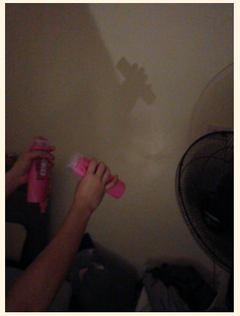
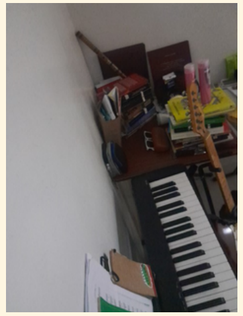
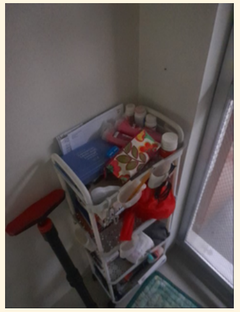
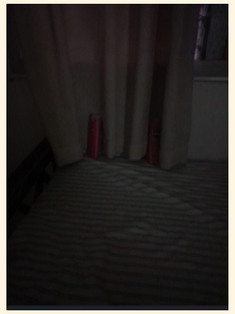
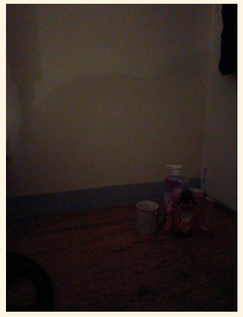
YOLO MODEL IMPROVEMENTS

DATASET

ADDED IMAGES (150 IMAGES OF SHAMPOO) WITH MORE COMPLEX ORIENTATIONS, OCCLUSIONS, DISTANCES, AND LIGHTING



MODEL DEVELOPMENT

USED LARGER MODEL (FROM YOLO11-NANO TO YOLO11-MEDIUM) AND DATA AUGMENTATION VIA ALBUMENTATIONS

```
model.train(
  data="dataset.yaml",
  project = "MEX6_runners",
  epochs=200,
  imgsz=640,
  batch=16,
  patience=10,
  name="run02",

  degrees=30,
  hsv_v=0.3,
  translate=0.4,
  shear=0.3,
  flipud=0.05,
  mixup=0.4,
  copy_paste=0.3)
```

MODEL DEPLOYMENT

WEB-BASED APPLICATION USING GRADIO AND WEBRTC

WEBRTC CONFIGURATION

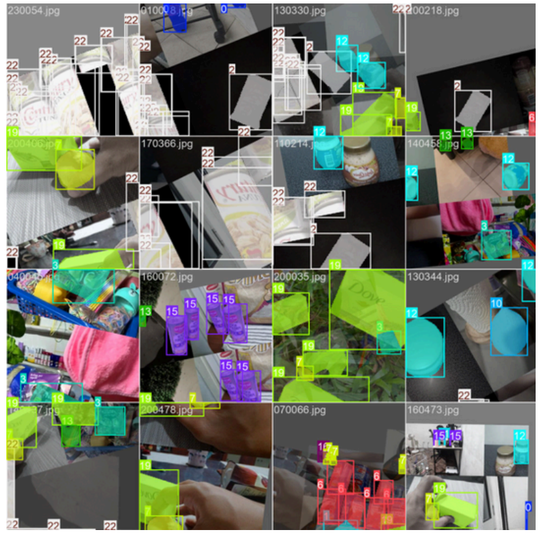
- Twilio API Tokens: Used to configure WebRTC ICE servers for smooth webcam streaming over the internet.
  - ICE servers help establish peer-to-peer (P2P) connections for real-time streaming.

GRADIO INTERFACE

- The app is initialized and launched using gr.Blocks with:
  - The WebRTC stream for live webcam feed.
  - A slider for adjusting confidence thresholds in real time.
- WebRTC Stream:
  - Captures live webcam input using the WebRTC component.
  - Passes the webcam stream to the YOLO model for detection and segmentation.
  - Uses a slider to dynamically adjust the confidence threshold for object detection (default: 0.3).

RESULTS

With Data Augmentations



HSV_V (0.4 -> 0.3)

- modifies the value (brightness) of the image by a fraction
- to detect under various lighting conditions
- **should have increased**

TRANSLATE (0.1 -> 0.4)

- translates the image horizontally and vertically by a fraction of the image size
- to detect partially visible objects

DEGREES (0.0 -> 30)

- rotates the image randomly within the specified degree range
- to recognize objects at various orientations

SHEAR (0.0 -> 0.3)

- shears the image by a specified degree
- mimicks the effect of objects being viewed from different angles

FLIPUD (0.0 -> 0.05)

- flips the image upside down with the specified probability
- increases data variability without affecting the object's characteristics.

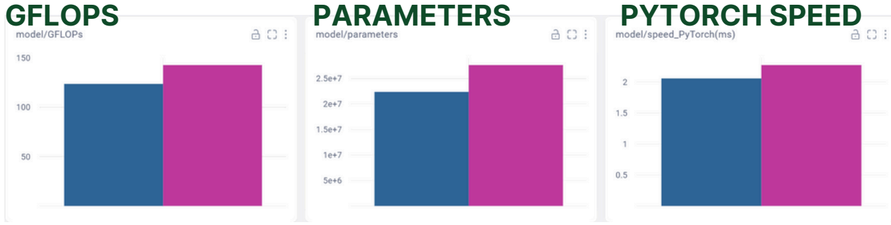
MIXUP (0.0 -> 0.4)

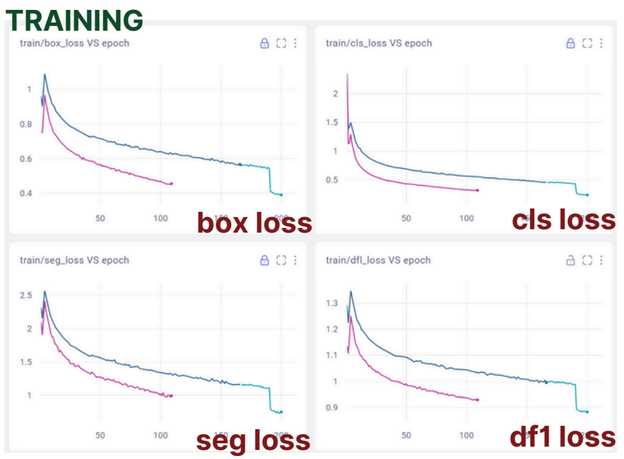
- blends two images and their labels, creating a composite image
- to generalize by introducing label noise and visual variability

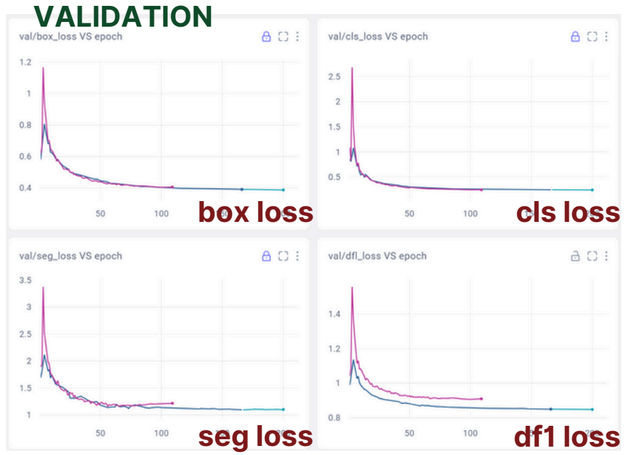
COPYPASTE (0.0 -> 0.3)

- copies and pastes objects across images
- to increase object instances and learning object occlusion

YOLO11-M w/ additional data augmentation
YOLO11-L w/o additional data augmentation







	BOX			
Model Size	P	R	mAP50	mAP50-95
Medium	96.9	92.7	95.6	88.9
Large	96.0	92.9	95.5	88.7

	MASK			
Model Size	P	R	mAP50	mAP50-95
Medium	96.6	92.2	94.8	84.9
Large	95.9	92.1	4.7	84.8

NEXT STEPS

Use ONNX

- inference with **onnxruntime.InferenceSession()**
- several pre-processing and post-processing
- not yet deployed via **webrtc**

Perform hyperparameter tuning with Ray-tune

- search space includes learning rate, momentum and data augmentation parameters
- computationally-intensive

Parameter	Value Range	Description
lr0	tune.uniform(1e-5, 1e-1)	Initial learning rate
lrf	tune.uniform(0.01, 1.0)	Final learning rate factor
momentum	tune.uniform(0.6, 0.98)	Momentum
weight_decay	tune.uniform(0.0, 0.001)	Weight decay
warmup_epochs	tune.uniform(0.0, 5.0)	Warmup epochs
warmup_momentum	tune.uniform(0.0, 0.95)	Warmup momentum
box	tune.uniform(0.02, 0.2)	Box loss weight
cls	tune.uniform(0.2, 4.0)	Class loss weight
hsv_h	tune.uniform(0.0, 0.1)	Hue augmentation range
hsv_s	tune.uniform(0.0, 0.9)	Saturation augmentation range
hsv_v	tune.uniform(0.0, 0.9)	Value (brightness) augmentation range
degrees	tune.uniform(0.0, 45.0)	Rotation augmentation range (degrees)
translate	tune.uniform(0.0, 0.9)	Translation augmentation range
scale	tune.uniform(0.0, 0.9)	Scaling augmentation range
shear	tune.uniform(0.0, 10.0)	Shear augmentation range (degrees)
perspective	tune.uniform(0.0, 0.001)	Perspective augmentation range
flipud	tune.uniform(0.0, 1.0)	Vertical flip augmentation probability
fliplr	tune.uniform(0.0, 1.0)	Horizontal flip augmentation probability
mosaic	tune.uniform(0.0, 1.0)	Mosaic augmentation probability
mixup	tune.uniform(0.0, 1.0)	Mixup augmentation probability
copy_paste	tune.uniform(0.0, 1.0)	Copy-paste augmentation probability

Use YOLO11-X

Incorporate more images that mimic real-world scenarios

Accumulate predictions across a buffer of frames and aggregate them to produce a smoother and more accurate output.