Real-Time Bottle Detection Using YOLOv8 and Roboflow
Improving Object Detection in Multi-Object Scene

Group 6

Leung Ho Cheung 13742912

Chris Limbu 13861857

To Siu Lap 13709398

Guo Kejian 12928966

Cheung Wing Choi 13714095

Background and Challenge

Background:

- YOLOv8:
  Real-time object detection, segmentation and classification
- Roboflow:
  Platform for training computer vision model datasets

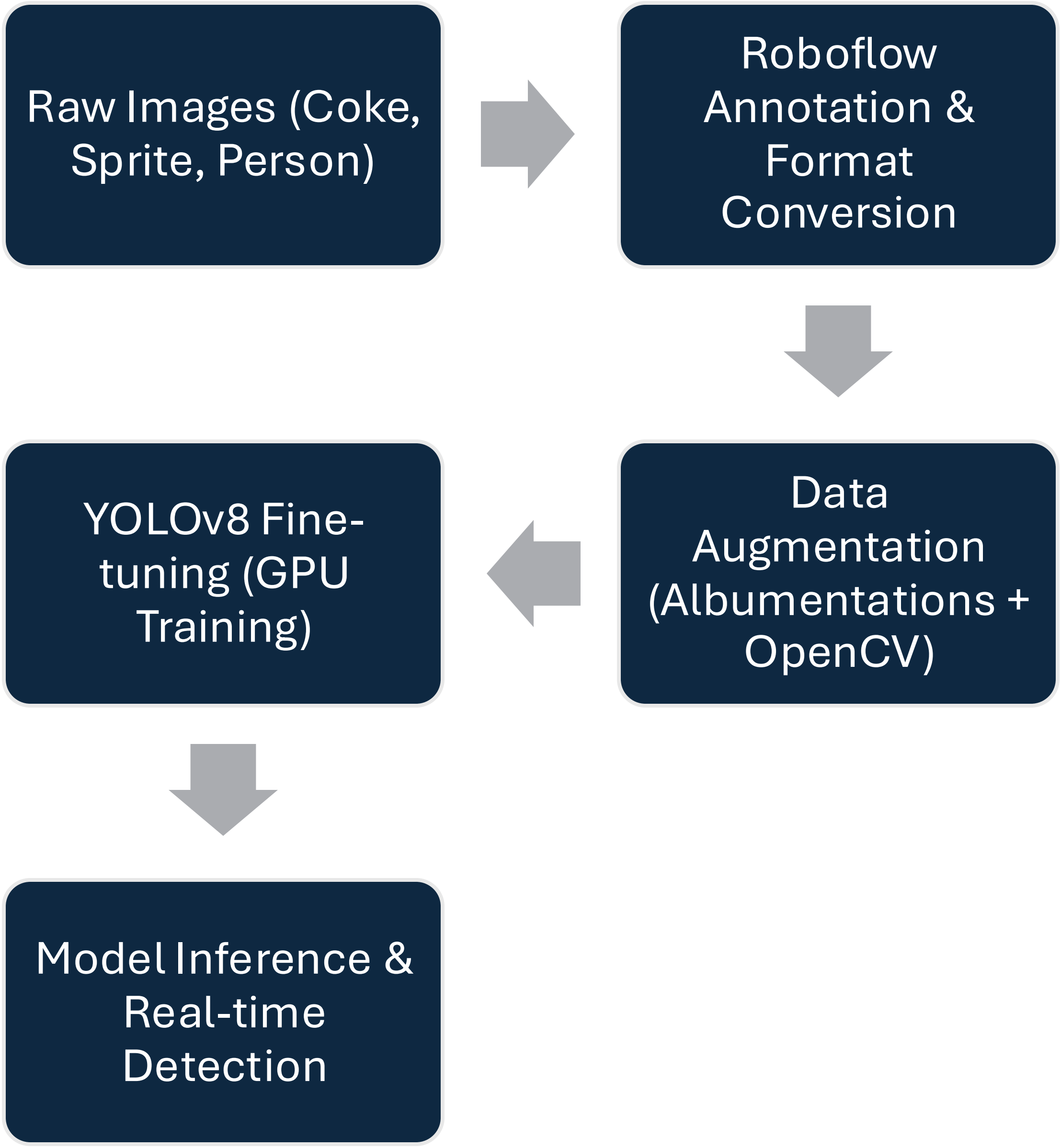
Challenges:

- Pre-trained YOLOv8 models online heavily biased toward human detection
- Models prioritize human detection, lowering detection scores of other objects

Solution Overview

- Collect and annotate bottle images using Roboflow
- Apply data augmentation (Albumentations + OpenCV)
- Fine-tune YOLOv8 on custom dataset with school GPU
- Adjust class weights to boost bottle detection priority

Technical Architecture & Training Pipeline



Technical Stack

- Libraries
  - Ultralytics YOLO v8
  - OpenCV
  - Albumentations
  - PyTorch
- Hardware
  - School GPU (CUDA enabled)
- Language
  - Python

Key Improvements & Optimization

- Dataset Optimization
  - Expanded sample diversity
    - Angles
    - Lighting
    - Blackground
  - Balanced class distribution
    - Equal bottles vs person samples
- Model Adjustments
  - Class weight rebalancing
    - Weighted loss function for bottle classes
  - Hyperparameter tuning
    - Learning rate
    - Augmentation intensity
  - Training epochs and early stopping strategy

Results & Evaluation

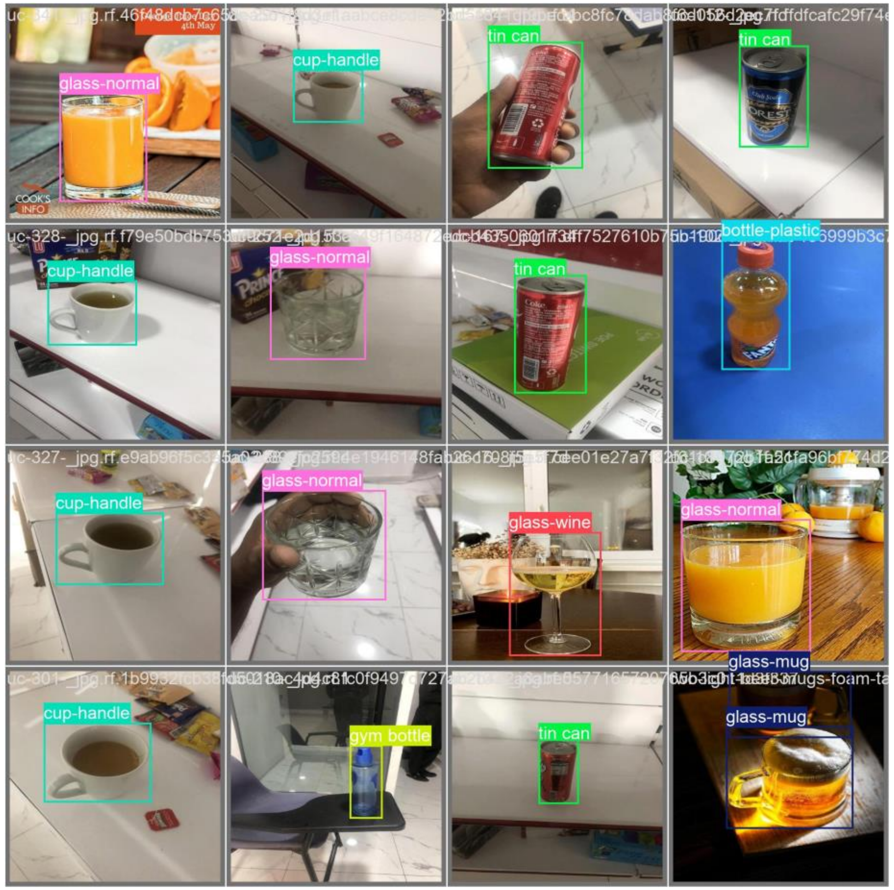
	Fine tune	Base (Original)
mAP@50	0.739	0.0067
mAP@50-95	0.572	0.0047
Precision	0.683	0.0124
Recall	0.661	0.0479

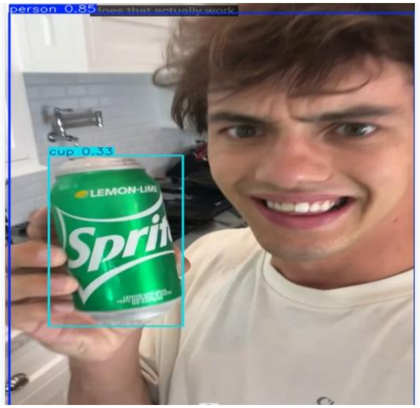
Conclusion: The base pretrained model cannot effectively detect bottles and cups due to class mismatch. Fine-tuning on our custom beverage dataset dramatically improves performance, confirming that domain-specific adaptation is crucial for reliable object detection.

Original model detection

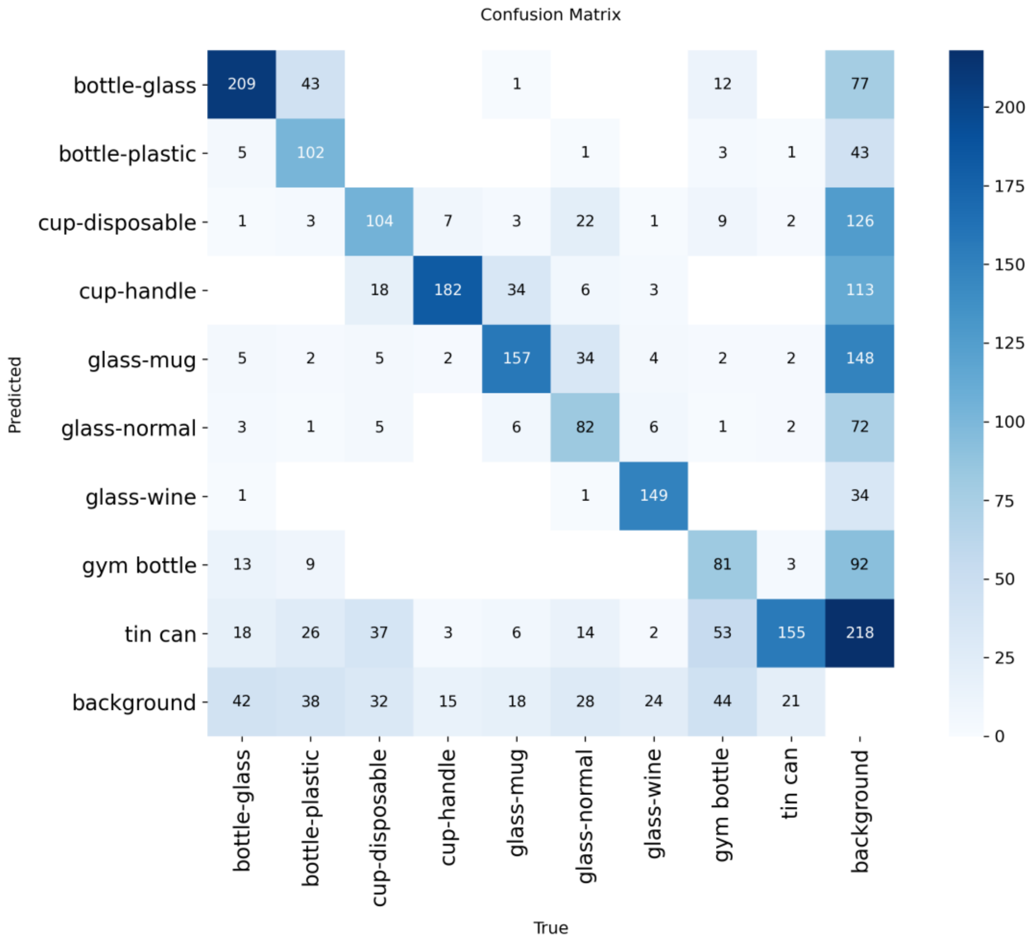
Fine-tuned model detection

VS





Fine-tuned model confusion matrix



Next Steps

Enhance the Dataset (Data Augmentation & Class Balance)

Certain categories (e.g., glass-normal, gym bottle) exhibit low recall and precision, with notable confusion observed in the confusion matrix.

It is recommended to collect more real-world images, especially for underperforming classes, and to include "hard cases," occlusions, or diverse angles to improve the model's generalization capability.

Apply data augmentation techniques (such as rotation, cropping, brightness adjustment, noise injection, and MixUp) to automatically generate transformed samples. This increases data diversity and helps alleviate class imbalance issues.