PROJECT H1: KAGGLE COTTON WEED DETECTION COMPETITION

MARKUS MARIPUU & SIIM ILISON

INTRODUCTION

The 3LC Cotton Weed Detection competition on Kaggle focuses on identifying three major weed species in cotton fields using image data. The task is a multi-class object detection problem where the goal is to detect: Carpet Weed, Morning Glory, Palmer Amaranth.

We have to use the YOLOv8n model and a dataset that is already split into three:

1. Training (542 images and labels)
2. Validation (133 images and labels)
3. Test (170 images)

GOALS

Our main goal is to train the YOLOv8n model so it is capable of detecting all three weed types as accurately as possible. The metric used to evaluate the model's accuracy is mAP@0.5 (Mean Average Precision with Intersection over Union ≥ 0.5). In order to achieve a higher mAP@0.5 score the model has to handle real field variability (lighting, perspective, overlapping plants). We also aimed to be in the top 50% in the leaderboard. Our initial hypothesis was that fixing the incorrect labelling in the training dataset would have the highest impact on our score.

PROCESS

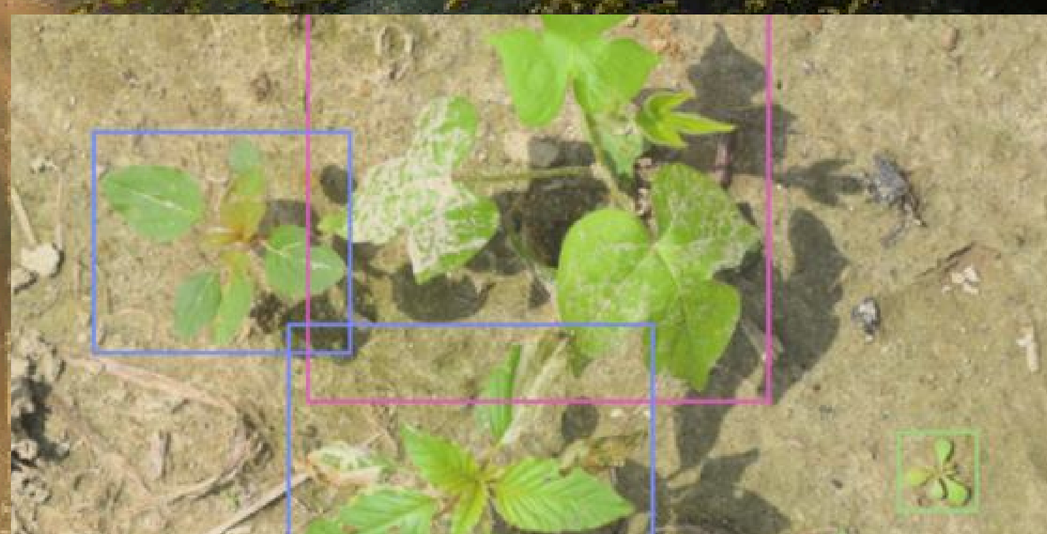
Since we were only allowed to use the YOLOv8n model our only options were to improve the data and change the hyper parameters of the model.

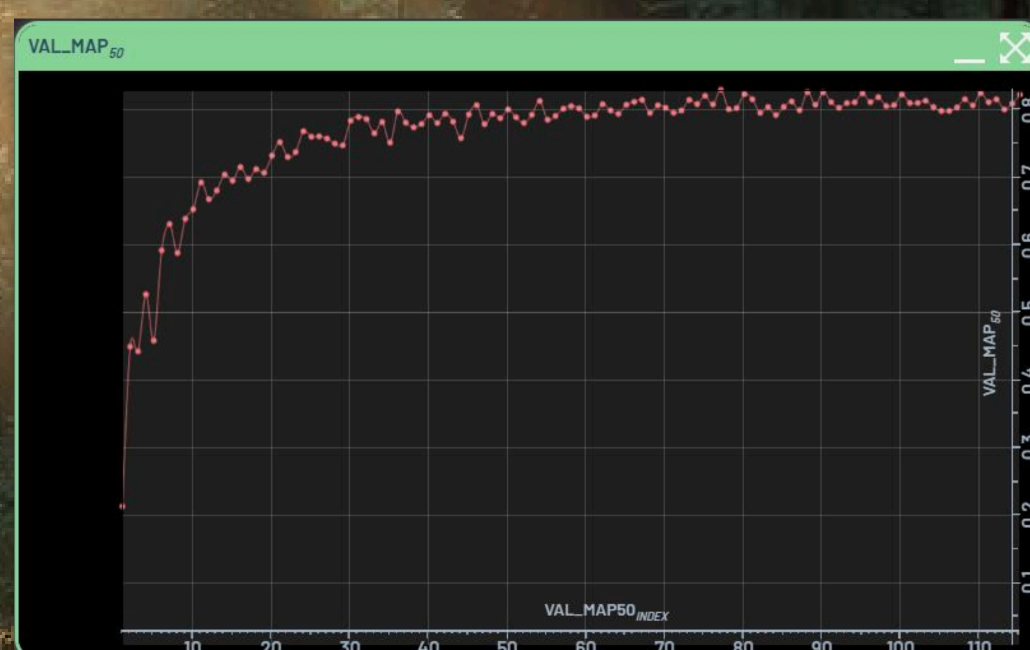
Methods used:

- Hyper parameter tuning
- Fixing the labelling on every image using the 3LC interface
- Data augmentation using Roboflow

Constraints:

- Compute (we were unable to train on a GPU)
- Time - due to lack of compute





RESULTS

The data augmentation didn't make a difference because there were all kinds of pictures in training dataset. Label fixing also didn't result in higher mAP@50 score and we assume it is because the test dataset is also poorly labelled. Throughout the whole process we achieved the best mAP@50 score with hyper parameter tuning. The mAP@50 of our best YOLOv8n model is shown in the graph. As we can see from the graph, the mAP@50 score has a major increase in the beginning and plateaus after about 50 epochs of training. The most crucial hyper parameters were epoch size and batch size. With 150 epochs, 4 batch size and after only 10 hours we achieved the best mAP@50 score of 0.837 on the test dataset. In the end we finished 60th.

CONCLUSION

Our YOLOv8n model successfully learned to detect the three weed species with strong accuracy despite dataset challenges. The most performance improvements were achieved by hyper parameter tuning.

When looking through the data, we discovered a data leak. Some of the pictures from the validation dataset were also found in the test dataset meaning that, if somebody used the validation dataset to train the model then they would get better results on the test dataset.