Individual Lab Report

Name:	Dung-Han Lee
Team Name:	Team E Wholesome Robotics
Team members:	Aaditya, Aman, John, Hillel
ILR Number:	05

October 1, 2019

1 individual progress overview

There were one primary focus area in the past 2 weeks, which is switching metric from area ratio to discrete severity level in evaluating Mask-RCNN's performance

1.1 Create a new ground truth for evaluating Mask-RCNN

1.2 Motivation

In order to evaluate how well the model is performing, there is a need for ground truth or basis to evaluate the performance. Since there is no standard, off the shelf data-set for our application, a custom test data-set must be built.

1.3 Create a reference scale from training data

Some training data is selected and fed into the current pipeline, which would generate fungus to leaves area ratio and holes to leaves area ratio. With these numbers, the selected training data can be sorted in ratios. As a result, a reference scale is created, and severity levels (mild, moderate, alarming) are assigned accordingly e.g. an image with hole to leaves area ratio ¿ 0.01 is labeled as alarming.

1.4 Sort test data

Some test data is manually selected from a different test data-set. The test data is sorted and labeled, by manually comparing test image to train image. Admittedly, This process is imperfect and subject to bias. (Fig 1)

2 Test Result

2.1 Quantify Evaluation

For a given test image, if the predicted severity is equal to ground truth, e.g. prediction is moderate and truth is moderate, then that would be counted as true positive i.e. true positive in moderate +1. Otherwise it would counted towards false positive and false negative respectively. e.g. prediction is mild, and truth is moderate, then mild false positive + 1 and moderate false negative +1.

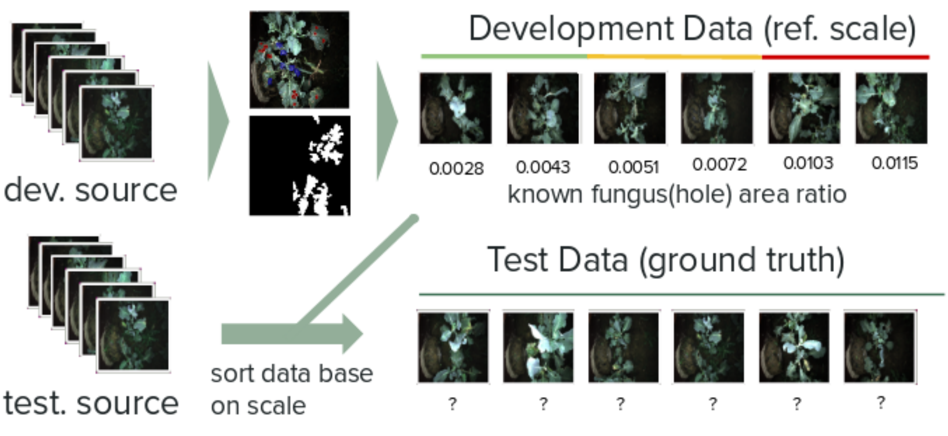

Figure 1: pipeline for labeling severity test data

3 Add extra 200 labels to handle the edge cases

3.1 Motivation

After evaluating the first test result, it appeared that the model failed to identify holes in low contrast areas, or holes surrounded with brown crown. This generally relates to insufficient training data i.e the model is not generalizing well to different variations. Therefore some efforts were put to increasing such variation by selecting and labeling such images.

3.2 Current Results

3.2.1 Holes

	mild:	moderate	alarming
true positive	19	4	2
false positive	2	3	0
false negative	3	2	0

3.2.2 Fungus

	mild:	moderate	alarming
true positive	14	5	2
false positive	2	4	3
false negative	4	4	1

3.3 Discussion of the results

Since there is a general bias distribution in severity, we are discussing the precision and recall for a category (hole/fungus) instead of severity (mild, moderate, alarming). It is observed that 66 - 75% precision and recall is obtained, which is the best result we achieved so far. However, it's still substantially lower than target 80%. After examining the data qualitatively, it appears that the model is still not working well with edge cases e.g. spurious fungus (dead leaf on the ground), spurious holes (intersection between branches and leaves)(Fig 2) .And such edge cases can sometimes increase or decrease severity by 1 level. In fact, half of the test result exhibit at least 1 error in either hole or fungus severity. It's worth mentioning though, since test data is labeled by human at the first place, some images may be mistakenly given wrong label.

Figure 2: hard cases

4 Challenges

The area ratio is still affected by spurious results from the deep neural network. A single false positive/negative can often increase or decrease severity level by 1 level. And increasing more labels in data does not seem to resolve this issue, at least not at the quantity that we are able to label given our time frame.

4.1 Future works

At this point, labeling more data does not seem to be a good approach. It's worth talking to our customer (farmers) about whether or not 80% is that important. Or even is severity is an important metric after all. Perhaps one way to work around this, is further simplify the problem into binary classification e.g. is there a fungus present in a given image. For the following week, we would be working toward using model outputs as features and try to use linear regression to predict the result.

5 Teamwork

In this week, I am working on severity predictions for the monitoring task. While John, Aaditya and Aman are working on in field navigation and localisation with Lidar, ZED and RTK GPS. All group members have visited Phipps for a simple field test. (Fig 3)Hillel has been working assembling the electronic subsystem for the platform.

Figure 3: in row navigation test