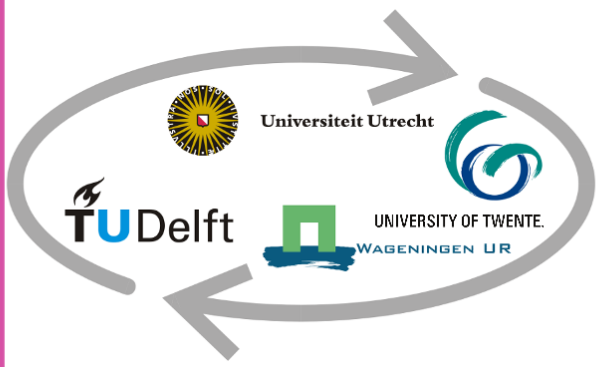
GIMA
Geographical Information Management and Applications

Thesis:

Automatic Spinach counting from UAV imagery using machine vision.

Name: Bilal Sari
E-mail: b.sari2@students.uu.nl

Supervisor: Dr. João Valente (WUR)
Responsible professor: Dr.ir. RJA (Ron) van Lammeren



Preface

After a long six months I am happy to present my Master's Thesis on Automatic spinach counting from UAV imagery using machine vision.

I would like to give special thanks to my supervisor Dr. João Valente for his incredible supervision and guidance throughout the whole thesis project. Without his devotion and involvement, this project would not have been possible. When I first contacted my supervisor João, during out first meeting he immediately started off by showing me the problem. He showed me an Ortho-mosaic of a spinach plant and asked me how I would approach counting every single individual plant in the image. From that moment on I knew that it was going to be a very fruitful 6 months.

In these six months, I have learned a lot of skills and new theory that is not typically part of a GIS curriculum, such as computer vision and machine learning. But I believe that through this type of multi-disciplinary research and application new and important insights can be shared and transferred between one field of study to the other.

I would also like to thank Dr.ir. RJA (Ron) van Lammeren for being the responsible professor and the GIMA Thesis coordinators ir. Edward Verbree and Dr. sci. nat. Frank O. Ostermann for their efforts during.

I hope you will enjoy reading my thesis,

Bilal Sari,

Amsterdam, 25-2-2019

Summary

Crop yield estimation has always been an important part of high precision agriculture. Knowing at any point in time how many of a farmer's crops are still healthy is key in optimizing labor and reducing waste in resources. This is not only beneficial for the farmers' own business but is also beneficial for the environment. With advancements in technology, more specifically Unmanned Aerial Vehicles getting better and cheaper it is a feasible option to deploy UAVs for data acquisition for agricultural purposes.

However data acquisition alone is not enough, interpretation of the data is just as important if not more important than raw data acquisition. Data needs to be converted into practical knowledge that can be further used by the farmers. One such practical knowledge is knowing how many plants a farmer has in his field at a specific point in time. However data interpretation by an expert can be time-consuming and costly. That is the reason that in this study the main goal is to develop a method to automatically count spinach plants by using machine vision.

In this study common machine vision image segmentation algorithms, such as the Excess Green Index and Otsu's method along with deep learning and convolutional neural networks will be used in order to create a fully automatic method of counting the number of plants.

This is achieved by segmenting and binarizing an ortho-mosaic of a spinach field. The result is a binary image, where all the true pixels represent pixels that belong to a spinach plant. By training a neural network to recognize individual spinach plants and classify them as such, the number of pixels per individual spinach plant can be automatically calculated. Afterward by diving the total amount of pixels by the average amount of pixels per plant the number of spinach plants can be calculated.

The outcome of this study is that the automatic algorithm performs is capable of taking an input image and returning the number of plants in that image. While there was no reliable ground-truth to validate the results of the used ortho-mosaics. Tests on smaller images where the plants could be counted by hand showed that the algorithm is capable of automatically counting the number of plants with an accuracy of 90%. The study also tested this approach on an ortho-mosaic of a smaller resolution and it still performs as expected. With the biggest error being 9.6% meaning that the algorithm is capable of counting plants from ortho-mosaics with different spatial resolutions.

Table of contents

1. Introduction

1.1 Research problem and context

Crop yield estimation and crop monitoring is a very valuable asset within agriculture. It is not only important for estimating the yield, weed control, disease detection. But it can also have a real impact on the economies of countries and the environment (Hayes & Decker, 1996) Improper crop monitoring can lead to a waste of valuable resources such as water and fertilizers.

Traditionally crop estimation and monitoring require manual labor, the field manager or the landowner still has to monitor the crops physically. This is time-consuming and can be prone to human error. Therefore there is a need in agriculture of more automation, that would not only take away the manual labor of land surveying but would also be more accurate, cheaper and more robust.

With recent advancements in technology, it becomes more feasible to solve these issues automatically and remotely. The increasing availability of the Unmanned Aerial Vehicle (UAV) is a potential solution to remotely and quickly acquire data on a plot of land without the manual labor that would be required traditionally(Rokhmana, 2015; Sarron et al., 2018). The land manager/owner does not have to survey the plot manually but can deploy a UAV in order to take aerial photographs from the crop that can be further analyzed.

The benefits of UAVs are that they can be flown at lower altitudes with greater safety than manned aircraft due to the absence of flying personnel, thereby increasing the resolution. UAV acquired data is also much more accurate than satellite imagery and can provide data very quickly, more than 500ha per day (Rokhmana, 2015). Deploying a UAV is therefore much more cost effective than a manned aircraft such as a survey plane or a helicopter. As Hunt et al. (2010) put it: "Low-cost, light-weight sensors are critical for the development of UAVs as a cost-effective platform for image acquisition" UAVs are already used in precision agriculture to improve profitability and productivity by providing data (Tokekar, Hook, Mulla, & Isler, 2016). However only image acquisition and remote sensing are not enough.

While using UAVs are a cheaper and faster way to collect aerial data, without a translation into information this data collection has very little added value. The land manager or owner has little benefit from aerial photographs without any translation to practical knowledge. However by applying machine vision, valuable information can be extracted from the photographs. For example, the yield can be automatically estimated by counting the number of plants or detect diseases by automatically tracking the growth of plants (Hunt et al., 2010). This type of analysis is often referred as Object-Based Image Analysis (OBIA) or in the case of georeferenced imagery Geographic Object-Based Image Analysis (GEOBIA) (Feng et al., 2015). The main difference between image segmentation or GEOBIA and more traditional GIS

applications such as Regionalization is the focus of the datasets. While Regionalization is more focused on Vector datasets image segmentation deals with raster-based imagery which is much easier to acquire. There are still other options in GIS to automatically classify raster images, such as performing clustering with the image classification tools like ArcGIS or Orfeo, among others. However type of unsupervised classification is often not as reliable for the identification of singular objects (Weih & Riggan, 2010).

While there are many different applications possible with image segmentation this research will focus on implementing machine vision in order to count the number of spinach automatically in a plot of land by using orthorectified aerial images acquired from UAVs. The choice for spinach is not only due to the availability of the data but also because spinach is highly consumed and produced crop in the Netherlands. In 2017 the Netherlands exported 46 million Euros worth of spinach (Statista, 2018). This research is part of a bigger project lead by Dr. Joao Valente of the Wageningen University & Research in the subject of Spinach Management. Very High Resolution (VHR) orthorectified aerial photographs of a spinach field was available. The dataset is made when the spinach plants were 10 weeks old at different resolutions.

1.2 Research objectives

The objective of this research is to develop an algorithm that automatically counts the number of spinach present in a plot of land from an ortho-mosaic built from aerial photographs acquired with a small quad-rotor.

In order to fulfill this goal these research questions have been made:

1. How can aerial photographs be automatically segmented in order to count the number of spinach plants by using machine vision?
2. How can the ground truth be calculated when there is no ground truth data?
3. How does the algorithm perform when using a dataset of a different spatial resolution?

2. Literature review:

This chapter will form the basis of theory that will be of importance in this thesis. This chapter will mainly focus on an exploration of the different segmentation algorithms and some general ideas and definitions of UAV and aerial photography will be given.

2.1 Unmanned Aerial Vehicles

An unmanned aerial vehicle (UAV), also known as Drone is an aircraft that is not operated by a human on board. A UAV can be either operated remotely or autonomously by using sensors(Al-Kaff, Martín, García, De La Escalera, & Armingol, 2017). UAVs were first developed and exclusively used for military purposes. With the developments in technology, the use of UAVs has become more and more popular among the research community. They are easier to operate, cheaper and smaller than regular aircrafts which make them a great substitute for aerial photography (Al-Kaff et al., 2017). Computer vision plays a big role in the current applications of UAVs. These applications can be as simple as aerial photography but can also be complex such as search and rescue missions. With the use of computer-vision UAVs can be used for a whole range of applications such as terrain mapping, exploration, and monitoring.

2.2 Aerial photography

Photo interpretation is an analytical tool and has an important value to research in the context of urban and landscape studies. The use of this method or technique is well-known and continue its growing (Gilliam, 1972). Another benefit of photos is that they are easy to interpret by humans. People with no expertise in a particular field are still able to interpret and analyze aerial photos because there is no need for special knowledge of photographic and photogrammetric processes. Aerial photos provide a basis for defining problems, are useful for knowing study areas, planning field trips for expeditions, mapping, as well as studying inaccessible areas. The photos are also of value as permanent records of continuously landscape changes in specific time and place.

The components of an aerial camera normally include a lens, inner cone, focal plane, outer cone, drive mechanism and magazine (Schenk & Quarter, 2005). There are many distinct configurations of cameras with different lenses, angles, focal plane distance, etc. In order to calculate the scale of the photos, one needs the focal length of the photograph (f) and the elevation difference between the flying height of the camera (H) and the height of the object above the datum (Figure 1) (Philpot and Philipson, 2012).

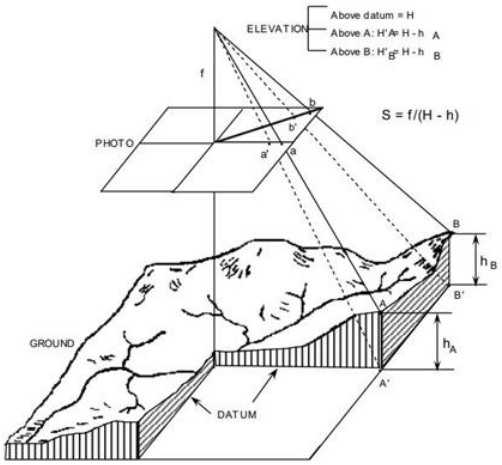


Figure 1: a schematic overview of the parameters and scale calculation of an aerial photograph (Philpot and Philipson, 2012).

2.3 Computer Vision Image segmentation algorithms for agricultural use.

A very important part of computer vision is being able to segment images into various meaningful parts that can be used for further analysis. There are many different types of algorithms for image segmentation. These algorithms can be broken down into one of three categories (Hamuda, Glavin, & Jones, 2016a; Romeo et al., 2013): Color Index based approach, Threshold-based approach, and Learning based approach. The main goal of many of these algorithms, in general, is to separate the plant material from the soil in the image. The result is that you often end up with two distinct classes that are either plant material or non-plant material.

The algorithms that have been selected for further explanations are based on the analysis of Hamuda et al. (2016). The selection has been based on the performance of the algorithms as evidenced by the survey that Hamuda et al. (2016) have done. The selected algorithms can be seen in figure 2. This table shows a summary of the strengths and weaknesses of each of the algorithms.

Algorithm	Type	Advantage	Disadvantage
NDI Woebbecke et al. (1993)	Color Index	-Easy to compute -Robustness to lighting	-Does not perform well when the light is very high or very low - A lot of false positives
ExG (Woebbecke, Meyer, Von Bargen, & Mortensen, 1995)	Color Index	-Easy to compute -Widely used -Low sensitivity to background errors and lighting conditions -Good adaptability for in- and outdoors	-Does not perform well when the light is very high or very low
ExR (Meyer, Hindman, & Laksmi, 1999).	Color Index	-Easy to compute -Segments soil texture	-Does not perform well when the light is very high or very low -Not as accurate as ExG
ExGR (Meyer, Neto, Jones, & Hindman, 2004)	Color Index	-Good adaptability for in- and outdoors -Can do both extracting green by ExG and eliminating noise by ExR	-Does not perform well when the light is very high or very low -Tends to segment shadow as plant
NGRDI (Hunt et al., 2010)	Color Index	-Reduces difference in exposure settings selected by digital cameras -Has two purposes: discriminates plants and soil and	-Does not perform well when the light is very high or very low -Limited use

		normalizes variations in light between different images	
MExG (Burgos-Artizzu, Ribeiro, Guijarro, & Pajares, 2011)	Color Index	-Good adaptability for in- and outdoors	-Does not perform well when the light is very high or very low
Ostu's method Otsu (1979)	Threshold-based	-Automatic method -Widely used	-Can produce under segmentation -Relatively Slow
Automatic Threshold (Kirk et al. 2009)	Threshold-based	-Good in handling light changes -Automatic method	-Longer computation time.

Figure 2: A table of the different Segmentation algorithms and their Advantages and Disadvantages (source: Hamuda et al., 2016)

Normalized difference index:

This algorithm has been developed by (Woebbecke, Meyer, Von Bargen, & Mortensen, 1993) and is a very classic segmentation approach. Woebbecke et al. (1993) used an index similar to the vegetation index that uses near-infrared and red light reflectance. The objective of this algorithm is to distinguish plant material from the soil in an RGB image. The algorithm can be expressed in the following formula:

$$\boldsymbol{NDI = 128 * \left(\left(\frac{(G - R)}{(G + R)} \right) \right) + 1}$$

Where G is the green pixel values, R is the red pixel values. However, the traditional NDI gives values ranging between -1 and 1. In order to convert these to RGB pixel values, the result is multiplied by 128 and added 1 to provide 256 gray scales. The resulting image is a near-binary image(David M. Woebbecke, Meyer, Von Bargen, & Mortensen, 1993).

Excess Green Index (ExG) :

The ExG is a simple algorithm that computes the amount of excess green in an image. This algorithm tries to separate the green from the bare soil. The ExG is a good choice for separating green plants from bare soil because it provides a good contrast between the plants and the soil. It also provides a near binary image (Woebbecke, Meyer, Von Bargen, & Mortensen, 1995). The Excess Green Index can be expressed as:

$$\boldsymbol{ExG = 2g - r - b}$$

where r, g, and b are the chromatic coordinates derived from:

$$r = \frac{R'}{(\mathrm{R}' + G' + B')} \quad g = \frac{G'}{(\mathrm{R}' + G' + B')} \; b = \frac{\mathrm{B}'}{(\mathrm{R}' + G' + B')}$$

R', G' and B' are the normalized RGB coordinates ranging from 0 to 1 and can be derived from:

$$R' = \frac{R}{\mathrm{Rmax}} \qquad G' = \frac{G}{\mathrm{Gmax}} \; B' = \frac{B}{\mathrm{Bmax}}$$

Where R,G,B are the actual pixel values and Rmax, Gmax and Bmax is the maximum value for the respective colors. (255 for 24-bits images).

Excess Red Index (ExR):

The excess red index is an alternation of the Excess green algorithm in which plant material is separated from the background. The separation of reds from the image was inspired by the fact that the human eye has more red cones in the retina than green and blue and therefore should yield better results when segmenting. However, the excess green algorithm outperforms this algorithm (Meyer et al., 1999).

The ExR can be expressed as:

$$ExR = 1.3 * \mathrm{R} - \mathrm{G}$$

Excess Green minus Excess Red Index (ExGR):

The ExGR is a combination of the ExG and the ExR, first used by (Meyer et al., 2004). The ExGR can be defined as follows:

$$ExGR = ExG - ExR$$

The objective of the ExGR is to isolate the plant material as well as to reduce the background noise that can be found in the excess reds.

Normalized Green–Red Difference Index (NGRDI):

The NGRDI is developed due to the fact that the Normalized Difference Vegetation Index (NDVI) cannot be used by digital cameras due to the fact that digital cameras often have filters that filter out near-infrared wavelengths (Hunt et al., 2010). This algorithm should also overcome the issue of differences in exposure time in digital cameras.

The NGRDI can be expressed as:

$$NGRDI = \frac{Green\ DN - Red\ DN}{Green\ DN + Red\ DN}$$

Where Green DN and Red DN are the digital values of the green and red bands of the image.

Modified Excess Green Index (MExG) Modified:

The MExG is a modified version of the Excess green method. The coefficients of the ExG have been optimized by using generic algorithm optimization and supposedly outperformed the original ExG (Burgos-Artizzu et al., 2011). The resulting coefficients are more robust during changing illumination conditions.

The MExG can be defined as:

$$MExG = 1.262G - 0.884R - 0.311B$$

Threshold-based approaches often solve the problem by reclassifying the image in two classes. The plant class and the soil class. Thresholding is often applied by transforming the original image in order to distinguish the desired classes (Hamuda, Glavin, & Jones, 2016b; Romeo et al., 2013). Selecting the right threshold is very important as a too high threshold will incorrectly classify plant pixels as non-plant and a too low threshold will incorrectly classify soil as plant pixels(Hamuda et al., 2016b).

Otsu's method:

Otsu's method was first proposed by Otsu (1979) and works by finding the threshold that minimizes the weighted "within class" variance. The first step is calculating the histogram and probabilities of each intensity level. The second step is setting up an initial weight and the initial class and the final step is to iterate through all possible thresholds until the threshold corresponds with the maximum "within class" variance (Otsu, 1979).

The maximum "within class" variance can be expressed as :

$$\sigma_w^2 = \omega_0(t)\sigma_0^2(t) + \omega_1(t)\sigma_1^2(t)$$

Where class probability $\omega_0(t)$ and $\omega_1(t)$ are calculated from:

$$\omega_0(t) = \sum_{i=0}^{t-1} p(i)$$

$$\omega_1(t) = \sum_{i=t}^{L-1} p(i)$$

Where L is the gray level, i is the pixel level.

Automatic Threshold selection (Kirk et al. 2009):

This algorithm introduced a new way for pixel classification of plant or soil by using the combination of the Red and Green pixel values. An automatic threshold was used, this threshold is based on the assumption that the distribution of the observed variables can be found by a mixture of two Gaussians with equal variances((Kirk, Andersen, Thomsen, Jørgensen, & Jørgensen, 2009).

The threshold d_t can be derived from the following formula:

$$d_t = \frac{2\ln\frac{p(s)}{p(v)}\sigma^2 + n_v^2 - n_2^2}{2(n_v - n_s)}$$

Where p(s) and P(v) can be derived from:

This algorithm can be expressed as:

$$p(d|s) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(d-n_s)^2}{2\sigma^2}} \qquad p(d|v) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(d-n_v)^2}{2\sigma^2}}$$

Where n_s and n_v are the soil and vegetation distributions respectively and σ^2 is the common variance.

Convolutional neural network

One of the learning-based approaches that is often used for image recognition is the convolutional neural network or CNN. Convolutional neural networks are a collection of high-performance classifiers with a large number of parameters that must be learned from training(Oquab, Bottou, Laptev, & Sivic, 2014). CNN image classification consists of two main steps: Feature Detection and Feature Classification. An input image is first put through a set of image filters, also called convolutional filters. Which each activates certain features of the image. After which the image goes through classification filters. That finally results in an output classification (MathWorks, n.d.)Figure 3 below shows a schematic overview of how this process looks like.

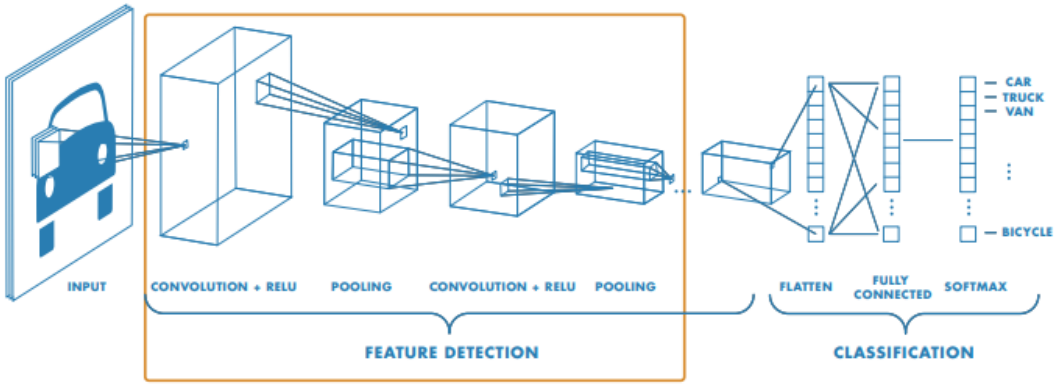


Figure 3: Schematic overview of Convolutional neural networks. (MathWorks, 2018)

Transfer learning

While a CNN can require a lot of training samples a pre-trained network can be used in order to overcome the challenges of limited resources. Transfer learning aims to transfer knowledge from a pre-trained network in order to repurpose this data to compensate for the lack of information that comes from a limited amount of training data (Oquab et al. 2014). With transfer learning, it is possible to retain a pre-trained networks knowledge while training it to a specific problem by providing training data of that specific problem (MathWorks, n.d.)

3. Methodology:

In this chapter, the used methods will be explained. First a flowchart of the whole process will be displayed, afterward the study area and the datasets will be explored and finally, each part of the methods will be explained. The flowchart in figure 4 shows the general outline of this research. It is a fairly linear process where the algorithm development is followed up by the validation and finally the sensitivity analysis will be performed afterward. Each block also represents a research question, once one of the blocks are finished the research question that goes with it can be answered.

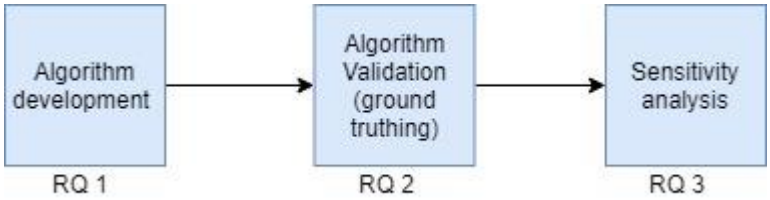


Figure 4: Flowchart of the general outline of this research

3.1 Study Area

The dataset used in this research is provided by Wageningen University and Research. The dataset consists of two TIFs one has a resolution of 16mm/pixel and the other of 8 mm/pixel. Both datasets have been captured by a UAV in June of 2018, with an average flying height of 40 and 20 meters respectively. Both datasets are of the same plot of land where spinach is being cultivated. The pictures have been taken when the spinach crops were approximately 10 weeks old. The original imagery that has been used to orthorectify the aerial photos into one TIF is also available. These pictures have been acquired by a drone of the brand DJI with serial number FC300X. The camera of this model has a focal length of 20mm.

The initial version is based on a plot of spinach field in the province of Flevoland, nearby Lelystad, in the Netherlands. This field has a size of approximately 30.500 m^2. The crops on the field are spinach that are approximately 10 weeks old. Figure 5 shows a map of the study area. Due to the confidentiality, the map's exact location and the coordinates have been left out.

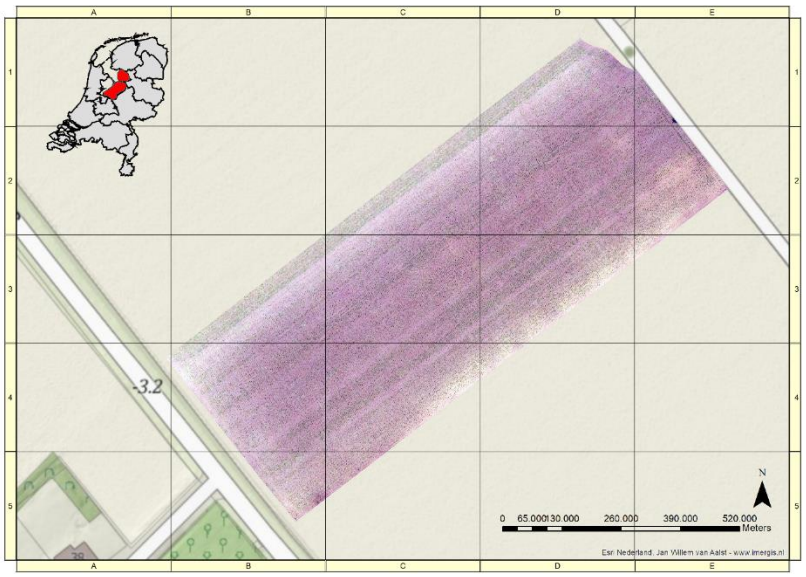


Figure 5: Map of the study area: The length and width of this field are approximately 285 meters long and 100 meters wide. This field has a size of approximately 30.500 m2 or about

3.2 Data explanation & Pre-processing

In this research, the data was provided by the courtesy of Wageningen University and Research. This data is a very high resolution ortho-mosaic. For the ortho-mosaics, there are two variant available: the 8mm resolution and the 16mm resolution version.

However, this high-resolution orthophoto is quite big in file size some pre-processing must be applied in order to make it manageable. The ortho-mosaic has to been tiled into several pieces in order to ensure smooth processing and preventing the computer from running out of physical memory. The tiles and the labels of these tiles can be seen in the figure below (figure 6):

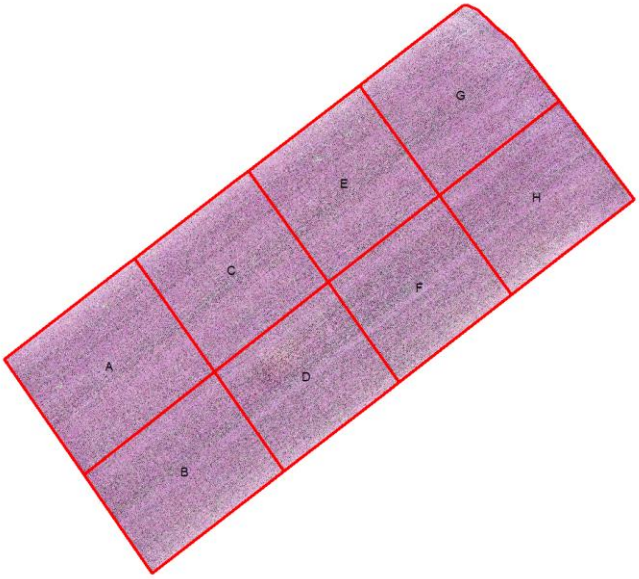


Figure 6: Tiling polygon used in order to create 8 pieces of the ortho-mosaic.

Each of these pieces will get an index A till H and will be processed separately. Piece E will be used to develop the algorithm due to its uniformity in regards to the number of plants on that piece of the picture. Figure 7 shows a close up of piece E and gives some more information on the characteristics of this piece.

Figure 7: Piece E of the ortho-mosaic

Figure 8: Zoomed in view of Piece E of the ortho-mosaic.

3.3 Ground Truth

In order to validate the algorithm, the accuracy of the resulting output needs to be calculated. This will be done by comparing the amount given by the algorithm to the ground truth. The ground truth refers to the information that has been collected in the study area. In this study, two types of ground truths are available: Amount of seeds and plant labeling from the ortho-mosaic.

Ground truth: Seeds

According to the experts that have planted the seeds, the field consists of 8-10 seeds per meter in each row. On average there are about 9 seeds per meter row. Each row is on a distance of 50 centimeters from the next row and has an average length of 283m (figure 9). The width of the field is approximately 100m. Which means there are 200 rows in the whole field and thus the entire dataset.

With this information, the number of seeds for the whole ortho-mosaic can be calculated. Each row contains on average 2547 seeds, which means there are a total of 509 400 seeds planted in the entirety of the field.

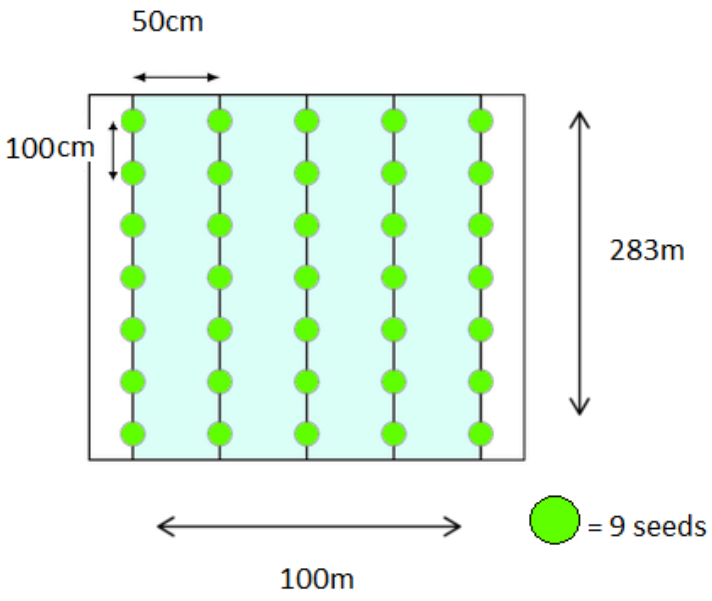


Figure 9: Schematic representation of the field information regarding the number of seeds planted.

Ground truth: Manual counting.

The second ground truth is the estimated ground truth that has been calculated by manually counting a small piece of the orthophoto completely.

For the calculation of the ground truth, a much an even smaller piece of the ortho-mosaic will be cut. This is because counting all the plants of the whole ortho-mosaic is a near impossibility, manually counting all the plants in a field this big is to labor intensive to consider. Figure 10 shows how this piece of image is made. By cutting Piece E of the original ortho-mosaic into a much smaller piece this image has been created. This piece also referred to as piece X, is shown in figure 11. This piece is small enough that manual counting is feasible and large enough to have enough plants to serve as a sample to perform accuracy tests.

Once Piece X is fully and manually counted, this information can be used to validate the algorithm, but also to estimate the ground truth of the other parts of the ortho-mosaic by extrapolating the results on a bigger area. This can be expressed as follows:

$$Ground\ Truth_i = \ Area_i * \frac{Plants_x}{Area_x}$$

Where $Ground\ Truth_i$ is the ground truth of the piece of image that the has to be calculated. $Area_i$, the area of that same piece of image. $Plants_x$ are the amount of plants in piece X and $Area_x$ the area of piece x

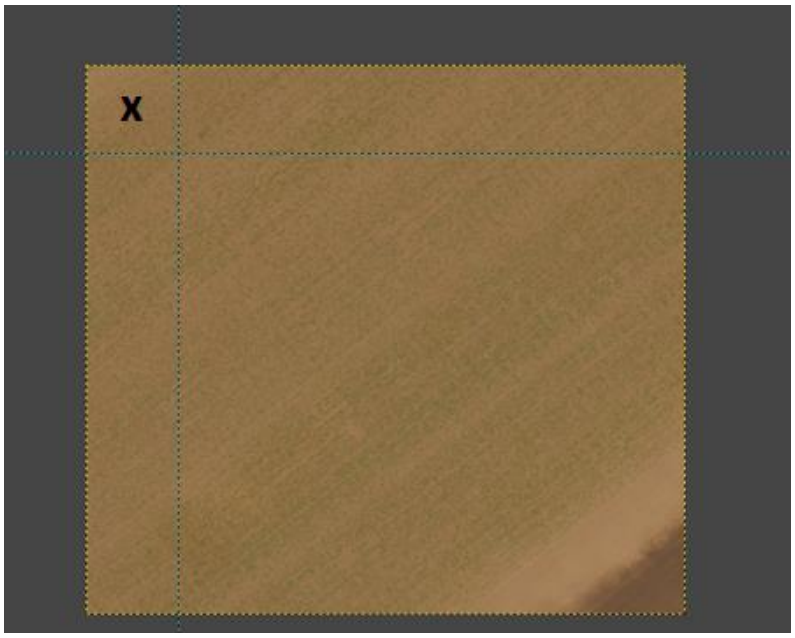


Figure 10: The lines that have been used to cut the image that created piece X.

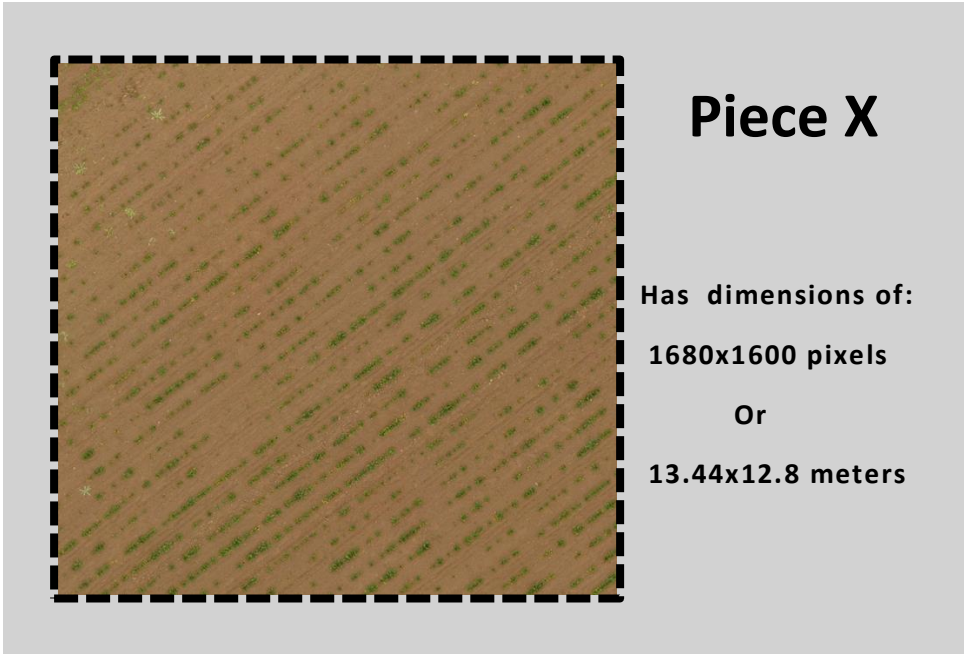


Figure 11: Full extent of piece X of the ortho

3.4 Automatic plant counting

In this study, an image segmentation approach will be used in order to calculate the number of plants in a given field. While six different image segmentation algorithms have been presented, in this research the combination of the ExG and Otsu's method has been selected as the best fit. This is partly based on the research results of Hamuda et al. (2016), in which the ExG + Otsu's method have an overall accuracy of 88%. The second reason for the choice of the ExG is the simplicity of the method. It is easy to understand and therefore also easy to develop with. The basic idea behind this approach is that the image will be converted into a binary image with the only pixels on that image being plant pixels. By summing the number of plant pixels there are and dividing this by the average amount of pixels that a spinach plant has, the amount of plants can be calculated. This approach can be seen in the following equation:

$$Amount\ Plants = \frac{Total\ amount\ of\ plant\ pixels}{average\ amount\ of\ pixel\ per\ plant}$$

While the total amount of plant pixels is a fairly straightforward procedure of applying the segmentation algorithm as described in the literature review. Computing the average amount of pixels per plant is a bit more complicated.

In order to calculate the average amount of pixels per plants, three approaches have been developed. The Automatic and the semi-automatic and the manual method.

In the manual method, the average amount of pixels per plant is calculated by manually labeling the UAV image. In this labeling, 200 individual plants are selected that are considered single individual plants. The average pixel size of these plants is then calculated.

In the semi-automatic method, the average amount of pixel per plant is calculated by performing a supervised classification. This classification is done in order to get a good representation of what an individual plant is.

And the fully automatic method uses transfer learning in combination with the pre-trained network of AlexNet to compute the average size of an individual plant.

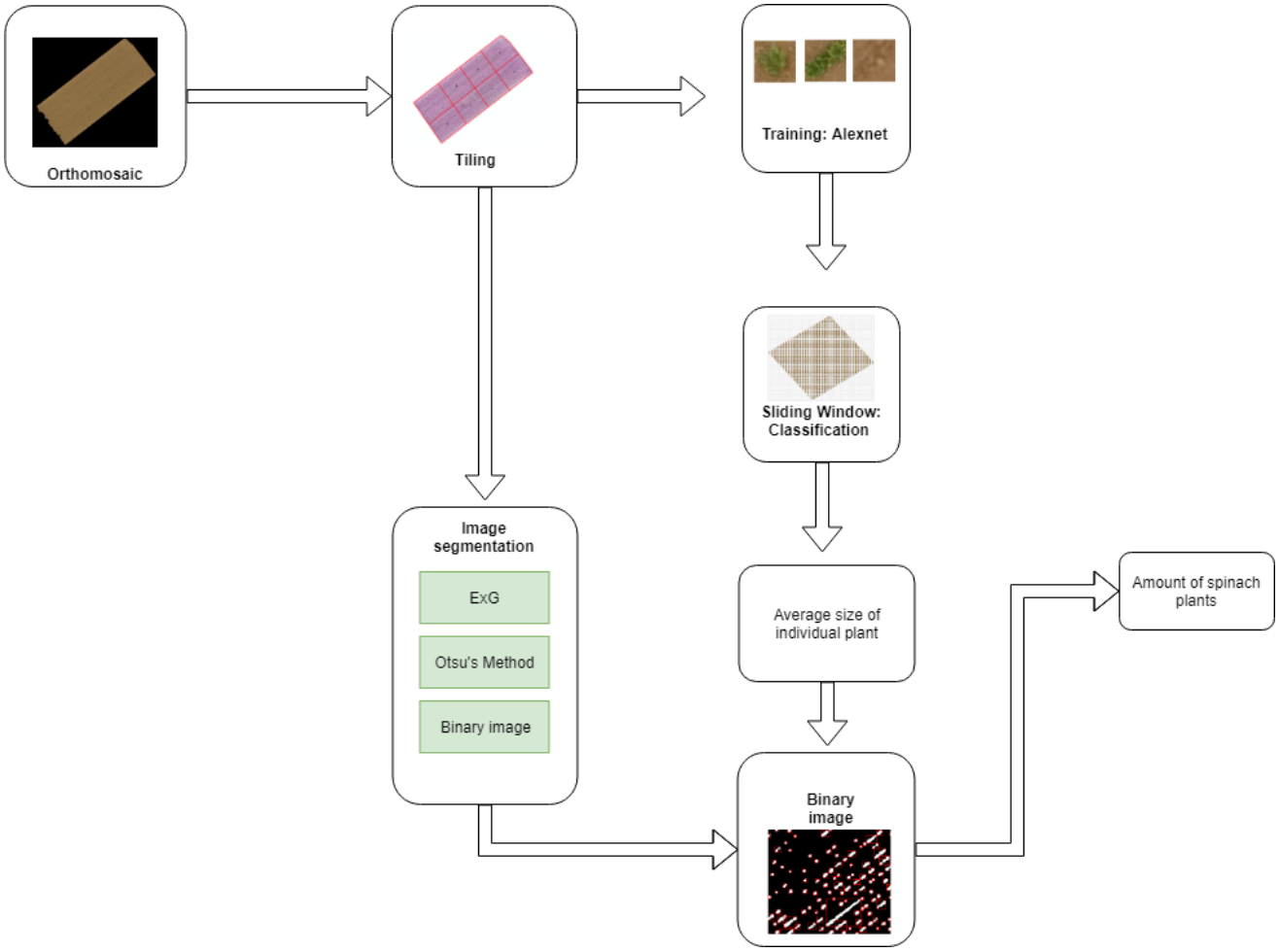


Figure 12: flowchart of the automatic plant counting algorithm

Figure 12 shows a flowchart with the outline of the automatic plant counting method. The blue squares indicate some sort of input data, this can be an image or a value and the green squares indicate a process and/or algorithm is applied.

Step 1 is importing the input image. This can be a TIFF or any other high-resolution orthophoto.

Step 2 is a tiling the input image into smaller tiles in order to ensure smooth processing.

The third step is the segmentation algorithm. Here the ExG and Otsu's method will be applied. The reason this algorithm has been chosen is due to its simplicity and effectiveness (Hamuda et al., 2016a). The image will be converted to a grayscale image and then converted to a binary image, with the green plants getting the value 1 and the rest getting the value 0.

Meanwhile, the input image will be processed by AlexNet and the individual plants will be identified. These will then be processed in the same manner and the amount average amount of pixels per individual plant will be computed.

Of this process, a semi-automatic and a manual method also exist. The semi-automatic method returns the histogram of each blob that has been found which then has to go through a supervised classification by iteratively selecting an upper and lower limit for the histogram and reviewing the output result. The goal here is to quickly find an upper and lower limit that will represent the pixel size of an average plant. Once the result is satisfactory the average of this new class is the average pixels per plant.

In the manual method, the input image first has to be manually sampled for individual plants. This is done by an image labeling program manually. Each individual plant has to be put in a bounding box that has to be drawn with the software. This will be done by using the labeling software LabelImg (Tzutalin, 2015), this is an open source tool that allows for easy labeling of pictures by bounding box and allows the labels to be exported to an XML format that can be read by other programs. Of these samples, the average amount of pixels will be calculated.

All three of these methods will result in an average number of pixels per plant. This number then is used to calculate the amount of image for the rest of the images. This is done by using the equation previously shown.

3.5 AlexNet training

In order for AlexNet to be able to detect individual plants, it has to be trained with a number of training samples. In this study, AlexNet has been trained to recognize three distinct classes. Individual Plants, Multiple Plants, and Background soil. The training set consisted of 200 images of individual plants, 130 images of multiple plants and 100 Images of background soil. However, of these training images, only 50% will be used to train the network the remaining 50% will be used to validate the network. Figure 13 shows one of each category of training samples.

After the training of AlexNet, in order to classify the bigger ortho-mosaic pieces some pre-processing needs to be done. AlexNet only accepts images of a specific size. The images have to be 227 by 227 pixels. On top of that individual images has to contain the plant, the image that has to be analyzed by AlexNet cannot be too different than the training data.

For example, if the Network has been trained with images of individual plants. It is impossible for AlexNet to classify an image that contains a whole row. So the ortho-mosaic has to be processed with a so-called sliding window. This means that the ortho-mosaic will be processed in small pieces of equal size, these windows will all be inputs for AlexNet which will then classify each small window as either an individual plant, multiple plants or simply background.

The individual plant images will then be further processed in the same manner as the other methods in order to compute the average amount of pixels per plant. This is done by automatically binarizing the individual plant images and counting the number of True pixels and calculating the average.

Figure 13: an example of the training data. Left is a single plant, the middle image is multiple plants and right is background soil.

3.6 Algorithm validation

In this stage of the research validity of the algorithm will be tested. This will be done in two ways, linear regression will be performed to see if the output results correlate with each other. The correlation coefficient will also be calculated in order to check for correlations.

The algorithm will also be tested by running it with a secondary ortho-mosaic of the same place with a different spatial resolution.

This will be done in order to check whether the algorithm still works and how this affects the overall accuracy. This is a fairly basic method of testing the algorithm but a crucial part of the research in order to ensure the usefulness of the algorithm.

The correlations will be calculated in order to test if the values that come out of the methods are random or not. If there is no correlation between the two different spatial resolutions that means that the methods are not reliable. Considering that the only difference between the two ortho-mosaic is nothing but the spatial resolution. More or less the same number of plants is expected because physically there are the same amount of plants present in the field.

4. Results:

In this chapter, the findings of various trials with the algorithm will be presented. First, the results of the ground truth will be discussed. Secondly, the results of the plant counting algorithm will be discussed.

4.1 Ground truth results

In order to be able to validate and make interpret the segmentation results into real numbers, some ground truth information is mandatory. As mentioned in the methods section for this study there are two types of ground truths: Amount of seeds planted and Manual counting.

Figure 14 shows the result of the manual counting, the left image shows the whole image with all the hand-drawn bounding boxes. The right image shows a zoomed in the part where these bounding boxes can be seen clearly. The resulting image contains 935 plants. This piece of the image has 1680x1600 pixels or 13.44x12.8 meters and an area of 172 m2. Based on these numbers the ground truths for all the pieces can be calculated with the formula described in the methods section.

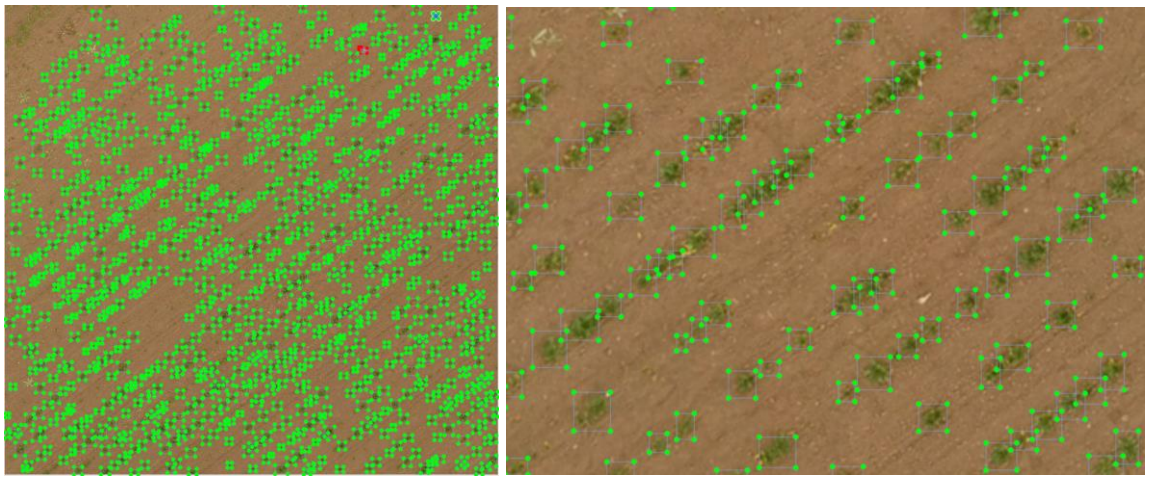


Figure 14: Example of a fully counted and labeled piece X. Contains 935 plants

	A	B	C	D	E	F	G	H
GT: Approximation	17696	15399	15307	13602	14715	13358	15351	14634
GT: Seeds	58662	51048	50742	45090	48780	44280	50886	48510
Area of piece (in m^2)	3259	2836	2819	2505	2710	2460	2827	2695

Figure 15: table of all the ground truths per ortho piece

4.2 Individual plant size: Manual

As mentioned in the methods. This method is the most reliable but also the most time-consuming method. This can take up an hour up to several hours of manual work in terms of drawing boxes in an image.

In this method, the Ortho-mosaic is being visually inspected and manually 200 individual plants will be sampled. In this case, the researcher labels plants that are clearly individual plants. The selected image for this method is one of the original aerial photographs. For this process Piece E as described in the methods sections has been used. While this method is more accurate the amount of

Figure 16 shows the labeled image. Each white box represents a box drawn around a plant that was deemed an individual plant. In total 200 individual plants have been identified. This data is then exported to an XML file and used to further analyze and compute the average area in pixels. Figure 17 shows the histogram of the areas of the labels that have been manually created. The mean pixels for all the labeled plants are 240 with a standard deviation of about 126.78px pixel. When converting this to centimeters the resulting size for an average plant is 644.26 cm^2 or approximately 23x28cm. Figure 18 shows each individual labeled plant cropped out to a separate image.

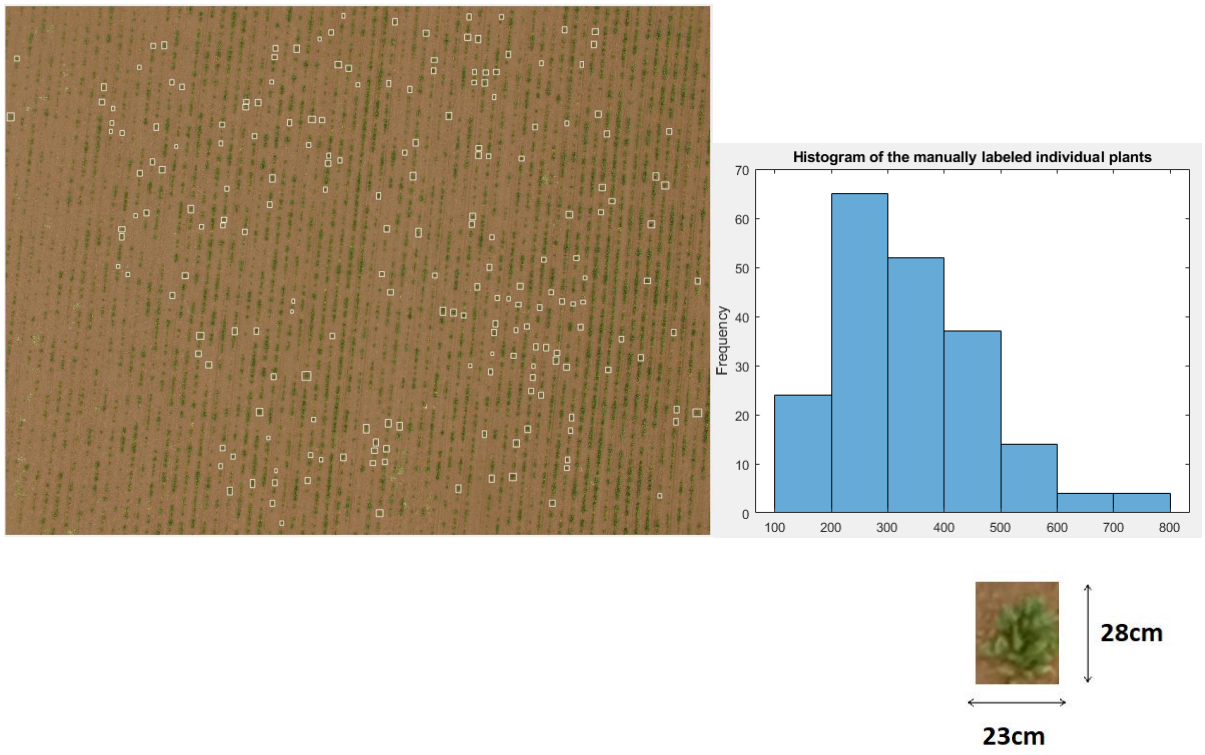


Figure 16: manually labeled aerial image. The top image shows all the labeled image. The bottom image shows an individual plant with the average dimensions in cm.

	px	cm
Mean in pixels	240	192
Standard Deviation	126.78	101.6
Sample size	200	

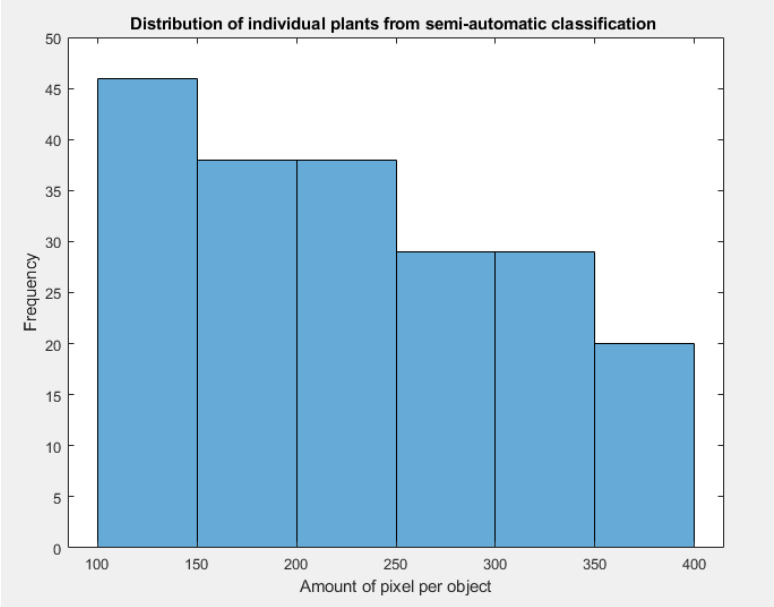
Figure 17: histogram of the manual single plant computation results

Figure 18 : Example of some cropped single plants obtained by manual labelling in the orthophoto.

4.3 Individual plant size: Semi-Automatic

The semi-automatic method has been performed as described in the methods.

By doing the supervised classification method as described in chapter 3.2, a class of pixel sizes has been determined in which individual plants are located. This class has an upper limit of 400 and a lower limit of 100 pixels. Using this range the labeled image has been filtered to only include this range and the histogram and its statistics have been calculated. Figure 19 shows the histogram of the piece E of the ortho-mosaic with the statistics. This distribution has a mean of 224 pixels and a standard deviation of 84 pixels. According to this distribution, the average individual plant is 224 pixels big. Figure 20 shows the labeled photo with the 200 automatically labeled plants.



	pixels	Cm
Mean in pix	224	179.6
Standard Deviation	84	67
Sample size	200	

Figure 19: Histogram of the range of individual plants. X-axis shows the number of pixels, Y-axis the frequency.

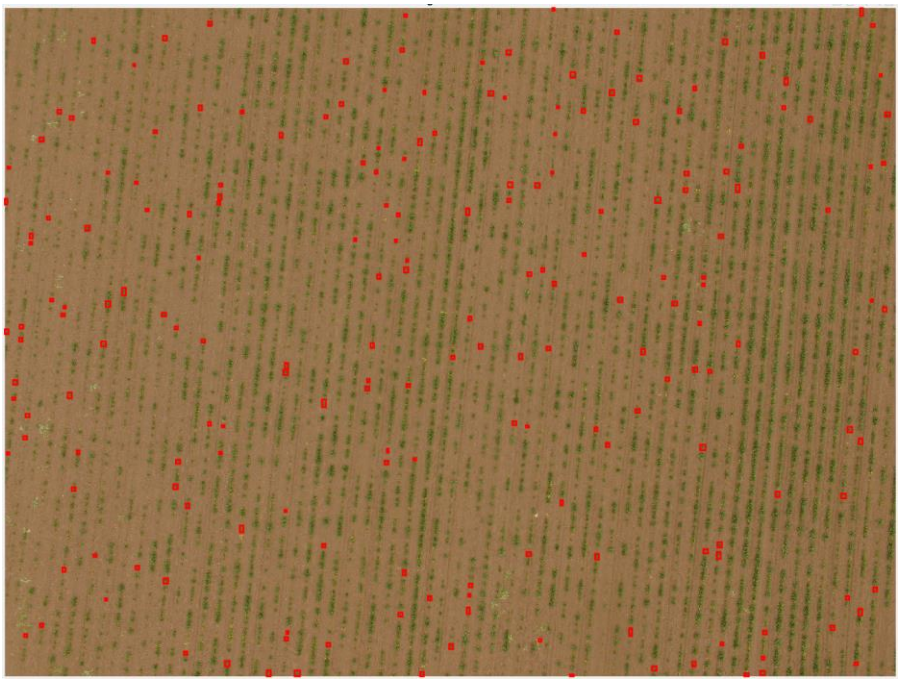


Figure 20: Figure 4.0: The automatically labeled image of Piece E of the ortho-mosaic. 200 random plants within the range of 100-400 pixels have been automatically labeled.

4.4 Individual plant size: Automatic

In this method, AlexNet has been used in order to find individual plants. The table in figure 21 shows the classification results of piece E. In the whole image AlexNet was capable of finding 64 individual plants. While it may seem a low amount, these are plants that are almost certainly individual plants. Figure 22 shows the individual plants found by AlexNet.

By applying the same segmentation as described in the methods these plants can be binarized and the average amount of pixels can be calculated. The result of doing so is 253 pixels. This means that the average amount of plants as calculated by the automatic method is 253.

Category	Amount
Single plant	64
Multiple plants	3145
background	25348

Figure 21: The classification results of piece E

Figure 22: The individual plants found by AlexNet.

4.2 Algorithm results:

The result of the segmentation and labeling can be seen in figure 21. The left image shows a zoomed-in version of the labeled binary image. The right image shows the labeling on the original RGB image. By comparing the two images it is possible to see that the algorithm works well in identifying the plants. However, the algorithm is not capable of distinguishing individual plants and therefore plants that grow very close to each other will be seen as one object. Each object in the binary image contains, therefore, one or more plants.

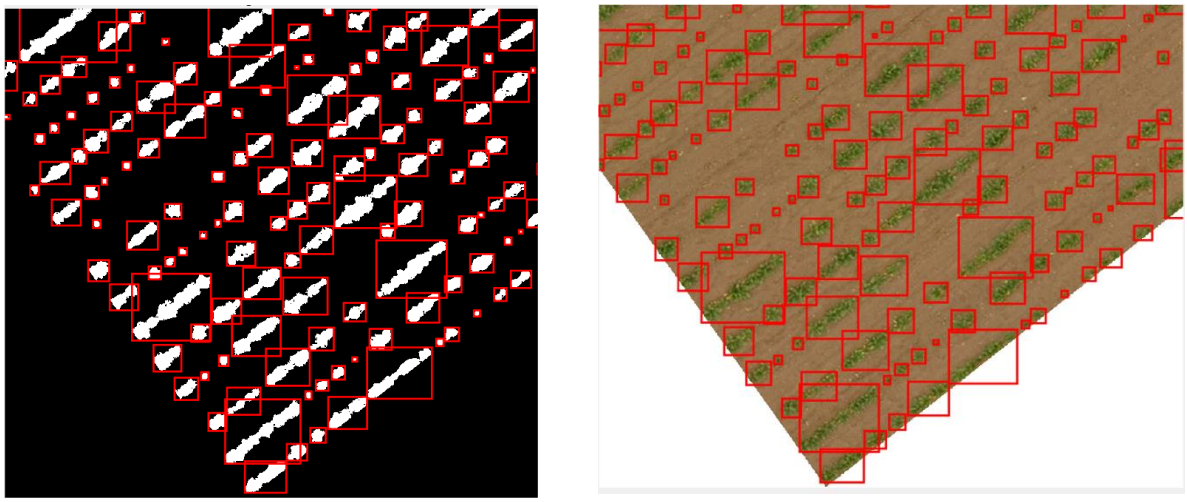


Figure 23: Results of the segmentation and labeling algorithm. Left shows the binary image and right shows the labeling on the RGB image.

This segmentation results in the total amount of white pixels. By dividing this number by the average pixel per plant the number of plants can be calculated:

$$Amount\ Plants = \frac{Sum\ of\ the\ amount\ of\ pixels\ per\ object}{average\ amount\ of\ pixel\ per\ plant}$$

Figure 24 shows the amounts of plants per ground truthing method for each of the pieces of the ortho-mosaic. The table shows for each of the pieces the number of plants that are a result of the semi-automatic method, manual method, full automatic method, ground truth: Manual and ground truth: seeds.

	A	B	C	D	E	F	G	H	Total
Manual Method	23310	24193	21119	24765	20930	24530	20314	21222	**180 382**
Semi-Auto	24975	25921	22627	26533	22425	26282	21765	22738	**193 267**
Full-Auto	22112	22950	20034	23492	19855	23269	19271	20131	**171 114**
GT: Approximation	17696	15399	15307	13602	14715	13358	15351	14634	**120 059**
GT: Seeds	58662	51048	50742	45090	48780	44280	50886	48510	**397 998**

Figure 24: the resulting amount of all the ortho-mosaic pieces with the number of plants along with the ground truths for the 8mm ortho-mosaic.

Figure 25 shows the errors between the different methods and the ground truths.

	A	B	C	D	E	F	G	H
Error Approx. – Manual (%)	31,7	57,1	38,0	82,1	42,2	83,6	32,3	31,7
Error Seeds - Manual (%)	60,3	52,6	58,4	45,1	57,1	44,6	60,1	60,3
Error Approx. -Semi Auto (%)	41,1	68,3	47,8	95,1	52,4	96,7	41,8	41,1
Error Seeds – Semi Auto(%)	57,4	49,2	55,4	41,2	54,0	40,6	57,2	57,4
Error Approx. – Full Auto (%)	25,0	49,0	30,9	72,7	34,9	74,2	25,5	25,0
Error Seeds – Full Auto (%)	62,3	55,0	60,5	47,9	59,3	47,5	62,1	62,3

Figure 25: Errors of the methods in percentages.

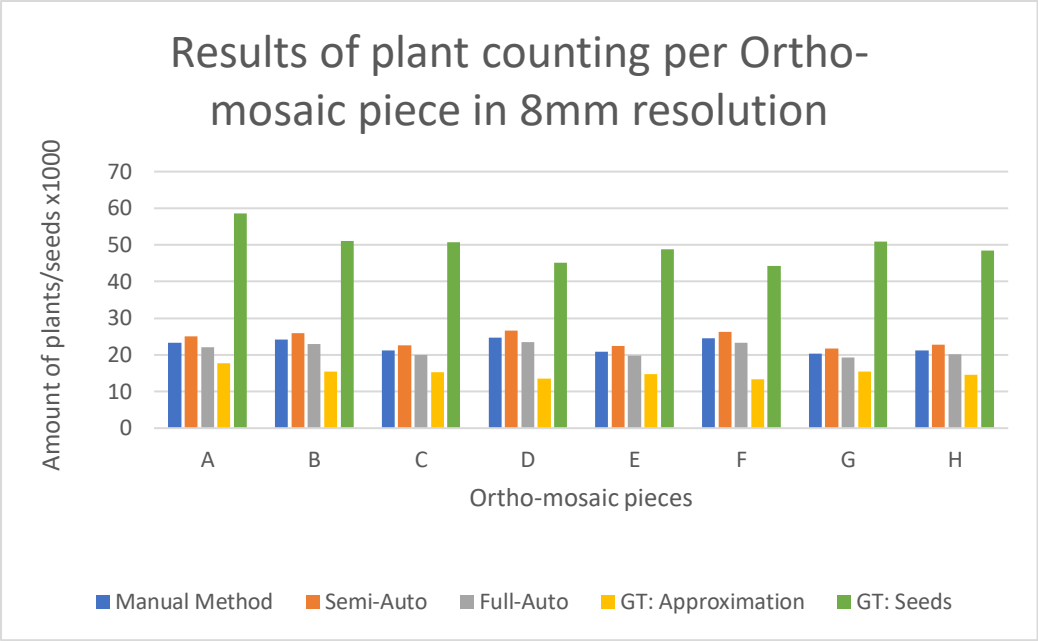


Figure 26: graph of all the ortho-mosaic pieces with the number of plants along with the ground truths.

5. Algorithm analysis

In order to make sure the results of the algorithm is not a result of pure randomness, a few tests will be done to prove this. There will be three major tests that will be done in order to shed more insight into the results these tests are: Running the algorithm on a piece with known parameters. Doing a statistical analysis of the results and finally running the algorithm on a different dataset with a different spatial resolution to ensure it also works on different datasets.

5.1 AlexNet training results

In order to evaluate how well AlexNet is trained the training results, as well as the testing results, will be discussed. The table in figure 27 shows the training results of AlexNet. This table shows along with how many iterations have been used to train AlexNet also the time elapsed during the training method. In total it took 1 minute and 7 seconds to train Alexnet with 3 classes and 150 iterations. However, this has been done with a fairly powerful GPU (Nvidia GTX 1060 3GB). The table also shows that after the 50th iteration the mini-batch accuracy reaches 100%.

Epoch	Iteration	Time Elapsed hh:mm:ss	Mini-batch Accuracy	Mini-batch Loss	Base Learning Rate
1	1	00:00:00	43.75%	1.0671	0.0010
10	50	00:00:21	100.00%	9.6754e-06	0.0010
20	100	00:00:44	100.00%	0.0002	0.0010
30	150	00:01:07	100.00%	3.4573e-06	0.0010

Figure 27: Table of the training results of Alexnet.

As mentioned in the methods section, 50% of the training sample has been used to test the accuracy of AlexNet after the training finished. The result of this is an overall accuracy of 0.98. Meaning all of the remaining images that AlexNet had for validation it was able to correctly categorize 98%. Figure 28 and 29 show the confusion matrix and a table of the Precision, Recall and F1-scores of this test. The F1-scores for this training result is 0.97, 0.98 and 0.97 for the Individual plants, background, and multiple plant classes respectively. This is an almost perfect f1-score meaning that the trained model is almost perfect in classifying these three classes.

Because the purpose of this classification is to get individual plants and calculate the average amount of pixels, the effect of AlexNet misclassifying an individual plant is minimal. Therefore false negatives, in this case, are not worrying. However false positives can affect the calculated average size of a plant, if the false positives are significantly larger than a single plant, the average size will also increase. But from the test results and the F1-scores, it's possible to conclude that the chance of this occurring is very minimal.

Figure 28: Confusion matrix of the three classes.

	Recall	Precision	F2-score
IndividualPlant	0,97	0,96	0,97
background	0,98	0,98	0,98
multplePlant	0,97	0,97	0,97

Figure 29: Recall, Precision and F2 scores.

5.2 Performance test

In the first test, small pieces of the image will be cut out and processed. The output results will then be compared to the number of seeds and the actual amount of plants counted by a human. In order to do this fives sets of five different pieces of the image have been cut out. These images consist of a single row with different lengths. The images have been cut into 1 meter, 2 meters, 4 meters, 5 meters, and 10 meters. There are in total of 25 images. Figure 30 shows one of each image.

Figure 30: The cropped row images.

These 25 images have been processed by the algorithm in the same way as the ortho-mosaic pieces. On top of that, the amount of plant in each image has been counted by a human. Figures 31 through 35 show the results per image size.

	Amount	Accuracy %	Amount	Accuracy %	Amount	Accuracy %	Amount	Accuracy %	Amount	Accuracy %
Manual	5,7	94,2	7,1	98,6	5,3	94,3	7,2	97,6	7,3	78,3
Semi	6,1	99,0	7,6	91,4	5,7	86,8	7,7	90,2	7,8	69,6
Auto	5,4	89,4	6,7	96,2	5,0	99,8	6,8	97,2	6,9	84,5
Count	6		7		5		7		6	
Seeds	9	66%	9	66%	9	55%	9	77%	9	66%

Figure 31: Results of 1-meter images.

	Amount	Accuracy %	Amount	Accuracy %	Amount	Accuracy %	Amount	Accuracy %	Amount	Accuracy %
Manual	11,2	88,0	13,7	85,8	15,5	44,9	11,2	93,4	13,2	98,2
Semi	12,0	80,0	14,7	77,6	16,6	33,8	12,0	100,0	14,2	90,9
Auto	10,6	93,8	13,0	91,7	14,7	52,9	10,6	88,6	12,6	96,6
Count	10		12		10		12		13	
Seeds	18	55%	18	66%	18	55%	18	66%	18	72%

Figure 32 Results of the 2-meter images.

	Amount	Accuracy %	Amount	Accuracy %	Amount	Accuracy %	Amount	Accuracy %	Amount	Accuracy %
Manual	33,3	85,1	28,2	99,3	26,0	86,9	30,3	95,4	22,3	98,5
Semi	35,7	76,8	30,2	92,1	27,9	78,8	32,5	87,9	23,9	91,3
Auto	31,6	91,0	26,7	95,5	24,7	92,7	28,8	99,3	21,2	96,2
Count	29		28		23		29		22	
Seeds	36	80%	36	77%	36	63%	36	80%	36	61%

Figure 33: Results of the 3-meter images.

	Amount	Accuracy %	Amount	Accuracy %	Amount	Accuracy %	Amount	Accuracy %	Amount	Accuracy %
Manual	31,5	87,5	39,3	87,7	35,6	98,4	34,6	95,2	35,7	95,0
Semi	33,8	79,4	42,1	79,7	38,1	91,1	37,1	87,7	38,3	87,5
Auto	29,9	93,3	37,3	93,5	33,7	96,4	32,8	99,4	33,9	99,6
Count	29		28		23		29		22	
Seeds	45	64%	45	62%	45	51%	45	64%	45	48%

Figure 34: Results of the 5m meter images.

	Amount	Accuracy %	Amount	Accuracy %	Amount	Accuracy %	Amount	Accuracy %	Amount	Accuracy %
Manual	73,2	96,9	50,4	98,8	78,6	87,7	77,8	96,3	69,5	97,9
Semi	78,4	89,5	54,0	94,1	84,2	79,6	83,3	88,9	74,5	95,1
Auto	69,5	97,8	47,8	93,7	74,6	93,4	73,8	98,4	65,9	92,9
Count	71		51		70		75		71	
Seeds	90	78%	90	56%	90	77%	90	93%	90	78%

Figure 35: Results of 10-meter images.

From these results, it is clear that the algorithm performs as expected. The manual and fully automatic methods both yield accuracies of over 90% consistently. However, the semi-automatic method performs the worst with an accuracy of 66% at best and 48% at worst. This can be explained by the fact that this method is very susceptible to human error during the supervised classification.

However, this test can be extended to piece X of the ortho-mosaic, considering the whole of piece X is also manually counted as well. The results of this test can be seen in figure 36. In this piece, the manual method has an accuracy of 90.8%. The semi-automatic has an accuracy of 72.0% and the fully automatic method an accuracy of 95.9%. These results are very much in line with the calculations of the single rows.

	Amount of plants	Accuracy
Manual	1028	90.8%
Semi-Auto	1299	72.0%
Auto	975	95.9%
Hand count	935	
GT Seeds	3096	30%

Figure 36: The accuracy numbers of Piece X of the ortho-mosaic

Maybe an even more important outcome of this analysis is that the number of seeds is a big overestimation in comparison to the actual amount of plants that are found in the different rows. The last row in figures 31 through 35 show the accuracy of the number of seeds in comparison to the amount of hand counted plants. This data shows that in the best case 80% of the seeds have germinated and grown into a plant but in the worst cases, about 48% of the seeds have become a plant. This loss of seeds is important because it renders the use of seeds a ground truth useless.

5.3 Sensitivity analysis

In order to check the sensitivity of the algorithm the same algorithm with the same methods has been applied to a secondary dataset of the same field at the same time, but with a different spatial resolution. The ortho-mosaic pieces, as well as piece X, will be compared with an ortho-mosaic and piece X of 16mm/pixel.

The results of this second run can be seen in figures 37 through 41.

	A	B	C	D	E	F	G	H	Total
Manual	24904	24562	21366	24240	21258	23901	22200	22875	**185 306**
Semi-Auto	27232	26785	23570	26770	23233	26593	24002	24930	**203 115**
Full-Auto	24151	23755	20903	23742	20605	23585	21287	22110	**180 138**
GT: Approximation	17696	15399	15307	13602	14715	13358	15351	14634	**120 059**
GT: Seeds	58662	51048	50742	45090	48780	44280	50886	48510	**397 998**

Figure 37: the resulting amount of all the ortho-mosaic pieces with the number of plants along with the ground truths for the 16mm ortho-mosaic.

	A	B	C	D	E	F	G	H
Error Approx. – Manual (%)	53,9	73,9	54,0	96,8	57,9	99,1	56,4	70,4
Error Seeds Manual (%)	53,6	47,5	53,5	40,6	52,4	39,9	52,8	48,6
Error Aprrox. -Semi (%)	36,5	54,3	36,6	74,5	40,0	76,6	38,7	51,1
Error Seeds – Semi (%)	58,8	53,5	58,8	47,3	57,8	46,7	58,2	54,4
Error Approx. – Full Auto (%)	40,7	59,5	39,6	78,2	44,5	78,9	44,6	56,3
Error Seeds – Full Auto (%)	57,5	51,9	57,9	46,2	56,4	46,0	56,4	52,8

Figure 38: Errors of the methods in percentages.

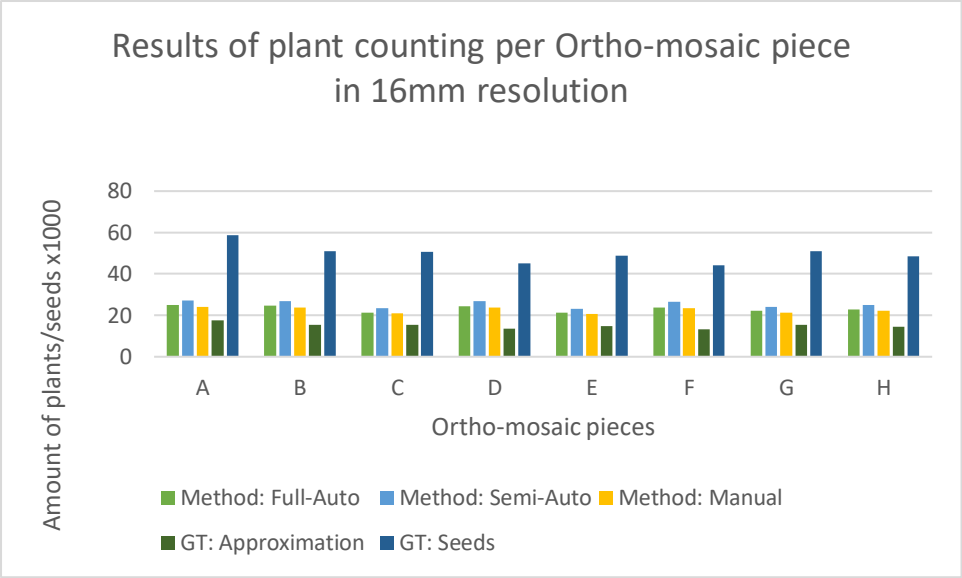


Figure 39: graph of all the ortho-mosaic pieces with the number of plants along with the ground truths for the 16mm ortho-mosaic.

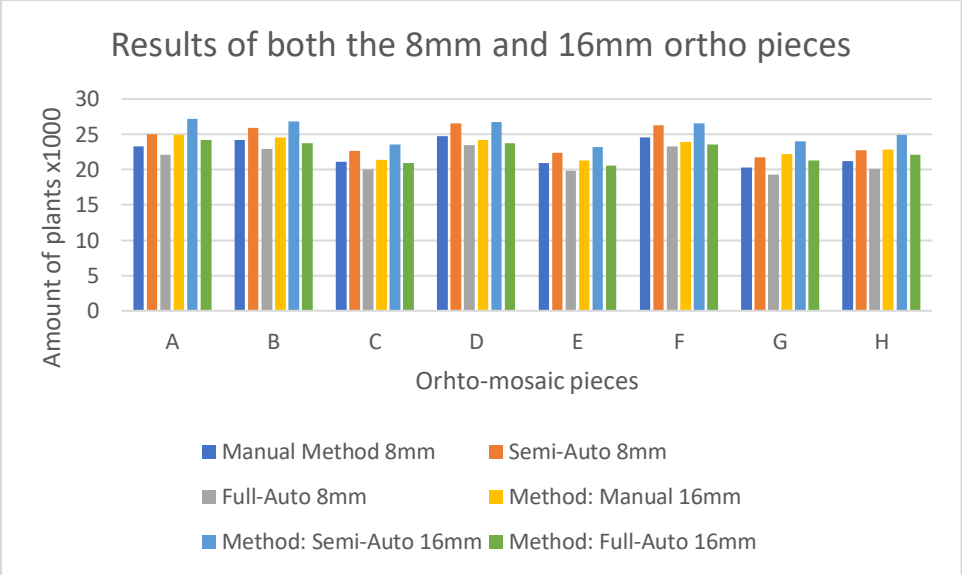


Figure 40: graph of all the ortho-mosaic pieces with the number of plants for both 8 and 16mm ortho-mosaics.

	A	B	C	D	E	F	G	H
Error (%) Manual 8mm-16mm	6,4	1,5	1,1	2,1	1,5	2,6	8,4	7,2
Error (%) Semi 8mm-16mm	8,2	3,2	3,9	0,8	3,4	1,1	9,3	8,7
Error (%) Auto 8mm-16mm	8,4	3,3	4,1	1,1	3,6	1,3	9,4	8,9

Figure 41: Error percents between 8mm and 16mm ortho-mosaic

When comparing the results of the 8mm to the 16mm ortho-mosaic pieces results, it is possible to conclude that the algorithm performs the same way in both datasets. In figure 34 both results are shown. The same patterns can be seen in both datasets. Figure 35 shows the errors between the two resolutions and from this can be concluded that the errors are relatively small with the biggest error being 9.4% and the smallest being 0.88%. This means that the algorithm is capable of counting the plants with input data that has half the spatial resolution.

Figure 41 shows piece X for the 16mm ortho-mosaic. By comparing the results of the 8mm piece X to the 16mm piece X it should also give an indication of how the algorithm performs on a piece of land with known parameters. Figure 42 shows the result of Piece X 16mm processed by the algorithm and figure 43 shows the errors between the 8mm Piece X and the 16mm piece X. The semi-automatic method, in this case, has the worst error of 23%, the manual and fully automatic methods perform much better with errors of only 13% and 9% respectively. This falls in line with the earlier findings where the algorithm performs with an error of approximately 10%.

Figure 41: Piece X 16mm version

	Amount of plants	Accuracy
Manual	904	96%
Semi-Auto	1001	93%
Auto	890	95%
Hand count	935	100%

Figure 42: Results of the algorithm for the 16mm version of PieceX

Method	Error
Error (%) Manual 8mm-16mm Piece X	13%
Error (%) Semi 8mm-16mm Piece X	29%
Error (%) Auto 8mm-16mm Piece X	9%

Figure 43: Error results between PieceX 8mm and

5.4 Correlation

In order to check if the results between the 8mm and 16mm are not a result of randomness, the Pearson's correlation coefficient will be calculated. The correlations between each method of a spatial resolution will be calculated. This means that the correlation coefficient for Manual Method 8mm and 16mm, Semi-Automatic method 8mm and 16mm and Automatic method 8mm and 16mm will be calculated along with the regression line.

Figures 44 through 46 show the results of the correlations. The manual, semi-automatic and automatic methods have a correlation coefficient of 0.846, 0.900 and 0.901 respectively. This means that the results between the different spatial resolutions have a high correlation. From this, it's possible to say that the output results from these methods are not just random numbers but are related to each other.

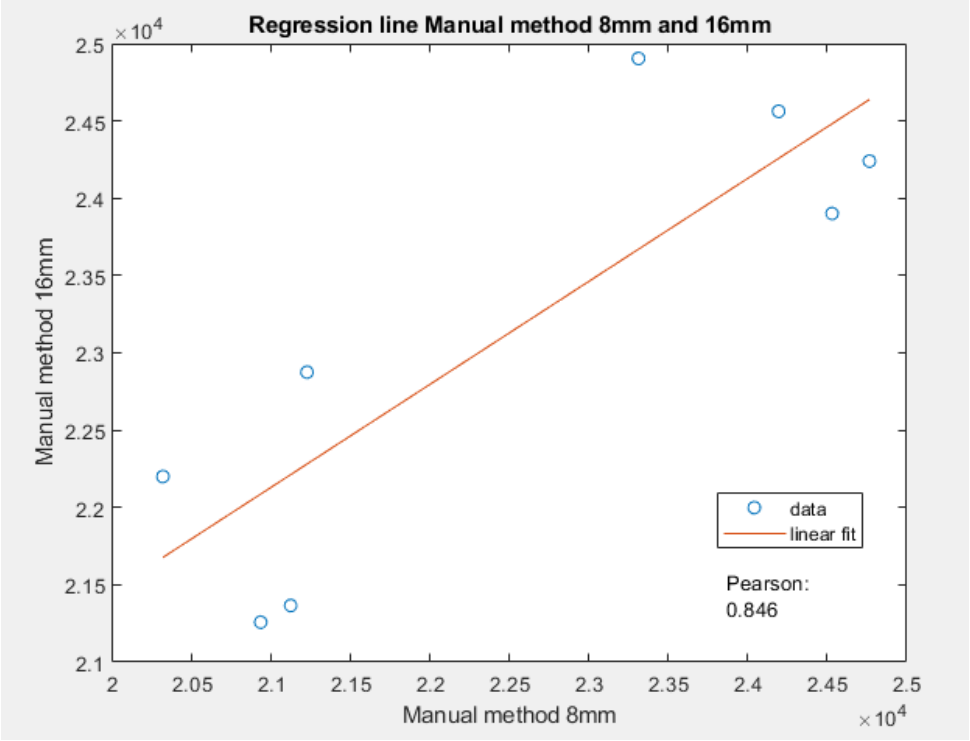


Figure 44: Regression line between Manual Methods 8mm and 16mm. With a correlation coefficient of 0.846

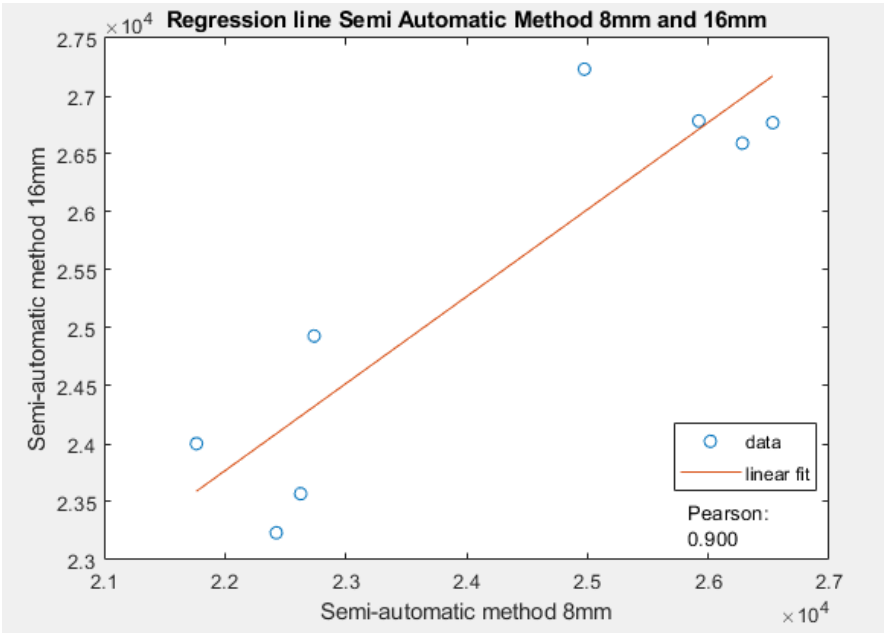


Figure 45: Regression line between Semi-automatic Methods 8mm and 16mm. With a correlation coefficient of 0.900

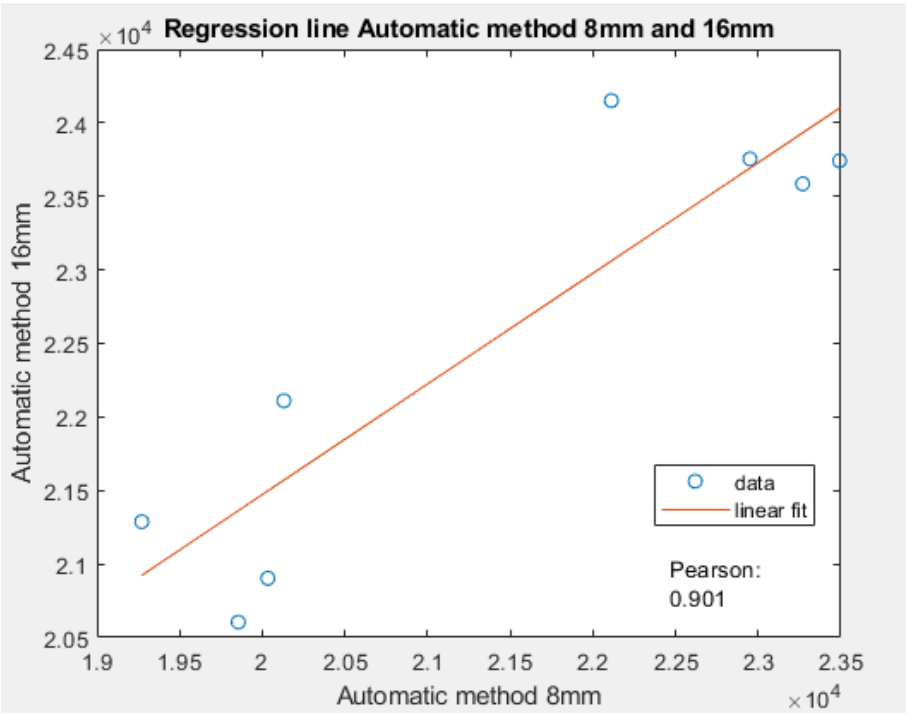


Figure 46: Regression line between Semi-automatic Methods 8mm and 16mm. With a correlation coefficient of 0.901

6. Discussion

In this section, the limitations of the research will be discussed. These can be either assumption that had to be made during the research or things that fell outside of the scope of the research that can be crucial to improving the methods used. Along with that some of the choices made will also be discussed.

6.1 Compared to other studies

Crop detection is a common application of machine vision. Being able to remotely distinguish plants from crops and the soil is a very valuable application for agriculture (Montalvo et al., 2012). Also, a lot of research has been done on how to detect crops and weeds. Most applications still rely on crop models that evaluate how a plant's responds to its genotype, environment and different cropping systems (Sarron et al., 2018).

Søgaard & Olsen (2003) attempted this by applying an RGB color transformation to turn the images into grayscale. And then applying a method to extract the green plant material.

Wheeler (2006) wanted to detect crop rows from an image and did this by transforming the original RGB to grayscale but afterward splitting the images into eight horizontal bands. These rows show a periodic variation in intensity due to the crop spacing, using the already known information about the camera and the fields, Wheeler could calculate the row spacing of each of the horizontal bands.

UAV imagery for crop detection is often used in combination with other remote sensing data. Senthilnath, et al. (2016) used a UAV to acquire aerial photographs of a tomato field and used spectral-spatial classification to classify the images in tomato and non-tomato.

Fontaine & Crow (2006) used the Blob method to search for areas within a grayscale picture of white pixels of a size equal to or greater than 200. This area, or blob, is then compared with the centerline and its center of gravity is compared to the whole picture. Then these blobs are categorized as plants.

Torres-Sánchez et al. (2015) used UAV acquired aerial imagery to segment herbaceous crops. In this research, they used various segmentation and thresholding algorithms, including Otsu's method, excess green (ExG) and the NDVI.

Sarron et al. (2018) used UAV captured imagery to use GEOBIA and combined this with a Digital Surface model in order to model the tree structure of mangoes and estimate the yield.

Ubbens et al. (2018) use deep learning to count the leaves of rosette plants. In this study Ubbens et al. (2018) use real data alongside a synthetic model of the plant in order to train a neural network, that is capable of counting on real plants.

Dobrescu, et al. (2017) use a similar method as Ubbens et al. (2018). In which they train a neural network to count the number of leaves of a rosette plant. However, in this study, they use data from real rosette plants and combine this with plants of other species.

Ribera et al. (2017) use deep learning in order to count crop plants from UAV images. In this study, Ribera et al. (2017) use regression to estimate the number of plants in an ortho-rectified UAV image.

However, research on Spinach plants (Spinacia oleracea) counting specifically is missing in recent literature. Most research on Spinach yield estimation is done on how the plant reacts to environmental factors and not on how spinach yield can be estimated. Therefore this research is one of the first researches that attempt to come up with an approach that fully automatically can count the number of spinach plants from UAV-imagery by using machine vision. While there are other studies that use deep learning to achieve similar goals, the major difference between this study and a study like Ribera et al. (2017) is that in this study deep learning is only used to compute the average size of a single spinach plant. While in the other approaches the neural network is doing the counting. The benefit of an approach like in this study is that the training samples can be very limited and still yield relatively good results.

6.2 Ground truth limitations

A very important metric in studies about yield estimation and machine vision is the ground truth. This number is crucial for validation purposes. In this study, the real ground truth was unfortunately absent. This is not just a matter of incomplete data but, especially for fields of this size having an accurate ground truth is an impossibility. This is not only very time consuming and resource demanding but often times also subjected to human error. Just manually counting a small piece of the image with about 950 plants took about 3 hours of manual work. Counting 8 pieces of the image with at least 20 000 plants in them accurately would have been an incredibly difficult and very time-consuming undertaking.

In substitute for this, there was the number of seeds planted by the landowners and estimation has been done by extrapolating the counting results of a small piece of land. While this gave some idea on how many plants there could have been it was far from a perfect metric. This can be seen back in the error numbers in figure 25. These numbers are relatively high and inconsistent between the pieces of the ortho-mosaic. This has two major reasons. The amount of seeds is a very high over-estimation of the number of actual plants that have successfully germinated and thus actually became plants. This can also be seen in the tests that have been done with a single linear meter of land. In these tests, the number of seeds is nearly double the number of plants that have been counted.

The other ground truth metric that has been used in this study is also not perfect. This is the result of an extrapolation of a small piece of land. But the problem with this is that the whole field is not one uniform whole but has a lot of difference between the pieces. Even within the same piece of land, there are a lot of differences. This could be mitigated by counting multiple

pieces of land evenly distributed all over the whole ortho-mosaic. But this is extremely time-consuming as just counting one piece took about three hours.

Because of these two unreliable ground truths, it is hard to say anything about the actual number of plants in the bigger ortho-mosaics. Even though the algorithm performs well, based on the tests done on small parts of the image. The only thing that can be said about the number of numbers calculated is that based on tests done on smaller, local images the algorithm has an accuracy of about 90%.

The most ideal situation would be redoing this study with a real-world ground truth instead of a calculated estimation or the number of seeds planted. However, unless its an experimental setting or a study with a lot of resources to allocate this metric will be missing, especially in a field of considerable size.

6.3 Counting algorithm limitations

In this study, the counting algorithm that has been used is a very simple approach that has some limitations. First of all, it assumes that all the plants in the field are spinaches. While this is true for the most part. It is inevitable that some unknown species of weeds can be in between the rows. The algorithm has no way of filtering this out. Secondly, the actual formula used to calculate the number of plants by diving the number of plant pixels by the average amount of pixels per plant has trouble dealing with closely growing plants. When two plants grow very close to each other there is bound to be overlap between the plants. This results in fewer plant pixels in the binarized image which in turn results into a lower amount total plant pixels in the image. Which then incorrectly results in a lower amount of plants than there would actually be in reality. This type of close vegetation also results in a difficulty counting for a human being as the overlap makes it hard to tell if a plant is one big plant or two very closely growing plants.

This approach also only works for spinach plants that have been planted in rows with no green vegetation in between the plants and in between the rows. If there were any other type of green vegetation in between the rows the algorithm would not be able to distinguish the spinach from the green vegetation and would render the approach useless. This could be solved by switching to a more learning based approach instead of a mostly image segmentation based approach

6.4 AlexNet limitations

While the usage of AlexNet in this study has been limited and only used to automatize the computation of the average amount of pixels per plant. The usage of AlexNet brings some limitations. The first limitation that there is, that is specific to this study is the relatively low, but still sufficient amount of training and validation data used.

But an even bigger limitation of AlexNet is the input requirements. AlexNet only takes images with a size of 227 by 227 pixels. When comparing this to the resolution of the ortho-mosaic this means that each image used as an AlexNet input has a real-world size of 1.81x1.81m with the 8mm resolution and 3.6x3.6m with the 16mm resolution. On these scales, it is impossible to get an image that only contains individual plants. So some clever image manipulation has to be done. In this study, the image has been processed by dividing it into small chunks of 50x50 pixels and 25x25 pixels for the 8mm and the 16mm resolution respectively. Afterward, each of these chunks has been resized to the appropriate input size. This approach works with the dataset used here, due to the very high resolution. But it would not work as well for lower resolution imagery due to the fact these input images would lose too much detail and become too blurry. Figure 47 shows an example of the original 227 by 227 ortho image alongside the resized image that is suitable for AlexNet to classify. Left is the original 8mm Ortho-mosaic image cut into 227 by 227. The middle is a 50 by 50 pixels image of the 8mm resized to 227 by 227 image. Right is a 25 by 25 pixels image of the 16mm ortho-mosaic and then resized to 227 by 227 pixels

In the most ideal case to have an ortho-mosaic in which 227x227 pixels correspond to a real-world size of approximately 40x40cm the spatial resolution of this ortho-mosaic would have to be approximately 0.7mm/pixel. Which is sub-millimeter level and currently not feasible with commercially available UAVs.

Figure 47: 227 by 227 pixels images. Left is original scale, the middle is 8mm Image magnified and right is 16mm image magnified.

7. Conclusion

This research has attempted to create a new method in which machine vision is applied to count the number of plants in UAV derived aerial imagery. While there were many other studies that have tried to estimate crop yield, not much had been done on the yield estimation of spinach plants. The goal of this research was therefore to create an algorithm capable of counting plants from aerial imagery.

When setting out to do this research three research questions were written:

1. *How can aerial photographs be automatically segmented in order to count the number of spinach plants by using machine vision?*
2. *How can the ground truth be calculated when there is no ground truth data?*
3. *How does the algorithm perform when using a dataset of a different spatial resolution?*

In this last section, these questions will be answered one by one based on the presented results in the previous sections.

7.1 Research question 1

The first question can be answered in many different ways, but the proposed method in this research is by using image segmentation techniques along with some statistical analysis. The general outline, as described in the methods section, is that the images should be segmented into a binary image. In this binary image, the true values in this image should represent the plant pixels and the false values anything else. This can be done by applying the ExG + Otsu's method to not only binarize the image but also to cluster each blob of True values to create objects that represent one or more plants. By counting the true pixels and dividing them by the average amount of pixels per plant the number of plants can be computed.

The difficulty in this method is not the segmentation or the clustering, but the computation of the average amount of pixels per plant. In this research, there are three methods on how the average amount of pixels per plant. However only one of them is fully automatic. In the so-called semi-automatic method, a supervised classification needs to be performed in order to come up with a class that represents the average amount of plants. In the manual method, sampling and labeling are used to get a sample of individual plants which then will be used to compute the average amount of pixels per plant. In the fully automatic method, AlexNet has been trained to automatically find individual plants that are used to calculate the average amount of pixels per plant.

By using the full-automatic method the total amount of plants in the study area is approximately 170 000 plants with an error of about 10%. Meaning that the algorithm is capable of getting 90% of the plants.

7.2 Research question 2

An even greater challenge can be determining the Ground Truth when no accurate field information is available. In this study, two types of ground truths have been calculated. One was an estimation that was derived through manual counting of a piece of the ortho-mosaic. And the second was using the number of seeds planted. However, in the end, it became apparent that the number of seeds used was not a good measure as a seed planted does not necessarily mean that a plant will grow per seed planted. In the validation part of the report, it can be seen that depending on the piece of land only 30% of the seeds actually grew into plants.

But to answer the research question, the most effective and accurate way of getting a ground truth from high-resolution aerial imagery is by manually counting it. But in cases, like these where there are too many plants to count effectively a second best option is to count a small sample and extrapolate the results to the rest of the ortho-mosaic. However, even this method has its limitations as seen by a large number of errors caused by this extrapolation method and heterogeneity of the crop in the field. It can be said that this is only effective for narrowing down the amounts, but for large areas, it is too unreliable to conclusively use it as ground truth.

7.3 Research question 3

An algorithm is only valuable if the results are correct and if it works on other datasets. For this reason, the algorithm has been tested on a different dataset with a different spatial resolution to ensure that it performs the same way for a different dataset. Both the 8mm ortho-mosaic as well as the 16mm ortho-mosaic both yield more or less the same results. Which makes a sense considering it is the same exact piece of land but only with a different spatial resolution. As seen as in figure 35 the errors between the two spatial resolutions are at worst 9%. Meaning that the algorithm works well for both 8 and 16 mm. However, as discussed in the AlexNet limitations the full-automatic method's performance decreases significantly when the spatial resolution gets smaller. This is due to the limitation of AlexNet only being able to use images of 227 by 227 pixels. While the algorithm still performed well with 16mm. It is hard to say if it would still work with an even lower resolution such as 32mm.

8. References

Al-Kaff, A., Martín, D., García, F., De La Escalera, A., & Armingol, J. M. (2017). Survey of computer vision algorithms and applications for unmanned aerial vehicles. *Expert Systems With Applications*, *92*, 447–463. https://doi.org/10.1016/j.eswa.2017.09.033

Burgos-Artizzu, X. P., Ribeiro, A., Guijarro, M., & Pajares, G. (2011). Real-time image processing for crop/weed discrimination in maize fields. *Computers and Electronics in Agriculture*, *75*(2), 337–346. https://doi.org/10.1016/J.COMPAG.2010.12.011

Dobrescu, A., Valerio Giuffrida, M., & Tsaftaris, S. A. (2017). Leveraging Multiple Datasets for Deep Leaf Counting. Retrieved from http://openaccess.thecvf.com/content_ICCV_2017_workshops/w29/html/Dobrescu_Leveraging_Multiple_Datasets_ICCV_2017_paper.html

Feng, Q., Liu, J., Gong, J., Feng, Q., Liu, J., & Gong, J. (2015). UAV Remote Sensing for Urban Vegetation Mapping Using Random Forest and Texture Analysis. *Remote Sensing*, *7*(1), 1074–1094. https://doi.org/10.3390/rs70101074

Gilliam, J. J. (1972). *Aerial photography and related products| Aids in expediting the construction and development of urban land-use maps*. Retrieved from https://scholarworks.umt.edu/etd/1482

Hamuda, E., Glavin, M., & Jones, E. (2016a). A survey of image processing techniques for plant extraction and segmentation in the field. *Computers and Electronics in Agriculture*, *125*, 184–199. https://doi.org/10.1016/j.compag.2016.04.024

Hamuda, E., Glavin, M., & Jones, E. (2016b). A survey of image processing techniques for plant extraction and segmentation in the field. *Computers and Electronics in Agriculture*, *125*, 184–199. https://doi.org/10.1016/J.COMPAG.2016.04.024

HAYES, M. J., & DECKER, W. L. (1996). Using NOAA AVHRR data to estimate maize production in the United States Corn Belt. *International Journal of Remote Sensing*, *17*(16), 3189–3200. https://doi.org/10.1080/01431169608949138

Hunt, E. R., Hively, W. D., Fujikawa, S., Linden, D., Daughtry, C. S., McCarty, G., … McCarty, G. W. (2010). Acquisition of NIR-Green-Blue Digital Photographs from Unmanned Aircraft for Crop Monitoring. *Remote Sensing*, *2*(1), 290–305. https://doi.org/10.3390/rs2010290

Kirk, K., Andersen, H. J., Thomsen, A. G., Jørgensen, J. R., & Jørgensen, R. N. (2009). Estimation of leaf area index in cereal crops using red–green images. *Biosystems Engineering*, *104*(3), 308–317. https://doi.org/10.1016/J.BIOSYSTEMSENG.2009.07.001

MathWorks. (n.d.). Introducing Deep Learning with MATLAB. Retrieved February 25, 2019, from https://nl.mathworks.com/campaigns/offers/deep-learning-with-matlab.html

Meyer, G. E., Hindman, T. W., & Laksmi, K. (1999). Machine vision detection parameters for plant species identification. In G. E. Meyer & J. A. DeShazer (Eds.) (Vol. 3543, pp. 327–335). International Society for Optics and Photonics. https://doi.org/10.1117/12.336896

Meyer, G. E., Neto, J. C., Jones, D. D., & Hindman, T. W. (2004). Intensified fuzzy clusters for classifying plant, soil, and residue regions of interest from color images. *Computers and Electronics in Agriculture*, *42*, 161–180. https://doi.org/10.1016/j.compag.2003.08.002

Montalvo, M., Pajares, G., Guerrero, J. M., Romeo, J., Guijarro, M., Ribeiro, A., … Cruz, J. M. (2012). *Automatic detection of crop rows in maize fields with high weeds pressure*.

https://doi.org/10.1016/j.eswa.2012.02.117

Oquab, M., Bottou, L., Laptev, I., & Sivic, J. (2014). Learning and Transferring Mid-Level Image Representations using Convolutional Neural Networks. Retrieved from http://openaccess.thecvf.com/content_cvpr_2014/html/Oquab_Learning_and_Transferring_2014_CVPR_paper.html

Otsu, N. (1979). A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, *9*(1), 62–66. https://doi.org/10.1109/TSMC.1979.4310076

Ribera, J., Chen, Y., Boomsma, C., & Delp, E. J. (2017). Counting plants using deep learning. In *2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP)* (pp. 1344–1348). IEEE. https://doi.org/10.1109/GlobalSIP.2017.8309180

Rokhmana, C. A. (2015). The Potential of UAV-based Remote Sensing for Supporting Precision Agriculture in Indonesia. *Procedia Environmental Sciences*, *24*, 245–253. https://doi.org/10.1016/J.PROENV.2015.03.032

Romeo, J., Pajares, G., Montalvo, M., Guerrero, J. M., Guijarro, M., & De La Cruz, J. M. (2013). A new Expert System for greenness identification in agricultural images. *Expert Systems With Applications*, *40*, 2275–2286. https://doi.org/10.1016/j.eswa.2012.10.033

Sarron, J., Malézieux, É., Sané, C., Faye, É., Sarron, J., Malézieux, É., … Faye, É. (2018). Mango Yield Mapping at the Orchard Scale Based on Tree Structure and Land Cover Assessed by UAV. *Remote Sensing*, *10*(12), 1900. https://doi.org/10.3390/rs10121900

Schenk, T., & Quarter, A. (2005). *Introduction to Photogrammetry*. Retrieved from https://pdfs.semanticscholar.org/7ed2/25c0799608539512fd84597892a5eb03e0b3.pdf

Senthilnath, J., Dokania, A., Kandukuri, M., Anand, G., & Omkar, S. (2016). Special Issue: Robotic Agriculture Detection of tomatoes using spectral-spatial methods in remotely sensed RGB images captured by UAV. https://doi.org/10.1016/j.biosystemseng.2015.12.003

Søgaard, H., & Olsen, H. (2003). *Determination of crop rows by image analysis without segmentation*. *Computers and Electronics in Agriculture* (Vol. 38). Retrieved from www.elsevier.com/locate/compag

Statista. (n.d.). • Netherlands: value import and export spinach 2008-2017 | Statistic. Retrieved December 4, 2018, from https://www.statista.com/statistics/576119/value-of-the-import-and-export-of-spinach-in-the-netherlands/

Tokekar, P., Hook, J. Vander, Mulla, D., & Isler, V. (2016). Sensor Planning for a Symbiotic UAV and UGV System for Precision Agriculture. *IEEE Transactions on Robotics*, *32*(6), 1498–1511. https://doi.org/10.1109/TRO.2016.2603528

Torres-Sánchez, J., López-Granados, F., & Peña, J. M. (2015). An automatic object-based method for optimal thresholding in UAV images: Application for vegetation detection in herbaceous crops. *Computers and Electronics in Agriculture*, *114*, 43–52. https://doi.org/10.1016/J.COMPAG.2015.03.019

Tzutalin. LabelImg. Git code (2015). https://github.com/tzutalin/labelImg

Ubbens, J., Cieslak, M., Prusinkiewicz, P., & Stavness, I. (2018). The use of plant models in deep learning: an application to leaf counting in rosette plants. *Plant Methods*, *14*(1), 6. https://doi.org/10.1186/s13007-018-0273-z

Weih, R. C., & Riggan, N. D. (2010). *The International Archives of the Photogrammetry*. *Remote Sensing and Spatial Information Sciences*. Retrieved from

http://depenv.ugent.be/geobia/proceedings/papers proceedings/Weih_81_Object_Based_Classification_vs_Pixel_Based_Classification_Comparitive_Importance_of_Multi_Resolution_Imagery.pdf

Woebbecke, D. M., Meyer, G. E., K. Von Bargen, K. Von, & Mortensen, D. A. (1995). Color Indices for Weed Identification Under Various Soil, Residue, and Lighting Conditions. *Transactions of the ASAE*, *38*(1), 259–269. https://doi.org/10.13031/2013.27838

Woebbecke, D. M., Meyer, G. E., Von Bargen, K., & Mortensen, D. A. (1993). Plant species identification, size, and enumeration using machine vision techniques on near-binary images. In J. A. DeShazer & G. E. Meyer (Eds.) (Vol. 1836, pp. 208–219). International Society for Optics and Photonics. https://doi.org/10.1117/12.144030