REAL-TIME OBJECT DETECTION AND SEGMENTATION OF COMMON GROCERY ITEMS

Prepared by Ryan Roi Cayas

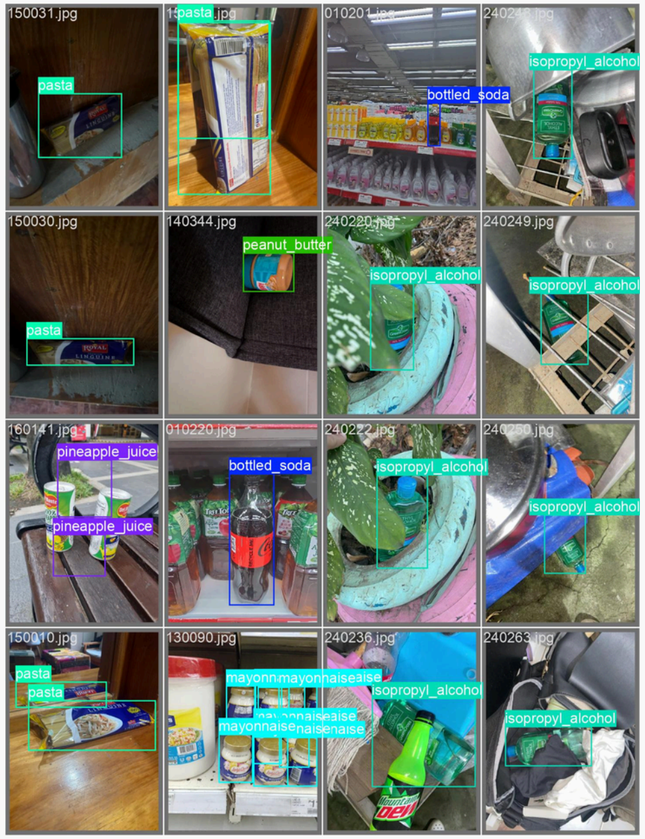
The Grocery Dataset

8,449
Images

24
Grocery Items

Training Images: 7604
Validation Images: 845

Preprocessing: Duplicate bounding boxes were removed and HD images were resized.



Model Training

Ultralytics was used to train YOLO11 models of varying sizes (nano, small, and medium). The following training approaches were performed:

Training of all parameters

All model parameters were retrained on the grocery dataset.

For models:
YOLO11n, YOLO11s

Freezing Backbone Layers

The first 11 layer blocks which served as the YOLO backbone were freezed.

For model: **YOLO11n**

Training the last layer only

The first 23 layer blocks of the YOLO model were freezed.

For models:
YOLO11n, YOLO11s, YOLO11m

Training Configuration: Epochs: 100, Patiance: 10, Batch Size: Auto
(other config set to Ultralytics default values)

Inference: Gradio was used to build an interface for real-time inference.

EVALUATION RESULTS

🔼 Higher mAP is desirable.
🔽 Lower Inference Speed is better.

BEST DETECTION MODEL

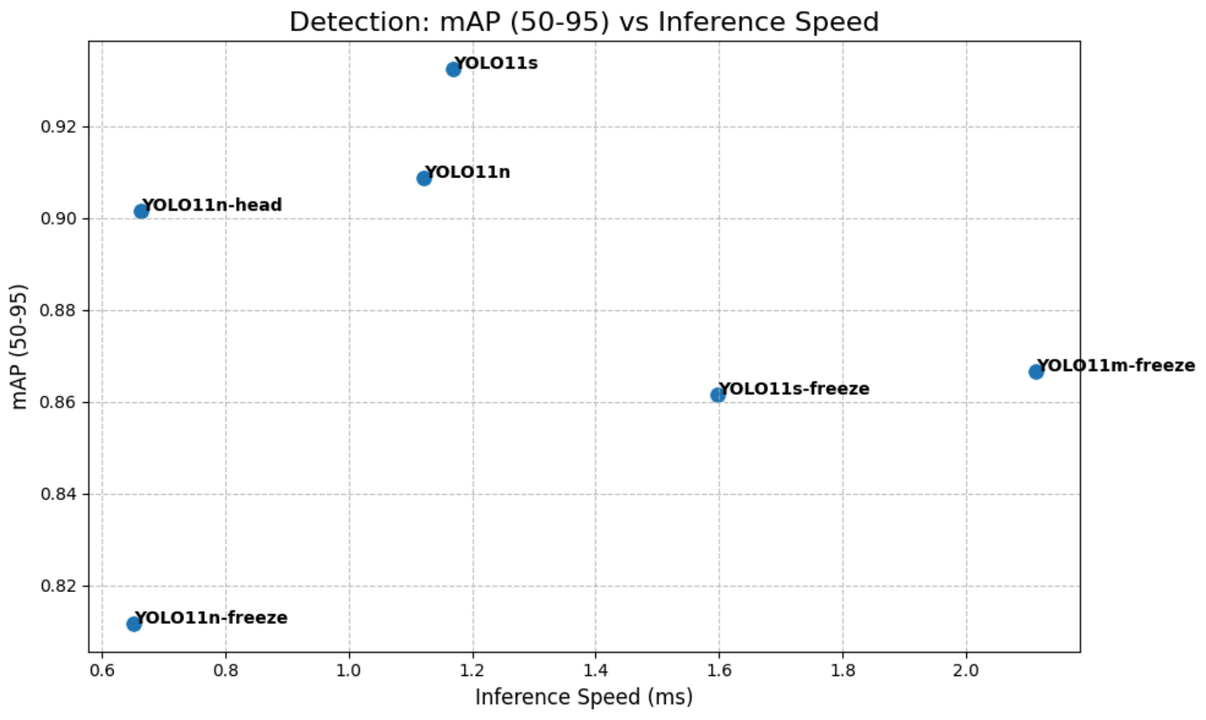
YOLO11-small

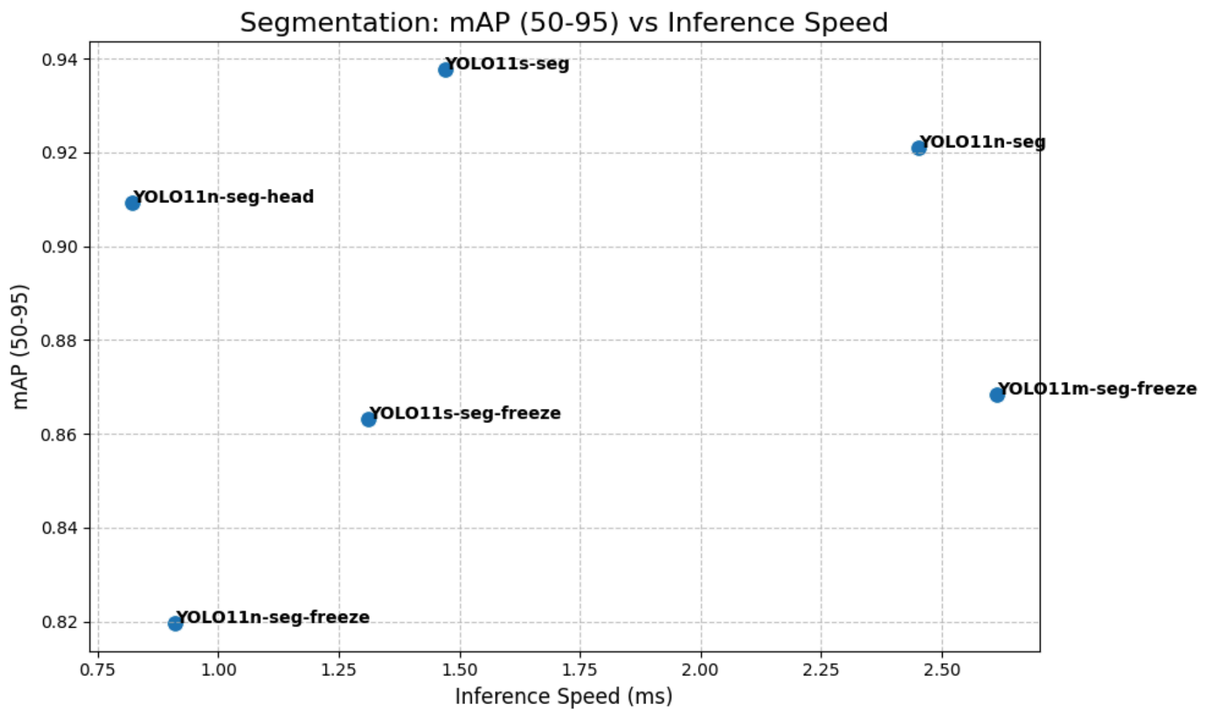
93.24%	1.17
Validation mAP (50-95)	A100 Inference Speed (ms/img)

BEST SEGMENTATION MODEL

YOLO11-small-seg

93.77%	1.47
Validation mAP (50-95)	A100 Inference Speed (ms/img)





Remarks:

- Two classes in particular have low mAP (50-95)
  - Canned tuna: 74.61% (Detection), 76.19% (Segmentation)
  - Soy Sauce: 76.87% (Detection), 77.70% (Segmentation)
- Real-time inference is poor for some items, notably for canned products.