REAL-TIME OBJECT DETECTION AND SEGMENTATION OF COMMON GROCERY ITEMS

Prepared by Ryan Roi Cayas

The Grocery Dataset v2

11,183
+2,734 from v1
Images

24
Grocery Items

Training Images: 10,064
Validation Images: 1,119

Improvements:

- More images taken in lowlight environments
- More images with multiple and occluded items

ORIGINAL IMAGES

ADDITIONAL IMAGES

DEVELOPMENT PROCESS

Aside from improving data quality, improvements were also made in training, evaluation, hardware, preprocessing, and real-time inference.

01 MODEL TRAINING

- Default and Segmentation variants of **YOLO11** were trained.
- All model sizes from **nano to extra large** were considered.
- Three training approaches were employed:
  - **End-to-End (Full):** all model parameters were retrained.
  - **Frozen Backbone:** only the “head” was trained.
  - **Detect-Only Fine-Tune:** trained the last detect layer only.
- YOLO11x was trained with a frozen backbone only.

02 MODEL EVALUATION

- The trained models were evaluated on unseen validation set based on their **mAP (50-95)** and **inference speed** on an A100 GPU.
- Segmentation variants of the YOLO11 models were also considered for object detection.
- Final models for object detection and segmentation were chosen among models with high mAP scores and reasonable speed.
- **Generalization** on real-time inference was prioritized.

03 HARDWARE AND PREPROCESSING

- A **high-resolution** (1080p @ 30fps) web camera was used for clearer image inputs.
- Camera comes with **software preprocessing** that adjusts image quality on lowlight environments.
- Image inputs retained their original dimensions **(1080x1920)** for inference.

04 INFERENCE APP

- **Gradio** was used to build the real-time inference app.
- Allows (almost) instantaneous switch between detection and segmentation tasks.
- Also allows user to instantaneously adjust **IoU threshold**.
- **“stream_every”** setting controls how often the image inputs are received (set at 0.075)

EVALUATION RESULTS

PREFERRED DETECTION MODEL

YOLO11-large -freeze-backbone

89.53% mAP (50-95)

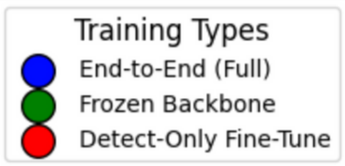
2.75 Speed (ms/img)

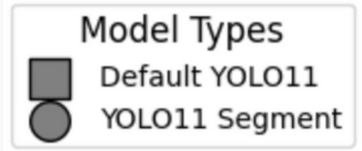
PREFERRED SEGMENT MODEL

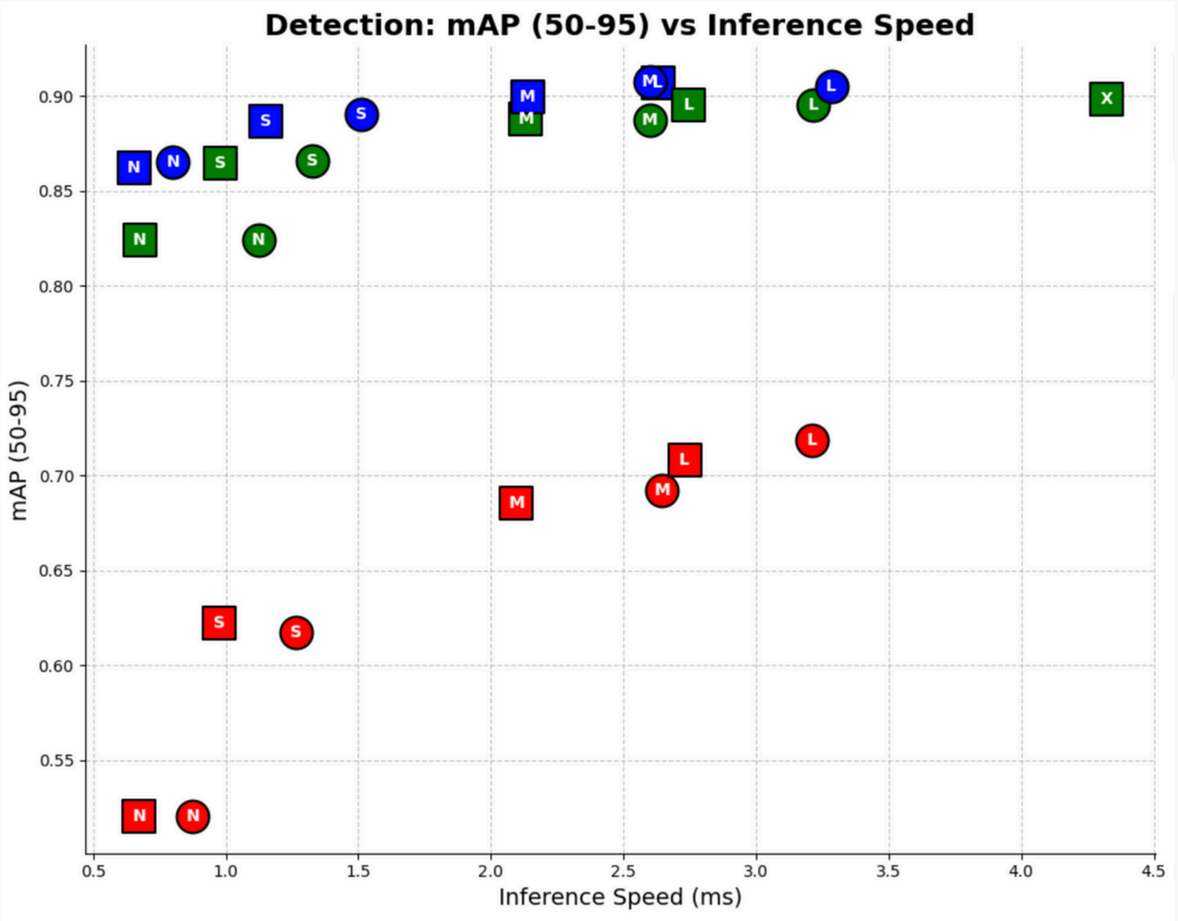
YOLO11-large -seg-freeze-backbone

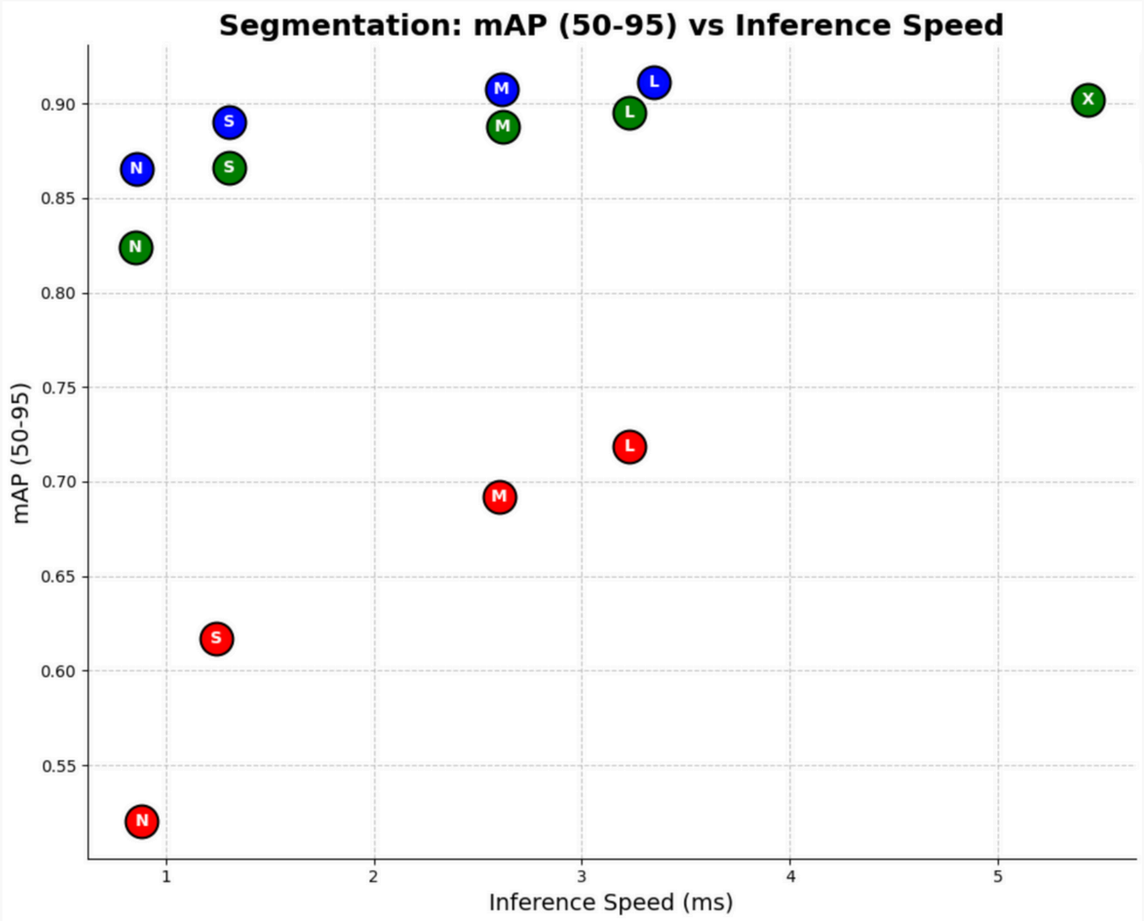
89.49% mAP (50-95)

3.23 Speed (ms/img)









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