COMP 4330SEF Group Project

Real-Time **Container Detection**  
Using YOLOv8 & Roboflow  
Improving Object Detection   
In Multi-Object Scene

**Team Members**

|  |  |
| --- | --- |
| Leung Ho Cheung | 13742912 |
| Chris Limbu | 13861857​ |
| To Siu Lap | 13709398 |
| Guo Kejian | 12928966​ |
| Cheng Wing Choi | 13714095 |

**Project Task Distribution**

|  |  |  |
| --- | --- | --- |
| Member | Task | Description |
| Leung Ho Cheung | Model Training & Hyperparameter Tuning | Develop training scripts, configure GPU environment, tune hyperparameters |
| Chris Limbu | Dataset Collection & Data Preprocessing | Assist annotation, handle data augmentation and dataset splitting |
| To Siu Lap | Model Training & Evaluation | Assist training |
| Guo Kejian | Dataset Collection & Annotation | Collect and annotate images using Roboflow, define 9 object classes |
| Cheng Wing Choi | Real-time Detection & Report Writing | Compile experimental results and analyze confusion matrix, evaluate mAP metrics and per-class |

Content Page

[Introduction 4](#_Toc216041068)

[Background 4](#_Toc216041069)

[Problem 4](#_Toc216041070)

[Objectives 4](#_Toc216041071)

[Scope 4](#_Toc216041072)

[Literature Review 5](#_Toc216041073)

[Methodology 6](#_Toc216041074)

[Datasets and Preprocessing 6](#_Toc216041075)

[Model training and hyperparameter optimization 7](#_Toc216041076)

[Results 10](#_Toc216041077)

[Experimental Setup 10](#_Toc216041078)

[Quantitative Results 10](#_Toc216041079)

[Inference Speed 12](#_Toc216041080)

[Per-Class Analysis 12](#_Toc216041081)

[Qualitative Analysis of Confusion Matrix 14](#_Toc216041082)

[Conclusion 15](#_Toc216041083)

[References 16](#_Toc216041084)

##### Abstract

This project develops a YOLOv8-based detection system designed to classify nine distinct types of beverage containers, including bottles, cups, glasses, and cans. Addressing the limitation that pre-trained YOLOv8 models often perform poorly on domain-specific objects due to underrepresentation in general datasets, we fine-tuned the YOLOv8n model using a comprehensive Roboflow dataset comprising 15,641 images and 20,444 annotations. By implementing rigorous data augmentation and leveraging mixed-precision training on a GPU, the system achieved a significant performance boost: the mean Average Precision (mAP@50) increased dramatically from a baseline of 0.0067 to 0.739, confirming that domain-specific adaptation substantially enhances detection accuracy for specialized object categories.

# Introduction

## Background

Real-time object detection is essential for automated inventory management and smart retail systems. **Among current solutions, YOLOv8 is widely recognized for its high efficiency**, offering **a compelling** balance between accuracy and inference speed.

## Problem

Despite the advancements in object detection, the feature extractors in standard YOLOv8 models are optimized to prioritize human features. Consequently, when detecting bottles results in poor performance.

## Objectives

The primary purpose of this project is to develop a robust object detection system that overcomes pre-trained class bias to accurately detect bottles in multi-object scenes. To achieve this, Roboflow has been used to aggregate and annotate a dataset, ensuring a high representation of bottle instances to counteract the pre-trained "person" bias. By training the YOLOv8 architecture, tuning hyperparameters to maximize recall for the "bottle" class while suppressing false positives from background noise.

## Scope

This project focuses on detection of beverage bottles. Additionally, performance is evaluated under different lighting conditions, angles and background.

# Literature Review

**YOLOv8**  
Released by Ultralytics on January 10, 2023, YOLOv8 (You Only Look Once Version 8) represents the eighth generation of the YOLO object detection architecture. While it retains the COCO dataset for training benchmarks, YOLOv8 demonstrates significant improvements in real-time object detection capabilities compared to its predecessors (Ultralytics, 2025). Its architecture is optimized for speed and accuracy, making it a state-of-the-art solution for computer vision tasks.

**COCO Dataset**  
The Common Objects in Context (COCO) dataset, introduced by Microsoft in 2014, is a large-scale object detection, segmentation, and captioning dataset. It comprises over 328,000 images categorized into 91 object types (Microsoft, 2014). Although "bottles" are included as a category, the dataset exhibits class imbalance. Models trained solely on standard COCO weights often develop a bias towards dominant classes, such as humans, resulting in lower confidence scores for specific objects like bottles.

**Roboflow**  
Launched in January 2020, Roboflow is an end-to-end computer vision platform designed to streamline the machine learning workflow. It offers comprehensive tools for data management, image preprocessing, and annotation, facilitating the creation and deployment of custom models (Huang, C., 2024). Its impact on the industry is significant; according to a 2024 report by GV (formerly Google Ventures), Roboflow has supported over 25,000 organizations since its inception (M, T., 2025).

**Existing Applications**  
Bottle detection using the YOLO architecture is an established research area, though specific implementations vary. A search for "YOLOv8" on GitHub yields approximately 27,000 results; however, refining the query to include "bottle detection" drastically reduces this number to 78 repositories. Notably, the majority of existing solutions still rely on older architectures, such as YOLOv5. Among the YOLOv8-specific implementations, the most prominent repository (by star count) was created by user "Iamm3taphorical" and last updated in July 2025. While this model achieves high confidence scores, its generalization capability is limited by the training data variety, classifying bottles only into basic categories such as water, glass, and plastic.

# Methodology

This project adopts a model fine-tuning strategy based on transfer learning, with the core goal of optimizing the general YOLOv8 model into a fine-grained classification detection model focusing on beverage containers. The methodology encompasses the complete process from data preparation and model initialization to training optimization, ensuring reproducibility of experiments.

## Datasets and Preprocessing

The success of this project relies on a high-quality, diverse domain-specific dataset.

* Data source: We use an open-source dataset directly from the Roboflow platform. This dataset is specific to beverage container inspection tasks and contains 9 finely defined categories: bottle-glass, bottle-plastic, cup-disposable, cup-handle, glass-mug, glass-normal, glass-wine, gym bottle, tin can.
* Data Preprocessing and Enhancement: All data preprocessing and enhancement work is done automatically by the ‘Roboflow’ platform. This includes, but is not limited to: pixel-level enhancements such as random rotation, brightness and contrast adjustments, blur, and noise injection using the ‘Albumentations’ library; At the same time, it is combined with OpenCV for geometric transformation. This automated process greatly increases the diversity of data, effectively simulating complex real-world scenarios, and its core purpose is to prevent the model from overfitting and improve its generalization capabilities.

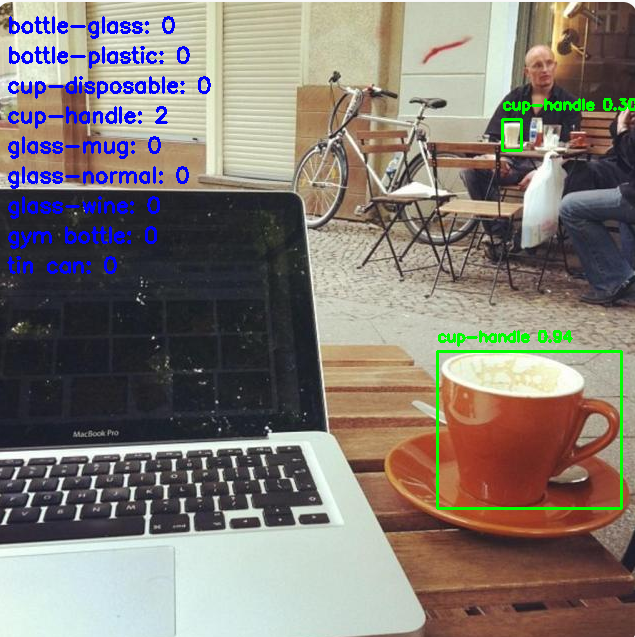
 

Fig 1: detection result on test images

## Model training and hyperparameter optimization

Model fine-tuning is achieved through the ‘Ultralytics’ framework, with the selection of key training parameters based on a balance between industry best practices and hardware resources. The detailed configuration and theoretical rationale are as follows: 一張含有 文字, 螢幕擷取畫面, 字型 的圖片

AI 產生的內容可能不正確。

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

AI 產生的內容可能不正確。

**Evaluation Metrics**

To comprehensively and objectively assess model performance, we adopt the following set of standard metrics:

|  |  |  |  |
| --- | --- | --- | --- |
| mAP@50 | mAP@50-95 | Precision | Recall |

After the completion of training and evaluation of model, a Flask-based web dashboard is built allowing upload of images or videos in order to view detection results. Additionally, the web dashboard provides live webcam streaming for real-time object detection.

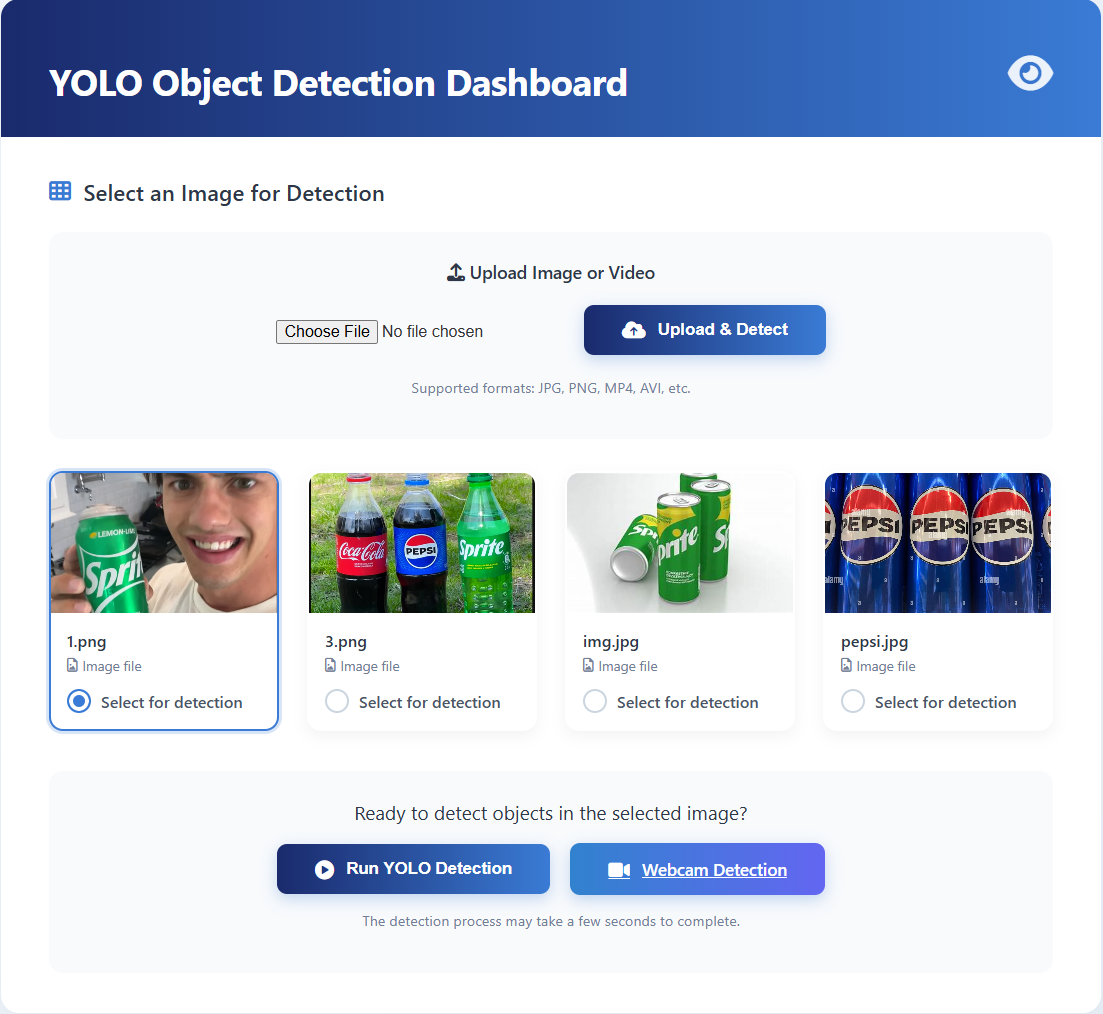


Fig 2: web dashboard UI

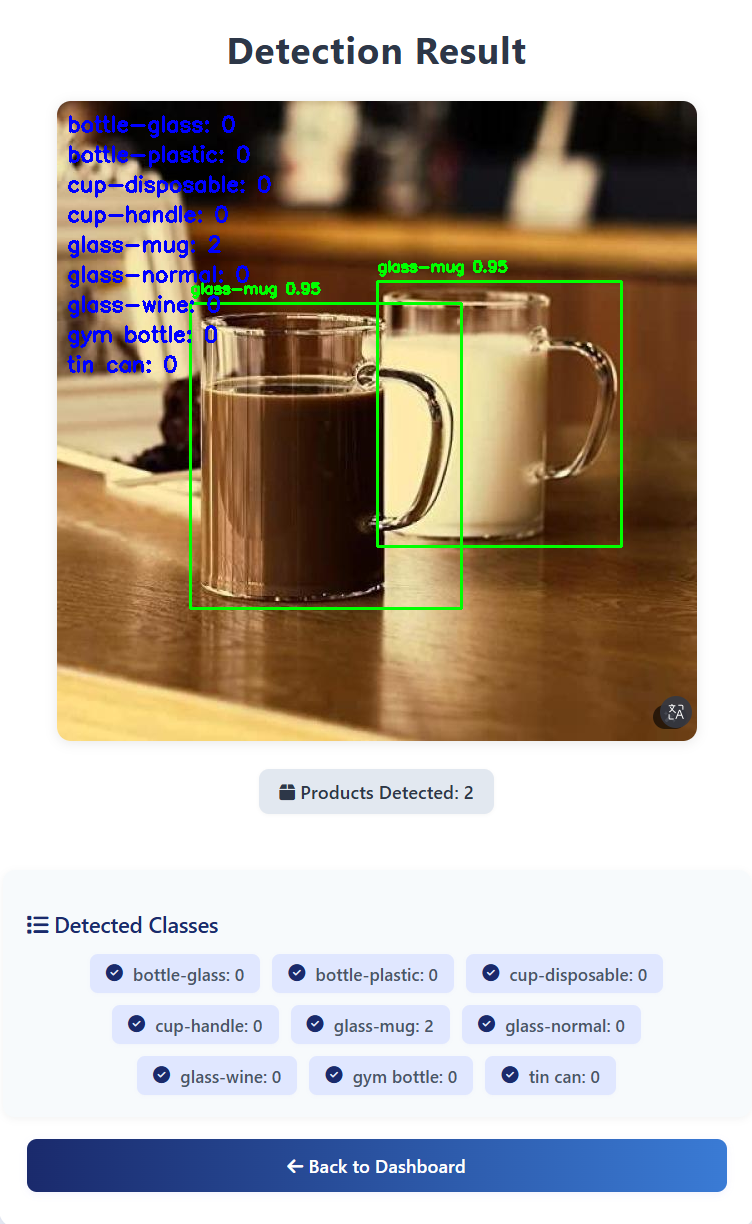
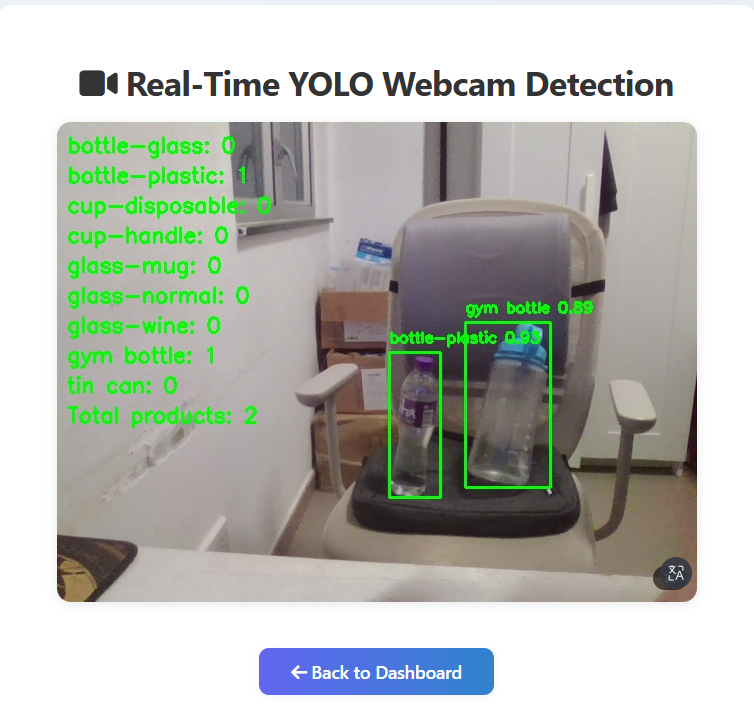
 

Fig 3: detection result UI for image/webcam

# Results

## Experimental Setup

Experiments were conducted on an NVIDIA RTX 4060 Ti GPU, training YOLOv8n for 50 epochs with batch size of 16 and input image size of 640×640. Mixed precision training (FP16) was enabled for faster computation.

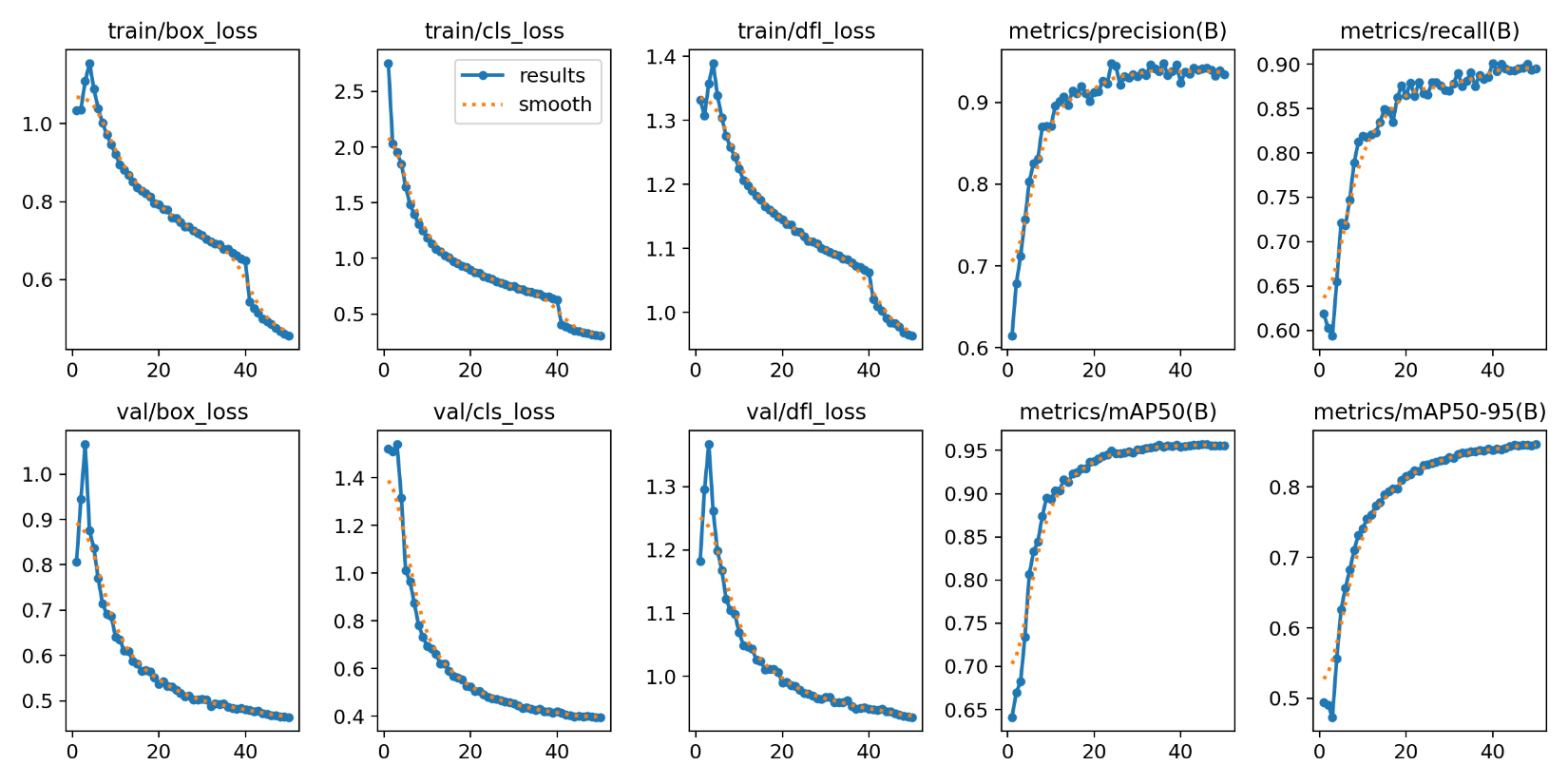
The dataset contains a total of 15,641 images split into training (13,685 images), validation (1,303 images), and test (653 images) sets across 9 beverage container classes. The training set

includes 20,444 object annotations with a relatively balanced distribution: bottle-glass (2,602), cup-handle (2,511), glass-mug (2,411), glass-normal (2,334), gym bottle (2,322), cup-disposable (2,157), bottle-plastic (2,145), glass-wine (2,025), and tin can (1,937).

## Quantitative Results

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Model* | *mAP@50* | *mAP@50-95* | *Precision* | *Recall* |
| *Baseline (yolov8n.pt)* | *0.0067* | *0.0047* | *0.0124* | *0.0479* |
| *Fine-tuned (Ours)* | *0.956* | *0.861* | *0.934* | *0.897* |

Fine-tuning achieves **significant improvement** in mAP@50, proving domain-specific training is essential.



|  |
| --- |
|  |
|  |

## Inference Speed

**Speed Comparison (GPU vs CPU)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Hardware | Preprocess | Inference | Postprocess | Total | FPS |
| RTX 4060 Ti | 0.2ms | 1.4ms | 0.9ms | 2.5ms | 400 |
| CPU(i5-13600kf) | 1.5ms | 85ms | 2ms | 88.5ms | 11 |

## Per-Class Analysis

**Class-wise Performance & Sample Distribution**

|  |  |  |
| --- | --- | --- |
| Class | mAP@50 | Train Samples |
| Bottle-plastic | 0.84 | ~2100 |
| Glass-normal | 0.95 | ~2300 |
| Cup-disposable | 0.84 | ~2100 |
| Bottle-glass | 0.93 | ~2600 |
| Cup-handle | 0.95 | ~2500 |
| Glass-wine | 0.95 | ~2000 |
| Glass-mug | 0.89 | ~2400 |
| Tin can | 0.95 | ~1900 |
| Gym bottle | 0.88 | ~2300 |

一張含有 文字, 螢幕擷取畫面, 圖表, 數字 的圖片

AI 產生的內容可能不正確。

|  |  |
| --- | --- |
| 一張含有 瓶子, 汽水, 飲料, 室內 的圖片  AI 產生的內容可能不正確。 | 一張含有 拼貼畫, 螢幕擷取畫面, 玩具, 設備 的圖片  AI 產生的內容可能不正確。 |

## Qualitative Analysis of Confusion Matrix

The confusion matrix reveals excellent classification performance across most categories, with **cup-handle, glass-normal, glass-wine, and tin can** achieve a **remarkable 95%** true positive rate. This indicates that the model effectively extracts unique features such as handles, stems, and metallic textures. Even challenging transparent classes like **bottle-glass (93%) and gym bottle (88%) performed strongly**, contradicting common issues with background interference. The primary source of error stems from material ambiguity: **bottle-plastic (84%)** shows the highest confusion rate, with 8% of instances misclassified as bottle-glass, suggesting the model struggles slightly to distinguish between clear plastic and glass textures. Overall, inter-class confusion is minimal, and background false positives remain low across all categories.

# Conclusion

This project successfully verifies the efficiency of targeted transfer learning to solve the adaptability problem in the field of pre-trained models. Faced with the performance limitations of the YOLOv8 model in the task of inspecting fine-grained beverage containers, we adopted a systematic optimization scheme: high-quality open-source datasets based on the Roboflow platform, combined with automated data enhancement processes, to fine-tune the model end-to-end.

Through a well-designed training strategy that includes conservative learning rates, early stop mechanisms, and mixed-precision training, we effectively optimized a universal object detection model into a specialized recognition tool focused on 9 types of beverage containers. The optimization model achieved a significant improvement in mAP@50, from 0.67% to 95.6%, demonstrating the method's strong potential in adapting general-purpose AI capabilities to specific application scenarios.

This methodology is universal and provides a clear and reproducible technical path for other inspection tasks that require fine-grained object classification, such as retail product identification, industrial parts sorting, medical image analysis, etc.

# References

Huang, C. (2024, November 19). *Roboflow: Making the (real) world programmable*. GV News. <https://www.gv.com/news/roboflow>

Iamm3taphorical. (2025). *YOLO-OpenCV-bottle-detection* [Computer software]. GitHub. <https://github.com/Iamm3taphorical/yolo-opencv-bottle-detection>

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft COCO: Common objects in context. In D. Fleet, T. Pajdla, B. Schiele, & T. Tuytelaars (Eds.), *Computer Vision – ECCV 2014* (pp. 740–755). Springer International Publishing. <https://doi.org/10.1007/978-3-319-10602-1_48>

M, T. (2025, July 9). *Best computer vision companies in 2025*. Roboflow Blog. <https://blog.roboflow.com/computer-vision-companies/>

Ultralytics. (2025). *Explore Ultralytics YOLOv8*. Ultralytics YOLO Docs. Retrieved October 28, 2025, from <https://docs.ultralytics.com/models/yolov8/>