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

Introduction
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. Total global production of apples increased from 46.47 million MT in 1994 to 86.14 million MT tons in 2018 (FAOSTAT).

Use of technology in the agricultural food chain had led to greater crop yields from smaller amounts of land.

Detection and localization of fruits can improve yield estimates as well as various automated picking [4][5] and pruning[6] systems.

We train and compare Faster R-CNN and Mask R-CNN models with various ResNet backbones to compare to current state of the art results on the MinneApple dataset. A collection of images of apple trees taken from the University of Minnesota’s Horticultural Research Center (HRC).

MinneApple Dataset
Previous datasets used either circles or boxes as ground truth labels. MinneApple gives ground truth labels as polygons, allowing for semantic segmentation.

Advantages of MinneApple Dataset
- Highest numbers of annotations per image
- High res images of full apple trees
- Large variety of species
- Different levels of illumination
- Many different sections of the HRC’s orchards

Methods
- For both Faster R-CNN & Mask R-CNN methods we use various ResNet backbones
- Residual Networks allow us to train deeper networks without the vanishing gradient problem
- We used a 50 layer and a 101 layer ResNet backbone
- Also a 101 layer ResNeXt backbone
- These backbones were pretrained on the ImageNet classification tasks

Evaluation & CodaLab Competition
We trained models with different ResNet backbones to compare the results.

Results
- Ours against MinneApple paper
- CodaLab Leaderboard

Discussion
Methods
- Benchmark states that Faster R-CNN is the best performer
- Interesting insight is that Mask R-CNN outperforms Faster R-CNN (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)

Dataset
- MinneApple still a small sample of total species and different conditions around the world significant difference between apples grown in different continents, as well as different countries
- New varieties constantly being bred
- Future datasets need to include these variations to train more general models

Conclusion
Detectron2’s Mask R-CNN with a ResNeXt-101 backbone achieves state of the art accuracy on the MinneApple Dataset.

It is currently in first place in the challenge (as of 5/11/2020)

<table>
<thead>
<tr>
<th>Method</th>
<th>Backbone</th>
<th>AP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Faster R-CNN</td>
<td>ResNet-50</td>
<td>0.398</td>
</tr>
<tr>
<td></td>
<td>ResNet-101</td>
<td>0.425</td>
</tr>
<tr>
<td></td>
<td>ResNeXt-101</td>
<td>0.436</td>
</tr>
<tr>
<td>Mask R-CNN</td>
<td>ResNet-50</td>
<td>0.386</td>
</tr>
<tr>
<td></td>
<td>ResNet-101</td>
<td>0.392</td>
</tr>
<tr>
<td></td>
<td>ResNeXt-101</td>
<td>0.441</td>
</tr>
</tbody>
</table>