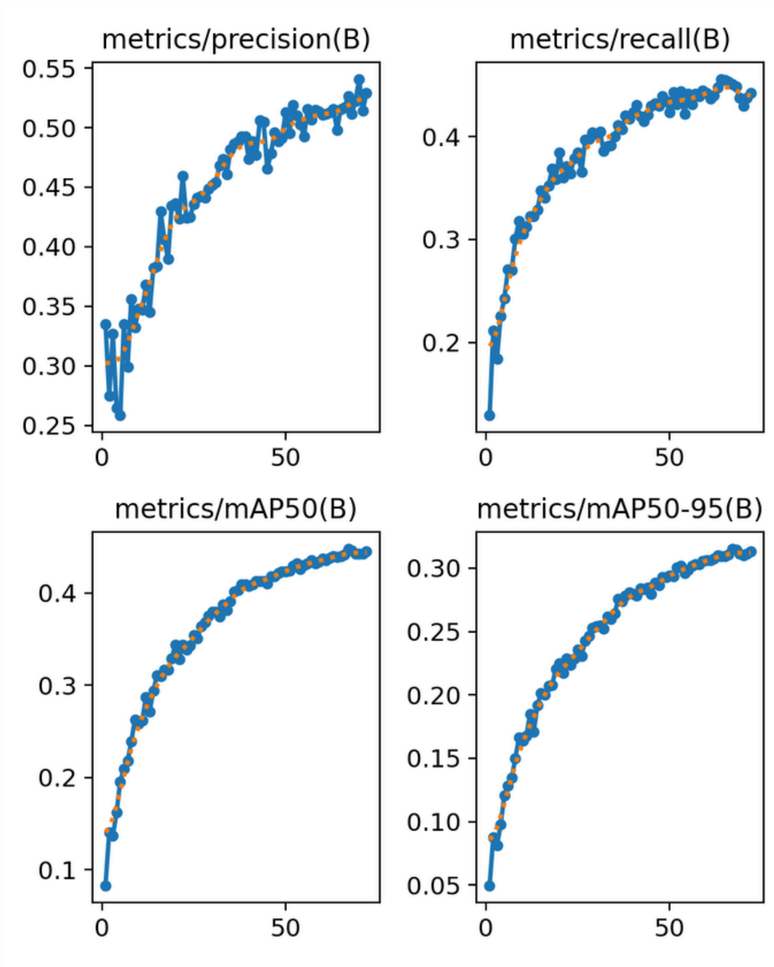
Data Augmentation

Albumentations

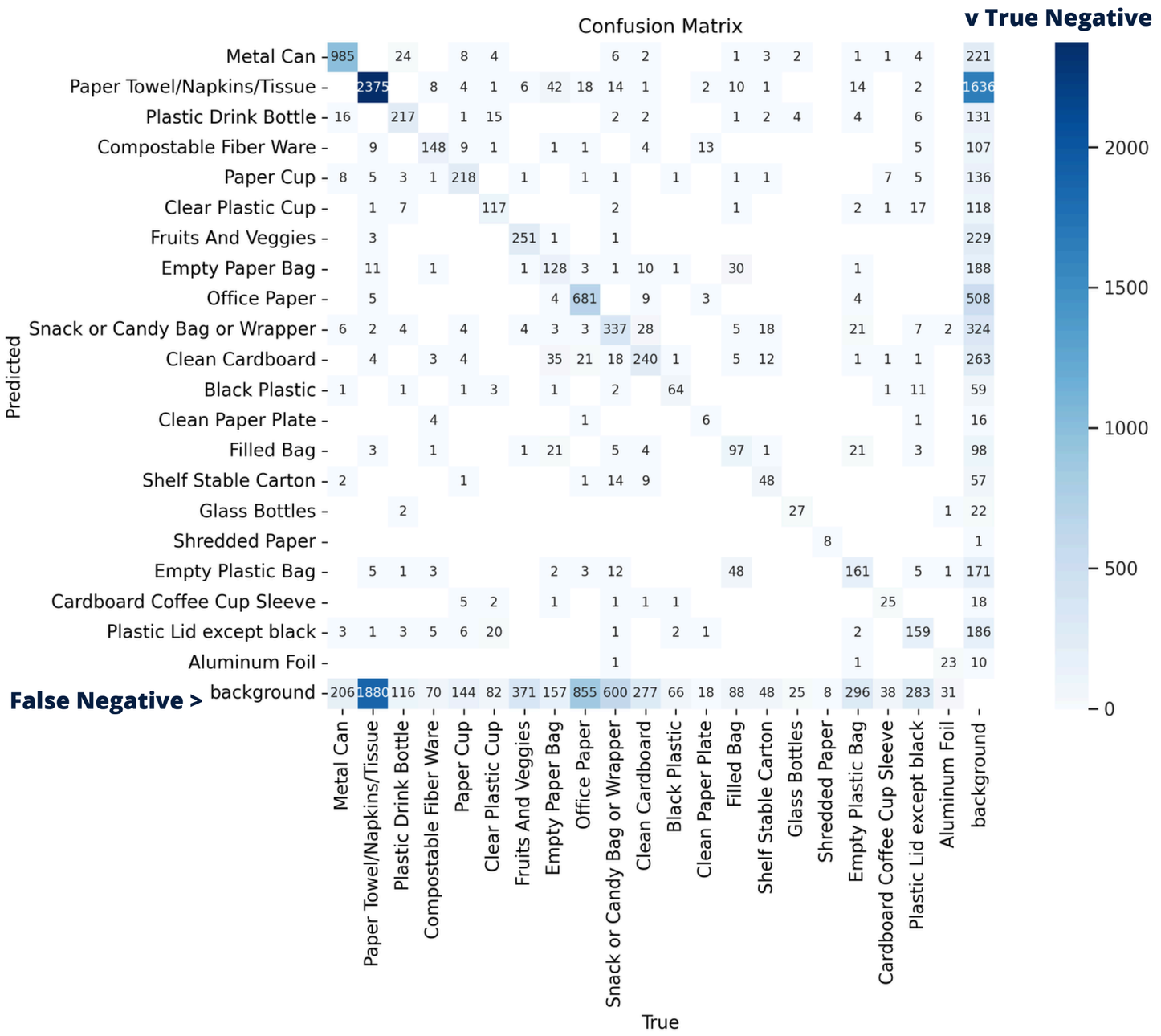
Results

52.7%

	Total Images	Samples	Precision	Recall	mAP50	mAP50-95
Aluminum Foil	1750	58	0.617	0.345	0.404	0.251
Metal Can	1750	1227	0.661	0.71	0.742	0.514
Paper Towel/Napkins /Tissue	1750	4304	0.426	0.518	0.438	0.233
Plastic Drink Bottle	1750	378	0.38	0.434	0.368	0.205
Compostable Fiber Ware	1750	244	0.355	0.471	0.355	0.269
Paper Cup	1750	405	0.47	0.383	0.402	0.273
Clear Plastic Cup	1750	245	0.347	0.299	0.236	0.14
Fruits And Veggies	1750	635	0.48	0.287	0.293	0.158
Empty Paper Bag	1750	396	0.211	0.447	0.226	0.135
Office Paper	1750	1588	0.423	0.305	0.292	0.164
Snack or Candy Bag or Wrapper	1750	1018	0.324	0.239	0.204	0.112
Clean Cardboard	1750	587	0.271	0.356	0.227	0.14
Black Plastic	1750	136	0.402	0.301	0.301	0.21
Clean Paper Plate	1750	43	0.229	0.209	0.154	0.119
Filled Bag	1750	287	0.31	0.293	0.226	0.128



Results: Confusion Matrix



Next Steps

As we continue to push the boundries of waste reduction through innovative data mangement and AI algorithms, our roadmap unfolds as follows:

01 Containerization and Deployment

02 Image generation and Model Development with DCGAN

03 Additional dataset exploration (Literati)

Containerization and Deployment

Ensure that the model can be reliably and consistently deployed across various environments, from local systems to cloud platforms.

CONTAINERIZATION

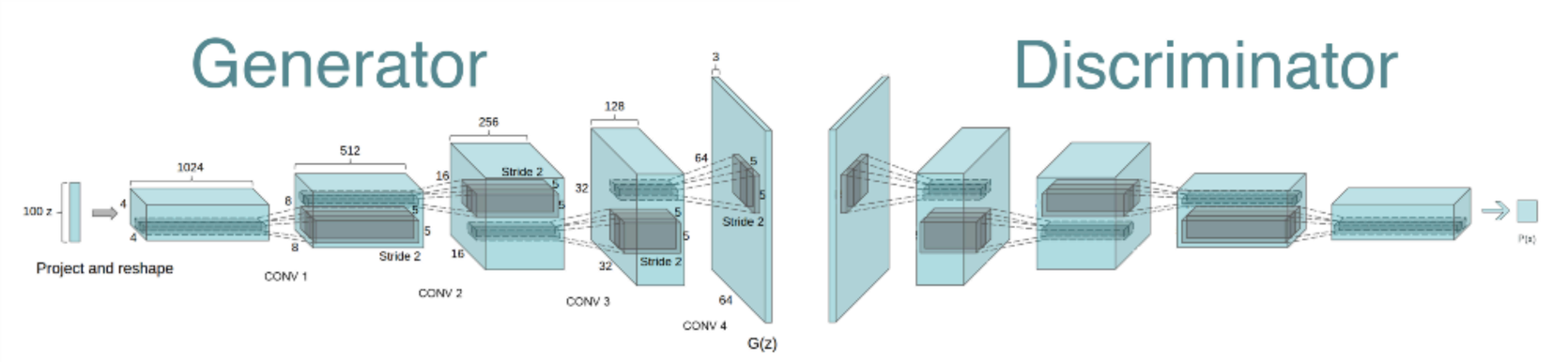
DEPLOYMENT

Docker Image
Deploy to Cloud

Image Generation

Address the challenge of limited waste classification data by generating high-quality synthetic images using DCGAN

Develop a DCGAN architecture tailored to generate waste images
Train the DCGAN with existing waste images
Use the generated images to augment the training dataset



Additional Dataset Exploration

Make the waste classification model more robust by incorporating diverse and extensive data sources.

- Expansion of the model’s training data beyond initial datasets.
- Enhancement of the model’s ability to recognize a wider range of waste types and conditions.

Our Team

Laurent Meunier
Client

Thomas Gardos
Instructor

Cheolmin Oh
Data Scientist

Mustafa Taibah
Data Scientist

Tessa Wu
Data Scientist

THANK YOU!