Segmentation Model Task Report

Author: Osama Al-Schameri

Date: 2025-04-28

1. Introduction

In this task, we develops a computer vision system to segment box-shaped objects in RGB images using state-of-the-art segmentation model.The task is aimed to delevlop a robust and efficient model that can accurately identify and segment these objects in various conditions.YOLOv11 is used to detect the objects, and the segmentation model is trained to predict the pixel-wise mask of the objects.

2. Objective

- This project develops a computer vision system to segment box-shaped objects in RGB images.

3. Dataset

Dataset Description

Source: OSCD Dataset

Size: 168758 instances

Classes: [Carton]

Images: 8401 images for trainng and validation / 27 images for testing

Train/Test Split: 80/20

Image Size: different sizes

Annotations: [boxes,masks]

Format: [COCO, Labelme]

Preprocessing Steps

1. Data Cleaning: Handling missing values, outliers, remove invalid labels, duplicate labels.

2. Convert labels to txt file to be accepted by YOLO such (class_id, x, y, w, h).

4. Training Setup

4.1 Model Structure

The model architecture is based on YOLOv11-S for Instance Segmentation. It is lightweight with deep feature extraction backbone. It decoupled detection head for classification and bounding box regression.The segmentation branch integrated with spatial attention modules and optimized for both speed and accuracy trade-offs.

4.2 Loss Function

Box Regression Loss : Binary Cross Entropy (BCE) + Dice Loss (for mask quality).

Segmentation Loss : Generalized Intersection over Union (GIoU) Loss.

Classification Loss : Binary Cross Entropy (BCE) Loss.

Distribution Focal Loss (DFL) : Used for precise bounding box localization.

4.3 Learning Rate & Optimizer

Optimizer: SGD (automatic set)

Initial Learning Rate: 0.01

Scheduler: Cosine learning schedule

Patience: 15

Augmentation: scaling = 0.5, hue = 0.015, saturation = 0.7, brightness = 0.4, translate = 0.1, flip right = 0.5

Epochs: 200 with early stopping

5. Performance Evaluation

5.1 Metrics

Metric	mAP@50 (IoU=0.5)	mAP@50:95 (IoU=0.5:0.95)	Precision	Recall	F1-Score
Validation	0.956	0.81	0.95	0.92	0.935

5.2 Model Result

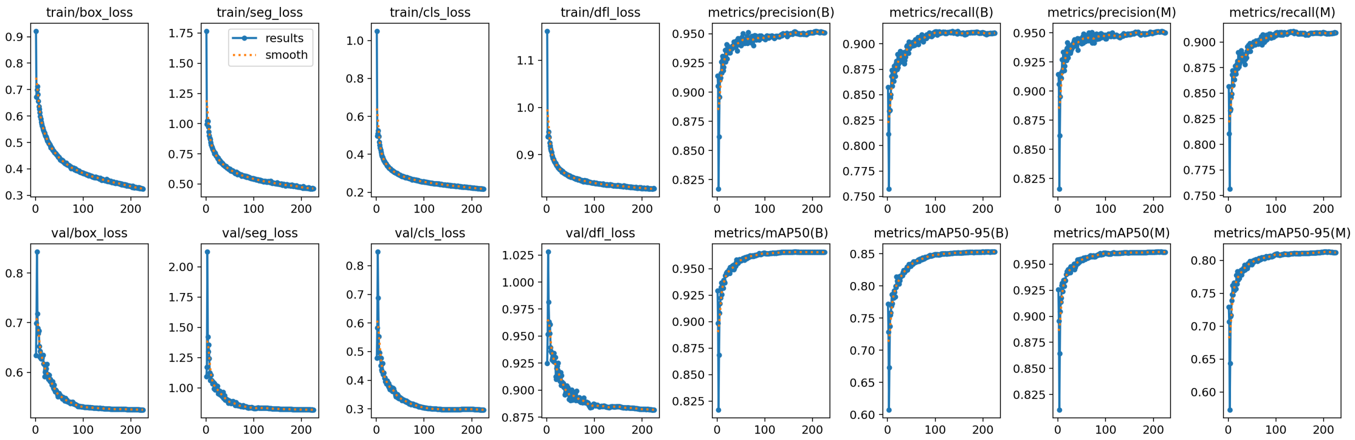

Fig 1: Model result including losses and evaluation metrics

5.3 Confusion Matrix

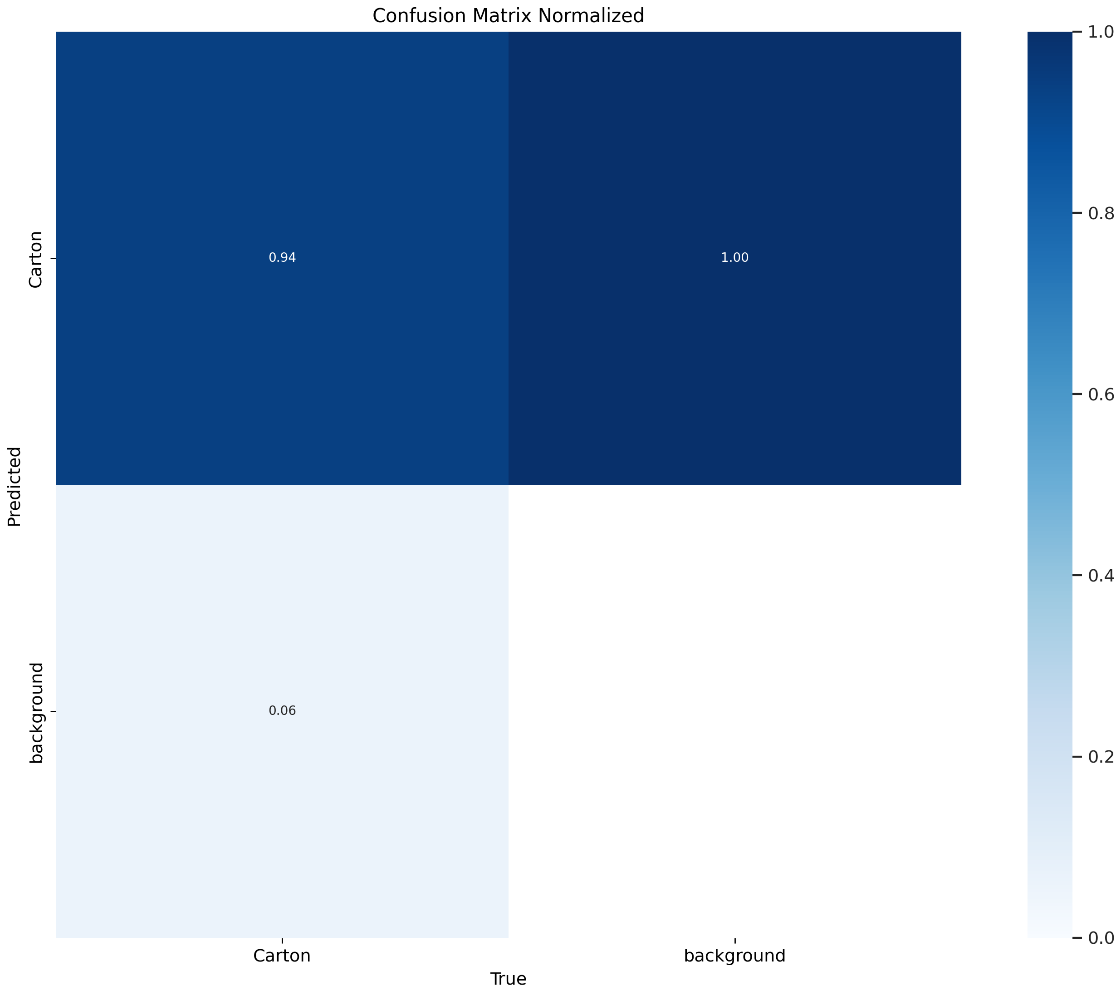

Fig 2: Confusion Matrix

5.4 Performance Analysis

Confusion matrix provides insight into the model's classification performance across two classes:

Carton and Background. True Positives (Carton correctly predicted as Carton): 0.94, the model correctly identifies 94% of cartons, indicating strong performance in detecting the target class. True Negatives (Background correctly predicted as Background): 1.00, the model perfectly identifies all background instances, showing excellent specificity. False Negatives (Carton incorrectly predicted as Background): 0.06, 6% of cartons are misclassified as background, suggesting a minor miss rate. False Positives (Background incorrectly predicted as Carton): 0.00, the model does not misclassify any background instances as cartons, indicating no false positives. Overall, the model demonstrates high accuracy, with a strong ability to distinguish cartons from the background. The 6% false negative rate for cartons is the primary area of concern, as it indicates some cartons are being missed.

All training losses decrease steadily and stabilize at low values (e.g., box_loss at ~0.4, seg_loss at ~0.2, cls_loss at ~0.8, dfl_loss at ~0.9).The smooth curves (orange) confirm consistent learning without significant fluctuations, indicating stable optimization. Validation losses follow a similar downward trend, stabilizing at slightly higher values than training losses (e.g., val/box_loss at ~0.5, val/seg_loss at ~0.3). The gap between training and validation losses is minimal, suggesting the model is not overfitting and generalizes well to unseen data. The model converges effectively, with low and stable losses. The absence of overfitting is a positive indicator of robustness.

The model achieves high precision and recall, with strong mAP scores, particularly at the 50% IoU threshold. The drop in mAP@50:95 suggests that the model's localization accuracy could be improved for more precise bounding box predictions. The carton class has a relatively consistent spatial distribution, which likely aids the model's high detection accuracy. However, the skewed width and height distributions suggest variability in carton sizes, which may contribute to the 6% false negative rate if smaller cartons are harder to detect.

6. Visualizations

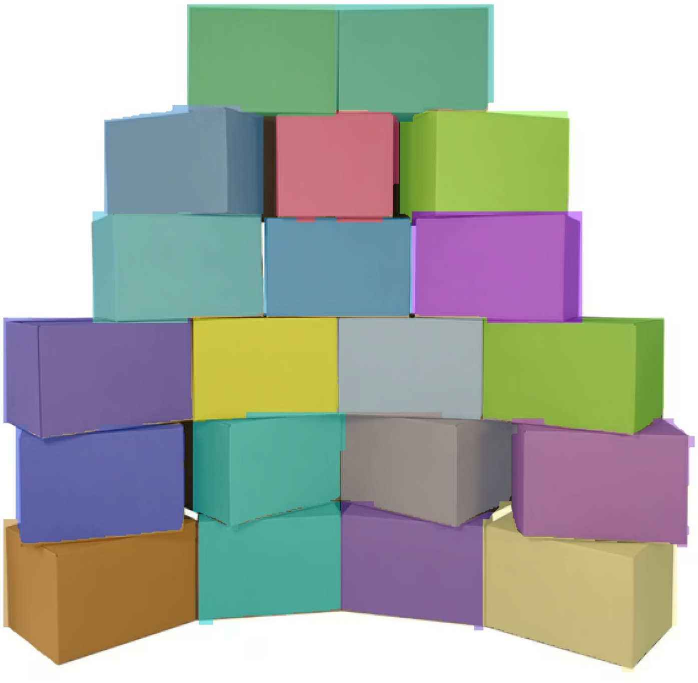

Fig 3: Model Performance on testing dataset

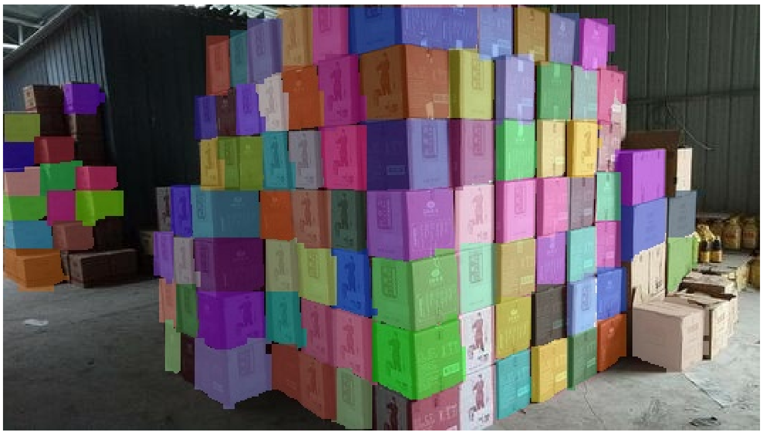

Fig 4: Model Performance on testing dataset

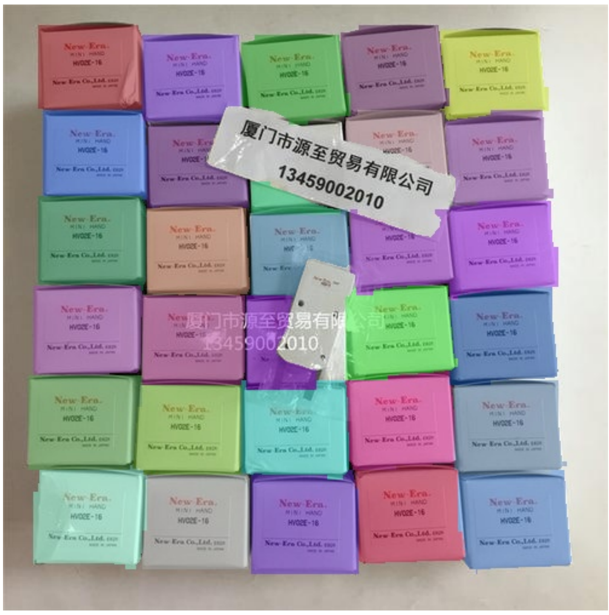

Fig 5: Model Performance on testing dataset

7. Conclusion

The YOLO model for carton detection performs well, with high accuracy, stable training, and strong metrics. However, addressing the false negative rate and improving localization precision could further enhance its effectiveness. The consistent data distribution supports the model's performance, but handling variability in carton sizes will be key to achieving near-perfect detection.

Recommendation

- Address False Negatives by augment the dataset with more examples of smaller cartons or apply data augmentation techniques such as scaling to improve detection of varied sizes.
- Improve Localization by fine-tune the model with a focus on stricter IoU thresholds to boost mAP@50:95 scores. Techniques like anchor box optimization or additional bounding box regression loss could help.

Appendix

Code Repository: https://github.com/osamaalschame/Estimate-the-3D-pose-of-a-known-box-shaped

Hardware Used: Google Cloud for the training