Applying Faster R-CNN and Mask R-CNN on the MinneApple Fruit Detection Challenge*

Callum Clark

Maastricht University, Paul-Henri Spaaklaan 1, 6229 EN Maastricht, Netherlands
c.clark@student.maastrichtuniversity.nl

1 Introduction

Technology has already had a major impact on the fruit industry. Image recognition for the detection and localisation of fruit is seen as a vital step to improve yield estimation[5] and therefore actual yields, as well as the automation of harvesting.

It is challenging to produce a generalised automation technique for all crop types because there are large differences in farming techniques across crop types. Previous research has been successful in developing specific techniques for individual crop types. In the case of apples, research has been done in autonomous apple picking[1][2], automated pruning[7] and yield estimation[8][5]. All of these tasks require the accurate detection and localization of apples. This is an important first step, since after detecting apples, one can estimate the size of the current harvest and make predictions on the final yield. Moreover, detecting and localizing apples can serve as a basis for assessing their health, detecting pests and thus support early intervention.

In this thesis we examine the problem of apple detection and localization as an Object Detection problem, applied to the challenging real-world dataset MinneApple[6].

We use the Facebook Artificial Intelligence Research's (FAIR) Detectron2 framework to train a Faster R-CNN and a Mask R-CNN on the MinneApple dataset and compare results to other state of the art methods.

Detectron2 provides a pretrained 101 layer ResNeXt network that we use as the backbone for both the Faster R-CNN and Mask R-CNN models. We then fine-tune these models by adjusting the learning rate and number of iterations to avoid overfitting.

2 Results

Our first set of results compares the AP scores of each model with various ResNet backbones. To identify the most suitable backbone network. Table 1 shows a

* This thesis was prepared in partial fulfilment of the requirements for the Degree of Bachelor of Science in Data Science and Knowledge Engineering, Maastricht University. Supervisor(s): Alexia Briassouli.

comparison of the AP@[.5:.05:.95] scores with different ResNet backbones. We see that networks with a ResNeXt-101 backbone score the highest. This is expected, as larger Residual Networks tend to be able to extract more salient features better than smaller ones.

Table 1. Comparison of different ResNet backbone networks

Method	**Backbone**	**AP**
Faster R-CNN	ResNet-50	0.398
	ResNet-101	0.425
	ResNeXt-101	0.436
Mask R-CNN	ResNet-50	0.386
	ResNet-101	0.394
	ResNeXt-101	**0.441**

Our next set of results presents our AP scores for all 6 COCO scoring categories, with AP@[.5:.05:.95] being our primary challenge metric. Table 2 compares our models' evaluation results to the benchmarked results, as well as the current second best score in the CodaLab competition. The competition server does not give information on what methods the other entries have used but we have included them for comparison.

2.1 Discussion

From our experiments, we can see that our proposed method has had success in beating other benchmarked scores. The use of a more accurate model can lead to a more accurate yield estimation with practical benefits. If scaled, this could have a major effect on reducing food waste in the agricultural production sector by increasing yield. As mentioned previously, this sector has the largest food waste in the food supply chain.

As expected, the deeper ResNet backbones provided a higher score. With the 101 layer ResNeXt network performing the best for both Faster R-CNN and Mask R-CNN.

The benchmark results provided by the challenge state that Faster R-CNN is the best performer. However, an interesting insight is that our Mask R-CNN outperforms our Faster R-CNN. This could be due to the Mask R-CNN using semantic segmentation and may have learnt to deal with clusters of apples better than the Faster R-CNN which only uses bounding boxes.

3 Conclusions

We can see that Detectron2's Mask R-CNN with a ResNeXt101 backbone achieves state of the art accuracy on the MinneApple Fruit Detection challenge. At time of publishing our technique sits at the first place on the challenge leaderboards.[3]

Table 2. Comparison with Benchmark

Author	**Method**	**AP@ IoU[.50:.05:0.95]**	**AP@ IoU=.50**	**AP@ IoU=.75**	**AP***small*	**AP***medium*	**AP***large*
Hani[6]	Faster R-CNN	0.438	0.775	**0.455**	0.297	0.578	**0.871**
Hani[6]	Mask R-CNN	0.433	0.763	0.449	0.295	0.571	0.809
Kuka[4]	NA	0.436	0.770	0.453	0.285	0.592	0.872
Ours	Faster R-CNN	0.436	0.791	0.436	0.291	0.590	0.848
Ours	Mask R-CNN	**0.441**	**0.801**	0.440	**0.300**	0.589	0.861

References

1. Baeten, J., Donné, K., Boedrij, S., Beckers, W., Claesen, E.: Autonomous fruit picking machine: A robotic apple harvester. In: Field and service robotics. pp. 531–539. Springer (2008)
2. De-An, Z., Jidong, L., Wei, J., Ying, Z., Yu, C.: Design and control of an apple harvesting robot. Biosystems engineering **110**(2), 112–122 (2011)
3. Haeni, N.: MinneApple Fruit Detection Challenge (2020 (accessed August 5, 2020)), https://competitions.codalab.org/competitions/21718
4. Haeni, N.: MinneApple Fruit Detection Challenge Results (2020 (accessed August 5, 2020)), https://competitions.codalab.org/competitions/21718results
5. Häni, N., Roy, P., Isler, V.: A comparative study of fruit detection and counting methods for yield mapping in apple orchards. Journal of Field Robotics **37**(2), 263–282 (2020)
6. Häni, N., Roy, P., Isler, V.: Minneapple: A benchmark dataset for apple detection and segmentation. IEEE Robotics and Automation Letters **5**(2), 852–858 (2020)
7. He, L., Schupp, J.: Sensing and automation in pruning of apple trees: A review. Agronomy **8**(10), 211 (2018)
8. Wang, Q., Nuske, S., Bergerman, M., Singh, S.: Automated crop yield estimation for apple orchards. In: Experimental robotics. pp. 745–758. Springer (2013)