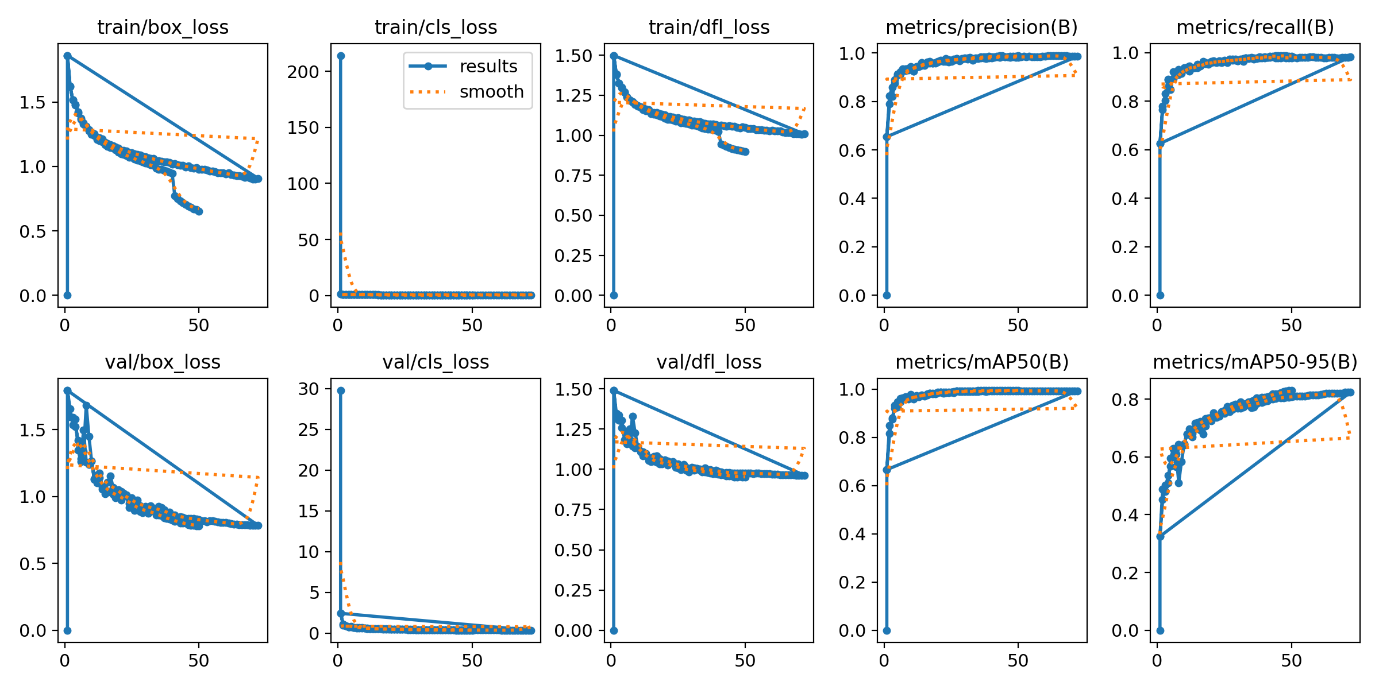
**YOLO Model Fine-Tuning & Inferencing Report**

**Name: Lo Zhi Hao**  
**Date: 21/2/2025**

**1. Model Training & Fine-Tuning Summary**

* **Base Model Used:** YoloV8
* **Dataset Used:** augmented\_dataset.zip
* **Training Parameters:**
  + Batch Size: 32
  + Epochs: 50
  + Image Size: 1280px (Width) x 720px (Height)
  + Augmentations Applied: Left-Right and Top-Down Mirroring
  + Optimizer & Learning Rate: ‘Adam’ Optimizer & LR = 0.001
* **Loss Reduction Over Epochs:**
  + Initial Loss: 1.862 (Box\_Loss) 1.466 (Cls\_Loss)
  + Final Loss: 0.6502 (Box\_Loss) 0.3058 (Cls\_Loss)
  + Graph of loss progression:

****

**2. Inference Performance**

**2.1 Dataset Information**

* **Test Dataset Used:** inferencing\_1.mp4
* **Number of Images Tested:** 721
* **Resolution of Images Used for Testing:** 1200px (Width) x 720 (Height)

**2.2 Inference Results**

|  |  |
| --- | --- |
| **Metric** | **Value** |
| FPS (Frames per Second) |  |
| Average Inference Time (ms) |  |
| Precision (%) |  |
| Recall (%) |  |
| mAP@50 (%) |  |
| mAP@50-95 (%) |  |
| Number of False Positives |  |
| Number of False Negatives |  |

* **Observations on Inference Quality:**  
  (Does it detect all expected objects? Are there misclassifications?)

Yes, misclassifications are noted, on quite a frequent basis. This is observed more frequently on SKUs that have fewer original data to train with. This may be related to the nature of how the model is being trained, as I modified all the labels to be a binary classification problem (Object/Background). Thus, the class imbalance may have led to disparity in terms of performance for different SKUs.

* **Example Detections (Screenshots with annotations attached)**
* 
* 
* 
* 

**3. Model Performance Comparison**

|  |  |  |
| --- | --- | --- |
| **Metric** | **Pre-Fine-Tuning** | **Post-Fine-Tuning** |
| Training Loss |  |  |
| mAP@50 |  |  |
| FPS |  |  |

* **Performance Gains Observed (if any):**

**4. Issues & Challenges**

* **Any anomalies in detection?**

Generally, the model is not very satisfying in identifying the object.

* **Dataset-related issues?**

Low number of background pictures in train dataset potentially causing high number of false positives, ie other objects (smartphones) may be identified as SKUs. A way of handling would be to increase number of background images as well.

* **Overfitting/underfitting observed?**

Potential overfitting issues based on eye checks, but both validation and train loss (class loss and box loss) decreasing at a consistent rate suggests otherwise (potential underfitting). Would need to experiment with different epochs and batch number to identify the issue. Generally however, the model results are not satisfying.

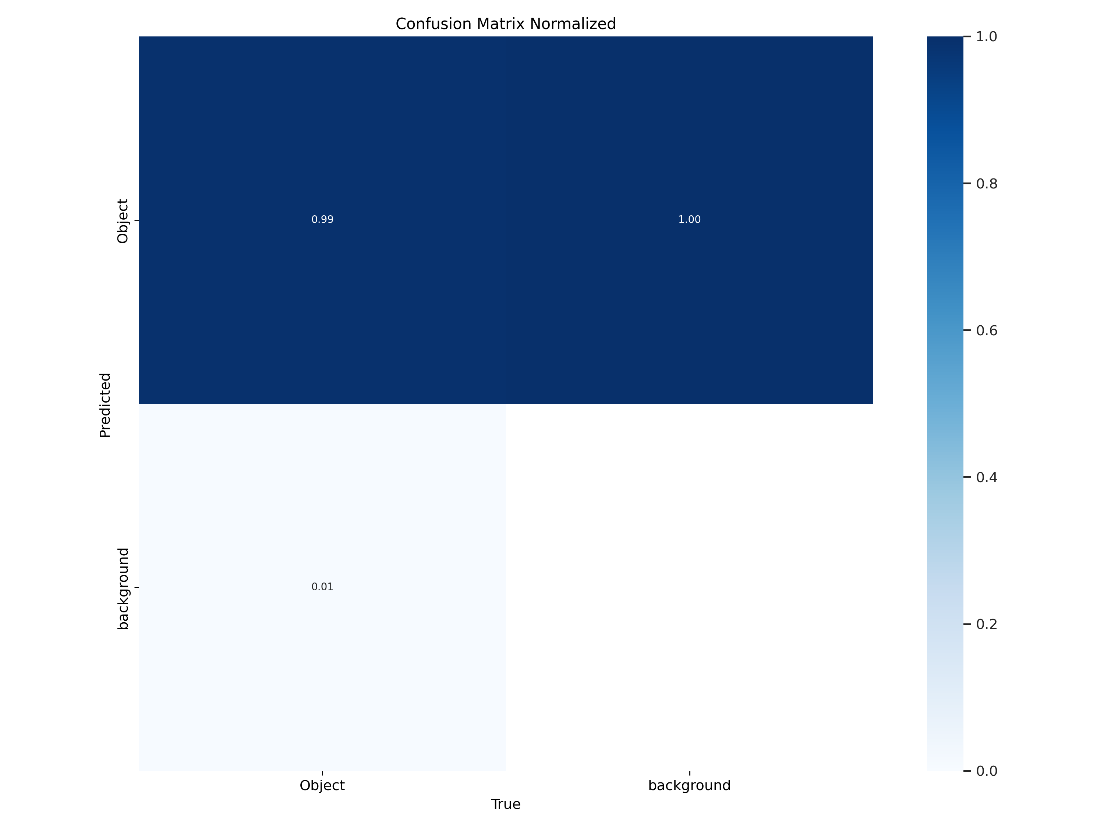
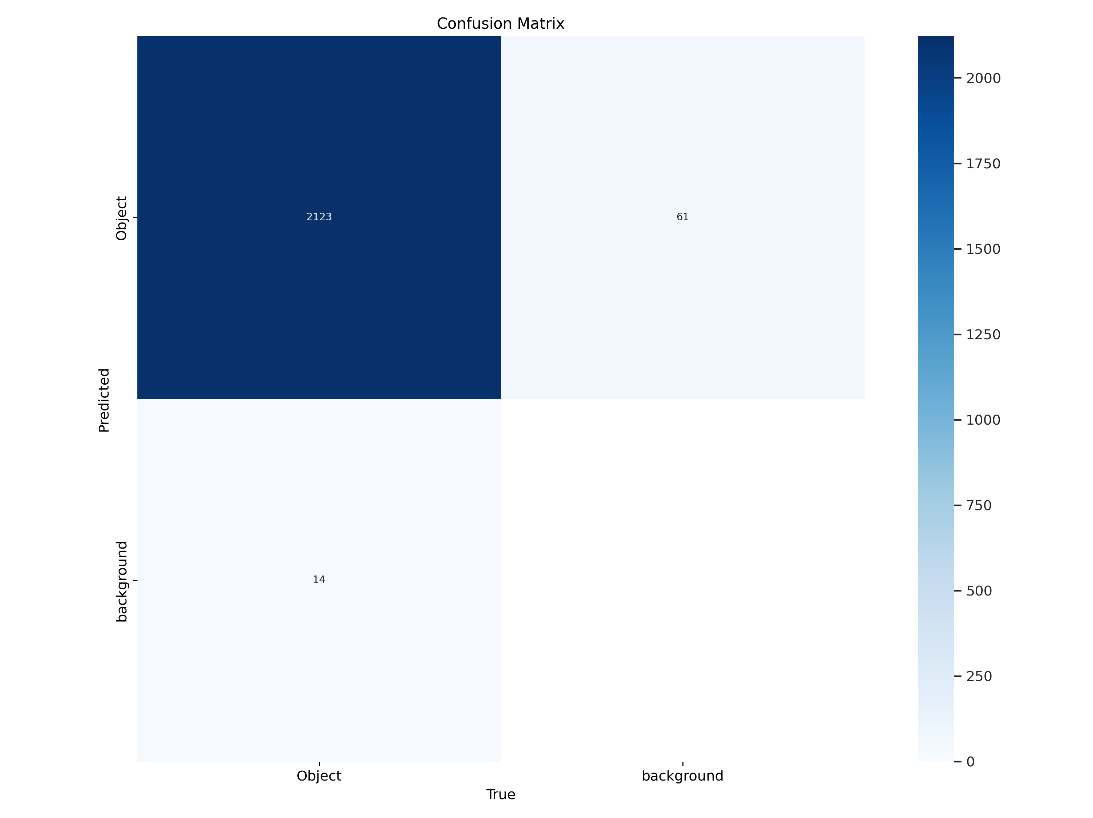
* **Any other difficulties faced?** N/A

**5. Next Steps & Recommendations**

* **Suggestions for further improvements:**

There are two parts to potentially work on and modify. Firstly, performances of the model across SKU classes varies. This may be because of the way the problem statement was framed (binary classification) and unbalanced dataset where number of frames collected for different SKUs differs, causing potential overfitting to the majority SKU class. This can be observed with the high confidence probability for MiniSalads & Salads (which is the SKU with the greatest number of frames in the dataset) as compared to the rest of the SKUs.

Second part may be a concern for high false positives in the background class, as shown in this confusion matrix. Perhaps due to the low number of background classes in training, the model has a tendency to identify background objects as SKUs. However, in this case, I do consider this as less urgent and costly issue as compared to the first part.



* **Need for additional training or dataset enhancement?**

More background data to supplement would be good for tackling the false positive issue.

* **Hardware or software limitations noticed?** N/A
* **Plan for the next iteration:**

The upcoming course of actions is as follows:  
1. Retrain the model with each SKU as individual classes. This is important to check if it tackles the performance variance across SKUs.

2. Add in more background photos. I will be trying on further augmentation techniques (blurring, gradient, etc) particularly for background photos.

3. Vary the number of epochs and batch size for training to verify the state the model is in (underfitted or overfitted). Primarily would be experimenting with epochs: 10/50/100 and batch size: 16/32/64

4. An important part of model training is to identify a qualitative metrics to optimise for in my opinion. Thus, I will work on manual labelling the current inferencing video, such that I can present a more qualitative performance review in the upcoming days. This is to come up with a qualitative review of the actual test set (misclassification, box loss, class loss) to make the results different models comparable.

**Notes / Additional Comments**

(Any extra remarks or insights)

N/A