PHENOCYTE: A PLANT PHENOTYPING TOOL VIA MACHINE LEARNING

By

Maxim Popov and Bradyn Beals

A Creative Component submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Major: Artificial Intelligence

Program of Study Committee:

Dr. Simanta Mitra, Major Professor

The student author, whose presentation of the scholarship herein was approved by the program of study committee, is solely responsible for the content of this creative component. The Graduate College will ensure this creative component is globally accessible and will not permit alterations after a degree is conferred.

Iowa State University

Ames, Iowa

2025

Copyright © Maxim Popov, 2025. All rights reserved.

TABLE OF CONTENTS

Page

LIST OF FIGURES 3

ACKNOWLEDGMENTS 5

ABSTRACT 6

CHAPTER 1. OVERVIEW 7

CHAPTER 2. REVIEW OF LITERATURE 9

2.1 State of the Art Solutions 9

2.2 Segment Anything 10

2.3 Evaluation Metric 10

CHAPTER 3. METHODOLOGY 12

3.1 Segment Anything 12

3.1.1 Raw SAM 12

3.1.2 SAM Adapter 14

3.2 SAM Prompt 15

3.2.1 Image Input 17

3.2.2 Preprocessing 18

3.2.3 Mask Finding 20

3.2.4 Evaluation 21

3.3 UNet Model 22

3.3.1 Data Augmentation 22

CHAPTER 4. EXPERIMENT AND RESULTS 24

4.1 SAM Prompt 24

4.1.1 Case Studies 28

4.2 UNet Model 33

CHAPTER 5. CONCLUSION AND FUTURE WORK 39

BIBLIOGRAPHY 41

LIST OF FIGURES

Page

3.1 Sample SAM2 Segmentation 12

3.2 Sample Raw SAM Segmentation 13

3.3 SAM Adapter Architecture 14

3.4 Sample SAM Prompt 15

3.5 Sample SAM Prompt Arabidopsis 16

3.6 SAM Prompt Pipeline 16

3.7 Sample Original Image and Respective Mask 17

3.8 Sample Obfuscating Line Disruption 19

3.9 Sample Obfuscating Line Detection and Removal 19

3.10 Sample Mask Finding Process 20

4.1 Leaf Results 25

4.2 Seed Results 25

4.3 Root Results 26

4.4 Upper Plant Combination Results 27

4.5 Coloration Success 28

4.6 Color Filtration Success 29

4.7 SAM Success 30

4.8 Color Filtration Failure 30

4.9 Color Bound Failure 31

4.10 Obfuscating Line Included Failure 32

4.11 Obfuscating Line Excluded Failure 32

4.12 Sample Model Output 33

4.13 Model IoU 34

4.14 Background IoU 34

4.15 Leaf IoU 35

4.16 Root IoU 35

4.17 Seed IoU 36

4.18 Stem IoU 36

ACKNOWLEDGMENTS

IWe would like to thank Professor Simanta Mitra for his mentorship and feedback throughout this project as an outside observer to the inner workings of our development procedure.

We would like to thank Professor Wallapak Tavanapong and PHD Student Abdurahman Mohammed for their continued guidance on the inner workings of the project and their helpful ideas for methods to improve our various strategies.

We would like to thank Professor Dior Kelley for her regular feedback on our work as a domain expert and as our client for the problem we aim to solve.

ABSTRACT

Agriculture is vital to human society and the science of improving the yields of various plants is of utmost importance to our species. Plant phenotyping, or the evaluation of traits relevant to plant yield, is an incredibly arduous and inconsistent task when performed by human domain experts. State of the art solutions have been developed for a variety of plants, but no solution exists capable of segmenting all parts of Arabidopsis Thaliana, a model plant in agriculture. This work seeks to automatically segment all relevant parts: the leaves, stem, seed, and root, of an arbitrary Arabidopsis plant image in order to easily phenotype the plant given the resulting masks. With limited temporal and monetary resources at our disposal, this work explores the use of the Segment Anything [3] model in a variety of methods to perform the segmentation. Upon finding high boundary precision but poor contour generation using these techniques, we train our own relatively small UNet model as per state of the art standard for related works. Our preliminary results are promising with regards to generating domain-expert level masks, but further training is necessary and currently in progress to determine if such is the case. In order to achieve superhuman levels of performance, we recommend merging the precision of Segment Anything with the benefits of model training while keeping within tight cost constraints by training a SAM Adapter [1] on our dataset using a similar set of data augmentations.

CHAPTER 1. OVERVIEW

Human society derives great benefit from the cultivation of plants for their use in agriculture and energy. For many thousands of years, humans have carefully selected for higher yield specimens. Today, this work happens at a much faster and deliberate pace than in times past, researchers seeking greater yields, resiliency against disease and climate, and other important qualities by means such as varying growth conditions or genetically modifying the organism. When doing such research, one must evaluate a variety of techniques or samples against one another to discover which produce the sought after traits.

Phenotyping, this evaluation of phenotypes, or in layman terms the evaluation of plant traits, is a difficult, tedious, and time consuming process which, as a result, is prone to inconsistent evaluation. As such, machine-learning based techniques have recently been proposed to automate this process, greatly increasing both the pace of progress and the consistency of evaluation by measuring clear cut plant phenotypes.

However, no adequate solution has been developed for the Arabidopsis Thaliana, a model plant in agricultural research who's tiny size makes it even more difficult to evaluate. This work seeks to address this by segmenting images for sample Arabidopsis photos into their components: leaf, stem, seed, and root, so that it is subsequently trivial to determine phenotypes such as root length, leaf area, or leaf color, allowing researchers to focus on developing techniques to optimize outcomes rather than tediously measuring these values themselves, while simultaneously increasing the consistency of these measurements via this automation.

The primary contributions to this work are:

1. We explore the various possible usages of the Segment Anything model toward our ends.
2. We develop a SAM Prompt technique to automatically find point prompts for the Segment Anything model to complete the segmentation.
3. We observe and analyze the shortfalls of the SAM Prompt technique.
4. We train a UNet-based model with promising results.

CHAPTER 2. REVIEW OF LITERATURE

This section explores existing literature related to our work including state of the art solutions to similar problems and other state of the art tools that we employ in our solutions.

2.1 State of the Art Solutions

The use of machine learning for the segmentation and phenotyping of plants, while novel within the past few years, is not new to this work. A CNN-employing technique that pairs the use of RhizoVision [4] and RootPainter [5] segments roots at an human expert level. However it is solely capable of processing herbaceous plant species, and does not work for the Arabidopsis family, nor is it capable of finding masks for other non-root plant parts.

Dobos et al. develop a “Deep Learning-Based Approach for High-Throughput Hypocotyl Phenotyping” [2], employing a U-Net architecture for segmenting plant images into background, stem, and non-stem segments. While domain experts are satisfied with the use of this tool for segmenting the stem, it does not solve the need for a unified tool capable of segmenting all Arabidopsis parts.

RootNav 2.0 [7], another CNN-based segmentation solution, is capable of segmenting the root for any plant, including Arabidopsis, however, like previously listed techniques, it does not address the other plant parts.

The last technique we found of note was RoAD: Robotic Assay for Drought [6], which employs a complex robotical system to highly accurately phenotype arbitrary

plants. However it is incapable of finding anything below the surface of the earth, such as the root, and is prohibitively expensive for many researchers.

2.2 Segment Anything

When examining the state of the art solutions described above, we notice that the training of CNN machine learning models is the go-to core for the typical solution. However, citing time and cost constraints, we were tasked by our client and advisors to avoid training our own model and instead find a different way to solve the problem.

Given our constraints, we explore the use of SAM2 [3], a Segment Anything model capable, as per the name, of segmenting arbitrary images either in full or by finding a single mask encompassing a set of given points on the image.

We proceed to describe the various implementations of SAM-based techniques later in this work in sections 3.1 and 3.2.

2.3 Evaluation Metric

When evaluating our automated Arabidopsis-part mask finding technique against expert-provided ground truth samples for given raw images, we seek to minimize the difference between the real ground truth mask and our computed mask for each plant part.

More formally, we seek to maximize the value IoU = $\frac{I}{U}$ where I is the intersection of and U the union of the real image mask R and the automatically computed one C of a given Arabidopsis plant part.

We are informed by domain experts that we should seek a 95% IoU for each plant part to be sufficiently powerful for domain use.

CHAPTER 3. METHODOLOGY

We propose four techniques for the segmentation of the Arabidopsis plant family into its components: leaf, stem, seed, and root. The first three techniques described in sections 3.1 and 3.2 employ the Segment Anything [3] model in various capacities. The final technique described in section 3.3 instead trains our own model without using SAM.

3.1 Segment Anything

3.1.1 Raw SAM

We notice that SAM2 is capable of segmenting arbitrary images with a high degree of perceived accuracy without any tailored prompting necessary. One can visually verify this by examining Figure 3.1 provided in [3] which shows the segmentation of a city block.

Figure 3.1 Sample SAM2 Segmentation

We proceed to use this Raw SAM technique on sample Arabidopsis images with promising initial results. As such, we conduct a search through the SAM hyperparameter space to maximize our results.

We vary the confidence thresholds necessary for a segmentation to be accepted to seek the necessary values to discover our plant segments with minimal background interference. We vary the overlap allowance between segmentation bounding boxes to find segmentations for parts that are physically close to one another, such as the seed and leaf. We vary the allowed size of a segmentation bounding box to better discover larger parts such as the root. Lastly, we vary the zoomed precision of the final segmentation to enhance the precise accuracy of our segmentation.

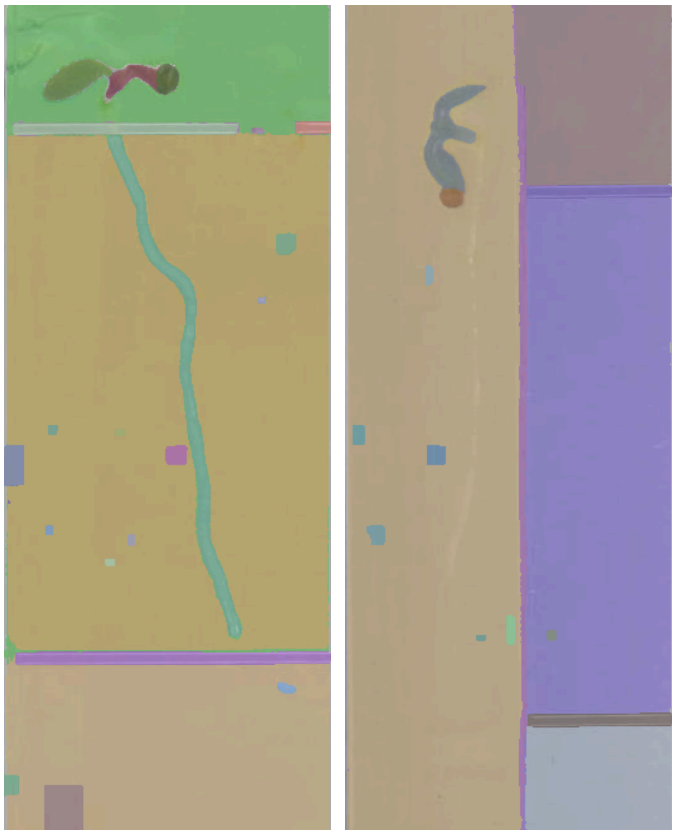

Figure 3.2 Sample Raw SAM Segmentation

While we can visually notice the high degree of precision in segmentation in Figure 3.2, this technique proves insufficiently powerful to move forward with. Even in some of the most successful segmentations which are shown in the figure, our hyperparameter search did not prove to find meaningful improvements, where our segmentations nearly always failed to discover some plant part or another. In the case of the given figure, the left sample failed to discover the stem while splitting the leaf into two objects and the right sample failed to discover the root while merging the stem and the leaf into one object.

3.1.2 SAM Adapter

We explore the possibility of training an adapter model on SAM2 as suggested by previous state of the art success in medical fields [1]. Such a technique could specialize the model for our field to achieve the level of results we see in Figure 3.1 from their work.

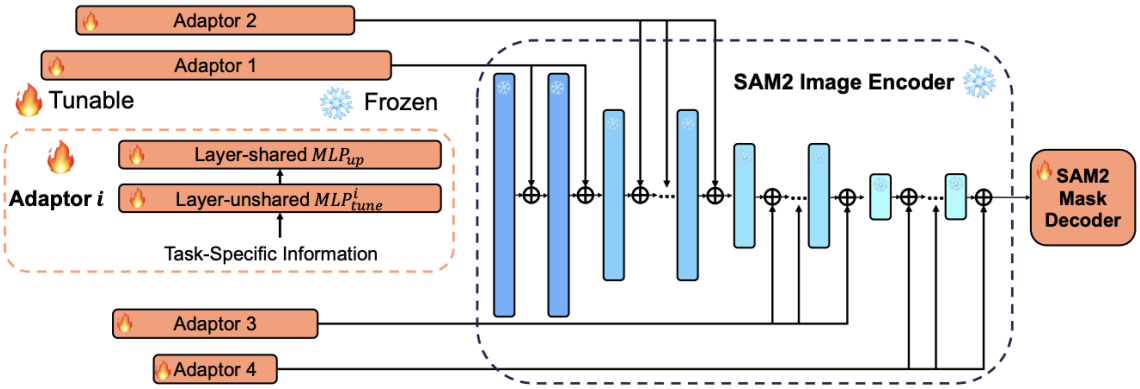

Figure 3.3 SAM Adapter Architecture

The use of an adapter allows for a greatly lessened number of parameters to train, decreasing the time and monetary costs cited for the reason to impose the initial no model training constraint on our work. The resources used in the original training of the technique [1] are available at our disposal in Iowa State's High Performance Computing cluster. However, our clients still deemed the time required to train the model too steep a cost and the technique was shelved.

We will return to discussing the SAM Adapter proposal in our conclusion after discussing the benefits and drawbacks of the forthcoming SAM Prompt technique and UNet Model in Chapter 4.

3.2 SAM Prompt

SAM also provides the ability to prompt the model with positive points and anti-prompt it with negative points as per Figure 3.4 from their work [3]. When employing this technique on our sample Arabidopsis images, we see promising results as depicted in Figure 3.5.

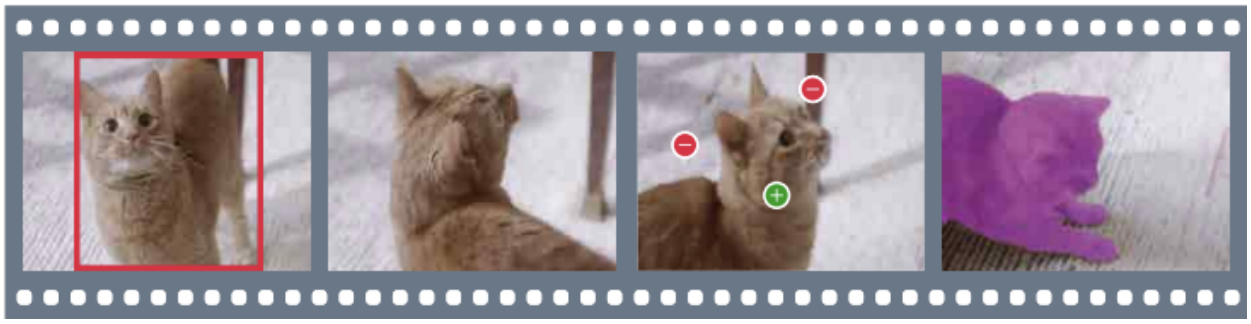

Figure 3.4 Sample SAM Prompt

Figure 3.5 Sample SAM Prompt Arabidopsis

Given this promising result in Figure 3.5 that detects the leaf, stem, and seed with high precision, we believe that with adequate prompting we can detect the plant itself and with further manipulation we can break that segmentation into its various plant parts.

We propose the pipeline displayed in Figure 3.6 to use the most striking aspect of the plant image to human eyes, the color, to both find good prompts for SAM and to manipulate the result to seek just the individual plant part in question.

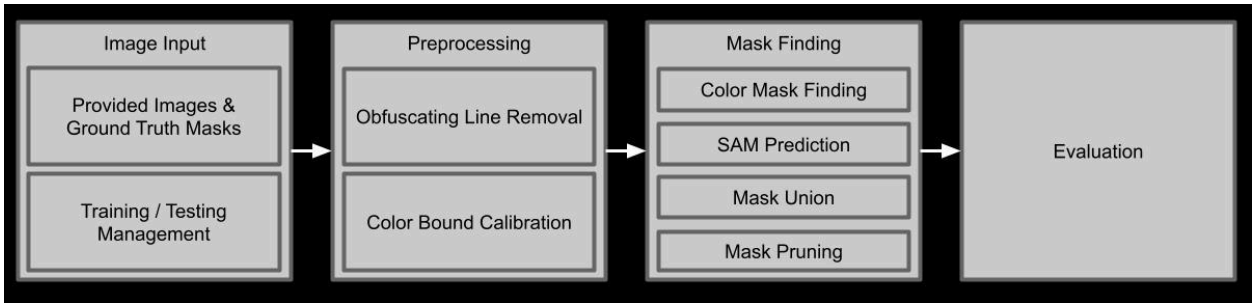

Figure 3.6 SAM Prompt Pipeline

3.2.1 Image Input

In the Image Input stage, we receive our original Arabidopsis plant image files as well as respective domain expert provided ground truth masks. In a real use case, we expect a small sample of ground truth masks provided alongside a large sample of original images for calibration. When training our color calibration, which is described in section 3.2.2, we use all of the given raw and ground truth image pairs to calibrate.

Given that the purpose of this work is to provide segmented masks, for the images where the ground truth masks are already provided, the masks are already there, so there is no loss to using those images for calibration. There is some loss in requiring users to provide masks for some of their images, however we save the grand majority of the expense by doing the remainder of the work.

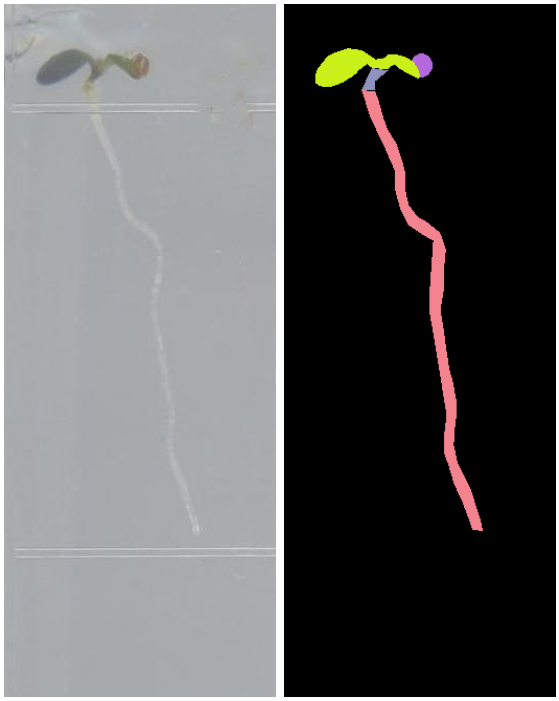

Figure 3.7 Sample Original Image and Respective Mask

3.2.2 Preprocessing

We notice that in our domain expert provided batch of images and ground truth masks, the plant coloration in the images appears consistent between images, such that the leaf always appears greener, the seed browner, the stem in between in terms of coloration, and the root white, all against a grey background with some artifacts present.

We are informed that this coloration may vary between different batches of images depending on the techniques used by the agricultural scientists, however the coloration within a batch is expected to remain similarly consistent.

Therefore, while we cannot set global color bounds that will always be consistent, we can calibrate color bounds within a batch of images given that a small subset of those images have ground truth masks provided.

We seek to find the color bounds in HSV (hue, saturation, value) format for greater contrast for the benefit of SAM. Given that the ground truth samples are not perfect pixel-by-pixel given natural human imperfection and inconsistency, we remove statistical outliers when we process the minimum and maximum HSV values, where hue, saturation, and value are all computed independently. As such, we find only the imperfectly extreme bounds within the plant part itself, not including adjacent plant parts or the background coloration.

The other component of our preprocessing consists of the removal of obfuscating vertical and horizontal lines on the glass, which serve as highly disruptive artifacts for SAM upon prompting, as demonstrated in Figure 3.8.

Figure 3.8 Sample Obfuscating Line Disruption

Given that the obfuscating lines are always horizontal or vertical, we detect long near-horizontal or near-vertical contours. We proceed to attempt to remove them by a mix of averaging nearby non-obfuscating-line pixel values and by adding Gaussian noise to avoid it still appearing as an object which confuses SAM. As per Figure 3.9, this appears somewhat but not fully successful, where a human and SAM both still see the obfuscating line and in the case of the latter, still includes it in the mask as per Figure 3.8.

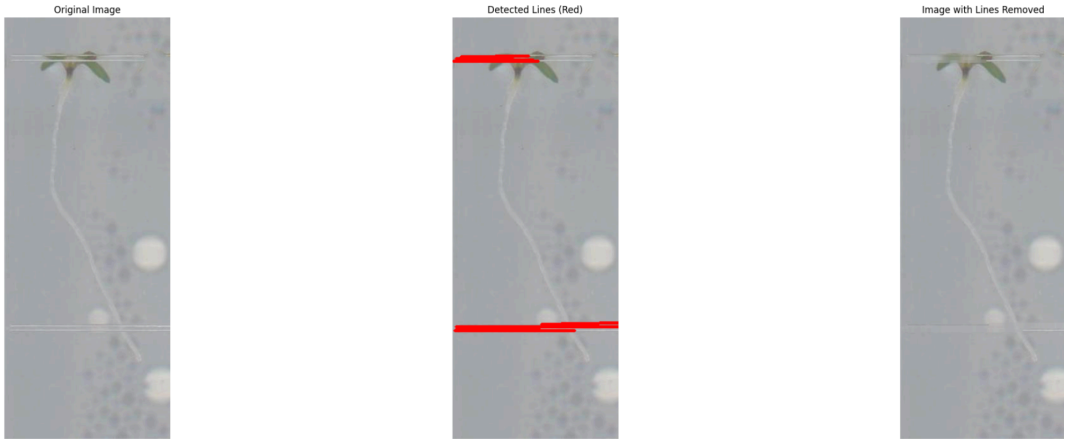

Figure 3.9 Sample Obfuscating Line Detection and Removal

We develop two other techniques for removing the obfuscating lines. The first continues along a similar method with an out-to-in boundary method rather than the preliminary top-to-bottom method used in Figure 3.9 above. The second seeks to find non-plant points in the obfuscating line to offer as negative anti-prompts to SAM. Neither of these methods are implemented due to color calibration failures described in Section 4.1.

3.2.3 Mask Finding

Our Mask Finding stage runs as a 7-part process as shown in Figure 3.10. During our color calibration stage, we found maximum and minimum HSV color bounds for each of our plant parts. We begin mask finding by finding all pixels in the image that are within our color bounds to compose our color-based mask.

We then find the largest single connected contour in our color-based mask and filter out all contours that are more than a threshold distance away from our largest contour. We end up with our filtered color mask.

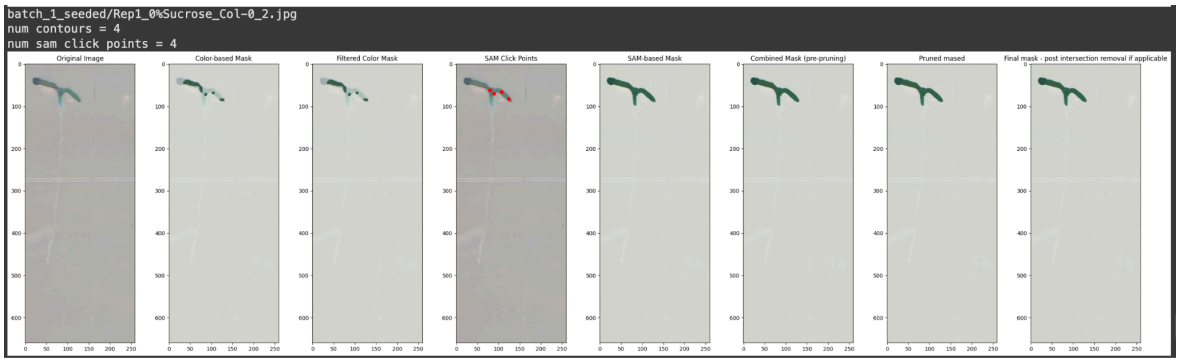

Figure 3.10 Sample Mask Finding Process

We proceed by taking the largest x% of our filtered color mask contours up to a maximum of y contours. We then find a point in each such contour such that the point is most deeply embedded within the contour, or equivalently furthest from a non-contour point, and collect that point to prompt SAM with.

For the seed, for the stem, and for the root $x = 100$ and $y = 1$, for the leaves $x = 50$ and $y = 5$. The reason for this is that when finding our color mask, our seed, stem, and root are generally consistent with SAM, while for our leaf we have more contours and this is a workaround to the line obfuscation problem.

We proceed to submit all found points to prompt SAM to generate the SAM-based mask. We then get the combined mask by finding the union of the SAM-based mask and the filtered color mask. This is because the SAM-based mask sometimes misses areas so the filtered color mask can prop up some of the missed area.

We then get a pruned mask by taking the single largest contour from the combined mask since all plant parts should be one single part. Lastly, we optionally generate a final mask by removing any masks for other plant parts that we have a higher success rate with. For instance, if the seed mask is highly accurate, as it was in our preliminary tests, while the leaf mask tends to find both the leaf and the seed in it, we can remove the seed mask from the leaf mask we find to increase accuracy.

3.2.4 Evaluation

We lastly evaluate our result by finding the IoU and related IoU statistics for all plant parts for all testing images. More specifically, we find the best, worst, Q1, Q2 (median), Q3, mean, and standard deviation IoUs for each plant part across all images.

3.3 UNet Model

Our last proposed solution is to train a small UNet machine learning model with a resnet34 encoder, to best match the typical state of the art solution for similar problems as was discussed earlier in this work. The scale of the model is the maximum scale we can train given free resources available via Google Colab. We choose UNet to best match related works such as [2] and [4].

We use Dice Loss with respect to the overall image IoU, where each pixel is treated with equal weight and the background being a legal classification in addition to our four plant part classes. We use an ADAM Optimizer with a default learning rate of 2e-4.

3.3.1 Data Augmentation

In order to maximize the consistency of our model, both within our provided image batch, and for arbitrary batches we may run on, we augment our data in several ways throughout training and validation.

We perform a horizontal flip with 50% likelihood given that a plant may grow to either side. We do not perform a vertical flip because plants are not expected to grow upside down.

We always perform a small scale, rotation, and shift. Our plants are expected to grow in different sizes and in slightly different directions, so this universalizes our training data in this respect.

We perform a random crop, constrained such that our entire plant remains in the frame, so that we don't rely on where in the provided images our plants are. We assume

that all images will contain the entirety of their plant as otherwise phenotyping that image will not be accurate.

We inject a small amount of gaussian noise with 10% likelihood with a standard deviation between 1% and 5% to allow for varying image quality.

We shift the perspective of the image by a small scale with a likelihood of 30% to allow for shifts in camera position, given that in our data our camera is always positioned in a given way that may not be consistent with other research. We likewise perform a slight elastic transform with 20% probability for the same reason.

Lastly, in order to protect against varying camera and more importantly color variations, we perform brightness contrast, sharpen, blur, and HSV color variation each with 90% likelihood.

Our expectation is that some of these augmentations may slightly decrease the quality of our work in our test batch, however they are expected to greatly improve them on other possible batches.

CHAPTER 4. EXPERIMENT AND RESULTS

In this chapter we evaluate and analyze the results of the SAM Prompt and UNet Model techniques described in sections 3.2 and 3.3 respectively. For both of these, we use our batch of 367 expert-provided Arabidopsis images and respective ground truth masks.

It is worth noting that the masks provided are imperfect, as it is incredibly tedious to draw them at the scale of number of images required to train a model. We believe that they themselves may sometimes fail to fall into the sought after 95% real IoU.

4.1 SAM Prompt

For our SAM Prompt technique evaluation, we remove the 2/367 images whose ground truth masks do not contain a seed, as is sometimes possible. We do this to keep evaluation consistent between all plant parts.

We proceed to split the images in a 20/80 train/test split. Note that this is a very small training set compared to the usual. This is because rather than having a complex training regime as per the typical machine learning model, we are only using this set for color calibration.

For our hyperparameters, we use a distance threshold of 100 pixels for color mask filtering described in Section 3.2.3, and we exclude the outer 90% outliers in our color calibration. This extreme value is justified shortly.

We first evaluate the leaf, seed, and root. We do not evaluate the stem due to time constraints and proceed to the implementation and evaluation of the UNet model upon poor results for the leaf, seed, and root shown in figures 4.1, 4.2, and 4.3 respectively.

Leaf Results:
Mean IoU: 0.03%
Standard Deviation: 0.16%
Best IoU: 2.08%
Worst IoU: 0%
25th percentile (Q1): 0%
Median (Q2): 0%
75th percentile (Q3): 0%

Figure 4.1 Leaf Results

Seed Results:
Mean IoU: 32.48%
Standard Deviation: 26.11%
Best IoU: 86.79%
Worst IoU: 0%
25th percentile (Q1): 7.11%
Median (Q2): 30.91%
75th percentile (Q3): 56.20%

Figure 4.2 Seed Results

Root Results:

Mean IoU: 0.36%

Standard Deviation: 0.78%

Best IoU: 4.59%

Worst IoU: 0%

25th percentile (Q1): 0%

Median (Q2): 0%

75th percentile (Q3): 0.26%

Figure 4.3 Root Results

We find incredibly poor results for the leaf and root, and simply poor results for the seed. Even in the best cases for the seed, we have a sub-90% IoU which was deemed poor by domain expertise.

The reason for this failure is clear in our following case studies. The color calibration is infeasible. Originally, we excluded the outer 10% of statistical outliers in calibration and, for both the leaf and seed, we would discover both the other and numerous background pixels. As such, we took a more extreme removal of 90% of outlier HSV values in calibration. Past this point, we would begin losing most of the sought after plant part pixels, while still finding some pixels in other plant parts and in the background. As such, we deemed color calibration to be infeasible for our batch of images and justify the results as such.

In our preliminary testing, we worked with a much smaller batch of images we were initially provided. In that batch, the coloration of the images was different, and we

performed better. However, given the expectation that in real usage we will sometimes receive batches such as our large batch we tested on, this technique is not a feasible solution.

We noticed that in many cases when seeking the seed we would find the entirety of the seed, and leaf, and stem. As a potentially useful mask to find to work off of in the future, we sought to evaluate how well we could find this leaf/stem/seed upper plant combination. If the results were very strong, we could work off of that.

In order to evaluate this combination, we combined the mask of our leaf, seed, and stem into a single mask to evaluate against, and ran the seed path against that for the results shown in Figure 4.4.

Combo Results:
Mean IoU: 27.11%
Standard Deviation: 31.61%
Best IoU: 88.34%
Worst IoU: 0%
25th percentile (Q1): 2.65%
Median (Q2): 10.18%
75th percentile (Q3): 53.68%

Figure 4.4 Upper Plant Combination Results

In our upper plant combination, we find similar poor but not abysmal results as we do for the seed. It appears that we find an unhappy medium between the seed and the upper plant combination and don't perform very well on either.

It is worth recalling that SAM provides much higher precision masks than the expert given masks given the impossibly tedious nature of providing even a single maximum precision mask. As such, our results are expected to be somewhat better than they are shown here, but that improvement is not expected to exceed 5-10% IoU, and as such the results are still very poor.

4.1.1 Case Studies

We explore several case studies on the successes and failures of our SAM Prompt solution. All of these case studies are on the upper plant combination technique which was easiest to understand given that the seed, as the only other reasonable performer, is much harder to view due to its smaller size.

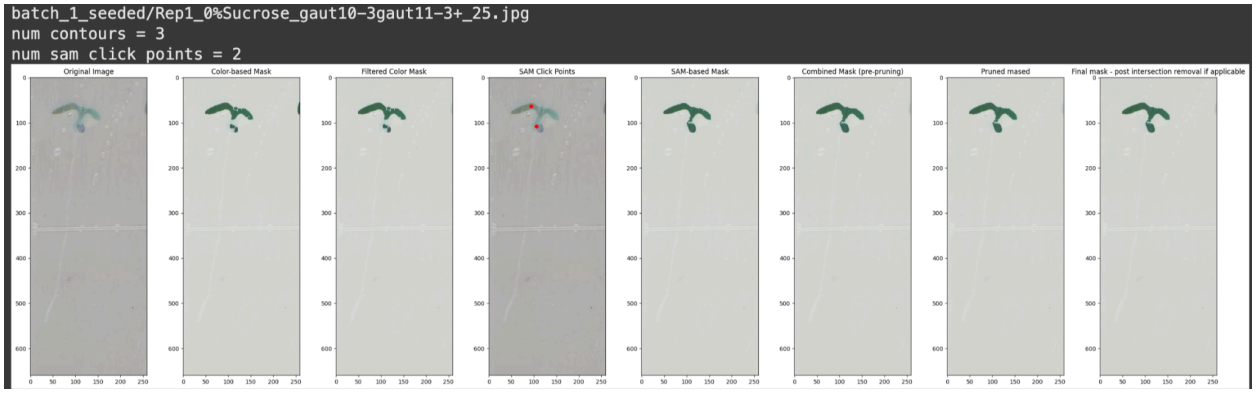

Figure 4.5 Coloration Success

Figure 4.5 shows a success story for the coloration bound story. The only pixels found within the color bounds are correctly part of the plant, although we do notice the edge of another plant appearing in this image on the right hand side that is also found. That contour is dealt with during the color mask filtration process. SAM successfully continues to find what appears to be a superset of the color mask, proving the combined mask to not be important in this case, but the original success of the color mask was vital for getting good SAM prompt points.

Figure 4.6 displays a strong success for the color mask filtration process. The original color mask found many background pixels accepting it, as is all too common in our method for the reasons discussed earlier, but the filtration process correctly removed all of those without removing the small color acceptance found on the edge of the right leaf. SAM proceeded to perform well like in the previous case and our resulting mask appears largely successful.

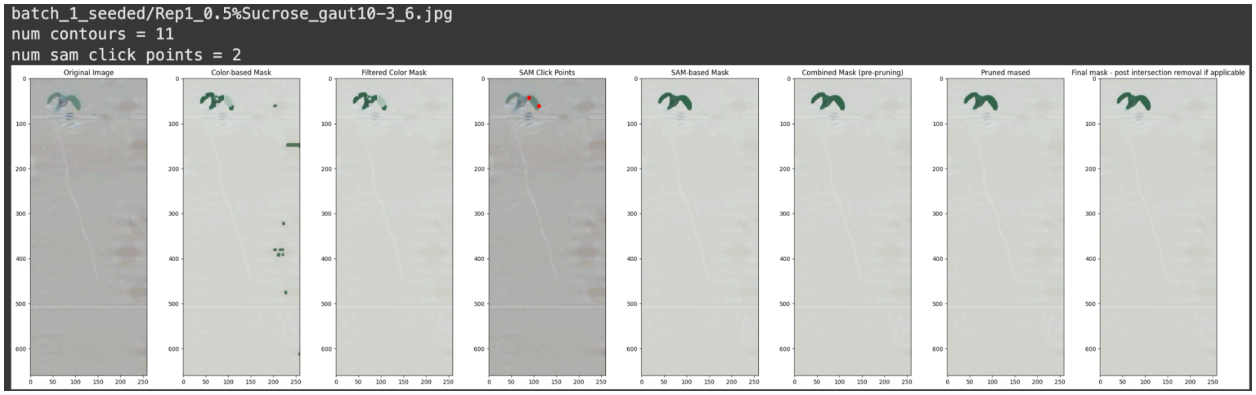

Figure 4.6 Color Filtration Success

In Figure 4.7 we see some trouble with the color-based mask missing much of the upper plant region, but we still find enough to give SAM prompting points and SAM manages to find the overall region successfully, even despite the horizontal obfuscating line in a rare care of not getting confused by it.

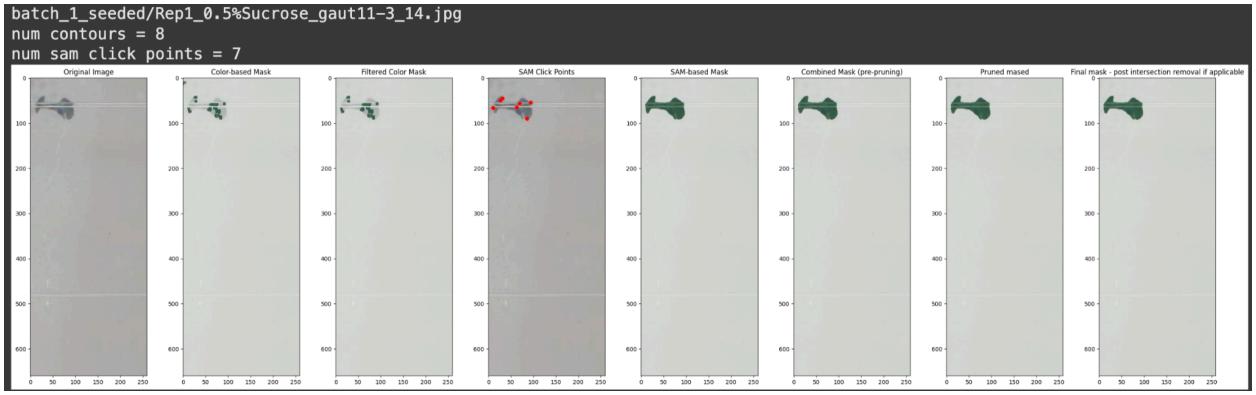

Figure 4.7 SAM Success

Figure 4.8 shows a case of failure for the color filtration process. While it successfully removes background pixels discovered within the color bounds, it also removes parts in the right leaf, which in combination with the obfuscating line results in a failure of discovering the entirety of that leaf.

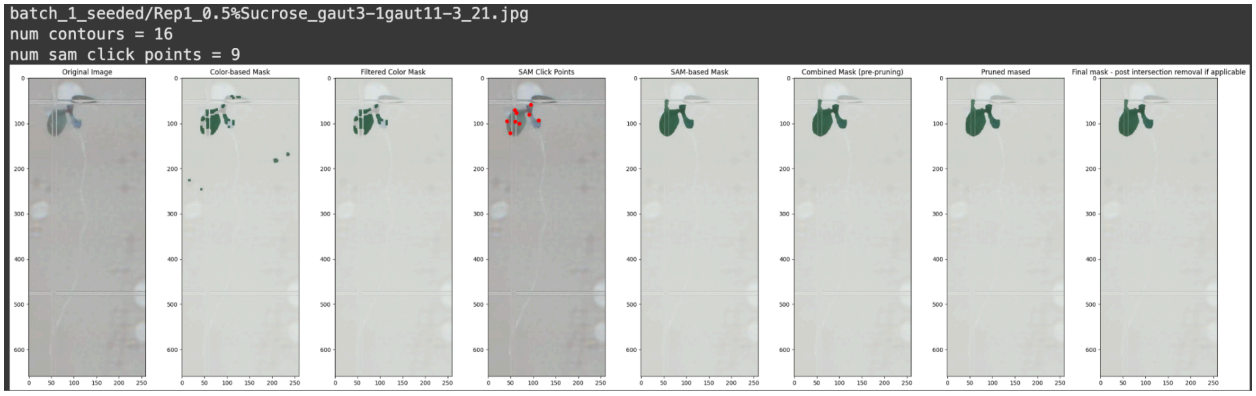

Figure 4.8 Color Filtration Failure

Figure 4.9 displays our most egregious sort of failure, one that is unfortunately not uncommon and the source of all of our 0% IoU cases, which is approximately the bottom quarter of our results. We find so many background pixels within our bounds that the filtration cannot aid us and we result with terrible SAM prompt points and a 0% IoU result.

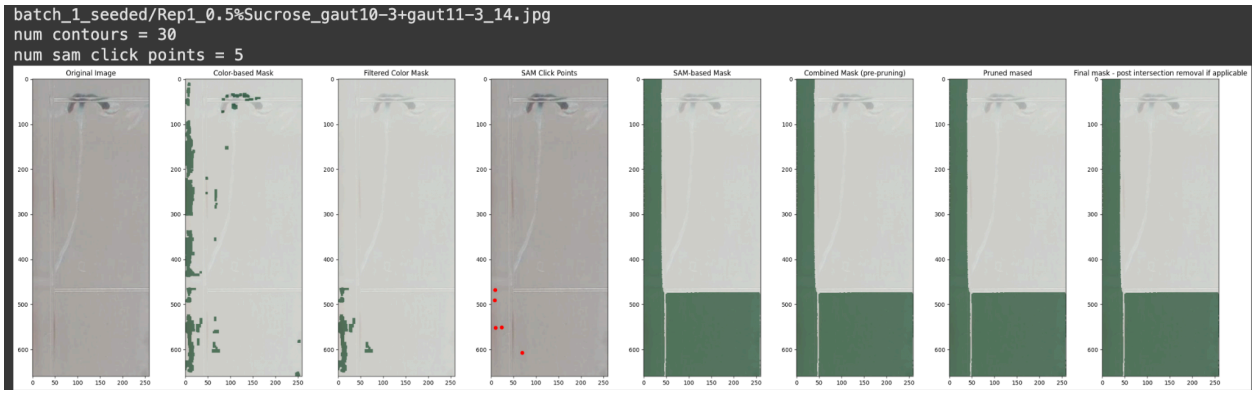

Figure 4.9 Color Bound Failure

Lastly, figures 4.10 and 4.11 display the two sorts of obfuscating line failures in otherwise successful processes. In the former, one of our SAM prompt points falls in the obfuscating line area, and as a result we prompt that area and the SAM mask includes it. In the latter, the points all fall outside of the area and SAM, considering the line as an object, does not include it.

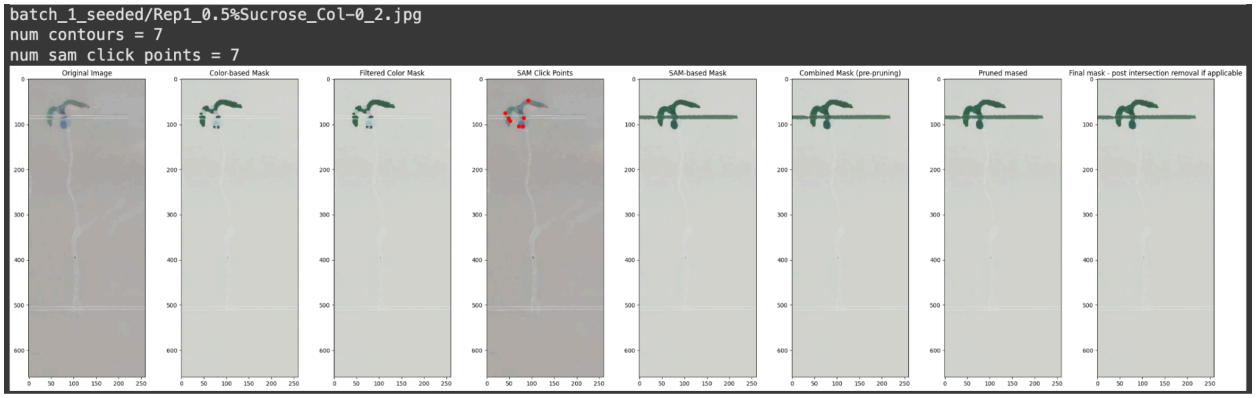

Figure 4.10 Obfuscating Line Included Failure

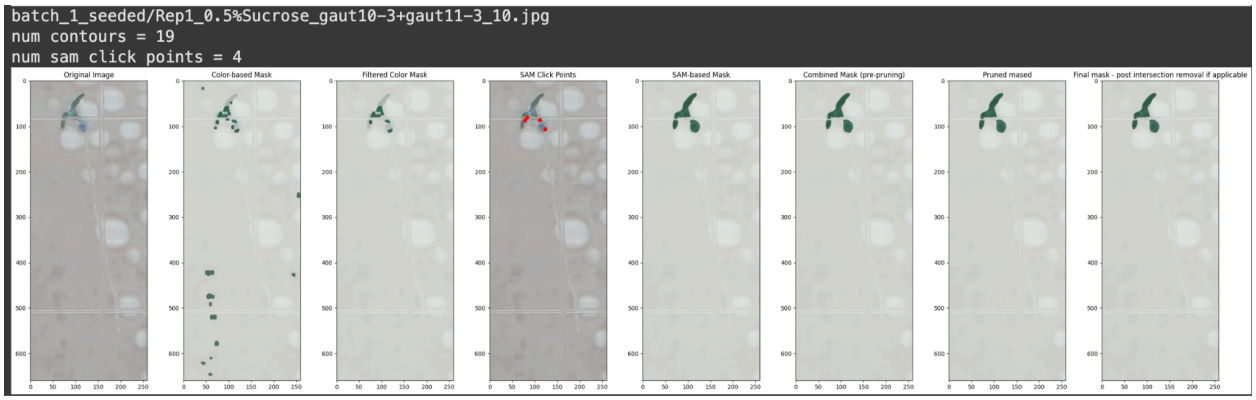

Figure 4.11 Obfuscating Line Excluded Failure

Throughout these case studies, we find that when we are successful we appear to do reasonably well, better in fact than the provided IoU's lead us to expect. Our belief is that a non insubstantial part of that difference is due to the ground truth itself being inaccurate for small objects, given the tediousness of providing ever more optimal masks, as we see very clean boundaries as a result of the SAM precision process, something far beyond what human experts provide in terms of accuracy on the fine boundary points.

Still, as per the infeasibility of consistent coloration leading all too often to cases such as the one in Figure 4.9, this SAM Prompt technique is not nearly consistent enough to be promising as a solution to our problem.

4.2 UNet Model

We run our UNet model for 50 epochs, with 8 steps per epoch, where the value was set as an estimate for the number of epochs we could get through before running out of free Google Colab T4 GPU resources.

In Figure 4.12 we see a random sample output from our model where, despite significant artifacts on the glass obfuscating the original image, our model is very near to the ground truth, much closer than any previous result.

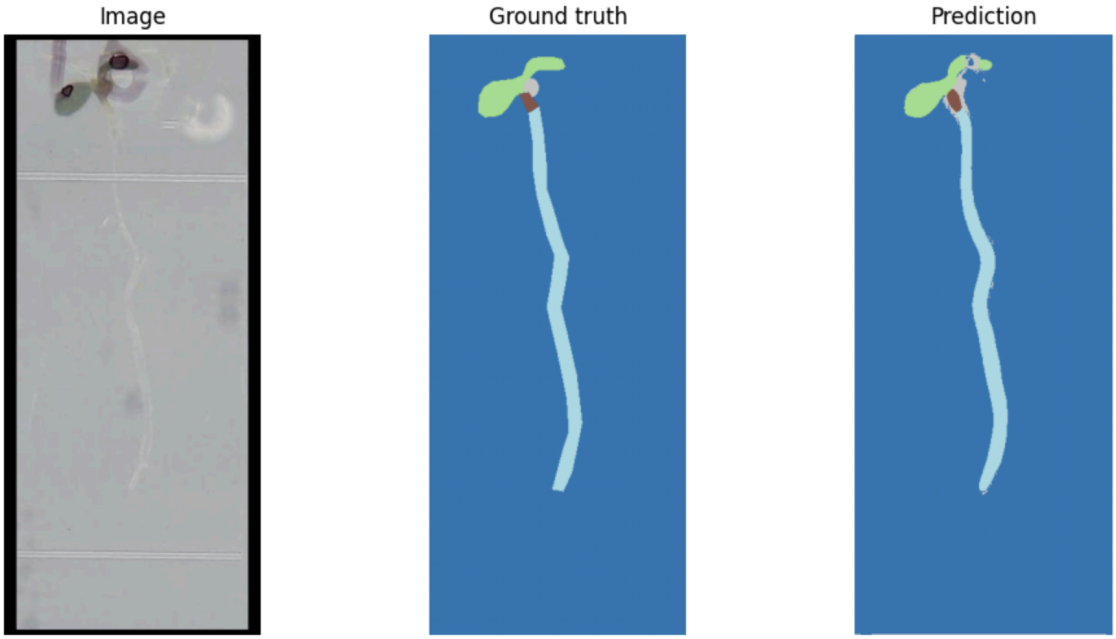

Figure 4.12 Sample Model Output

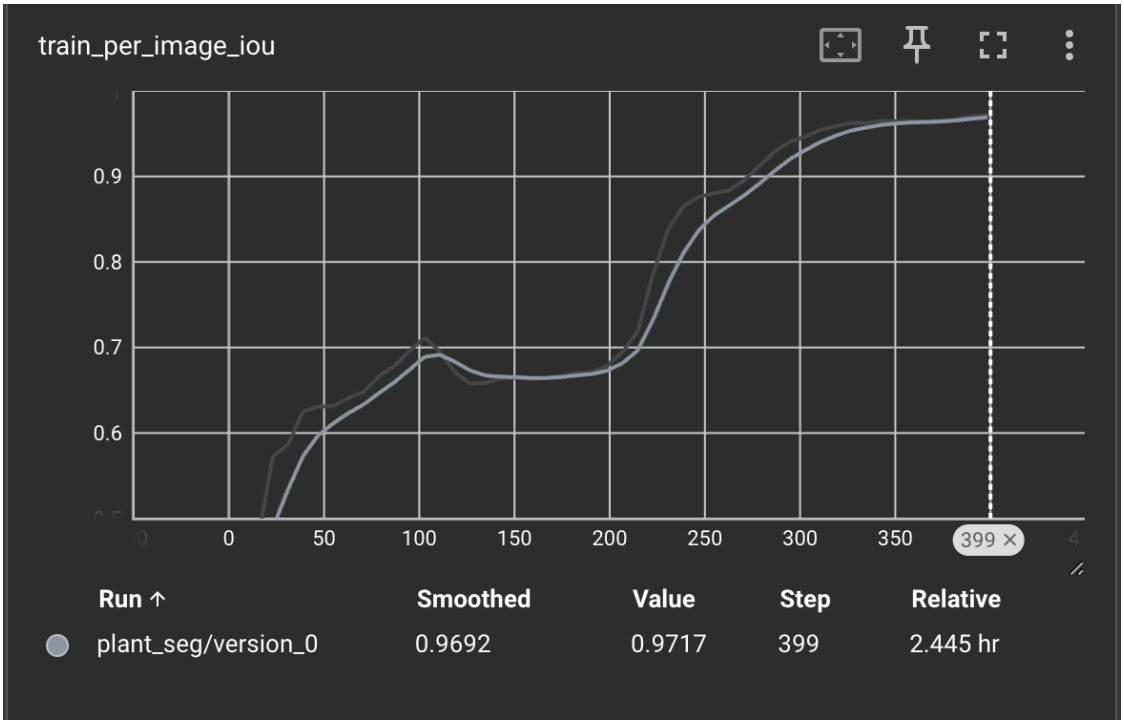

Figure 4.13 Model IoU

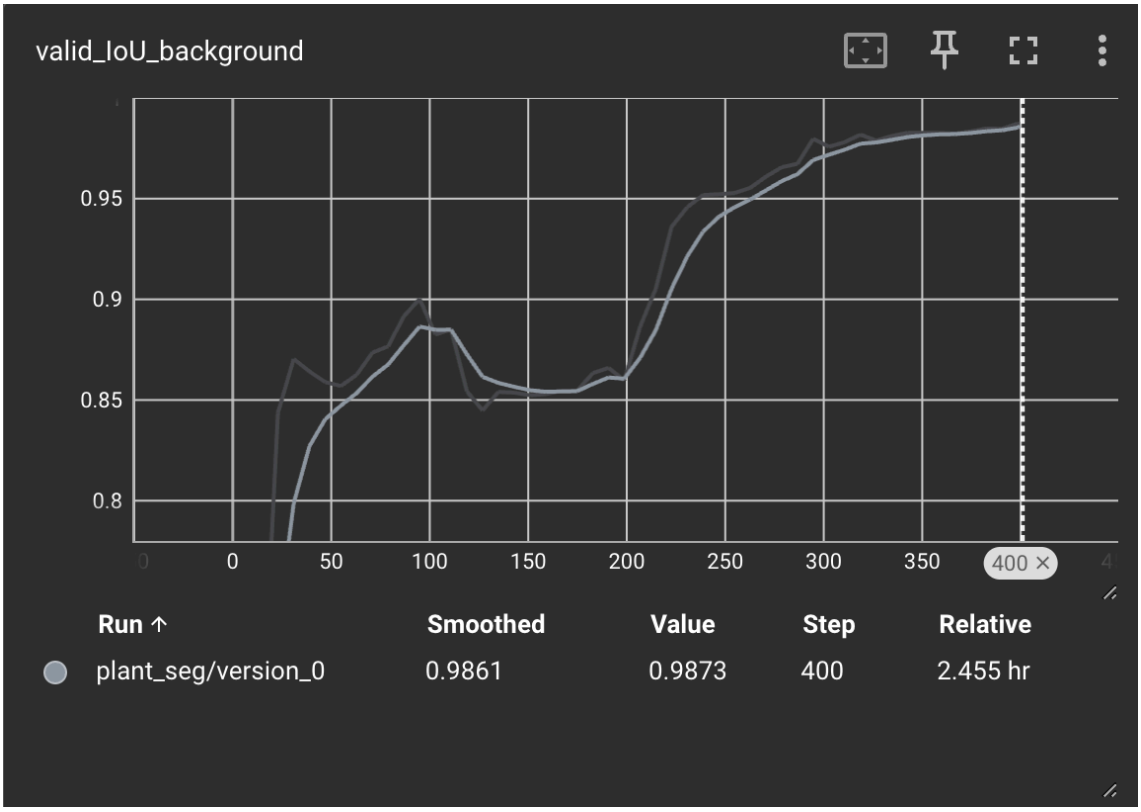

Figure 4.14 Background IoU

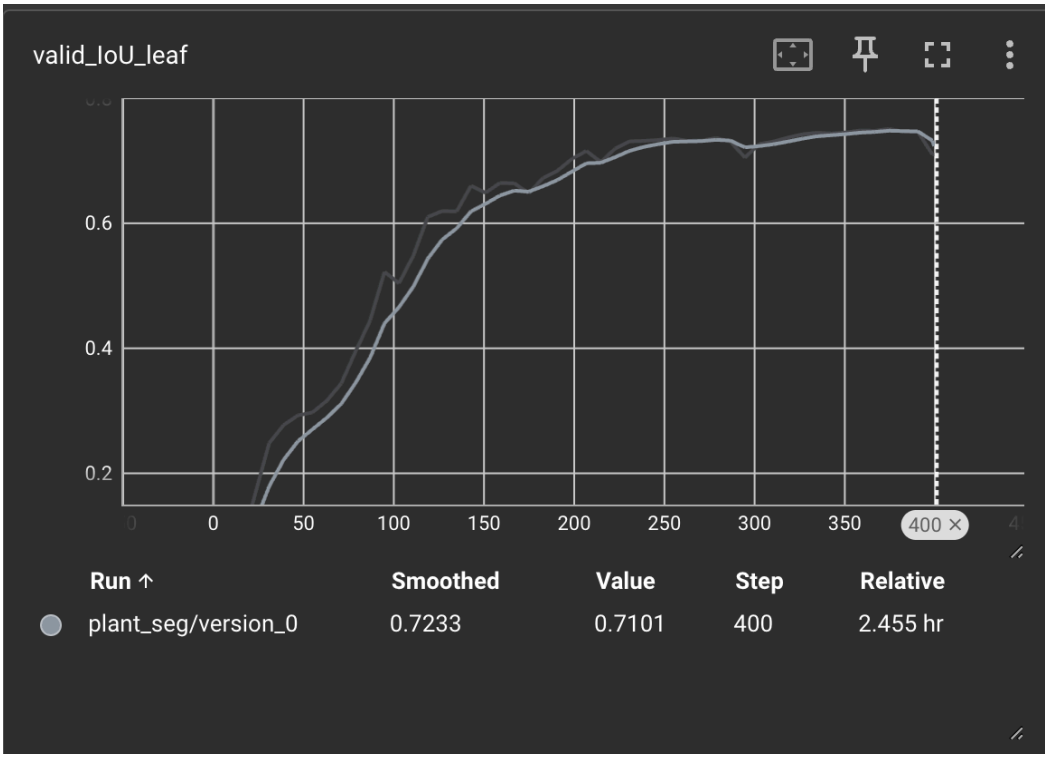

Figure 4.15 Leaf IoU

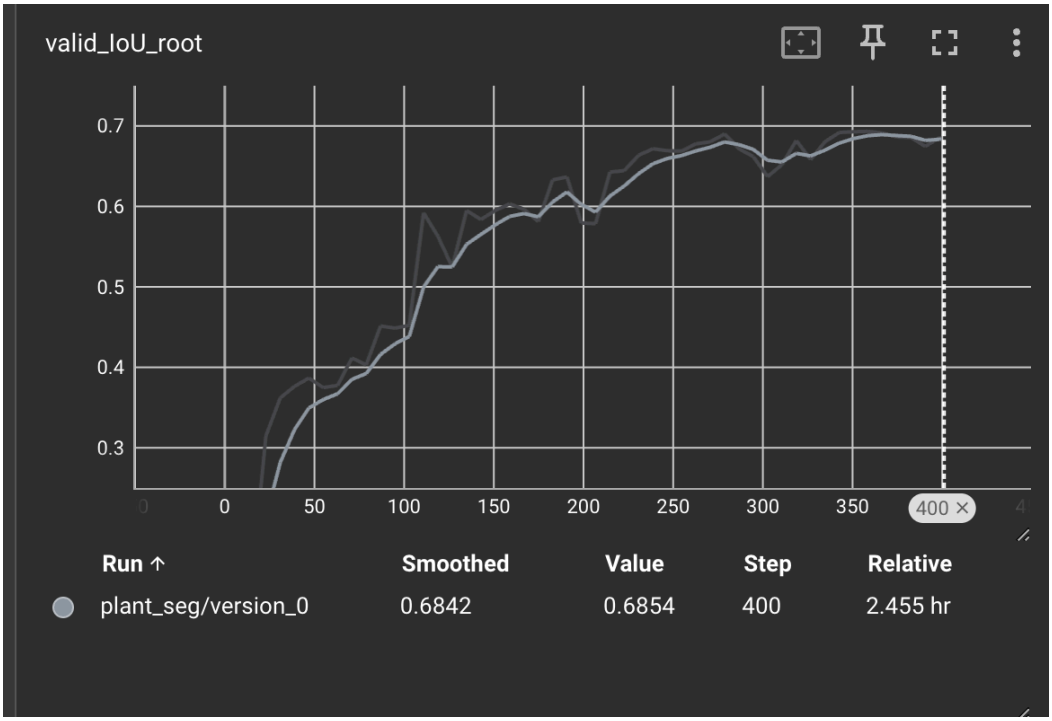

Figure 4.16 Root IoU

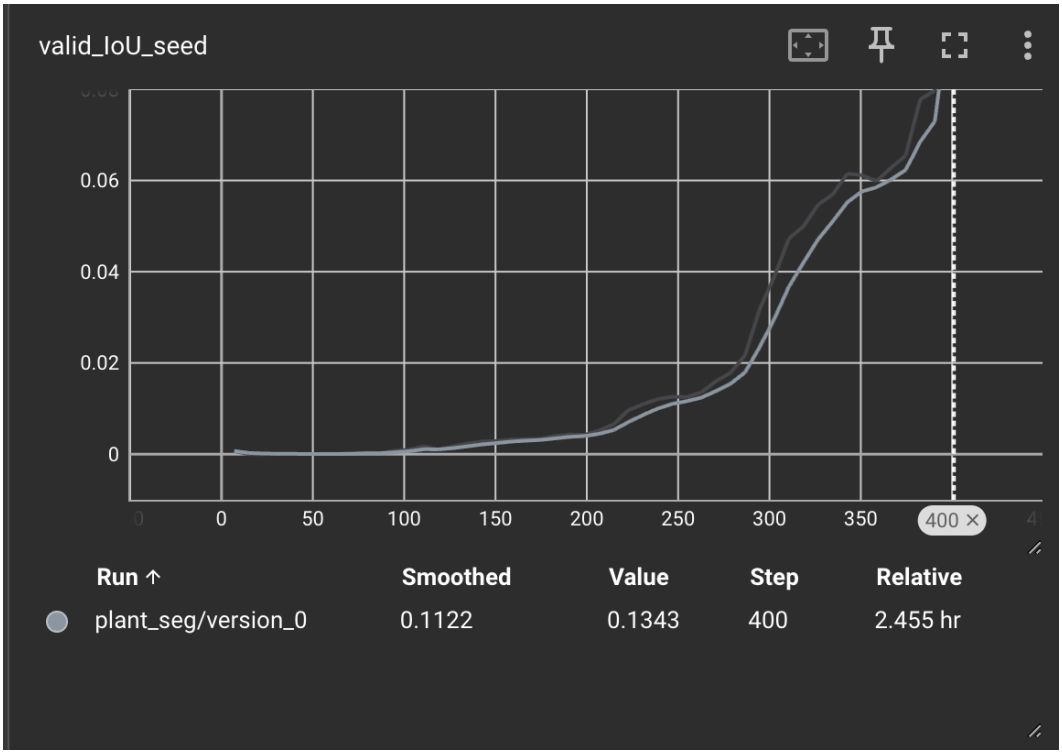

Figure 4.17 Seed IoU

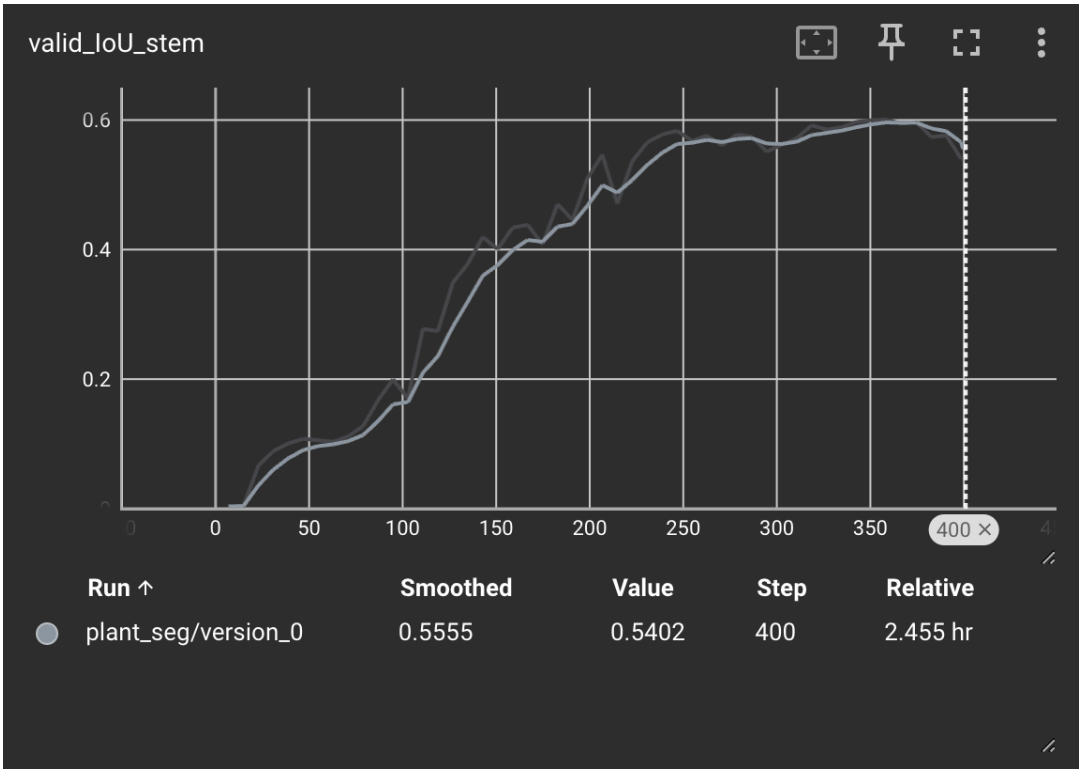

Figure 4.18 Stem IoU

Note that the figures above, while technically the results of the validation set rather than the test set, are effectively a second test set as they were not used in any way to affect training. The test set results were slightly stronger than the validation set, but those images were lost when our Colab instance died so we cannot provide those figures.

In Figure 4.13 we see that our overall model IoU is highly performant at over 97% IoU, which is beyond the 95% that our domain expert required. The driving force behind this is our background IoU at 98.83% IoU in Figure 4.14. Given that the grand majority of pixels in the image are background pixels, the high accuracy here has a strong impact on our overall IoU.

The plant parts do not fare quite as well, but still perform significantly better than the SAM Prompt technique. The Leaf IoU is over 71%, the root IoU over 68.5%, the stem IoU over 54%, and the seed IoU much weaker at 13.43% in figures 4.15, 4.16, 4.18, and 4.17 respectively.

We notice in our sample outputs such as the one in Figure 4.12, that our model still has extreme trouble with the seed. However, we also notice that in Figure 4.16 that our seed IoU is rapidly improving and expected to perform much better with more training. As a very small part of the image, but with our overall IoU so close to perfection, we believe that despite the slowed curvature of our overall IoU we will reach strong results for all plant parts with more training, of which we have a larger run in progress at the time of this writing.

When showing our randomly selected sample results to our domain expert client, we were informed that our leaf and root results are generally satisfactory, despite being significantly lower IoU than initially sought. The reason behind this is that the most important sought phenotypes can be calculated with a relatively high degree of accuracy

in our output. For instance, our root outputs tend to be wider than the actual root, in part because they are so more often in the given ground truth, but the lengths are quite accurate. The lengths turn out to be far more important phenotypes for plant success than the widths according to the domain expert.

We proceed to measure our root lengths against the ground truth, given that root length is one of the most important features we are looking for, where our measurement defines the length as the maximum euclidean distance in pixels between any two points in the root mask.

We find that mean relative error is 7.06%, with the best relative error at 0%, the Q1 error at 2.00%, Q3 error at 9.79%, and worst error at 35.75%. Conducting this at the very end of our project with no time remaining, we did not have a chance to examine individual case studies. However, there are points worth noting that suggest that the real error is less than that we find here.

Our means of measurement are imperfect. The actual length of the root is curved, not linear, meaning the error will be a smaller percent given the error occurs primarily due to the extreme ends of the ground truth mask. This is because it is not reasonable to expect the human expert ground truth mask labeler to provide a sufficient number of points on the root polygon to match the real shape of the root perfectly and it most prominently shows at these far corners and thus appears in our measurements. A better measurement may be to take the middle point of each far edge and find their distance traversed only within the root mask. Our expectation is that our mean relative error would be within the 5% sought after error margin.

CHAPTER 5. CONCLUSION AND FUTURE WORK

We conclude, based on the results of our experiments, that training a CNN-based model as per the typical state of the art technique for similar problems, is the most promising solution for the phenotyping of Arabidopsis.

Our UNet-based model preliminarily provides satisfactory results in many, but not all cases for the leaf and root and fewer and unsatisfactory number of cases for the stem and seed respectively. However, as discussed in Section 4.2, further training is expected to alleviate that and bring about stronger results for all parts in question and especially the poorest performing parts.

Given domain expert satisfaction with the leaf and root in many of our randomly sampled test cases, we believe that our model with further training resources is sufficient for a real world use case. However, we also believe that given the resulting IoU evaluation, we can still greatly improve the results to outperform human domain expert masks.

Our model is bounded by the quality of our data. Since a core aspect of our problem is the tediousness of gathering the necessary data, we believe our model can reach domain expert levels of precision with more resources. However, we also expect it to be bounded by the drawbacks of the given data, which as mentioned is incredibly tedious to collect to an ever more accurate extent. Despite seeing higher accuracy in cases such as softer curves on the root in Figure 4.12, we still see the root drawn too widely in the same case.

In order to achieve superhuman levels of performance, we should integrate SAM into a trained solution. At every step of our use thereof, SAM provided incredibly high

precision contours. The trouble with it was that the overall areas it discovered were wrong frequently. The reason for that is that it didn't have any related ground truth to work from.

SAM is a significantly larger model than what we can reasonably train ourselves, which is the reason for its high precision. However, as shown in the SAM2 Adapter [1], we could likely train it to achieve superhuman results on our Arabidopsis use case. We have confirmed that the necessary resources are available for general research use at Iowa State University.

Ultimately, we believe that a larger training run with free Google Colab resources is likely to be sufficient for our client's use case, on par with domain-expert labeled masks. However, we believe that superhuman performance is close within reach by training a SAM2 Adapter model on the data that we have already been provided and we recommend that the time be invested in order to do such with the free resources at our disposal at our institution.

BIBLIOGRAPHY

[1] Chen, T., Lu, A., Zhu, L., Ding, C., Yu, C., Ji, D., Li, Z., Sun, L., Mao, P., & Zang, Y. (2024, August 10). Sam2-adapter: Evaluating & Adapting Segment Anything 2 in downstream tasks: Camouflage, Shadow, medical image segmentation, and more. arXiv.org. https://arxiv.org/abs/2408.04579

[2] Orsolya, D., Horvath, P., Nagy, F., Danka, T., & Viczián, A. (2019, December). Deep learning-based approach for high-throughput Hypocotyl Phenotyping. https://academic.oup.com/plphys/article/181/4/1415/6000539

[3] Ravi, N., Gabeur, V., Hu, Y.-T., Hu, R., Ryali, C., Ma, T., Khedr, H., Rädle, R., Rolland, C., Gustafson, L., Mintun, E., Pan, J., Alwala, K. V., Carion, N., Wu, C.-Y., Girshick, R., Dollár, P., & Feichtenhofer, C. (2024, October 28). Sam 2: Segment anything in images and videos. arXiv.org. https://arxiv.org/abs/2408.00714

[4] Seethepalli, A., Guo, H., Liu, X., Griffiths, M., Almtarfi, H., Li, Z., Liu, S., Zare, A., Fritschi, F. B., Blancaflor, E. B., Ma, X.-F., & York, L. M. (2020). RhizoVision crown: An integrated hardware and software platform for Root Crown phenotyping. Plant Phenomics, 2020, 3074916. https://doi.org/10.34133/2020/3074916

[5] Smith, A. G., Han, E., Petersen, J., Olsen, N. A., Giese, C., Athmann, M., Dresbøll, D. B., & Thorup‐Kristensen, K. (2022). RootPainter: Deep learning segmentation of biological images with corrective annotation. New Phytologist, 236(2), 774–791. https://doi.org/10.1111/nph.18387

[6] Xiang, L., Nolan, T. M., Bao, Y., Elmore, M., Tuel, T., Gai, J., Shah, D., Wang, P., Huser, N. M., Hurd, A. M., McLaughlin, S. A., Howell, S. H., Walley, J. W., Yin, Y., & Tang, L. (2021). Robotic assay for drought (road): An automated phenotyping system for Brassinosteroid and drought responses. The Plant Journal, 107(6), 1837–1853. https://doi.org/10.1111/tpj.15401

[7] Yasrab, R., Atkinson, J. A., Wells, D. M., French, A. P., Pridmore, T. P., & Pound, M. P. (n.d.). Rootnav 2.0: Deep Learning for automatic navigation of Complex Plant Root Architectures | gigascience | oxford academic. https://academic.oup.com/gigascience/article/8/11/giz123/5614712