Non-Destructive Biomass and Relative Growth Rate Estimation in Aeroponic Agriculture using Machine Learning

Oskar Åström

Master's thesis
2022:E75

LUND UNIVERSITY

Faculty of Engineering
Centre for Mathematical Sciences
Mathematics

CENTRUM SCIENTIARUM MATHEMATICARUM

I would like to extend my sincere thanks to
Alexandros Sopasakis and Henrik Hedlund,
without whom this thesis would not have been possible.

I would also like to thank Alovivum,
for allowing me to use their aeroponic cultivation pilot
and helping me collect the data used in this paper.

Abstract

Optimising plant growth in a controlled climate requires good measurements of both biomass (measured in grams) and relative growth rate (measured in grams of growth per day and gram of plant). In order to do this efficiently and continuously on an individual level during plant development, this has to be done non-destructively and without frequent and labor intensive weighing of plant biomass. This thesis compares the ability of two machine learning methods, Multi-Variate Regression and Neural Networks, to estimate the biomass and relative growth rate from images of plants. The plant data set consists of images of 57 plants from two angles taken on 1-hour intervals during a 5 day period. The results show that images taken from a top-down perspective are best used with multi-variate regression, while images taken from the side are better when used with neural networks. In addition, using images from both cameras improved the biomass estimates from the neural network, but not those from the multi-variate regression. The predictions were improved in all cases when a moving average was taken of consecutive predictions, which likely reduced short-time variance in the data set. For both methods, the relative growth rate estimates were greatly improved by using estimates from both cameras. The low number of individual plants and high image capture frequency created a lot of correlation within the training set, which likely decreased generalization and lowered accuracy of the predictions on the test set. The best biomass estimates were made using multi-variate regression with images from the top camera and a moving average filter, resulting in an RMSE of 0.0391 g. This corresponds to a relative RMSE of around 11% which is comparable to previous studies. The relative growth rate estimates were not very accurate, but the best method used a neural network with both cameras, resulting in an RMSE of 0.1767 $g/(g \cdot day)$. This corresponds to a relative RMSE of over 100%. A bigger data set with measurements from a larger set of individual plants during a longer time interval within the cultivation period would likely improve these estimates.

Contents

1 Introduction

Food and water security is a crucial topic for ensuring the well-being of the human population, as noted by the UN, see Sustainable Development Goals 2 "Zero Hunger" and 6 "Clean Water and Sanitation"[1]. With the growing threat of global warming and its effects on arable land, safe drinking water, weather, and climate, it is of utmost importance to take sustainability into account when discussing our agricultural practices.

The current emergence of hydrocultural practices, such as *Aeroponics*, is one of many possible paths toward securing a sustainable food production in the future. With controlled growth conditions that such practices entail, it is possible to optimize the cost-to-yield in whatever cost metric is of importance. Be that the cost in terms of water, nutrition, money, emissions, or land-use. However, in order to make these optimization, it is necessary to determine the effects that parameters such as temperature, pH, lighting, space, CO_2, and nutrient concentration. have on the the status, quality, growth and development of plants. These experiments often require *destructive* measurements of the plants characteristics by harvesting them. This, of course, prevents the plant from growing further and only results in one data point per plant. This method is not only labor intensive, but results in poor estimates on an individual level and can only contribute to an estimate of the population as a whole. In order to estimate the effects on the scale of individual plants, continuous and non-destructive biomass estimates is a preferred choice.

The goal of this thesis is to evaluate two different image-based machine learning methods for estimating plant growth in *aeroponic* farming. These two methods are; Multi-Variate Regression (MVR) and Neural Networks (NN). The aim is for these methods to estimate the biomass (g) of a given plant, as well as the Relative Growth Rate (RGR) measured in $g/(g \cdot day)$.

The methods will be evaluated on a curated data set of spinach (*Spinacia oleracea*) grown in an *aeroponic* setup at an indoor farming facility in Gödelöv, Skåne. The data set used contains images of plants from two different angles taken at one-hour intervals. The aim is to estimate the biomass of the depicted plant given an image. The accuracy of the estimate is compared between using images from just one angle, multiple angles, or at multiple adjacent time points. These biomass estimates are then used to estimate the RGR.

With an accurate enough estimate of RGR, a *virtual sensor* could be developed, that can measure RGR continuously. This would allow for real-time optimization of plant growth, which can decrease the amount of unnecessary resources spent, including water, nutrients, lighting, electricity, and labor.

2 Background

2.1 Aeroponic Hydroculture

In conventional farming, nutrients and water are supplied to the plant roots through soil. However, these kinds of cultivation methods comes with some issues and limitations. For example, a lot of the water dispersed on the field will either trickle down into the ground water or evaporate due to the large exposed surface area, meaning that a lot of the water is lost to the environment. The fertilizers used together with soil or other solid substrates can also be carried down with the water, leading to nutrient waste and water eutrophication. In addition, outdoor farming leaves the plants exposed to a number of factors that can harm the plants, such as droughts, soil-borne diseases, bad weather, pest attacks, and floods.

An alternative to soil-based farming is hydroculture, where the nutrients are carried to the plants using water as the nutrient carrier. In particular, hydroponic farming has seen a large increase in commercial use in the last couple of years, Nordic Harvest, Swegreen, and Grönska being recent Nordic examples.

Aeroponics is a variant of hydroculture where the nutrient-rich water is dispersed as an aerosol around the roots. This method has a number of benefits compared to soil cultivation and hydroponics. The first benefit is reduced water usage. In comparison, the water cost of 1 kg of tomatoes is 200-400 L in conventional soil farming, 70 L in hydroponics, and 20 L in aeroponics [2].
Another advantage is that a limiting factor of hydroponic farming is the oxygen content in the water, which can be a maximum of 8 ppm [2]. This means that the aeration of the plants roots is much better in aeroponics, leading to a higher growth rate.
A third and major benefit of aeroponics is the effective distribution of water and nutrients as droplets to the roots, which together with an adaptive branching strategy of plant roots (a dense network of fine roots and root hairs) creates an optimized supply and a large uptake area for water and nutrients. If the fluid and the droplets in the aerosol contains the necessary nutrients (N, P, K, Ca, S, and Fe) in optimal proportions, they will be supplied in a non-limiting capacity, where uptake just is determined by the root uptake capacity.

2.2 Factors in Plant Development

There are several factors or conditions in the environment that will determine plant growth and development. Two factors are mentioned in section 2.1, namely water- and nutrient supply. Three other main factors are

- supply of carbon as CO2 for uptake by plant leaves and used in photosynthesis,
- supply of light as photons, captured by chloroplasts in plant leaf cells and tissues and used in photosynthesis, and
- a number of climate factors and environmental conditions, such as temperatures (air, leaf, fluid, root), air humidity, CO2 concentration in air, nutrient concentration (measured as ions with electrical conductivity, EC), pH and dissolved oxygen (DO).

The main strategy in controlling plant growth and development is to quantify and control these factors to levels that induce the desired plant behavior.

The research conducted at Swedish University of Agricultural Sciences (SLU), during a long period from the 1970s to the millennium shift, has created a broad and solid scientific framework for controlled plant growth- and development and, thereby, highly efficient and productive plant cultivation. This research is well documented in scientific papers, dissertations and plant development databases [3–8] and provides a solid ground for the hypotheses and assumptions to be tested within the scope of this master thesis.

One main assumption behind the goal of this thesis is that there exists a strong correlation and coherence between relative rates of growth for biomass and leaf area, respectively. This has been verified in different experiments and is documented in [3, 7], where biomass growth rates have been determined by weighing of plants, roots and leaves and leaf area growth rates by measuring of projected area via a scanner/copier.

The main idea is that these correlations between the plant's total biomass and leaf area is a good motivation for using image analysis. Since the leaf area is likely extractable from images, it should mean that the biomass is as well.

2.3 Cultivation Setup

The setup used for plant growth and development consists of four *growing beds* and one *reservoir*. The reservoir contains the nutrient rich water along with sensors for temperature, pH, and electrical conductivity, and a heater to keep the water at a preferred set temperature. The water from the reservoir is then pumped to the four growing beds.

Each growing bed consists of a container with a removable lid. Inside the container, there are two *sonicators*, which contain membranes that vibrate at an ultrasonic frequency that agitate the water into an aerosol, in the form of a dense fog. The lid of the container has 24 holes with small baskets (plant holders), in which the plants are placed such that the roots hang down and are immersed in the aerosol, continuously replenishing the supply of water and nutrients. The bottom of the container has a drainage pipe which returns the water back to the reservoir, as shown in Figure 1.

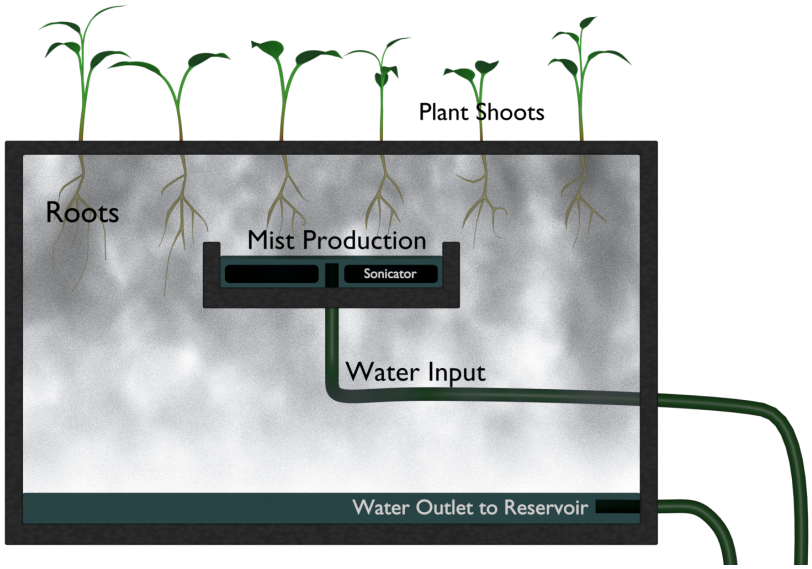

Figure 1: Growing bed schematic from the side.

Above each growing bed is an array of LED lights. These can be either set to full or half intensity. The full intensity assures that the plants receive saturated light levels, measured as Quantum Flux Density (QFD) in the range of 350-400 $\mu mol/m^2s$. At half intensity the QFD is in the range of 150-200 $\mu mol/m^2s$. At this level, the light conditions become a limiting factor, which according to theory and experiments [8] will lower the relative growth rate (RGR).
Two of the growing beds received optimal and saturated light conditions, while the other two were grown under limiting light conditions. This introduces variance in the RGR between the plants, and thus increases the variance in our data set. This will, most likely, improve the generalization of our estimates, but the effect of this is not evaluated in this thesis.

2.4 Camera Rig

The camera rig consists of 8 cameras (2 per bed) of the model *PlexGear WC-800* set up as shown in Figure 2. Four of the cameras capture images from a top view, placed on the long edge of each bed. These cameras are referenced to as the *top cameras*. The other four cameras are placed further down and over the adjacent bed, so that they capture images from a lower angle. These cameras are referenced to as the *angled cameras*.

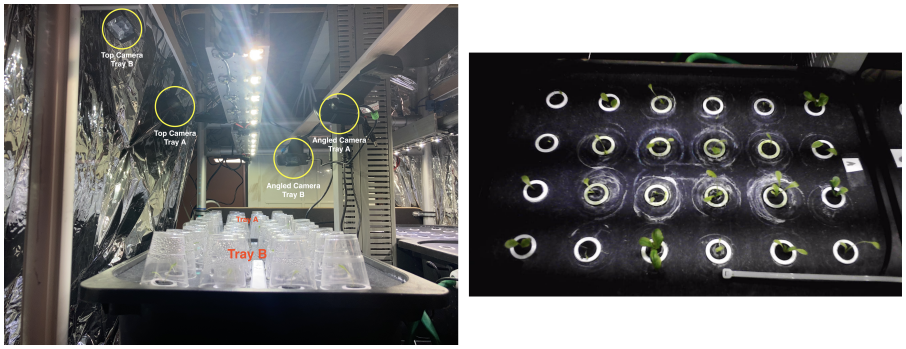

Figure 2: Camera rig for image capture (left) and example of captured image (right).

The cameras are connected to a computer next to the rig, which runs a script to capture images, shown in Figure 2, with each camera on 1 hour intervals. These images are then stored locally and sent directly to a cloud sharing service so that the camera status and plant development can be monitored remotely.

3 Method

The goal of this thesis is to compare the precision of two methods for predicting the biomass based on a given image or set of images. The first method is to use Multi-Variate Regression (MVR). This essentially means that a number of potentially interesting features are extracted from each image and these features are then used to estimate the biomass. This is similar to a methods employed by *Jung, Dae-Hyun, et al.* [9] for estimating biomass in lettuce. In their paper they used several methods for pixel classification between plant and background. The number of pixels were then used to estimate the biomass using polynomial regression. The structure for my method uses similar manually curated features in order to estimate the biomass, which should be promising based on their findings.

The second method is to use Neural Networks (NN) in order to predict the biomass. A pre-trained network base called *ResNet-50* is used to extract a feature vector which is then fed into a regression head which aims to predict the biomass. This method is based on previous research by *N. Buxbaum et al.* [10].

In order to predict biomass from images we will need *ground truth* data and *input data*. The input data consists of images of plants and the ground truth data consists of values for the biomass of the corresponding plant. These image-value pairs constitute our data set, which is used as training data for the two methods.

The estimates of the biomass in the images are then compared using three *View Types* in three *Tasks*. The first two view types are predictions made using the images from the top and angled camera respectively. The third view, called *Dual View*, is the average of these predictions for a given plant and time point. The purpose of evaluating this is to determine if there is a benefit to having both cameras.

These three *views* are used to complete three tasks. The first task, *Single Image*, is simple biomass prediction using a single plant and time point. The second task, *Moving Average*, uses the average of three consecutive time points to estimate the biomass. This task should remove high frequency variations in the estimates since the real biomass has had minimal change during this 3 hour period. The third task, *Relative Growth Rate*, uses three random time points of a plant to estimate the RGR. By sampling this 100,000 times, a good estimate of the method's ability to estimate RGR from biomass is evaluated.

3.1 Trial Experiment

In studies analyzing plant growth using machine learning, it is common to use a classification output where the plant growth is lumped into a number stages of progression. For example, growth stages of wheat is typically grouped into the

stages; Tillering, Jointing, Booting, Heading and Ripening [11, 12], corresponding to different visual and biological processes in the plants. This makes the learning easier since the targets are limited to discrete values, and usually one is only interested in a rough estimate of the plant's state of growth. However, for the purpose of estimating the RGR, a numeric value of the biomass is necessary. Therefore, the commonly used classification is insufficient.

3.1.1 Trial Data Set

To see the effect of using a numeric output instead of a classification, a trial experiment was conducted on a public data set before shaping our own experiment and data set. The data set used for this is outlined in Beck et al. [13]. The authors of this paper released an open-access sample of this data set containing 14 different plant species with 1000 images each. The data set also contains information about the date and time at which the images were taken. As a trial experiment, a neural network of the same structure as the one used on our own data set was used to predict the age of soybean plants, examples of which are shown in Figure 3.

Figure 3: Example images from the trial data set.

The dates were converted to hours and normalized such that the earliest image had the age *0 hours*. In the interval 600-1100 hours, no images seemed to exist as shown in Figure 4. The images were then split into two classes, those taken before this period and those taken after.

The *young* age group was much larger than the *old* age group, with a split of 861 vs 139 samples. Therefore, *augmentation* was introduced to *balance* the sizes of the two age group. Each image added was rotated by a random angle and resized to 64x64 pixels, as this was about the size of the smallest images. The *old* age group could therefore be filled with rotated copies of itself so that both sets were of the same size. This experiment was conducted both with and without the class balancing in order to evaluate its effect. Finally, the data set was divided into a training and validation set with a 80-20 split.

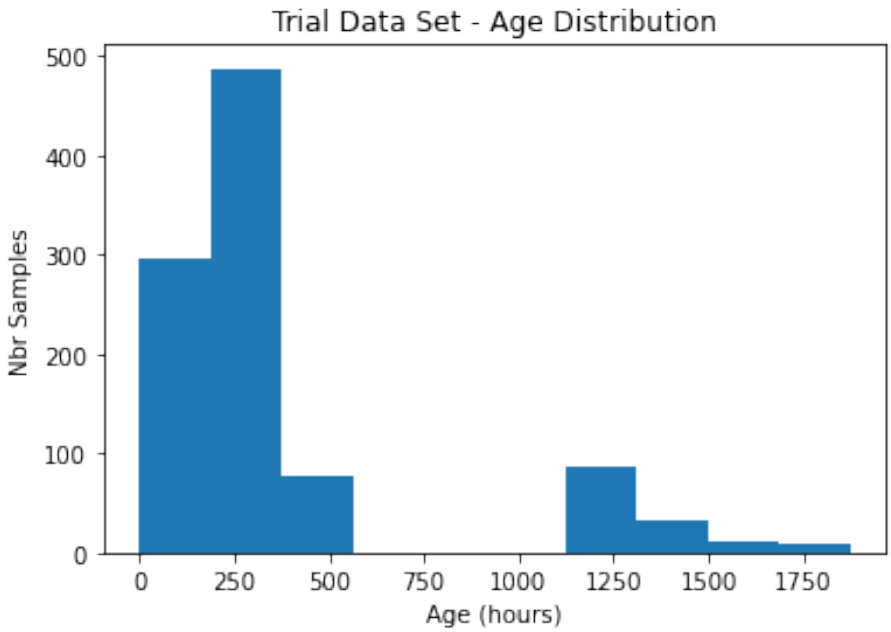

Figure 4: Age distribution in the trial data set.

3.1.2 Trial Outline

The goal of this trial is to classify the plants in two ways, with a classification network, that classifies what age group the plant belongs to, and a regression network that estimates the age of the plant in a given image. The output of the regression network can then be used to classify the plants into the two groups. This is done by finding an optimal age threshold in the training set where the network output can be split into the two age classes.
The ability of these two networks to classify the plants are then compared, and the accuracy of the regression network to predict the direct age is analyzed.

Usually, when the data has a clear split similar to this, it could introduce biases into a classification network through features such as background shifts, camera equipment, or human input that happen in the period between 600-1100 hours. This is also the case for this purpose, however, these biases should not affect the accuracy of the regression network within the classes. So if the accuracy of the two networks to classify the plants are comparable, and the accuracy of the regression network is good within the two groups then this indicates that a regression network could perform well for the purpose of predicting the direct biomass in our experiment as well.

Both networks used an architecture similar to the one used for our main experiment (see Figure 10), using a pre-trained ResNet-50 network base with a 1000 feature vector output followed by a regression head (further explained in Chapter 3.5). The regression head consisted of two dense layers of size 512 and 64 with $ReLU$ activation, followed by an output layer that differed between the two network types. The classification network used an output layer of size 2 with *Softmax* activation. The regression network used an output layer consisting of a single node with *Linear* activation.

3.2 Ground Truth

The ground truth data is gathered by sampling the plant weight during the growth period. This was done at two points during the growth period. At each measurement, all plants are weighed non-destructively in their baskets. Since the plants are weighed in their baskets, the basket weights are measured as well so that these can be subtracted from the total measured weight.

Since images are taken on one hour intervals, these two measurements have to be interpolated in order to get ground truth data for images in between measurements.
A paper by O. Hellgren and T. Ingestad [6] has shown that in controlled cultivation experiments with a constant relative addition rate of nutrients, plants have a *constant* relative growth rate. To achieve this for the thesis experiment, nutrients and water are supplied in free access, which imply non-limiting addition rates. This should ensure that the plant's RGR is close to constant, meaning that each gram of plant increases its mass by a constant amount per day. This holds true in the early stages of a plant when all plant cells are dividing and the whole biomass is productive and growing, called *vegetative growth*, which is the case for leafy green plants such as spinach. This is not necessarily true for plants which produce structural cells that are not dividing (such as woody plants), or for example plants which have started flowering. Since our harvest period only covers the vegetative state, this assumption should be a good approximation. However, since the plants experience an environmental shift when they are placed in the containers, there is an initial period of plant acclimatisation, where the RGR increases to a stationary level. This means that the estimates through our interpolation might differ from the true biomass of the plants. A larger data set over a longer period should give a clearer and more accurate *ground truth* data, but given the limited data set this should at least give a close approximation.

Given this assumption of a constant RGR, that means that the *change* in biomass is proportional to the biomass itself. This can be expressed as a differential function $\dot{w} = a \cdot w$, where w is the biomass and a is the constant RGR. This differential function is solved by a generic exponential function

$$w(t) = C \cdot e^{at},$$

where C is a constant. By taking the logarithm of the biomass measurements, we get

$$ln(w(t)) = log(C) + a \cdot t,$$

which is the equation for a line. By fitting a line to the logarithm of the weight measurements as a function of time, we can estimate the relative growth rate and the exponential function which the biomass follows. This function is then used to get ground truth data for intermediary time points.

3.3 Input Data

The input data to the model consists of image-biomass pairs. At each time point, 8 images are captured, each of which cover an entire tray.

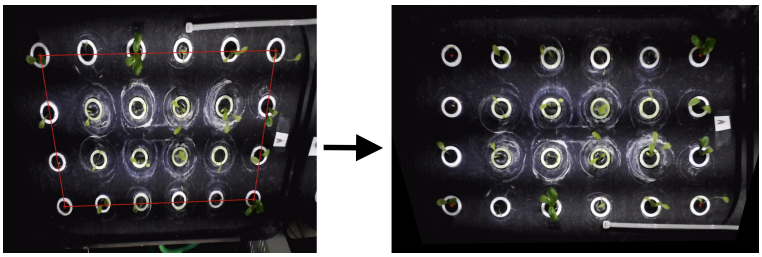

Figure 5: Example of input image transformation.

In order to convert the images of an entire tray to images of individual plants, the image has to be segmented. This is done by manually marking the positions of the four corner baskets of the tray in each camera. From these points, the positions of all plant containers can be inferred. The four corner points will make up a distorted rectangle, which can be transformed into the correct proportions since the aspect ratio of the tray is known. This is done using a projective transformation, an example of which is shown in Figure 5. After this transformation, a square region is cut out around each plant basket's position. Finally, each plant image is resized to so that all images are of the same dimensions, 64 by 64 pixels. An example of the resulting images for one plant is shown in Figure 6.

As can be seen, the top camera captures the image more clearly, while the angled camera becomes slightly distorted to accommodate the larger distance between the closest and furthest plants. This is, however, important to make sure that the pixel scale is proportionate for all plants, independent of their distance to the camera. The angled camera does capture more complexity, such as height and leaf slope, which is lost in the top view.

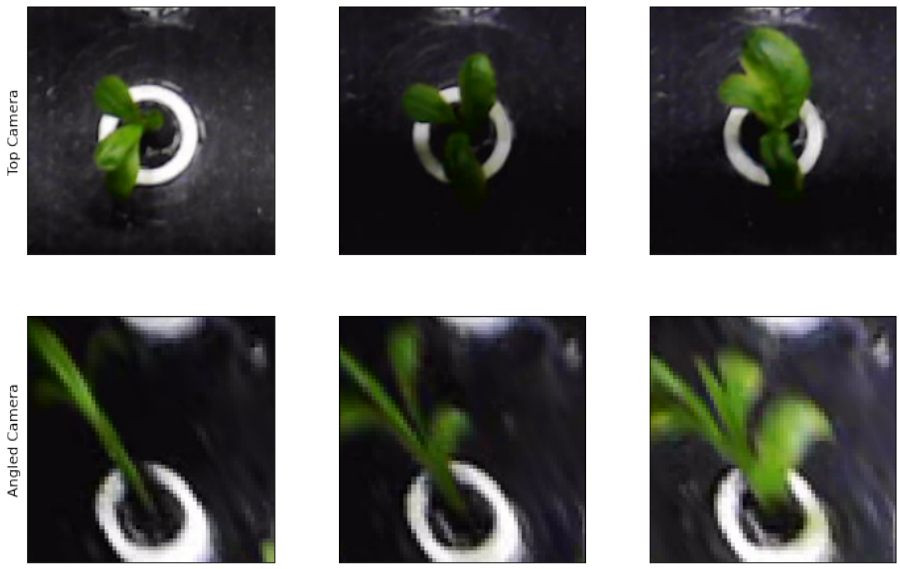

Figure 6: Example of plant development as input images from both the top and angled camera.

In total the data set is over a 5 day period with a total of 96 plants. However, due to some acclimation issues, some of the plants had to be replanted. These plants had a shorter time period between the two biomass measurements which resulted in a high variance. Therefore, these were excluded from the data set. In total 57 plants were in the data set, of which 5 days of images were taken. This resulted in 10,197 images of individual plants.

The data set was split into training, validation and test sets. Since both the biomass and the images are highly correlated between time points, the data splitting was done based on plant individuals as opposed to a random splitting. When a random splitting was tested, the accuracy was extremely high with almost no error. However, since potential future plants will not be represented in the data set, this high accuracy cannot be said to reflect the generalization of the problem. Therefore, the split was done on an individual basis. The validation and test set were created with 6 plants each, this resulted in a split of [77,4%, 11,3%, 11,3%] with 7894, 1153, and 1150 samples in each set respectively. The reason that there is a difference in sizes between the validation and test set is that some images were lost due to camera malfunction. Half of the images in each set was from the top view, and the other half from the side view. Therefore, these sets were further split based on the camera angles. The result is two data sets with the two view angles consisting of image-biomass pairs. Each data set has a training, validation and test set that were split using the same individuals.

3.4 Method 1: Multi-Variate Regression

The first method for estimating biomass is a multi-variate regression based on manually decided features, similar to what is explored in, [14]. Specifically, linear regression of *pixel-wise* features is used. These types of features do not capture structure which differentiates it from the NN method which is highly dependent on structure. These features $F_i : \mathbb{R}^{64x64x3} \to \mathbb{R}$ can be expressed as

$$F_i(I) = \sum_{x=1}^{64} \sum_{y=1}^{64} f_i(I_{x,y}), \qquad f_i(p) : \mathbb{R}^3 \to \mathbb{R},$$

where I is an image and $I_{x,y}$ is a pixel in that image. These features are based both on the RGB and HSV values of a given pixel $p = I_{x,y}$. These are expressed as p_r, p_g, p_b, p_h, p_s, p_v below. The features included were inspired by previous research of image based plant segmentation [14–16]. Due to the principles of linear regression, the features that could be written as a linear combination of others were excluded. In addition to these features, some hue and saturation-based features were added.
The features can be grouped into six groups. The first group is just summations of each of the six color channels: Red, green, blue, hue, saturation and value. The next group is ratios between the three colors. The third group is Thresholding Features which count the number of pixels within certain hue ranges; Red-Green, Green-Blue, Blue-Red. The fourth group is Highlighting features, which operates similarly to the Thresholding Features but weights the pixels by their saturation so that more vibrant pixels count for more. The fifth group is a single feature VEG which represents the Vegetative Index [17].

- Red: $f_1(p) = p_r$
- Green: $f_2(p) = p_g$
- Blue: $f_3(p) = p_b$
- Red/Green: $f_4(p) = p_r/p_g$
- Green/Blue: $f_5(p) = p_g/p_b$
- Blue/Red: $f_6(p) = p_b/p_r$
- Red-Green Hue: $f_7(p) = 1$ *if* $p_h < 1/3$ *else* 0
- Green-Blue Hue: $f_8(p) = 1$ *if* $1/3 < p_h < 2/3$ *else* 0
- Blue-Red Hue: $f_9(p) = 1$ *if* $2/3 < p_h$ *else* 0
- Red-Green Highlight: $f_{10}(p) = p_s$ *if* $p_h < 1/3$ *else* 0
- Green-Blue Highlight: $f_{11}(p) = p_s$ *if* $1/3 < p_h < 2/3$ *else* 0
- Blue-Red Highlight: $f_{12}(p) = p_s$ *if* $2/3 < p_h$ *else* 0
- Vegetative Index: $f_{13}(p) = p_g p_r^{-2/3} p_b^{-1/3}$

These 13 features were normalized to the interval [0,1] based on the training set and exported, along with the ground truth values, as an *csv*-file to the programming language **R**.

3.4.1 Feature Evaluation in R

In order to determine what features are used in the final model, a step-wise iteration of models is implemented. This is done by starting with an empty constant model containing only the *intercept*. Features are then added to the model depending on the *Akaike information criterion* (AIC). The AIC of a model is defined as

$$AIC = 2k + 2ln(\hat{L}),$$

where k is the number of predicted parameters and $\hat{L}$ is the maximum value of the *Likelihood function*. When the AIC cannot be improved by adding or removing a feature, the iteration is done and a final model is given.

This iteration is also done using the BIC value, which is a variant on the AIC value where the function is changed from $2k + 2ln(\hat{L})$ to $ln(n)k + 2ln(\hat{L})$, where n is the number of samples in the training set. This change punishes large models, typically resulting in a smaller model.

To compare the AIC model to the BIC model, their corresponding R^2_{adj} values are calculated. This is a variant on the R^2 value which is a measurement of how much of the variance is explained by the model. The R^2 value is calculated through

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}},$$

$$SS_{res} = \sum_{i=1}^{N}(y_i - \hat{y}_i)^2, \qquad S_{tot} = \sum_{i=1}^{N}(y_i - \bar{y})^2,$$

where SS_{res} is the sum of squares of the residuals, SS_{tot} is the sum of squares between the data points and their mean, $\hat{y}_i$ is the predicted value and $\bar{y}$ is the mean of the data. To get the R^2_{adj} value, the R^2 value is adjusted through

$$R^2_{adj} = 1 - \frac{(1 - R^2)(n - 1)}{n - k - 1},$$

where n is the number of data points and k is the number of features.

In addition to performing this fitting to the biomass, a similar method is performed for the log-biomass. The selected features for the biomass model and log-biomass model is then used on the validation set. The model with the lowest *Root Mean Squared Error* (RMSE) on the validation set is used as the final model for the MVR method.

For this method, two of these models are made, one for the top-view cameras, and one for the angled-view cameras.

3.5 Method 2: Neural Network

The second method used is a *Neural Network* (NN), inspired by previous research in biomass prediction [10]. The network utilizes a pre-trained image recognition network called ResNet-50, trained on data from ImageNet[18], as a base. This network returns a feature vector. These features are then fed into the second part of the network, a regression head, with untrained weights, which predicts the biomass.

3.5.1 Pre-Processing of Training Data

The image data of the plants are highly correlated with each other. This means that the network can easily overfit to the training data. In order to combat this augmentation is applied to the images. This is done by rotating each image by a random amount. This resulted in a massive decrease in the validation loss, at the cost of increasing the training loss.

In addition to this, different color spaces were explored as shown in Figure 7. The first color space was the conversion to HSV. The background mainly consists of black values with some white from the plant holder and water residue. In the RGB colorspace, black and white are [0,0,0] and [1,1,1] respectively, while they are [0,0,0] and [0,0,1] respectively in HSV. This means that a HSV color space could be beneficial since the difference between black and white is much smaller. This color space also has a basis in previous research for plant segmentation [16, 19]. In addition, two personally curated color spaces were tested.
The second color space explored was one where the Value channel of the HSV color space was set to 0.5 and converted back to RGB. The idea behind this was to keep the color space RGB, while removing the difference between black and white.
The third color space also took the Saturation of the HSV color space into consideration. Instead of setting the Value channel to a constant 0.5, it was set to $(0.5 - |0.5 - V|) \cdot 2S$, where V is the Value channel and S is the saturation channel. This meant that pixels that are dark, white, or lack color are dimmed. The resulting image was then turned back to RGB. This highlighted the plant parts very well as seen in Figure 7.

However, the different variations of color space did not outperform the simple RGB. This is likely because our network uses a pre-trained ResNet-50 network which has been trained on real images. By modifying the color space, the network receives input that is further away from the data set it has been trained on, thereby decreasing the accuracy. It is therefore possible that a network built from scratch could benefit from these other color spaces while ours didn't.

A pre-processing of the biomass was also explored. Since the plants grow exponentially, the use of the log-biomass as target was tested. In addition, the

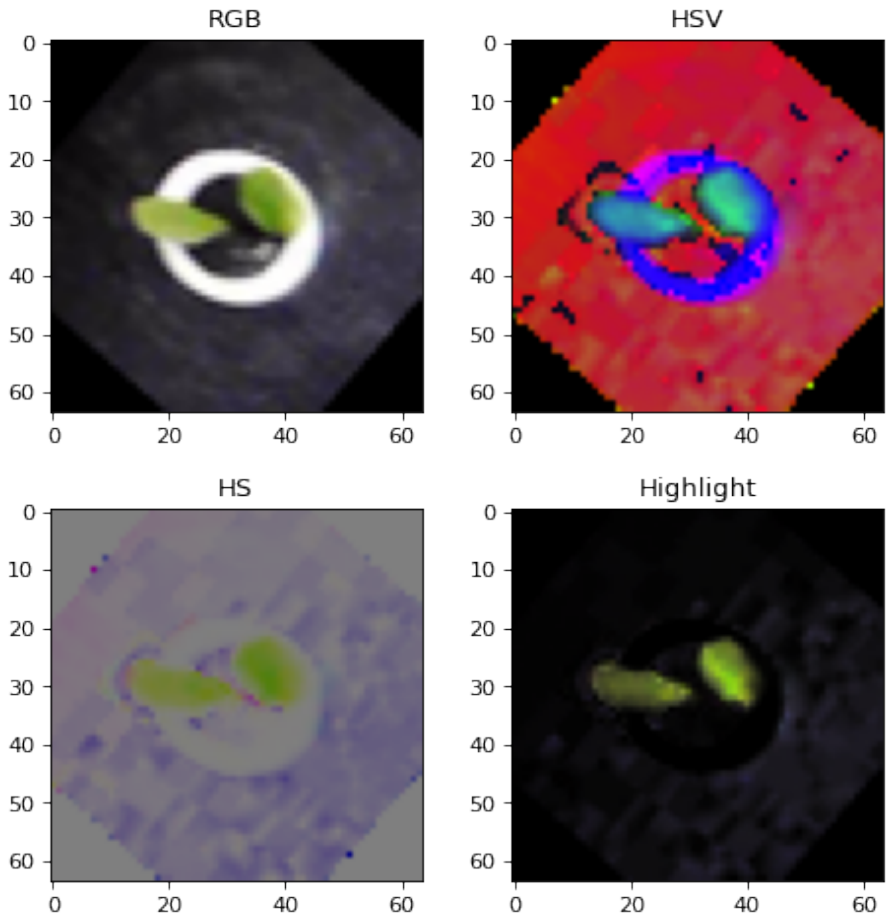

Figure 7: The different explored color spaces.

biomass and log-biomass was normalized based on the maximum and minimum values of the training set. The distribution of both these targets in the training set is shown in Figure 8. For the final model, the log-biomass was used since it was found to perform much better on the validation set. This could be because the log-biomass is much more evenly distributed than the linear biomass.

In conclusion, the data used to train the NN was randomly rotated RGB images of the plants with normalized log-biomass targets.

3.5.2 ResNet-50

Deep neural networks are common in the field of Image Recognition. The depth, specifically, is important to capture the complex structures of images. There

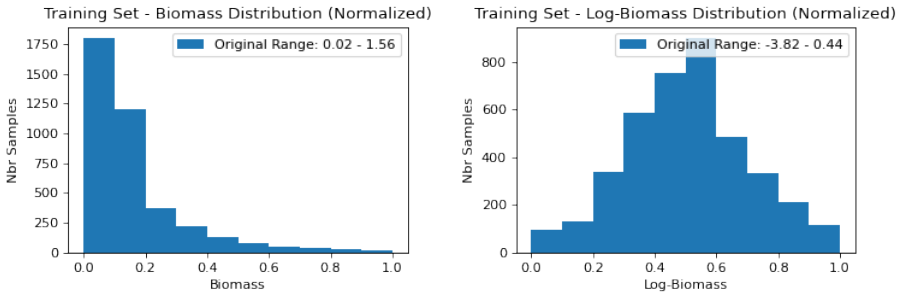

Figure 8: Distribution of Biomass and Log-biomass in Training set.

are, however, some issues when increasing the depth [20].

The first issue is Vanishing Gradients. When back-propagating the feedback from training to a weight, the gradient of the loss with respect to that weight is utilized. If the network is deep, the deep inner derivatives result in a product of many factors. If these factors are small the gradient quickly becomes small, meaning that the network learns very slowly on the deeper weights.

The second issue is an observed degradation in the accuracy. As explained in [20], increasing the depth of a network increases the accuracy to a point, after which the accuracy stagnated before degrading quickly. As they point out, this degradation is not due to overfitting.

Deep neural networks have many layers. These extra layers are noted as $g(x)$ in Figure 9. If the shallow layers, noted as $f(x)$, of the network are accurate, then $f(x)$ is a good approximation of the desired output y. By adding the deep layers, the output of the network becomes $g(f(x))$. The network can achieve the same accuracy if $g(f(x)) \approx f(x)$, so the training accuracy should not decline with more layers. However, this requires that $g(x) \approx x$, but convergence to the identity function can be difficult to achieve in training according to [20].

In order to combat this, the authors created a *Residual Neural Network* (ResNet). In the Residual Block the result from previous layers are added to the output, changing it to $g(f(x)) + f(x)$. In order for this to achieve the same accuracy as the shallow network, we instead get

$$\begin{aligned} g(f(x)) + f(x) &\approx f(x) \Rightarrow \\ g(f(x)) &\approx 0 \Rightarrow \\ g(x) &\approx 0. \end{aligned}$$

This is much easier to achieve, since this simply involves setting the weights to 0. This means that adding the Residual block should never decrease the training accuracy, since it can easily default back to the result of the previous

layers.

The number of layers inside such a residual block is typically two or three, and stacking multiple of these after each other allows for very deep networks that minimize the risk of accuracy degradation.

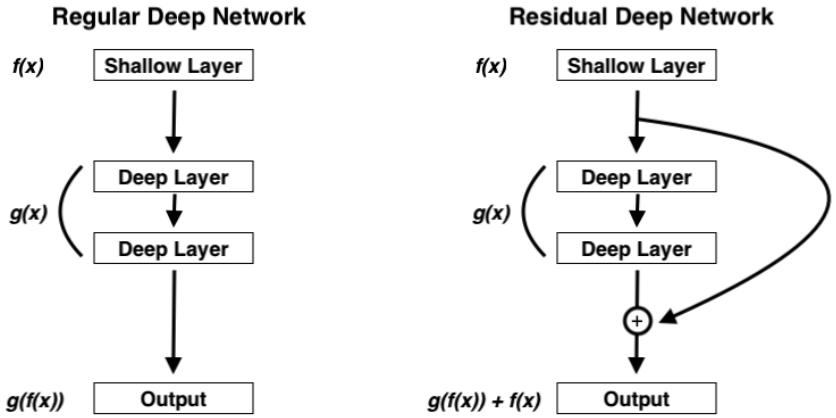

Figure 9: Residual Block used in ResNet.

The ResNet architecture is used as the first part of my network. In particular, ResNet-50 is used, where 50 denotes the number of layers. The network has been pre-trained on the ImageNet data set [18] and should therefore allow for advanced image feature recognition without the need for a large data set.

3.5.3 Network Architecture

Two networks were employed, one for each camera angle. The network architectures were identical and consisted of a ResNet-50 network base that resulted in a feature vector of length 1000. This vector was then fed through a Regression Head with 3 layers of size 512, 128, and 1 respectively as shown in Figure 10. These layers were densly connected and used $ReLU$ activation, except for the final output layer which used a *Linear* activation function. Since the problem is rather complex, but has a lot of redundancy in the training set due to the short times between images, the network complexity is operating on a fine line between capturing complexity while not overfitting, as is often the case with tasks in image analysis. So a smaller regression head, with fewer and/or smaller layers, increased the error of both the training and validation set, while a larger one decreased the error for the training set but increased it for the validation.

The learning rate had to be set very low, at 0.0005. A higher learning rate lead to instant collapse where the network would output a constant value. A lower learning rate slightly increased the error for both the training and validation set. The networks were trained for 20 epochs.

In order to combat overfitting, L2 Regularization was employed. The L2 value was set to 0.1, which is on the high side. But a lower level of regularisation lead to increased overfitting. A higher level of regularization didn't have that much effect on the network, so it was kept at 0.1.

Another very important aspect was to allow the ResNet-50 weights to be trainable. This meant that we would take advantage of the large network, while allowing for slight adaptation to our problem. This greatly decreased the error.

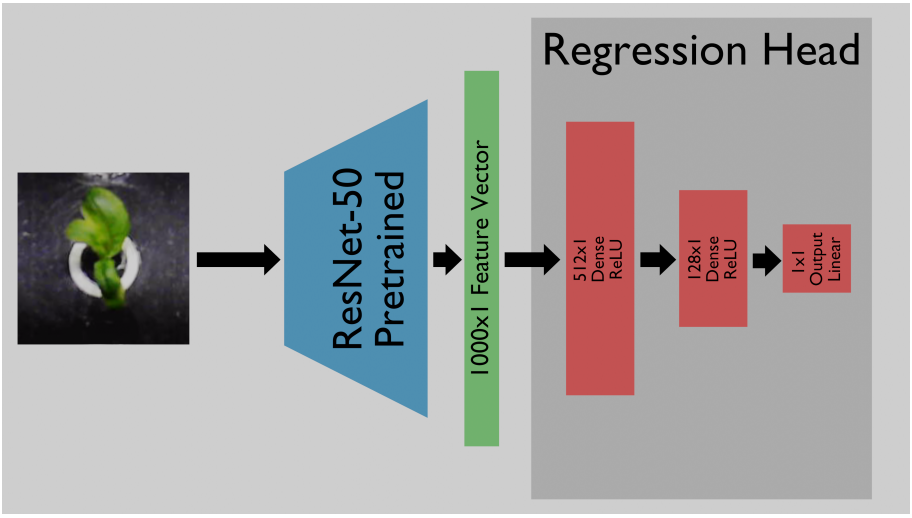

Figure 10: Architecture of the neural networks.

3.6 Method Comparison

For both the MVR and the NN approach, the result comes in the form of two vectors of predictions on the *test tet*, one for each camera angle. We call the predictions for the top view $\hat{y}_t$ and the predictions for the angled view $\hat{y}_a$. These can be compared to the ground truth values y. The comparison between the estimate and the true value is expressed using the *root mean squared error* (RMSE), calculated using $\sqrt{\frac{1}{N}|\hat{y} - y|_2^2}$, where N is the number of elements in the set.

As mentioned the two methods are evaluated on the accuracy of the estimates on a single image. This is done simply by comparing $\hat{y}_t$ and $\hat{y}_a$ with their true counter parts. In addition to this, a new vector $\hat{y}_d$, representing the dual view, is constructed by taking the average of $\hat{y}_t$ and $\hat{y}_a$. This vector is then

compared to the true biomass at each time point. The purpose of this is to check if there is any benefit to having multiple cameras.

Secondly, the moving average of the biomass estimates are calculated by grouping the data points into six groups based on the six plant individuals in the test set. The estimated biomass vector is then sorted based on the true biomass vector. A Moving Average filter with window size 3 is then applied to this vector. So each biomass estimate also uses the estimates for the two time points before it. The purpose of this is to minimize noise in the estimates. This resulting vector is then compared to the true biomass. This is done three times, once for each camera angle and once for the dual view.

Finally, the possibility of estimating relative growth rate is explored. This is done by sampling three random time points in the biomass estimate and fitting an exponential curve to the biomass, similar to how the ground truth values was created. The estimate for the relative growth rate is then compared to the real value for that plant.

4 Result

4.1 Trial Experiment

In the trial experiment, the age of images of Soybean plants from a public data set were analyzed using two networks, a regression network and a classification network. The training progress of the two networks when using *unbalanced* and *balanced* age classes are shown in Figure 11 and 12 respectively.

The predictions of the regression network is shown in Figures 13 and 14 for *unbalanced* and *balanced* data sets respectively.

The threshold used for the regression network to classify the age group of plants were 790 hours for both the *unbalanced* and *balanced* data sets.

Figures 15 and 16 show the ground truth age against the predicted ages along with fitted lines using different three different groupings of the validation data; *all* plants, only *young* plants, or only *old* plants. These lines show how correlated the predictions are with the true age.
For the *unbalanced* data set, the slope of the lines were; 0.427, 0.709 and -0.227 for the groups *all*, *young* and *old* respectively.
For the *balanced* data set, the slope of the lines were; 0.767, 0.755 and 0.597 for the groups *all*, *young* and *old* respectively.

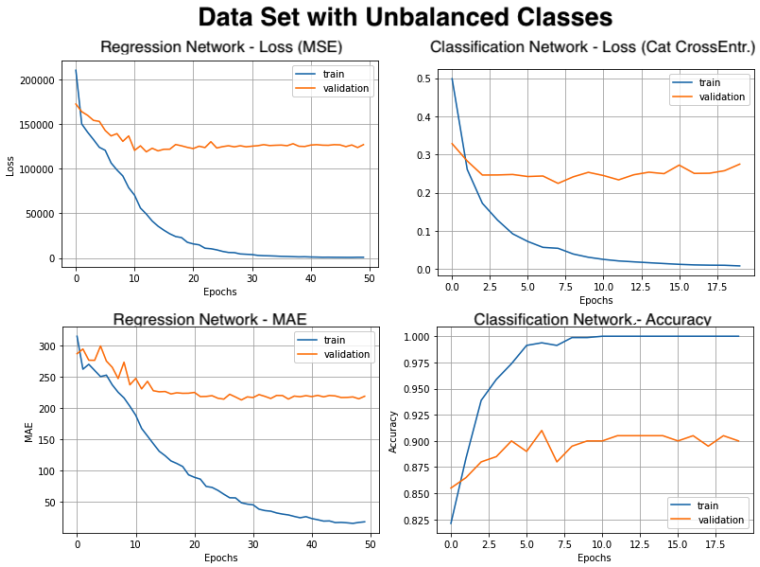

Figure 11: Training progression for the regression and classification network when using a data set with unbalanced age classes. This is an acceptable result.

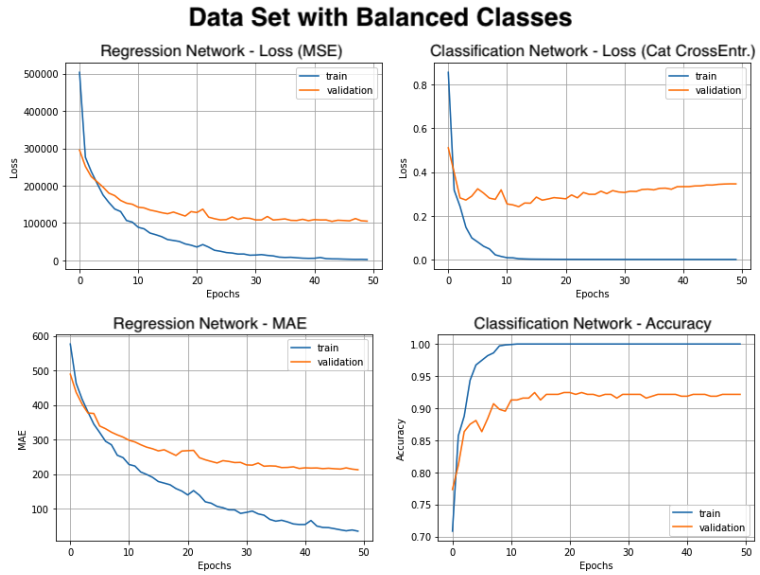

Figure 12: Training progression for the regression and classification Network when using a data set with balanced age classes. The balancing improves the result, but leads to a slight over-fitting for the classification network.

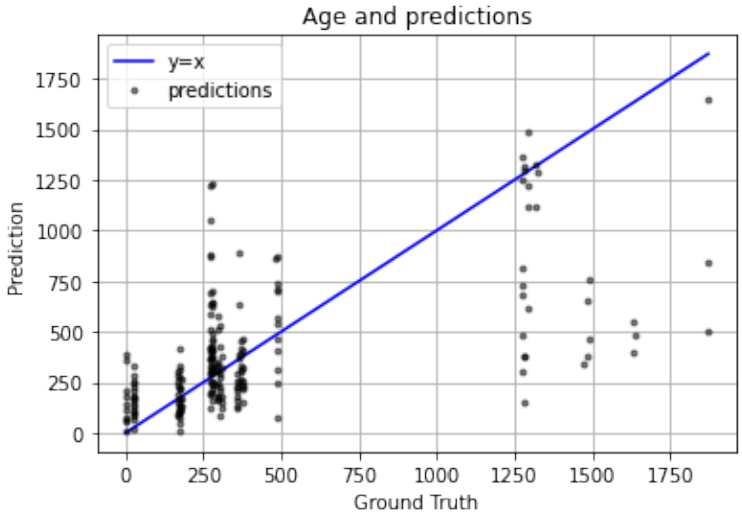

Figure 13: Ground truth age against predictions of the regression network for the unbalanced data set. Blue line shows optimal correlation $x = y$. The predictions are fairly good in the young age group, but are not well fitted in the old age group.

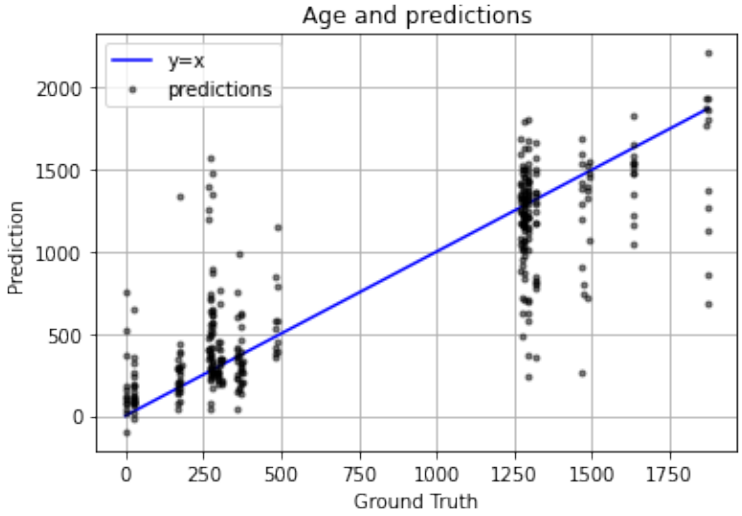

Figure 14: Ground truth age against predictions of the regression network for the balanced data set. Blue line shows optimal correlation $x = y$. The predictions are fairly good in all age groups.

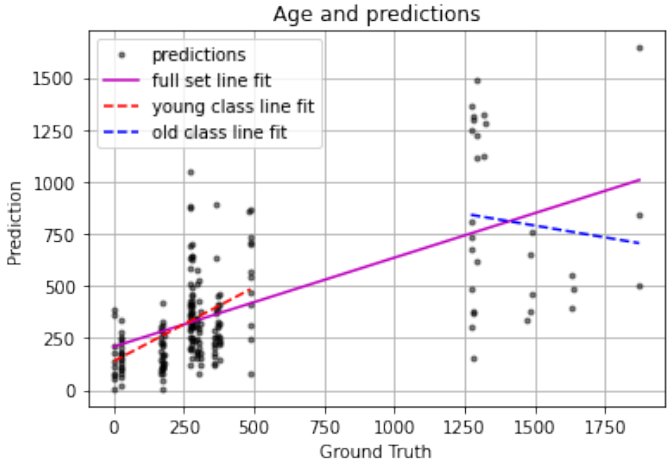

Figure 15: Fitted lines with ground truth age vs predictions on the unbalanced data set for; entire validation set (Purple= $209 + 0.427x$), young class (Red= $139 + 0.709x$), old class (Blue= $1130 - 0.227x$). The regression can distinguish quite well between ages in the young age group, but fails to do so within the old age group.

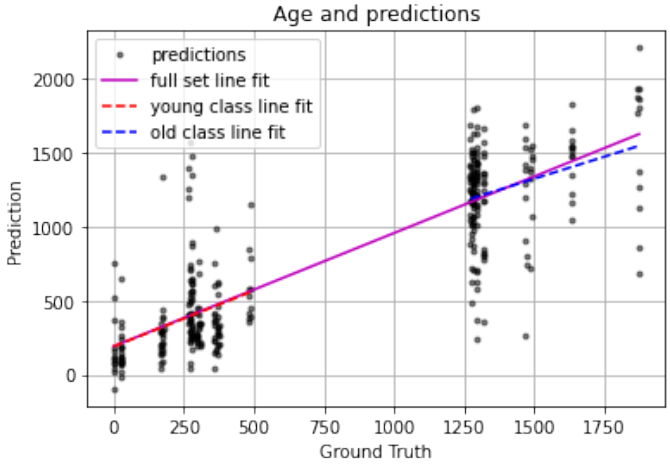

Figure 16: Fitted lines with ground truth age vs predictions on the balanced data set for; entire validation set (Purple= $191 + 0.767x$), young class (Red= $189 + 0.755x$), old class (Blue= $403 + 0.596x$). The regression can distinguish between ages within all age groups.

4.1.1 Confusion Matrices for Trial Experiment

Tables 1-4 show the confidence matrices for the classifications of the two networks and two data sets, along with *true positive/negative rates*, *positive/negative predictive value* and *accuracy*. The confusion matrices are very good with a high accuracy.

	Pred. Young	Pred. Old	True Pos/Neg Rate
True Young	**168**	2	0.988
True Old	18	**12**	0.400
Predictive Rate	0.903	0.857	Acc: 0.900

Table 1: Confusion Matrix for classification using the classification network on the *unbalanced data set*.

	Pred. Young	Pred. Old	True Pos/Neg Rate
True Young	**162**	8	0.953
True Old	17	**13**	0.433
Predictive Rate	0.905	0.619	Acc: 0.875

Table 2: Confusion matrix for classification using the *regression network* on the *unbalanced data set*.

	Pred. Young	Pred. Old	True Pos/Neg Rate
True Young	**161**	17	0.904
True Old	10	**156**	0.940
Predictive Rate	0.942	0.902	Acc: 0.922

Table 3: Confusion matrix for classification using the *classification network* on the *balanced data set*.

	Pred. Young	Pred. Old	True Pos/Neg Rate
True Young	**166**	12	0.933
True Old	16	**150**	0.904
Predictive Rate	0.912	0.926	Acc: 0.919

Table 4: Confusion matrix for classification using the *regression network* on the *balanced data set*.

4.2 Main Result

4.2.1 Multi-Variate Regression

Tables 5 and 6 show which features are used in the models created by the iterative model generation. The tables show all 8 models that were created with aspect to AIC vs BIC metric, Linear vs Logarithmic Biomass and top vs angled camera view.

Feature/Model	Lin-AIC	Lin-BIC	Log-AIC	Log-BIC
R	X	X	X	X
G	X	X	X	X
B				
R/G	X	X	X	X
G/B			X	
B/R	X	X	X	X
Hue_{RG}	X	X	X	X
Hue_{GB}	X		X	
Hue_{BR}				
HLRG	X	X	X	X
HLGB	X		X	
HLBR	X	X	X	X
VEG				

Table 5: Features included in the models for the top camera view.

Feature/Model	Lin-AIC	Lin-BIC	Log-AIC	Log-BIC
R	X			
G	X		X	X
B			X	X
R/G	X		X	
G/B	X	X	X	
B/R	X			X
Hue_{RG}	X	X	X	X
Hue_{GB}	X			
Hue_{BR}			X	
HLRG		X	X	X
HLGB			X	
HLBR	X	X	X	
VEG	X		X	

Table 6: Features included in the models for the angled camera view.

Table 7 shows the R^2 and R^2_{adj} values for each of the 8 created models. Bold numbers show the best model with respect to AIC vs BIC metric. In each case the model generated using the AIC metric outperformed the BIC model with

respect to both the R^2 and R^2_{adj} values. Therefore, the AIC versions of the models were passed on to be compared using the RMSE of the validation set.

View	Target	Measure	R^2	R^2_{adj}
Top	Biomass	AIC	**0.315**	**0.314**
Top	Biomass	BIC	0.314	0.312
Top	Log-Biomass	AIC	**0.298**	**0.296**
Top	Log-Biomass	BIC	0.294	0.293
Angled	Biomass	AIC	**0.344**	**0.343**
Angled	Biomass	BIC	0.334	0.333
Angled	Log-Biomass	AIC	**0.295**	**0.293**
Angled	Log-Biomass	BIC	0.281	0.280

Table 7: R^2 and R^2_{adj} of models generated though AIC and BIC iteration on the different biomass targets. The AIC was found to be superior in all cases.

For the top camera, the model that aimed to predict log-biomass performed better than the model for untransformed (linear) biomass. The RMSE for the log-biomass model had an RMSE of 0.1182g compared to 0.1374g for the untransformed biomass.

On the other hand, for the angled camera, the untransformed model performed better, with a RMSE of 0.1015g compared to 0.1098g for the Log-Biomass model. Therefore, the Log-Biomass model was selected for the top camera, while the untransformed biomass model was selected for the angled camera. The weights for the two models are shown in Tables 8 and 9 below.

Feature	Intercept	R	G	R/G	G/B	B/R
Coeff.	0.1427	-30.8498	29.9778	-4.4577	-0.4435	-4.3437
Feature	Hue_{RG}	Hue_{GB}	HL_{RG}	HL_{GB}	HL_{BR}	
Coeff.	0.8625	1.1628	-1.3134	-2.0212	6.2662	

Table 8: Top camera - MVR Model Coefficients. The highest effect comes from the red and green features.

Feature	Intercept	R	G	R/G	G/B	B/R
Coeff.	0.4567	-4.0938	3.9349	-0.8282	1.0791	-0.7736
Feature	Hue_{RG}	Hue_{GB}	HL_{BR}	VEG		
Coeff.	0.6988	0.4656	1.0448	-0.6336		

Table 9: Angled camera - MVR Model Coefficients. The highest effect comes from the red and green features.

4.2.2 Neural Network

Figures 17 and 18 show the training process for the NN trained on images from the top camera and angled camera respectively.

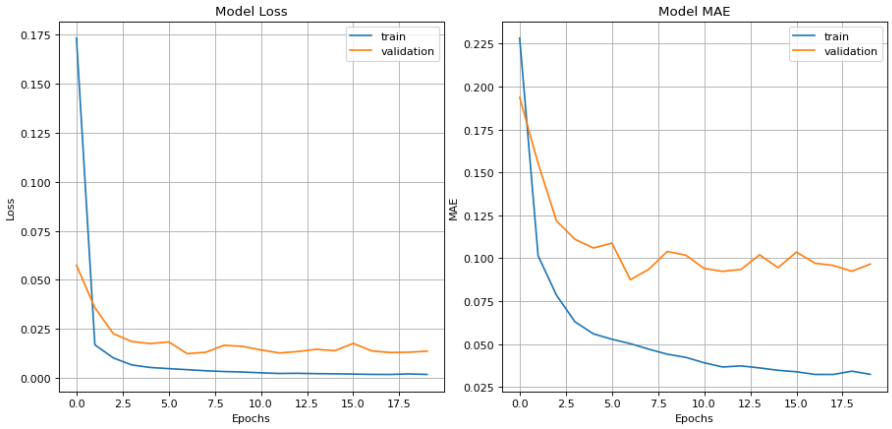

Figure 17: Training Process for the neural network using images from the top camera. The loss (MSE) and MAE are at an acceptable level.

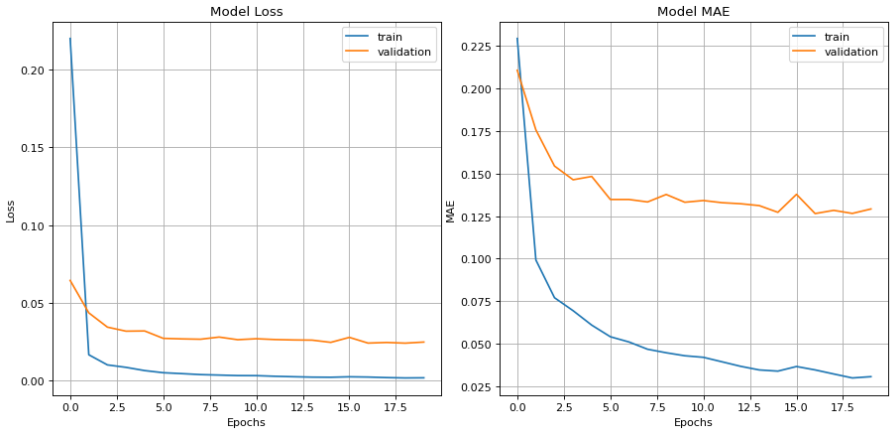

Figure 18: Training Process for the neural network using images from the angled camera. The loss (MSE) and MAE are at an acceptable level.

4.2.3 Comparison

Tables 11 and 10 show the RMSE on the test set for the MVR and NN respectively. These tables show the quality of the method on; Single image biomass prediction, Average of three chronologically consecutive biomass predictions, and RGR prediction using three random data points. Each of these tasks are performed using the model for the top camera, angled camera, and the average between their outputs (*dual view*). In addition, the RMSE is shown for the full test set and on the individual plants in the test set.

Table 10: RMSE on test set as a whole and for individual plants, when using multi-variate regression. This shows the results for all tasks (SI = *Single Image*, MA = *Moving Average*, RGR = *Relative Growth Rate*). The best method for a given task on each individual is highlighted in bold. The top view seems to perform best when predicting biomass, while the dual view performs the best for RGR.

Task - View	All	#24	#33	#42	#64	#65	#78
SI - Top	**0.0466**	**0.0316**	**0.0260**	**0.0448**	**0.0573**	**0.0643**	0.0447
SI - Angled	0.1227	0.0924	0.0640	0.1347	0.2173	0.0960	0.0585
SI - Dual	0.0734	0.0521	0.0406	0.0882	0.1151	0.0751	**0.0363**
MA - Top	**0.0391**	**0.0253**	**0.0200**	**0.0426**	**0.0382**	**0.0614**	0.0332
MA - Angled	0.1178	0.0875	0.0598	0.1336	0.2089	0.0922	0.0479
MA - Dual	0.0703	0.0491	0.0380	0.0875	0.1096	0.0729	**0.0287**
RGR - Top	0.2268	0.2088	0.2038	0.1911	0.2413	0.2573	0.2500
RGR - Angled	0.3669	0.2816	0.2684	0.2261	0.2889	0.5410	0.4788
RGR - Dual	**0.1984**	**0.2049**	**0.1736**	**0.1658**	**0.2019**	**0.2383**	**0.1974**

Table 11: RMSE on test set as a whole and for individual plants, when using a neural network. This shows the results for all tasks (SI = *Single Image*, MA = *Moving Average*, RGR = *Relative Growth Rate*). The best method for a given task on each individual is highlighted in bold. The dual view seems to perform the best for both biomass and RGR predictions.

Task - View	All	#24	#33	#42	#64	#65	#78
SI - Top	0.0862	0.0339	0.0301	0.0261	0.1833	**0.0553**	**0.0744**
SI - Angled	0.0653	0.0265	0.0332	0.0284	0.0823	0.0809	0.0992
SI - Dual	**0.0550**	**0.0260**	**0.0228**	**0.0233**	**0.0735**	0.0639	0.0840
MA - Top	0.0812	0.0291	0.0278	**0.0166**	0.1752	**0.0502**	**0.0689**
MA - Angled	0.0637	**0.0214**	0.0304	0.0271	0.0804	0.0792	0.0985
MA - Dual	**0.0523**	0.0230	**0.0204**	0.0205	**0.0669**	0.0620	0.0829
RGR - Top	0.2647	0.2195	0.1982	0.3260	0.2444	0.2483	0.3243
RGR - Angled	0.1837	0.1910	0.1490	**0.1314**	**0.1818**	0.2294	0.2022
RGR - Dual	**0.1767**	**0.1434**	**0.1427**	0.1687	0.1867	**0.2078**	**0.2000**

Figure 19 shows the resulting predictions of the two methods, MVR and NN, on the two tasks, *Single Image* and *Moving Average*. The predictions are plotted against the ground truth value for biomass in that image.

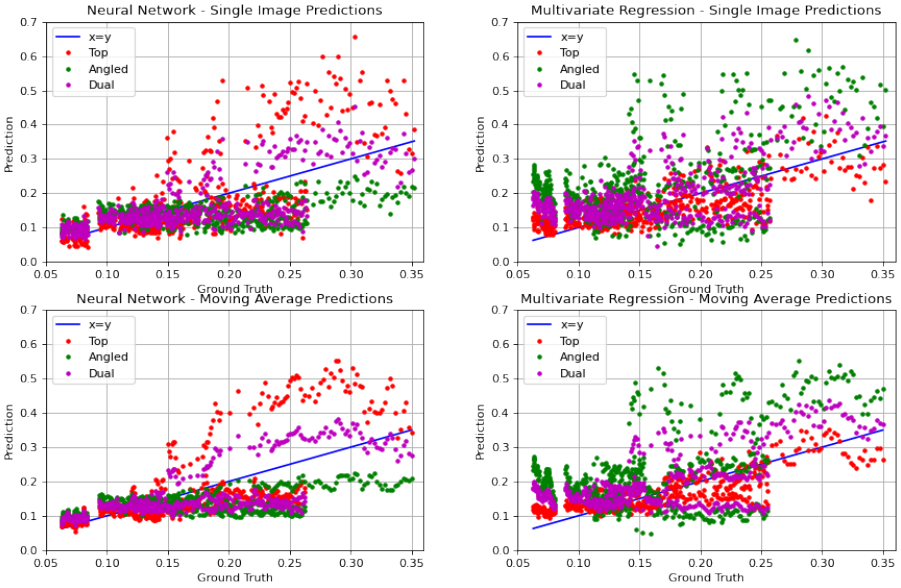

Figure 19: Predictions of the MVR and NN on the *Single Image* and *Moving Average* tasks. The blue line shows the ideal correlation, which the points follow to an acceptable degree.

5 Discussion

5.1 Trial Experiment

5.1.1 Classification Accuracy

The results of the trial experiment was very promising. The categorisation accuracy of the regression network was similar to the accuracy of the classification network, as shown in the confidence matrices in Tables 1-4. This result shows that estimating growth progress as a number instead of categorically keeps the precision high. The accuracy is a little lower when the classes are unbalanced, but doesn't fall too low. However, the *true old rate* is very low, which is expected since that class doesn't contain much training data.

Another interesting point is that the difference in accuracy between the two network types is much higher for the *unbalanced* data set. This could just come down to random chance, but it could also point to the regression network suffering from the *unbalance* more than the classification network. However, this cannot be determined for sure.

5.1.2 Training Progression and Loss

The training process was quite different between the regression and classification networks. The classification network was at a much higher risk of overfitting, so for the *unbalanced* data set the training only ran for 20 epochs instead of 50. Even for the larger *balanced* data set, the classification network was slightly overfit, as can be seen in the top-right Categorical Cross-Entropy graph in Figure 12, however, this didn't affect the accuracy too much. This indicates that a regression network both needs to, and can, be trained for a longer period. This increased complexity might come from the fact that the targets are binary in the classification network, while continuous in the regression network.

5.1.3 Regression within Age Classes

As shown in Figures 15 and 16, the regression network had a strong correlation between the real and predicted age, meaning that it has learned something. One could imagine that this comes from the two age classes being very separated, therefore, these figures also show a fitted line using only members of each class. When the age classes are *balanced* (Figure 16), we can see that the slope is quite high even within the classes. Therefore, the regression network has learned both to distinguish between the classes, but also learned what age means for a plant within the two classes. This shows us that the potential for predicting the biomass of a plant is promising.

However, with the *unbalanced* data set (Figure 15), the fitted line for the *Old* class has a negative correlation. This means that within the *Old* class, the network has not learned what an increase in age looks like. This is most likely due to the small training data available in this class, but it shows us the importance of having training data for the whole time period that we are interested in.

5.2 Method 1: Multi-variate Regression

The MVR models that were generated using the AIC measure performed better than the BIC models in every case, as seen in Table 7. In the book *Model Selection and Multimodel Inference: A practical information-theoretic approach* by K. P. Burnham and D. R. Anderson [21], a comparison between AIC and BIC is conducted. Here, they conclude that BIC has an advantage to AIC if the model used to generate the data set is included in the set of possible models. However, they note that since the real world is too complicated, "the primary foundations of the BIC criteria do not apply in the biological sciences and medicine and the other "noisy" sciences.".
This holds true for our case, since we can with confidence say that the biomass data was not generated using the pixel values of the image. This could be the reason that the AIC performs much better.

From the result, we see that the MVR performed better using log-transformed biomass for the top camera, while performing better with un-transformed biomass for the angled camera. The reason for this is unknown, and might come from the random variation since difference wasn't that large.
Furthermore, for the top camera, the RMSE was much lower on the test set compared to the validation set. The RMSE did not change considerably between the validation and test set for the angled camera. This could just be down to random chance since the images in the test set comes from just six individuals and is therefore highly correlated with itself. Which individuals are in the test and validation set therefore has a large impact on the performance. The large shift in the performance for the top camera between the sets could therefore indicate that the log-transformation is less stable, but it is difficult to draw such conclusion.

Overall, the MVR performs substantially better on the top camera compared to the angled camera, as can be seen in Table 10. This is true, not just on the set as a whole, but also for every individual in the test set. This makes sense since the top camera captures more of the leaf area and is therefore a better representation of the biomass. In addition, the nature of the angled camera causes parts of other plants be included in the images. Since the MVR has no way of distinguishing between these pixels, there is an inherent flaw with applying this method on the images from the angled camera.

In Table 10 we can also see that the **dual view** had a higher quality com-

pared to the angled view, but was worse than the top view. This makes sense since the dual view is just the average between these two estimates. Because the quality differs so much between the top and angled view, the benefit of reducing noise seems inconsequential compared to the averaging effect between the two camera views.

In every case, the *moving average* estimate had a lower RMSE than the corresponding RMSE for the *Single Image* task. This shows that there is a benefit to capturing multiple images on a short time span, even though the biomass increase is minimal. This benefit probably comes from the fact that the flickering of the LED lights were noticeable in the camera. The noise that comes from this likely leads to some variation in the estimates which is reduced by the moving average filter.

Similar to the biomass estimate, the RGR estimate was significantly better on the top camera compared to the angled camera. An interesting point is that, on an individual basis, a poor estimate of biomass did not necessarily lead to a poor RGR estimate. This could mean that some estimates for the individuals follow the slope of the curve but with an offset, while others stay close to the true value while not following the slope.

A very interesting point is that the estimate for the RGR is substantially increased for the *dual view*. In fact, this is true on an individual basis as well. This shows that there is a benefit to having two camera views when estimating the RGR, even if there isn't a benefit for estimating biomass.

5.3 Method 2: Neural Networks

For the NN method, the top camera performed worse than the angled camera, as seen in Tables 11. However, this only holds true when looking at the overall RMSE. If you consider the RMSE on the individual plants, the top camera performs better on four out of six plants. Plant #64 has a surprisingly large RMSE in the top view, which inflates the overall RMSE. The low number of individuals could be the reason for this, as this increases the influence of potential outliers. It cannot be said for sure if plant #64 is an outlier, but I would hesitate to state conclusively that the angled camera is preferable to the top camera.

As opposed to the result from the MVR, the *dual view* performed better than both the top and angled views. This indicates that having two perspectives is beneficial for the NN method. The reason for this could be that the difference in RMSE between the top and angled view is much smaller, making the variance-reducing effect of their average much clearer.

Just as for the MVR, the *moving average* estimate performed better on every view and on all individuals. This, again, shows the effect of averaging the estimates on a short time span.

Again, the RMSE for the *RGR* estimate did not correspond to the RMSE of the biomass estimate. For example, plant #64 in the top camera had an incredibly poor estimate for the biomass, but had the second best *RGR* estimate.

Another pattern that shows up in the NN result is that the *RGR* estimate was better when using the *dual view* compared to the top or angled view.

5.4 Method Comparison

Comparing the RMSE between the two methods shows that the MVR performed best on the top camera, while the NN performed the best on the angled camera, as can be seen in Tables 10 and 11.
As mentioned, the reason that the MVR has a poor result on the angled view could be that a smaller fraction of the image is plant, and that other plants can show up and distort the estimate.
Even if we exclude plant #64 from the RMSE of the top camera in the NN, it has a higher RMSE compared to the corresponding MVR model. Since the biomass is correlated with the size of the plant which is a highly visible quality, the conditions for a Regression model are good, while the relatively small amount of data could be making it hard for the network to learn.

When it comes to the *Moving Average*, the same patterns hold true, but with overall higher quality predictions. This result shows that a continuous image capturing method should result in higher accuracy of biomass estimation.

In both models, the best *Relative Growth Rate* estimate was made when using the dual view, showing the benefit of having multiple cameras for this purpose.

5.5 Prediction Quality

The quality of the result is comparable to previous studies, however comparison is difficult since most operate on different mass scales. Our MVR model performed the best, with an RMSE of around 0.04g using the moving average filter. With biomasses up to 0.35g this represents a relative RMSE of 11.4%. In the paper by Wenjian Liu et al. [14], they achieve an RMSE of around 0.32g on fresh biomass samples up to 3g, which corresponds to a relative RMSE of 10.7%. However other papers have achieved greater accuracy on larger data sets. The paper by N. Buxbaum et al. [10] used images of 3,888 individuals as their data set and got an RMSE of between 1-2g on biomasses up to 40g, corresponding to a relative RMSE of 2.5-5%.

The quality of the RGR estimates is quite poor however. As the average RGR in the test set was $0.1173 g/(g \cdot day)$, even the best RMSE of 0.1767 is larger than the average RGR. This means that the estimate is having a hard time detecting if the plants were growing or not. The reason for this could be

the large variance in the estimates of time-adjacent images, as the flicker of the LED-light created some noise between the images. Having a setup where the collected images are not affected by external factors should therefore improve the estimate of both the biomass and RGR.

While our data set contains many images, the number of individuals is quite small, and are taken over a short amount of time. Detecting relative changes in biomass could be easier for larger plants, as the visual difference is larger the further into the growth period the plants are. Further research should therefore be done on longer growth periods with more plants.

6 Summary

The highest quality biomass estimates were made using the moving average filter on the top camera with multi-variate regression. The biomass estimate was improved greatly in all models when applying the moving average filter over the neighbouring time points. Having multiple cameras did not improve the quality of the estimates for the multi-variate regression, but did do so for the neural network.

The highest quality *Relative Growth Rate* estimates were made using data from both cameras with a neural network. Using the estimates from both cameras improved this estimate for both the multi-variate regression and the neural network.

The quality of the biomass predictions is comparable to other studies, but leaves plenty of room for improvement. In addition, the RGR estimates are quite poor which indicates the need for larger data sets and longer growth periods where the plants have more time to visually change. In addition to this, the fact that the data set only contained two real measurements of the biomass makes the ground truth interpolation very susceptible to both measurement errors and deviations in the growth rate. Further more, by fitting the exponential function to only two points there is no way to estimate if the assumption of a constant growth rate holds true. Further analysis using data sets of more individuals over longer periods of time and using more real measurements is required to verify these results.

References

[1] *THE 17 GOALS — Sustainable Development*. URL: sdgs.un.org.

[2] R. Ziegler. "The Vertical Aeroponic Growing System". In: *Synergy International Inc.* (2005).

[3] H. Hedlund. "Temperature distribution and plant responses of birch (Betula pendula Roth.) at constant growth". In: *Acta Universitatis Agriculturae Sueciae Agraria* (1999).

[4] O. Hellgren and T. Ingestad. *Data base for birch plants at steady-state - Methods and performance of birch plants (Betula pendula roth) under non-limiting conditions and under limitation of nitrogen and light*. 1994.

[5] O. Hellgren and T. Ingestad. *Data base for tomato plants at steady-state - Methods and performance of tomato plants (Lycopersicon esculentum Mill cv Solentos) under non-limiting conditions and under limitation of nitrogen and light*. 1994.

[6] O. Hellgren and T. Ingestad. "A comparison between methods used to control nutrient supply". In: *Journal of Experimental Botany* 47.294 (1996), pp. 117–122.

[7] O. Hellgren and T. Ingestad. *Responses of birch (Betula pendula roth) and tomato plants (Lycopersicon esculentum Mill cv Solentos) to CO2 concentration and to limiting and non-limiting supply of CO2*. 1996.

[8] A. J. S. McDonald, T. Lohammar, and T. Ingestad. "Net assimilation rate and shoot area development in birch (Betula pendula Roth.) at different steady-state values of nutrition and photon flux density". In: *Trees* 6.1 (1992), pp. 1–6.

[9] D.-H. Jung et al. "Image Processing Methods for Measurement of Lettuce Fresh Weight". In: *Journal of Biosystems Engineering* 40.1 (2015), pp. 89–93.

[10] N. Buxbaum, J. H. Lieth, and M. Earles. "Non-destructive Plant Biomass Monitoring With High Spatio-Temporal Resolution via Proximal RGB-D Imagery and End-to-End Deep Learning". In: *Frontiers in Plant Science* 13 (2022).

[11] D. Gao et al. "Estimation of spectral responses and chlorophyll based on growth stage effects explored by machine learning methods". In: *The Crop Journal* 10.5 (2022), pp. 1292–1302.

[12] P. Kumar et al. "Estimation of winter wheat crop growth parameters using time series Sentinel-1A SAR data". In: *Geocarto International* 33 (Apr. 2017), pp. 1–24.

[13] M. A. Beck et al. "Presenting an extensive lab- and field-image dataset of crops and weeds for computer vision tasks in agriculture". In: (2021).

[14] W. Liu et al. “Estimation of Plant Height and Aboveground Biomass of Toona sinensis under Drought Stress Using RGB-D Imaging”. In: *Forests* 12.12 (2021), p. 1747.

[15] R. N. Lati, S. Filin, and H. Eizenberg. “Robust Methods for Measurement of Leaf-Cover Area and Biomass from Image Data”. In: *Weed Science* 59.2 (2011), pp. 276–284.

[16] W. Yang et al. “Greenness identification based on HSV decision tree”. In: *Information Processing in Agriculture* 2.3 (2015), pp. 149–160. ISSN: 2214-3173.

[17] T. Hague, N. D. Tillett, and H. Wheeler. “Automated Crop and Weed Monitoring in Widely Spaced Cereals”. In: *Precision Agriculture* 7 (Mar. 2006), pp. 21–32.

[18] *ImageNet Database*. URL: `https://www.image-net.org/`.

[19] J. Praveen Kumar and S. Domnic. “Image based leaf segmentation and counting in rosette plants”. In: *Information Processing in Agriculture* 6.2 (2019), pp. 233–246. ISSN: 2214-3173.

[20] K. He et al. “Deep Residual Learning for Image Recognition”. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 322.10 (2016), pp. 770–778.

[21] K.P. Burnham and D.R. Anderson. *Model selection and multimodel inference: a practical information-theoretic approach*. Springer Verlag, 2002.

Master's Theses in Mathematical Sciences 2022:E75
ISSN 1404-6342

LUTFMA-3492-2022

Mathematics
Centre for Mathematical Sciences
Lund University
Box 118, SE-221 00 Lund, Sweden

http://www.maths.lth.se/