Viability of Avocado Yield Quantification using Machine Vision and Machine Learning

A thesis presented in partial fulfilment of the requirements for the degree of Master of Engineering at Massey University, Manawatū, New Zealand.

KYLE MACADAM

2022

Abstract

The avocado is the green celebrity of the fruit world with NZ exports more than doubling since 2006. Quantitative crop estimation is a critical piece of information in a horticultural supply chain, as it heavily dictates planning of harvest timing, labour, and marketing. Like all fruit crops, the avocado industry has a need to estimate their harvest volume each season. And so arises the question many growers ask; how can they accurately estimate their harvest? However, factors like the close colour matching of the fruit with the canopy, and the large canopy of the production trees, create some challenges for crop estimation. This work provides an insight into the current methods that have been explored and what direction growers should move towards, to increase their ability to estimate crop load. Research solutions options were rated using a weighting matrix that considers accuracy, availability, cost, and relevance in order to justify technologies that are most suited to further development. A novel labelled avocado dataset was used to train a convolutional neural network called YOLOv5. The trained YOLOv5 model showed promising results, with a real time detection accuracy of 88%, and a model mAP of 86.9% at 0.6 IoU threshold when trained at an image size of 512x512 pixels. Yolov5 predicted fruit per tree counts were within 5% compared to the harvest fruit per tree count.

Acknowledgements

I would like to express my deepest appreciation to my immediate supervisors, PhD Gabe Redding and PhD Sebastian Rivera Smith, for the many patient hours they invested into this project. Their encouraging guidance and professionalism will remain with me far beyond the bounds of this thesis.

I would also like to extend my deepest gratitude to Massey AgriFood Digital Labs (MAFDL) for providing a cultured, lively and supportive work environment. Their facilities and people were outstanding. Special thanks to Prof. Johan Potgieter for offering the chance to work within MAFDL. I also recognise PhD Andrew East of MAFDL, who provided the opportunity for this project and administered it. Alongside this, I would like to thank Andrew Vivian of OVĀVO for organising with Callaghan Innovation to provide the funding of this project.

I am extremely grateful to the staff of King Avocado, especially Claudia Madrid, the orchard manager, who provided an incredibly hospitable experience on visits to gather data on her orchard. Claudia has an extensive knowledge and passion for Avocado fruit and this insight shaped the direction of this work.

I am also grateful to Damian Hansen and Peter Jeffery who volunteered to help collect data on the King Avocado Orchard. Without their diligent help, this work might not have happened.

Thanks also to my friends and colleagues for their support and help during my time at MAFDL, namely Daniel Foster, Chris Hamling, Nitin Bhatia, Heath Ascott-Evans, Stefan Carter, Ruishu Cao, Carlos Lopez-Lozano and Raquel Lozano.

I cannot begin to thank my family who reared me, ensured my wellbeing throughout my life and encouraged me throughout the completion of my studies.

I am deeply grateful to Katianne Balmer, my beautiful and caring partner. Your encouragement, support, and understanding during the onerous stages of this Master's project, shall never be forgotten.

Finally, I am indebted to my Father for His counsel, direction and education, without which, I would be a very different person.

List of Abbreviations

AOV	Angle of View
AP	Average Precision
CNN	Convolutional Neural Network
FN	False Negative
FP	False Positive
FR-CNN	Faster Region Based Convolution Neural Network
GC	Google Colab
HD	High Definition
LiDAR	Light Detection and Ranging
MAE	Mean Absolute Error
mAP	Mean average Precision
MD	Mean of Difference
MP	Mega Pixel
NZ	New Zealand
OGR	orchard gate returns
R-CNN	Region Based Convolution Neural Network
RGB	Red, Green, Blue
RMSE	Root Mean Square Error
SS	Sample Size
STD	Standard Deviation
TP	True Positive
UAV	Unmanned Aerial Vehicle
YOLO	‘You Only Look Once’ software
YOLOv3	‘You Only Look Once’ software - version three
YOLOv5	‘You Only Look Once’ software - version five

Table of Contents

List of Figures

List of Tables

CHAPTER 1
INTRODUCTION AND BACKGROUND

1.1 Introduction

Persea americana, more commonly known as the avocado has been around since ~ 5000 B.C. with evidence of cultivated seeds found in various archaeological explorations in Peru and Mexico [1]. In 1871, the avocado began its journey of cultivation from Mexico to California thanks to Judge R.B. Ord. Back then, the major market contributor was Fuerte, taking up two thirds of the market [2]. In 1919, the first tree was introduced into New Zealand. Around the 1970s the Hass avocado, the variant most consumers are familiar with, established dominance in the market. Despite this, avocados had limited appeal to consumers until their rise to fame within the American market in the early 1990s. This was when Mexico decided to market the product at the American Superbowl, promoting guacamole as the perfect dip for any chip [3]. Gwyneth Paltrow helped boom the market in 2013 with her 'avocado on toast', featuring as a key component in her cookbook. The book surged the 'clean eating' movement around the world as people started to recognise the avocado as a superfood packed full of nutrients and healthy fats.

The increase in demand led to a sharp rise in price and New Zealand growers began to expand their operations to keep up supply levels [4]. By 2016, the avocado had reached celebrity status in New Zealand, encouraging many to venture into growing the profitable fruit. But growers encountered the problem of irregular fruit bearing and an inability to accurately predict their yield, resulting in losses from unpicked fruit, unproductive labour, and unutilised or overloaded storehouse space. Hence, the investment in these areas grew, with the low number of scientific publications, as listed by Scopus, relating to 'tree AND yield AND estimation AND fruit' increasing sharply this past decade [5]. It is clear that counting and estimating the yield is of interest to the horticultural industry.

Every year, an average avocado weight, is used to calculate the number of avocados from the total weight of a fruit batch. The count estimates are converted into a kilogram estimate, which is calculated from the estimated average fruit mass. This kilogram estimate is then compared to the harvest mass and a percentage deviation is found between the two, shown in (1.1). Note that (1.1) was provided by King Avocado for use in this thesis.

$$\%\, Deviation = \frac{(Harvest\ Mass - \ Estimated\ Mass)}{Estimated\ Mass} \times\ 100 \qquad (1.1)$$

The orchard this study worked with deviation is the percentage difference of kilograms between the orchard's eye count and the measured weight of the batch (e.g., bins the harvested avocados were collected into). The orchard deviation data, collected over the past several years using the process described above is presented in (**Figure 1.1)**, in which the ideal deviation is 0.

By looking at the medians, 2017 had the least deviation and was therefore the best year for orchard estimation, with all medians near or touching the ideal centre line. By contrast, 2018 had the most deviation, with two of its three zones having medians furthest from the ideal. Looking at the other years shows that

the consistency in 2017 does not hold true for any other year. 2016 and 2019 each had one zone that was near ideal, but all other medians of other years and zones display deviation from the ideal.

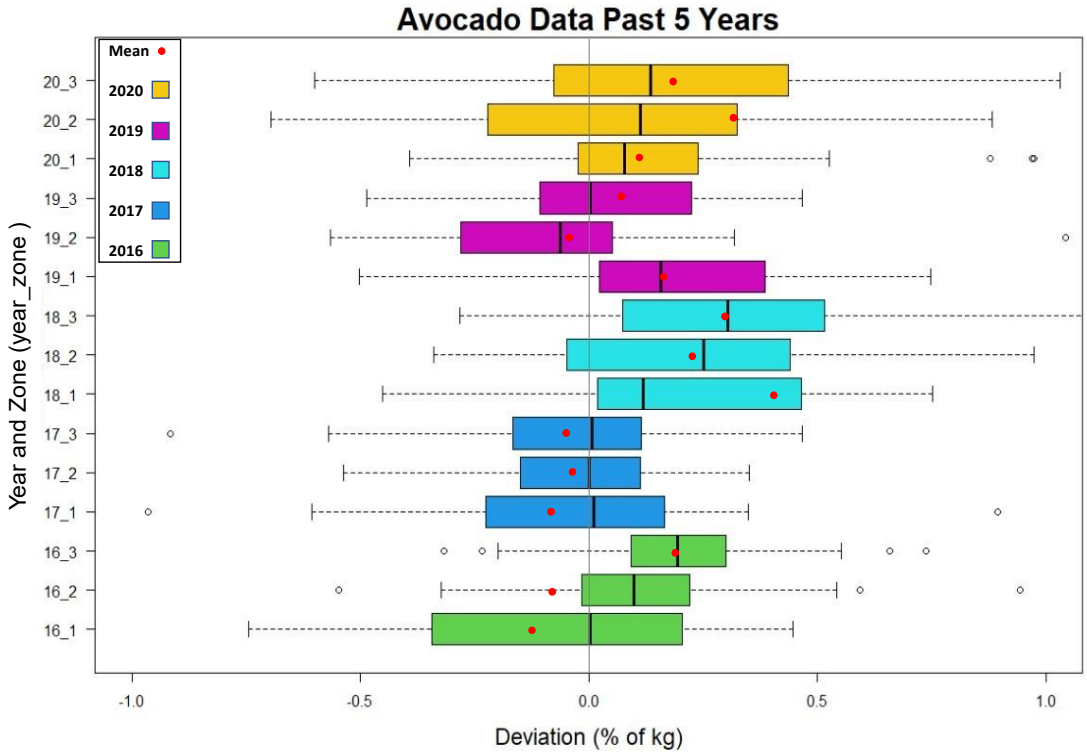

Figure 1.1. Box and whisker plot of avocado orchard count deviation throughout different zones from 2016-2020; means (red circles) alongside the medians with a centreline showing the ideal deviation

The means reinforce this variation with each year varying from the ideal. The year 2017 is again the least deviant, with a best mean deviation of -2.64%, and 2018 being the most deviant with a mean deviation of +42.6%. By looking at the consistency of the data, every year shows large amounts of deviation not only within zones but also within the estimated data.

The estimates in the best year, 2017, range from around -60% to +40%. The estimates in the worst year, 2018, range from around -50% to 110%. The explanation for 2018's high deviation is likely due to the high over estimations made by the counters that year, most likely due to the issue of human error that needs to be addressed.

It is important to have a reliable estimation strategy as the data has a flow on effect to NZ Avocado's estimates, and the harvest and post-harvest procedures of King Avocado. A simple example of an avocado cold chain is provided in Appendix A. Harvest is the first procedure in the avocado cold chain and the King Avocado orchard requires a large amount of labour as there are over 110 hectare of fruiting avocado trees.

If the avocado estimates for orchard blocks are heavily variant, then the hired labour may not suffice to harvest the variant block within the desired time frame, or the opposite may occur, where too much labour is harvesting too little fruit. This would lead to a drop in harvest productivity, and complications in storage,

logistics, and export procedures. Fruit unable to be harvested in time becomes susceptible to fruit drop from either wind, or as the tree prepares to flower for the next season.

Poor estimation may lead to an unproductive harvest, which would undoubtedly disrupt the rest of the cold chain (Appendix A, **Figure A.1**) and may cause a drop in fruit quality for the consumer, and lower profits for the supplier. Alongside this, crop load estimation assists with agronomical management such as soil nutrition or irrigation. The orchard therefore needs more consistent, reliable, counting strategies than what they are currently implementing. An assistive machine vision based solution, such as that investigated in this work, is one approach which may reduce count variability and potentially also reduce the human resource requirements.

The time required to count a single tree using the orchard's existing methodology was estimated by timing how long it took orchard counters to count 62 trees. This process took approximately two hours, giving a per tree time of ~2 minutes which serves as a target and point of comparison for the methodology developed in this work (since it is considered unlikely that the orchard will adopt a new methodology if it is more time intensive than current practice).

Current methods used to measure yield, count and sizing within the industry have led to inaccurate predictions above and below the actual value.

The aim of this Master Thesis is to provide an insight into the current methods that have been explored and what direction growers should move towards, to increase their ability to accurately estimate crop load in a timely and cost effective manner.

1.1.1 Experimental Objectives

1) Establish an avocado orchard as a basis for the research
2) Develop a novel, labelled 'on tree' avocado training dataset
3) Investigate the viability of a modern convolutional neural network as an identification tool for avocados
4) Investigate the viability of a modern convolutional neural network for the yield quantification of avocados

Alongside the literature review in Chapter 2, the experimental objectives were explored in two research chapters of this Master Thesis. Chapter 3 details the exhaustive process of gathering the novel image dataset from a NZ avocado orchard and the materials and methods used. The viability of the machine vision and machine learning model is investigated in Chapter 4.

If the objectives are successful a novel labelled training dataset would facilitate future investigation by other researchers, providing a major backbone for the advancement of automated machine vision and machine learning applications regarding avocados.

1.2 Background

The following section provides background information which is intended to assist readers from a variety of different disciplines in interpreting the content of this thesis. This section gives enough background information and core concepts related to this field so that a reader may feel more confident with the terminology and concepts used throughout the rest of thesis, especially the vision related concepts in section 1.2.2.

1.2.1 Avocado related

1.2.1.1 Location and supply

Northland, Whangarei and Bay of Plenty host New Zealand's industrial avocado growers. The climate of these regions are warmer and drier, suiting the physiological needs of the Hass avocado trees and fruits, especially in the northern areas like Northland, where temperatures are higher [6, 7]. It also benefits this study that one of New Zealand's largest avocado growers, King Avocado, is in Northland.

Avocado New Zealand announced their aim to quadruple sales to $280 million and triple productivity by 2023. Of the world's current Hass avocado suppliers, New Zealand ranks 9th largest, contributing 2% to global exports. Irregular swings in volume put strain on the New Zealand avocado industry and its economic impact within the country, contributing $122 million to export GDP in 2020 [4, 6]. It is important to understand the objectives of the NZ avocado industry to scope solutions from the research.

1.2.1.2 Considerations for Avocado Growers and Researchers

Avocados trees prefer northeast or north facing slopes for best light conditions [6]. New Zealand growers need to be aware of waterlogging, high wind, irrigation, soil condition, root health, temperature, light, soil nutrition and fertiliser effects. Of the aforementioned variables, a critically low temperature frost took place during the data collection of this study and may have impacted the counts in section 3.4.1.

1.2.1.3 Orchard operations

Most trees will be seedlings of Hass grafted onto Zutano rootstocks; however, Clonal rootstocks are better suited to the New Zealand climate. Normal start-up costs for an avocado orchard range from $45,000-$61,000, with each tree expected to cost an average of $94 to maintain per year. This, of course, is balanced out by the orchard gate returns (OGR). An average avocado grower in NZ will expect 8.56 tonnes of produce each season. This results in an OGR of around $30,000 ($12-$20 per tray). The harvesting season lasts around 8 months in New Zealand, with dips from May-August. It is important to know this for data collection purposes within the methods section as this study would only deal with the Hass variety on Zutano rootstocks.

1.2.1.4 Maturing and ripe

Interestingly, avocados do not ripen on the tree, they only mature. This is the act of increasing in size and developing the contoured features of its skin. If it is picked too early, the avocado remains hard and will never ripen post-harvest. This lack of colour change is going to be problematic for on tree identification.

1.2.1.5 Distribution

Approximately 60-65% of the avocados grown in New Zealand are exported, 25-30% remain for local retailers and consumers, and the remaining 8-10% is sent to processing facilities where they harvest the oil from the avocado to use in healthcare products.

1.2.1.6 Risk of disease

There are two main diseases that affect avocados in New Zealand. The first is root rot which is caused by *Phytophthora cinnamomi*, a type of mould that attacks root systems. The second is fruit rot, which is caused by a variety of fungi, mainly *Anthracnose*.

1.2.1.7 Avocado Tree and Fruit Types

There are 5 different types of avocados in New Zealand: Hass, Reed, Fuerte, Maluma and Carmen. Hass avocados are the only variety with export standards in New Zealand, therefore 95% of the total avocado trees in New Zealand are of the Hass variety [8].

1.2.1.8 Zutano seedling

Most avocado trees in New Zealand are grafted onto seedling rootstocks. The most common source of seed for rootstocks is Zutano. There is a high variance in the seedling rootstocks. For example, some may be more vigorous than others or exhibit more/less tolerance to *Phytophthora cinnamomi*. These rootstocks can produce reasonable yields [8].

1.2.1.9 Clonal rootstocks

Clonal rootstocks (vegetatively propagated) can overcome some of the variability in yield, vigour and tolerance to disease. They are commonly used in several countries to produce uniform trees with desirable characteristics (e.g., *P. cinnamomi* tolerance, high yield, tolerance to saline conditions, etc). Many of the imported clonal rootstocks are covered by Plant Variety Rights (PVR, similar to a patent) and purchase of these requires a non-propagation agreement [8].

1.2.2 Vision related

1.2.2.1 Machine vision

Computer machine vision is an interdisciplinary scientific field that deals with how computers can gain high-level understanding from digital images or videos. From the perspective of engineering, it seeks to understand and automate tasks that the human visual system can do [9].

1.2.2.2 YOLO

'You Only Look Once' (YOLO) is a deep learning single stage algorithm that uses a convolutional neural network (CNN) to detect objects. The open source software was released in 2016 [10]. The original YOLO model achieved 63.4% mAP while processing at 45 frames per second, more than double the mAP of other real-time detection algorithms at the time [11]. YOLOv5, YOLO's 5th iteration, consists of a 'Focus structure' with 'CSPdarknet53' as a backbone. The repetitive gradient information in large backbones is resolved by this backbone, which incorporates gradient change into feature maps to increase accuracy, decrease inference speed, and shrink the size of the model by reducing the parameters [12].

1.2.2.3 How a CNN 'sees'

Since computers do not have physical eyes, they are not able to physically see, so they view images differently to a human.

While a colour image to a human is viewed holistically (as a whole), a computer commonly breaks it down into layers of digits, each layer representing a band of that spectral image [13]. Commonly this is red, green, and blue (RGB) or hue, saturation, and value (HSV). Each layer can be represented as a matrix of numbers (**Figure 1.2**), with the dimensions of the matrix giving the width and height of the image, and the individual elements within the matrix, referred to as pixels, contain 1 byte (8 bits) of information and are used represent colour as an integer ranging between 0 and 255, with 0 being the lowest intensity and 255 being the highest [14]. Intensity is also commonly normalised to be in the range of 0 to 1.

A matrix is a rectangular array of quantities or expressions in rows and columns that is treated as a single entity and manipulated according to particular rules [15]. Each cell is filled with information. In noughts and crosses, there are 3 rows and 3 columns. If represented in computer code, it becomes an array containing 9 characters of information, and the information for every space would be 'O', 'X' or blank. But for images, the row number is the total pixels that make up the height of the image, and the column number is the total pixels that make up the width of the image. So, an RGB image with dimensions 640x480 would become three integer matrices, each having 480 rows and 640 columns. Each pixel can be exactly located according to its unique row and column combination.

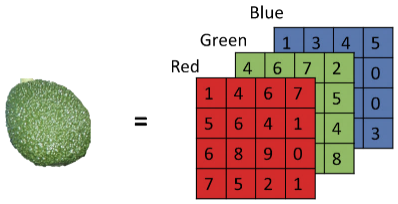

Figure 1.2. Visual aid for understanding how an image is viewed in a CNN.

It is important to understand how a computer views an image because just as humans can be taught to identify and locate objects in a picture through repetition and example, computers can also be taught to identify and locate objects in an image. The only exception is that they do not train in the same way that humans do, but instead look for certain patterns or repetitions of numbers within the matrices that can then be linked to objects of interest. This is the basis of CNNs.

Convolution is a function derived from two given functions by integration which expresses how the shape of one is modified by the other [15]. Within CNNs, the convolution is performed during the step process across the integers of an image matrix. The act of 'viewing' an image is where a 'filter' or 'kernel' traverses across the image and acts as a vision window. Kernels are functions that take input vectors in the original space and return the dot product of the vectors in the feature space [16]. Kernel size is always in odd numbers, 1x1, 3x3, 5x5 etc. Put simply, the values of the kernel matrix are multiplied with the corresponding values of the image matrix [17]. By modifying the values of the kernel that is convolved across the image matrix, the output image matrix will be affected. This creates a convolved image with

standout features, and these are called feature maps. Images can have thousands of pixels, by modifying a kernel to identify features, it allows computer 'vision' of that feature (**Figure 1.3**).

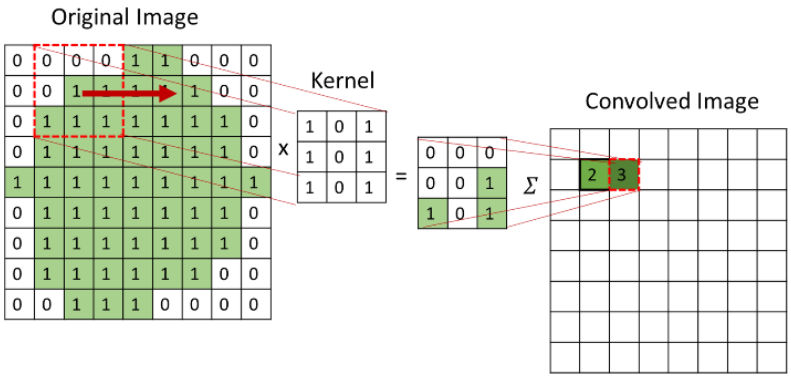

Figure 1.3 Red-dashed 3x3 matrix is the current position of the kernel traversing an image matrix. The convolved image values are the corresponding pixel values of the original image multiplied by the kernel values.

With more kernels, more feature maps are made to identify more features or objects within an image. Ideally, through the learning process the CNN will combine feature maps to best achieve the desired objective, such as identifying an object of interest. The downside is that every extra kernel requires additional computation, taking up more random-access memory (RAM), a form of computer memory in which data can be changed, removed, and can be looked at in any order [15]. So, there are limitations, but over time these limitations are lessened as computational hardware and software advances.

1.2.3 Agriculture related

1.2.3.1 Precision agriculture

Precision agriculture is a farming management concept based on observing, measuring and responding to inter and intra-field variability in crops. The goal of precision agriculture research is to define a decision support system (DSS) for whole farm management with the goal of optimizing returns on inputs while preserving resources [18]. Within the scope of this thesis, the observing and measuring methods of precision agriculture are useful as they may allow for accurate quantification of on tree avocados.

1.2.3.2 Biennial Bearing

Biennial bearing is a concept that describes a bumper crop in a particular year, where the tree produces a staggering amount of fruit, so much that the energy from the tree either cannot support the enormous quantity, or the resulting yield is high, but fruit are small. In the following year, the tree's energy is so depleted that it barely fruits, if at all [19]. This may potentially be an issue for counting methods associated with individual avocados as each tree may perform differently each year. Luckily, warmer locations like Northland are less prone to this effect, see section 2.3.6.

CHAPTER 2

REVIEW OF LITERATURE

This literature review covers relevant literature on avocados relating to machine vision, machine learning, detection, counting, sizing, yield estimation and other related areas. Particular results, equipment and methods will be highlighted. Due to the few works that deal specifically with avocados, similar crops like mangos were considered where methods may be transferable.

2.1 Estimation

This section will review relevant literature of avocados, and similar fruit, relating specifically to the estimation, what may affect it, and the quantification of it through a variety of methods.

2.1.1 Satellite Imagery for Avocados

Yield estimation tends to be inaccurate within the avocado orchard sector as the most common estimation technique is to physically count the fruits on selected trees throughout the orchard. Unfortunately, this presents multiple drawbacks: poor accuracy due to human error, occlusion and similar leaf colouration, unproductive time and labour usage and misrepresentation of the whole orchard due to limited sample sizing. By using the relationship between the measured yield and the optimal vegetation indices, an estimate for the yield was calculated [20]. Understanding of the relationship was further enhanced by sampling five avocado blocks in the 2016 and 2017 growing seasons. When compared to the actual yield, grower estimates provided considerably lower accuracy compared to the prediction accuracies achieved for each block [21]. The same group went on to do a similar experiment regarding mangos [22] and achieved similar accuracy and results. The algorithm R^2 indicated that there was definite influence on location and management practices regarding different orchards. They concluded that their ability to predict the yield parameters was better than current practices across the mango industry, which like the avocado industry, also used manual counting. It even showed the ability to quantify yield loss as a result of delayed harvest, (Appendix A, **Figure A.2** &**Figure A.3**).

2.1.2 Yield Mapping of Mangos using UAV

In Australia Sarron et al. in Australia developed an estimation model for the yield of three different mango cultivars through the use of tree structure and UAV photogrammetry [23]. Like the avocado industry, the current estimation methods are based upon visual inspection on a limited tree count and are manually counted. In the study, they introduced the idea of low-cost imagery for estimation, using an RGB camera, which had not been explored in this field before (but it is common knowledge that in recent years there has been exponential increase in camera resolution, allowing for better image capture). The developed estimation model had an R^2 of 0.77 with an RMSE of 20-29%. Low-cost imagery like this may be a viable method to use for avocado quantification.

2.1.3 Estimation of Mangos with Machine Vision and Satellite Imagery

Anderson et al. [5] also expressed that the current methods of pre-harvest fruit load estimation were based on manual counting from about 5% of the trees in the orchard. This manual counting led to tree fruit

variance across the orchards of about 27-93%. Avocado estimation shares a similar variance (see **Figure 1.1**).

2.1.4 Quantifying Root Rot using Image Analysis

Phytophthora heavily affects the tree's ability to yield by infecting the roots which intake nutrients and water, eventually leading to canopy/foliage loss and the inevitable death of the tree. This study by Salgadoe et al. [24] focused on segmenting the image background from the foreground using Otsu's method [25], which uses histogram peaks to differentiate the two, and Canny edge detection thereafter. The image analysis was automated and required no user input. This research could be useful to the counting of fruit in the sense of the image analysis and methods, i.e., automating the processing of the images, but also if the background leaves/branches were removed and only the fruits are left, then fruit counting would become elementary.

2.1.5 Alternate bearing

California growers of Hass avocados [19] showed how 66% of their trees had a factor of alternate bearing from 0.5-1.0 (50%-100% yield difference). This was the variance between the on and off crop years. They also stated that 47% of the alternate bearing trees had a factor between 0.75 and 1.0. Despite this, most of the avocado trees produced 28kg/tree per year over a 20-year season. This research could be useful in developing models which quantify avocado yields, in the sense that it accounts for year-to-year variance of avocado yields due to biennial bearing.

2.1.6 Bavendorf Method

The Bavendorf Method is a model of yield estimation. The growth in an orchard is largely based around 3 metrics: yield capacity, average fruit density and average weight of the fruit at harvest. Yield capacity can be expressed in a formula, C = d x h, where 'C' is the yield per hectare, 'd' is average canopy diameter and 'h' is average canopy height. Avocado estimation of the two modelled orchards showed results that exceeded the actual yield in both cases, 14% and 16% respectively [26].

2.1.7 Bavendorf Method Adjusted

The Bavendorf Method, normally used for apples, was adjusted for avocados in a South African study [27]. They came up with an avocado factor, α =0.47, based upon the properties of the trees. The formula was adjusted in order to obtain the crop prediction.

$$CROP = CAP * FSD * \frac{1}{AFW} * \propto \qquad \text{[27] (2.1)}$$

Where,

CROP = in boxes at 18.5kg netto

CAP = tree capacity

FSD = fruit set density

AFW = average number of fruit per lug box

$\propto$ = avocado factor

This model had a 10% variance from the actual crop number at harvest. The area covered was 200 hectares.

2.1.8 UAVs and Precision Agriculture

Unmanned aerial vehicles (UAVs) and drones have begun to change the agricultural industry. The important economic selling point for these devices is their ability to facilitate precision agriculture across large areas. In the past, growers have treated their fields homogeneously, but now they can selectively target areas which require attention, therefore saving resources, time and money while also limiting their negative impacts on the environment. 'Perspectives for Remote Sensing with Unmanned Aerial Vehicles in Precision Agriculture' shows the different ways in which the devices can be used (**Figure 2.1**) [28].

Application		Type of sensor/camera			
		RGB	Multispectral (broad band)	Hyperspectral (narrow band)	Thermal
Drought stress	Detection in early stages	–	–	S[b]	HS
	Long-term consequences	–	HS	HS	S
Pathogen detection	Detection in early stages	–	–	HS	HS
	Severity of infection	HS	HS	HS	S
Weed detection	Spectral discrimination	–	S	HS	–
	Object based	HS	HS	–	–
Nutrient status		S	HS	HS	S
Growth vigor	Growth stage	HS	–	–	–
	Canopy height and biomass	HS	HS	–	–
	Lodging	HS	–	–	S
Yield prediction		S	HS	–	–

Figure 2.1. Overview of Applications and Suitability of Different Sensors where S = suited, HS = highly suitable [28].

2.1.9 Methods for Automated Monitoring enabling Precision Agriculture

Das et al. [29] demonstrated the ability to map an orange orchard with LiDAR cameras. Light Detection and Ranging (LiDAR) is a method used for measuring distances by illuminating the target with laser light and measuring the reflection with a sensor. Differences in laser return times and wavelengths can then be used to make digital 3D representations of the target. It has terrestrial, airborne, and mobile applications [30]. The devices could be mounted onto UAVs, drones or handheld by orchard scouts. The algorithm utilised was able to observe, locate, and count the fruit and then predict the total amount in that section, as some would be occluded by leaves (**Figure 2.2** & **Figure 2.3**). There was a correlation between measured true leaf area and LiDAR area index (R-squared value of 0.82). Limitations included occlusion, illumination and false positives from canopy features.

Figure 2.2. Shows the developed algorithm which uses support vector machines to locate and count the fruit and provide bounding boxes according to their locations [29].

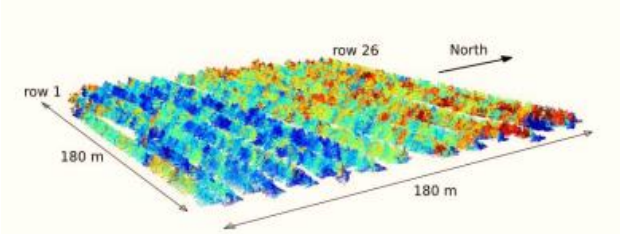

Figure 2.3. Showing the distribution of fruit in the field and how productive certain areas will be. Blue are low fruit count areas, red are high fruit count areas [29].

2.2 Detection, Sizing, Counting

This section will review relevant literature of avocados, and similar fruit, relating specifically to the detection, sizing, and counting through a variety of methods.

2.2.1 Faster R-CNN with Inception V2 and SSD with MobileNet

J.P. Vasconez et al. trained two networks to detect avocado fruit. They had an average precision (AP) of 90% for their validation set of images and 84% for their test set of images [31]. Their method involved a handheld commercial camera with video images taken of the sides of the corridors of the grove. Their conclusions highlighted that the effectiveness of convolutional neural network (CNN) architectures are dependent on training set quality (e.g. image blurriness). This is important regarding the objectives of this thesis because if the dataset collected is of poor quality then it will have a flow on effect to the CNN used to quantify the fruit within this data.

2.2.2 Deep Learning for Fruit detection and Yield Estimation

Koirala et al. investigated multiple CNN models (YOLO, YOLOv3, R-CNN, etc) across multiple fruits in their literature with results summarising their accuracy. They bring attention to a variety of available deep learning models used in counting, detecting and estimating yield. The key learnings are that orchards can be mapped digitally; occlusion can be accounted for, and yield estimation algorithms are viable for pre-harvest counting. One relevant example would be the yield estimate of 16 mango trees using a multi-view method. This involved taking 37 images around the tree and creating a detection model on those images [32]. The F1 score was > 0.8 for 6 fruit categories.

2.2.3 Automated Avocado Yield Forecasting Using Multi-Modal Imaging.

Within Multi-Modal imaging, Woodson and Ross [33] discovered that thermal imaging at the canopy level is a viable option for automated counting of the avocados. Since the avocados appear white in the thermal image, they are far easier to separate from their background. The group also developed a prediction model based upon the Bavendorf model. The drawback of this study is that they did not adequately present enough evidence, either through calculations, graphs and/or data, to affirm the results they came to. The only figure presenting data analysis was a histogram, which did not explain what was achieved or how successful the hypothesis was.

2.2.4 Fruit detection using LiDAR with forced air flow

By detecting fruit through LiDAR [34], Gené-Mola et al. developed a system that was able to detect more than 80% of the visible apple fruit with an RMSE of less than 6% (**Figure 2.4**). They mounted a LiDAR camera to the top of an automated vehicle which navigated using satellite GPS tracking. They also attached a machine that provided forced airflow onto the trees, reducing the effect of occlusion. While this study was relevant towards apples, the LiDAR and airflow methods used alongside the automated GPS driven vehicle showed promising results. LiDAR should be considered as a detection option in this study, while the self-driven vehicle could be used in future work for autonomous quantification of avocados.

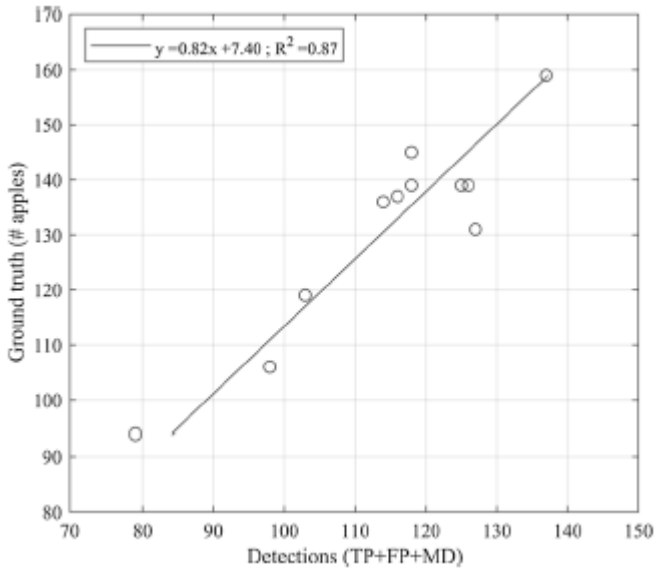

Figure 2.4. Showing the correlation between the detected fruit and the ground truth, R^2 = 0.87 [34].

2.3 Increasing yield

2.3.1 Avocado Ripeness, Smartphone Image and Machine Learning Model

The post-harvest ripeness of Hass avocados was determined using images from a smartphone and support vector regression (**Figure 2.5**). Hass avocados are one of the only varieties to change colour when ripe, hence this study took advantage of the gradual colour change and through the use of machine vision techniques, the model could accurately predict the ripeness with an R^2 of 0.92 and a RMSE of 7.54 [35]. Image capture methods using low-cost smartphones alongside the machine learning methods presented in

this avocado research showed promising results. While the results of avocado ripeness are not within the scope of this thesis, the image capture and machine learning methods could be transposed into avocado quantification methods to assist the quantification of the avocado fruits.

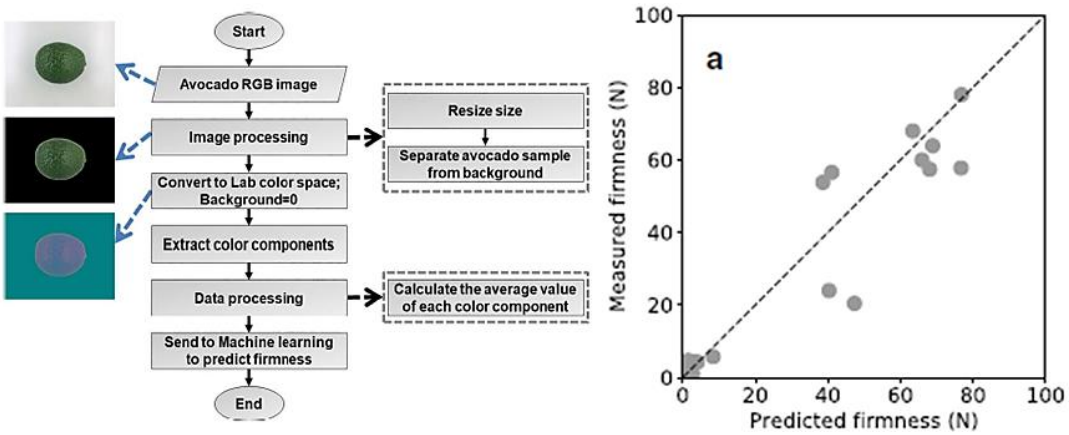

Figure 2.5. Flow chart demonstrating the process followed by [35] along with the correlation between predicted and measured firmness.

2.3.2 Light interception modelling using unstructured LiDAR

This study by Westling et al. [36] looked at overseeing the light diffusion through canopies, stating it as a key influencing factor as the quantity of sunlight gathered by each tree impacts the yield and quality of fruit produced. It is subsequently an important characteristic to quantify and eventually control with pruning. They demonstrate a solar-geometric model that evaluates the light interception with an R^2 of 0.85 with future work looking to corelate this to yield. The methods used in this research assist in providing the predictive parameter of light interception for avocado quantification on an orchard scale and would be useful in future research.

2.3.3 Avocado trees under different regimes of water

In this study by Moreno-Ortega et al. [37], water irrigation methods were investigated to determine their effects on avocado size, yield and water productivity. One idea was to identify the viability of low water scenarios and their effect on the avocado trees and fruit. Another was to over supply water to the trees and investigate the effects on the trees and fruit. They discovered that about 7500 $m^3\ ha^{-1} year^{-1}$, representing 30% over current water allowances in the Spanish region, yielded larger fruit size and quantities. The methods used in this research assist in providing the predictive parameter of irrigation for avocado quantification on an orchard scale and would be useful in future research.

2.3.4 Salinity and Water Effects on Avocado

Oster el al. [38] utilised observed salinity levels to try to find the optimal values in which to grow Hass avocados. They found that the yields of the crops increased when increasing the amount of supplied water, however, there was a limit reached in which the yield declined 65% above that limit. This limit was said to be 0.57 $dS \cdot m^{-1}$(decisiemens per meter, a measure of salinity). Their results indicated that the maximum yield for Hass avocados on Mexican rootstock were unable to be achieved when the level of average salinity exceeded 0.6 $dS \cdot m^{-1}$. The results of this research affirm the conclusions drawn from 2.3.3.

2.3.5 Biochar

Biochar in avocado plantation soils has not been explored thoroughly. In this study [19], the researchers determined the impact of wood biochar on avocado development, yield and financial gain. Biochar was used in a 0%, 5%, 10% and 20% volume by volume basis. Biochar contributed to strong development of avocado seedlings and increased the quantity of avocados produced after the third year of planting. There was a measured rise in soil carbon, product yield, tree circumference and height across all biochar controls over the seasons. Trees planted with biochar had 18–26% more noteworthy development rates (in terms of tallness and stem distance across) than the control. The trunk diameters were also increased with biochar, 145.4 ± 3.3 mm, relative to the control treatment, 125.0 ± 2.7 mm. Tree stature was improved as well with biochar, 3.7 ± 0.1 m, relative to control treatment 3.4 ± 0.1 m. Yield from the biochar was improved with a 97% increase in the 2018 growing season compared to that of the control. Biennial bearing trees regularly have a lower yield within the ensuing year but in spite of this, the 2019 natural product tallies were higher in total for the biochar corrected trees by 20% relative to the control. In terms of costing, if a grower used Biochar over a hectare, their net benefit would sum to US$8581, or US$105 per ton of biochar [39].

2.3.6 Irregular Bearing

Avocado trees have an issue in bearing which is related to on and off seasons. New Zealand's Bay of Plenty region, which provides 60% of total avocados in NZ, suffers the most from this effect (**Figure 2.6**). After a good year of produce, the following year could taper off, resulting in a fluctuating supply of fruit. There are considerations of what factors might cause the irregular bearing, but applied treatments were unable solve the problem, they only reduced the effect [40, 41]. This research also shows greater consistency in the "Far North" regions like Northland, which is also the location of King Avocado, affirming the decision in using King Avocado as the orchard of data collection for this thesis as it will assist the objectives of this thesis.

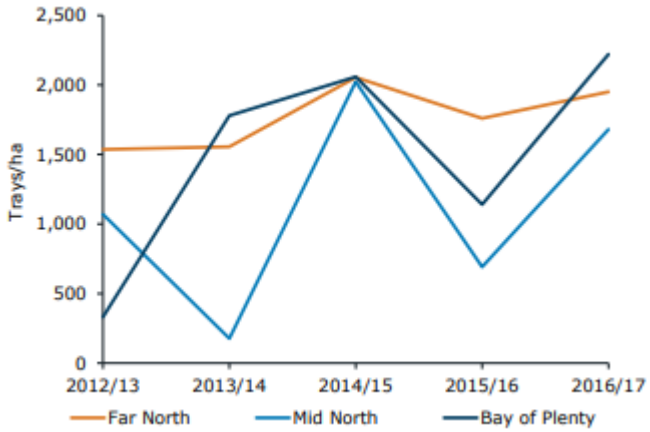

Figure 2.6. Differing avocado yield patterns of each region in NZ [40, 41]

2.3.7 Ripening of Avocado by Impedance Spectroscopy and Support Vector Machine

Analysing the fruit of avocados post-harvest to detect their stages of ripeness could be something used within storehouses or within supermarkets. This study [42] used impedance spectroscopy alongside a support vector machine learning approach to accurately predict the stages of avocado ripening; firm

breaking, ripe and overripe. Their model was able to predict the current state with an accuracy of 90% and a precision of 93%. The benefit of this method is low cost and its non-destructive nature.

2.4 Machine Learning predictions and other Avocado related research

This section reviews relevant literature of avocados relating to machine learning and other avocado related research which may be helpful to the objective of this thesis. Notably, the machine learning techniques used to create prediction models around avocado research objectives could be useful for predicting future counts of the fruit. Alongside this, the research in this section also aligns with Avocado NZ's sales and productivity objectives (see section 1.2.1.1).

2.4.1 Pre-ripening

There has been research done which shows that consumers are more likely to purchase an avocado if it is in a ready to eat state. This has led to an increase in pre-ripening of fruit before shelved at retailers. The way in which this fruit is maintained is by exposing it to ethylene and maintaining the temperature at 21 degrees Celsius until its key indicators, colour and firmness, proved its ripening [43].

2.4.2 Price Forecasting by Flock Consulting

Flock consulting is a data science company which provides price predictions based on provided data. For avocados, they were given the retail price of avocados in New Zealand prior to 2020 and were asked to provide a prediction of what was to come. Worthy observations were as follows: prices are lowest in summer and highest in winter which coincides with the quantity supplied being highest in summer and lowest in winter. The supply within New Zealand seems to be increasing over time, however there is only a certain amount of land available to grow avocados, therefore the supply is expected to plateau, so instead of linear growth, they predict logistic growth. Most interesting of their predictions was trend data provided by Google being compared with New Zealand demand for avocados, as the trend of veganism increased, avocados demand began to see incredible spikes (**Figure 2.7**) [44]. The machine learning methods used to predict the price of avocados could potentially be used to predict the yield year to year in an avocado orchard. King Avocado has years of data history already, potentially all that is needed are the orchard parameters like weather conditions, fertilizer frequency, irrigation quality, etc. It would take time to tune parameters, but this research shows promise in this regard for yield prediction.

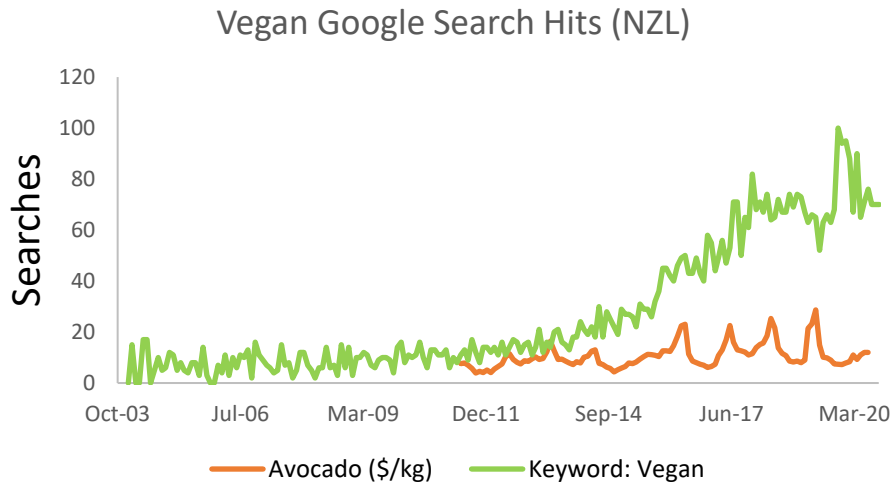

Figure 2.7. Line graph showing the Google search data for keyword 'vegan' (green) overlaid on top of New Zealand Avocado ($/kg) (orange) [45].

2.4.3 Estimating Avocado Sales Using Machine Learning Algorithms and Weather Data

Machine learning algorithms were used to produce a prediction model for the United States avocado market [46]. A support vector machine regression and multivariate regression prediction model had correlation coefficients of 0.995 and 0.996 and a relative absolute error of 7.97 and 7.81, respectively, when compared to the actual data. The model takes in weather and sales data to produce accurate predictions. By knowing the data of the preceding year, they could predict the fluctuations of the United States avocado market (**Figure 2.8**). This research is like that of 2.4.2 in which machine learning was used for price prediction of avocados. The benefits of the research in 2.4.2 toward the objectives of this thesis also apply in this section.

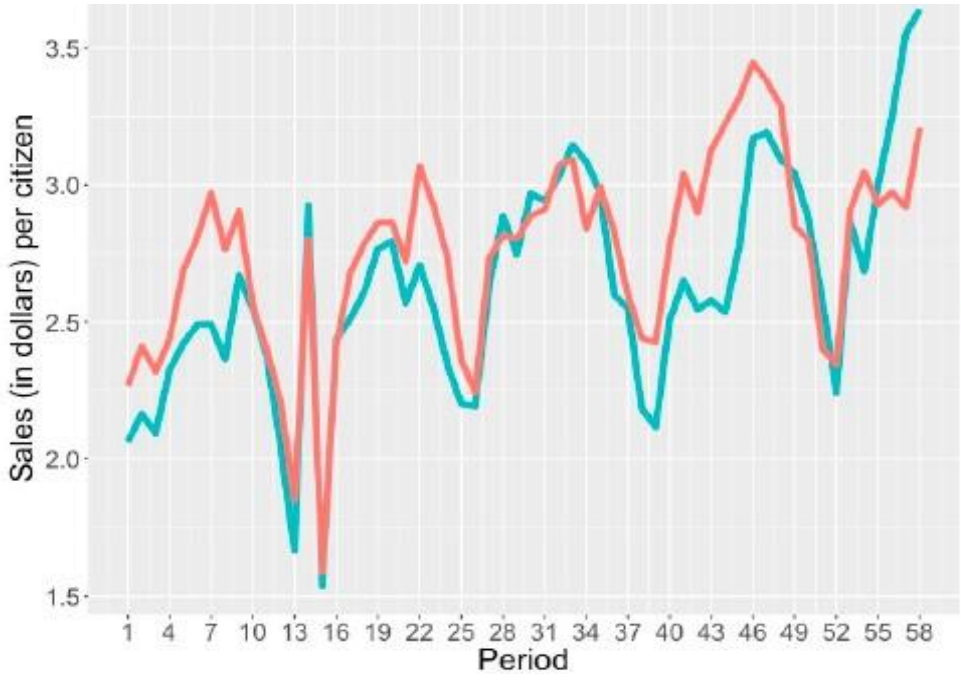

Figure 2.8. Los Angeles market for avocados on average per individual, predicted (red), real (turquoise) [46].

2.4.4 Non-destructive discrimination of avocado fruit ripeness using laser Doppler vibrometry

A study evaluated non-destructive laser Doppler vibrometry (LDV) to assess avocado ripeness [47]. All fruit measured showed a significant decrease in RF during ripening (**Figure 2.9**). Predictions explained 90.7% variability, proving the model successful and can be used to determine when to eat the avocado (**Figure 2.10**), benefitting supply chain logistics.

This research would indirectly benefit the objectives of this thesis because the quantification of avocado fruit has a flow on effect to the post-harvest processes. If orchards are more able to quantify their fruit, they may also be inclined to invest in the post-harvest processes too which may increase their sales and may achieve the goals of Avocado NZ. For example, orchards could invest in non-destructive vibrometry alongside machine vision and machine learning to create a system that identifies an avocado on a supermarket shelf, measures the ripeness of it, and records and displays that information for both consumer and supermarket benefit. This would quantify a yield of fruit available to eat in the present moment.

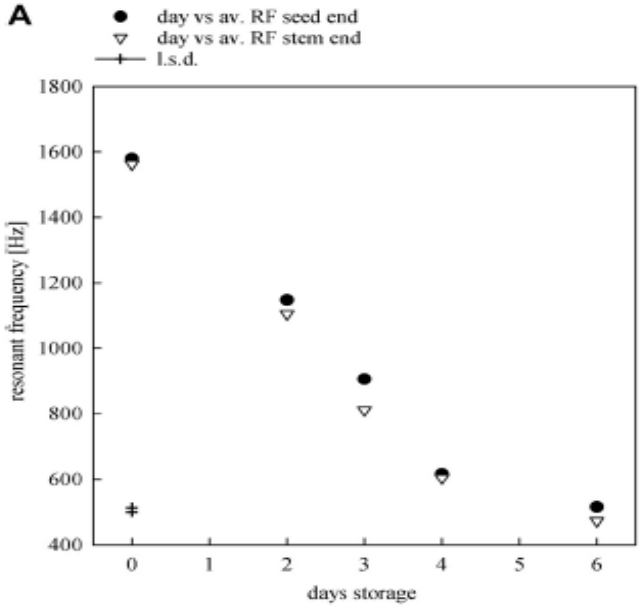

Figure 2.9. Showing the decrease in resonant frequency during storage days [47].

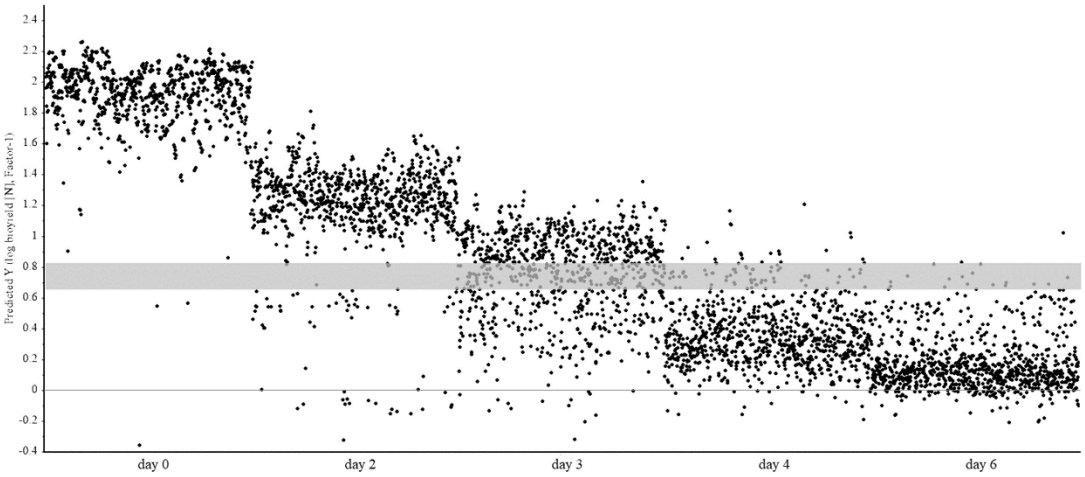

Figure 2.10. Scatter plot of the predicted force break values of Hass avocados, during storage, which are repeatedly measured using laser doppler vibrometry. The grey bar indicates the zone where the fruit is good to eat [47].

2.5 Existing methods in other fields

This section will review relevant literature of similar fruit regarding detection and quantification techniques, with the aim to transfer methods to avocados, notably the methods and metrics used are relevant to the objectives of this thesis.

2.5.1 Quantifying the Crop Load on Apple Trees using Hyperspectral Snapshot Imaging

Hyperspectral imaging was taken within an apple orchard to detect and quantify green apples [48]. Classification methods involved Support Vector Machines, k-Nearest-Neighbour, Neural Networks and Random Forest and a mix of combinations to analyse the data. Total accuracy for the project was 93.5%, indicating a good classification result. The collection method could be improved; a singular tree was looked at with a tripod camera. However, the detection accuracy was very good and should be considered as a possibility for identifying and locating the fruit.

2.5.2 Automated Crop Yield Estimation for Apple Orchards

An autonomous vehicle drove through an apple orchard collecting images using two Nikon cameras [49]. The system processed the apples from the background using segmentation and specular reflection (**Figure 2.11**). It could scan both sides of each row of trees. Their estimate errors were 3.2% and 1.2% for red and green apples, respectively. The benefits of this research are within the nature of the autonomous collection to the unique detection method of specular reflection. Similarly, avocado detection may benefit from specular reflection as the fruits do have round portions in their shape.

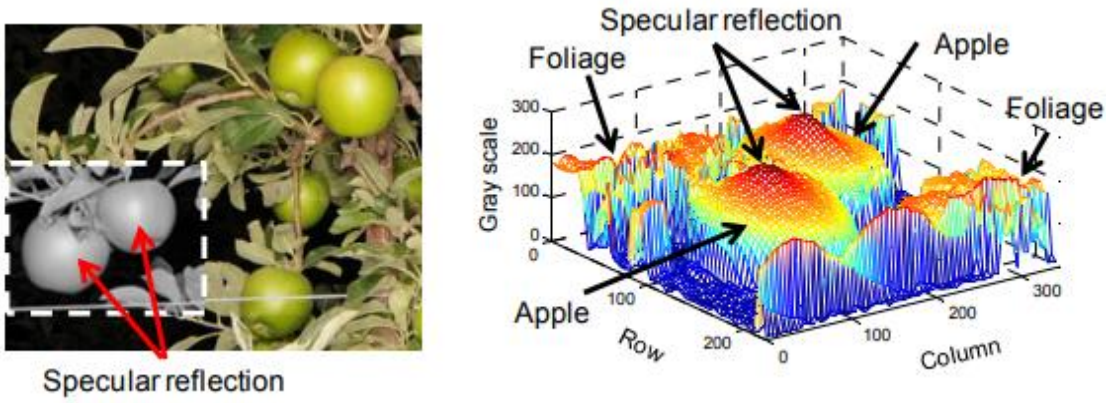

Figure 2.11. Example of specular reflection where the regions of the light intensity diminishes radially for the apples but randomly for the foliage, hence allowing them to be detected [49].

2.5.3 Almond Yield Forecasting by Drone

Synthetic modelling an actual orchard has been achieved within the almond industry. Drones were paired with RBG cameras in order to digitally capture the orchard [50]. The aim was to investigate accurate forecasting of yield. The results showed an average standard deviation of 207 lbs/acre, proving better than current methods. While this method may seem complicated, the simulated crop load would save thousands of hours of manual inspection, and data for models could be collected autonomously. A promising solution to consider within the objectives of this thesis.

2.5.4 Grape Detection with Neural Networks

Santos et al. found in their study on grapes that image segmentation using Mask R-CNN presented promising results with an average precision (AP) of 80% [51]. They advocated the use of R-CNN but also used YOLO networks. Santos et al. also made a valid point that agricultural detection systems such as yield prediction need low-cost inference systems, such as mobile phones which if used along with the free (or relatively cheap) detection software, could be a progressive and affordable way forward for agricultural detection systems. This research, both its methods of data collection through image capture to processing, and results of precision, recall, and average precision, are incredibly useful for the objectives of this thesis and may be transferable to avocados.

2.5.5 Image based detection of Mangos

Stein et al. used not only R-CNN but also a novel LiDAR (Light Detection and Ranging) component, allowing each fruit to be associated to its tree [52]. This means further information can be gathered to help the farmer. By associating the crop size to a particular tree, the farmer can then look at the health of the trees by the fruit they produce, potentially spotting diseases, or lack of nutrition in particular canopies which need attending to by the farmer, to increase yield. The combination of imaging and depth measurement could be an important element to include in yield estimation detection and a way for the future. This research made use of a CNN in order to detect fruits on tree which is incredibly useful for the objectives of this thesis and may be transferable to avocados.

2.6 Discussion

Assuming the market goals of Avocado New Zealand remain consistent, sales will be expected to quadruple from $70 million to $280 million and productivity will triple by 2023. In order to achieve these goals, it is vital New Zealand growers have access to accurate and low-cost models that count and size their crops, with more emphasis on the counting so yields can be estimated, and labour can be distributed more productively during harvest. As seen in Chapter 2, there are a multitude of viable models to estimate and count avocados. However, it is not feasible for companies to opt for multiple options due to costs, resources and time. A decision matrix (**Table 2.1**) will narrow the scope and allow the most appropriate options to move forward [53].

2.7 Decision Matrix for estimation

Within engineering design, Ullman describes adequate ways to make focussed decisions when presented with multiple solutions [54]. To give preference to one solution, criteria are weighted so that the chosen solutions are the most appropriate to the needs of the objective. In this instance, an interview with Claudia, the King Avocado orchard manager, provided relevant scope, with accuracy being a particular focus point. The criteria presented in (**Table 2.1**) are weighted in accordance with what was least/most valuable to the orchard on a scale of 1 to 10, with 10 being excellent and 1 being poor according to that criterion, therefore solutions with the higher ratings will be considered over those with lower ratings. The solutions criteria are defined as:

- Capital Cost Estimate
    - The estimated capital costs associated with this solution. This criterion is important as the solution must be affordable to the orchard will therefore be given a moderately high weighting.
- Labour Intensity
    - The amount of labour required to maintain and or implement the solution. This criterion is a ‘nice to have’ for the orchard as less labour-intensive solutions may be more accurate but is also not essential and will therefore be given an average weighting.
- Feasibility
    - How relevant and feasible is the solution to the objectives of this thesis and the objectives of the orchard and by extension Avocado NZ? This criterion is important and will be given a moderately high weighting
- Modern
    - Does the solution make use of beneficial modern technological advancements? This is not an essential factor but should be considered when choosing a solution, therefore this criterion will have a below average weighting
- Low Risk
    - How much of a risk does the solution bring with it? A good solution should have a low systematic error, equipment failure should not significantly hinder results, etc. This is important but should not draw away from the objectives of the thesis, therefore the criterion will be weighted just above average.

- Industry Ready
    - If the orchard were to ask for the solution to be implemented as soon as possible how much work would be needed to do this? Considering the time constraints of this research, this is a crucial criterion and will be given a very high weighting.
- Accuracy
    - The solutions ability to accurately quantify, detect and or count the objects of interest. This is critical to the objectives of this thesis and the orchard. Therefore, this criterion will be given a maximum weighting.

Table 2.1 Weighted Decision Matrix

	Weighting Multiplier							
	8	5	8	4	6	9	10	Total
Criteria	Capital Cost Estimate	Labour Intensity	Feasibility	Modern	Low Risk	Industry Ready	Accuracy	
Satellite Imagery yield prediction using NDVI	7	7	9	8	8	8	9 (± 5%) [21]	405
Bavendorf Model yield prediction	10	3	6	2	5	8	2 (± 15%) [26]	273
Bavendorf Model Adjusted yield prediction	10	3	7	3	6	10	4 (± 10%) [27]	329
Overhead yield prediction w/ drones or UAVs	8	8	8	7	8	9	7 (± 5-10%) [28]	395
Machine vision and Machine Learning yield prediction w/ low-cost imagery	9	9	7	8	7	9	8 (± 4-15%) [31] [46] [51]	408
Thermal Imaging w/ yield prediction model	8	7	7	7	6	6	4 (± 20%) [55]	313
LiDAR yield prediction w/ GPS robot	6	7	5	9	5	8	4 (± 10%) [34]	301
Hyperspectral Imaging yield prediction w/ GPS robot	6	8	9	9	5	8	8 (± 4-7%) [48]	378
Specular reflection yield prediction w/ GPS robot	6	9	6	6	4	7	9 (± 4%) [49]	342
Drone forecasting yield prediction using synthetic 3D map	7	9	7	10	5	10	7 (± 5-10%) [50]	387

Highest scoring options were Machine vision and Machine learning with low-cost image capture, Satellite imagery yield prediction using NDVI, Overhead yield prediction using drones or UAVs, Hyperspectral imaging with a GPS robot, and Drone forecasting.

2.8 Comparing and Contrasting

2.8.1 Aerial Candidates

Of the five best candidates from **Table 2.1**, two are similar: satellite imagery yield prediction using NDVI, overhead yield prediction using drones. Aerial technology such as drones, UAVs, satellites and planes can be used to capture the field of interest in varying spectral bands [18]. The benefit of these strategies is that the use of aerial technology reduces the labour and effort required during maintenance and harvesting. The grower can easily identify areas of interest through the different spectral bands and make decisions based on the photosynthetic activity of those areas.

2.8.2 Satellite Imagery vs Overhead yield prediction using Drones

Both options capture overhead data which can be used for precision agriculture [18] but there is a difference in the collection method. The grower would have to purchase the images directly from satellite imagery providers, or purchase a UAV or drone to capture their own data which may be expensive.

The downside of purchasing images from satellite providers is that this will be an ongoing cost. Alongside this, increasing the image quality, spectral bands, size, and quantity will result in higher prices, as shown on the apollomapping.com website. The minimum area to buy an image is quite large and the image quality may be limited somewhat based upon distance, camera quality and atmospheric conditions. The upside is that satellite imagery is cheaper compared to a UAV or drone cost. The capital required and labour intensity is low while the research done around this imagery shows it can be accurate despite the limited spectral bands [21].

The downside of the UAVs and drones: The capturing must be done by the grower, the initial capital required to set up and maintain them is costly, and the price increases relative to the level of technology incorporated, hence higher-cost drones will provide higher quality data.

2.8.3 3D map forecasting using drones vs Overhead yield prediction using drones or UAVs

Both options use drones, but quantifying yield using drones like in [28] focusses on overhead flights [28, 29], neglecting the lower ground / under canopy structure mapping. In this way, a synthetic 3D model may provide more accurate mapping of the orchard data and may allow for machine learning algorithms to estimate the yield synthetically like in [56].

The downsides of synthetic modelling: the model researched was designed for almonds and would need to be adjusted for avocados, occlusion of fruits, lighting conditions and quality of camera. The downsides of quantifying yield with overhead drones: the orchard can only be viewed from a top-down vantage point; weather conditions affect flight times and sensors and lens would be priced higher for overhead drones compared to low flying drones.

2.8.4 Aerial vs Grounded

The fourth candidate, hyperspectral imaging with a GPS robot would use a grounded robot to automatically move around the orchard and collect data via cameras and sensors, similar to that in [49]. This model is similar to what the grower would experience while surveying their own orchard. The robot would be fitted with state-of-the-art equipment to provide accurate data collection and prediction models. Their drone counterparts would struggle to do similar due to weight limitations of equipment mounted on the drone. More expensive drones may support heavier equipment, but the costing and singular vantage point of these drones may be undesirable to some growers [18]. The downside of an autonomous robot is the cost. Developing such a machine may exceed the budget of some growers. A relatable cost example came from an interview with Professor Johan Potgieter of MAFDL, who said the prototype automated GPS robot his team developed cost $50k-$80k. Alongside this, grounded surveying would take longer than that of a drone. Potentially the use of both aerial and grounded techniques could be used in tandem, but this comes with the increase in capital costs due to purchasing more than one model.

2.8.5 Bavendorf vs Machine vision and Machine Learning with low-cost imagery

The Bavendorf model and the adjusted Bavendorf model are capable of estimating based on a human eye count within a defined area but are prone to error, both human and model error. Machine vision and machine learning models, even with poor quality images, are capable of learning over time and reducing error, which the Bavendorf model lacks. Alongside this, machine vision and machine learning could easily replace and improve upon the Bavendorf model as a CNN was seen to accurately count avocados in [31].

2.8.6 Machine vision and machine learning with low-cost imagery vs Aerial Candidates vs Hyperspectral imaging with a GPS Robot

The main upside of machine vision and machine learning with low-cost imagery over aerial candidates and hyperspectral imagery with a GPS robot is that it can be relatively cheap to implement operations. Low-cost smartphones or cameras are capable of capturing imaging datasets to use for detection [35, 57], counting and estimation. Due to the low cost, multiple smartphones could be purchased and used in parallel to collect data. To contrast this, aerial candidates, and hyperspectral imaging with a GPS robot present expensive upfront capital costs, but their data collection productivity would likely be greater than manual image collection. Hyperspectral cameras also have the benefit of extra spectral bands, but the scope of the information the camera provide may be an information overload for yield quantification of avocados. The cost is also far higher than smartphones, meaning data capture would be limited to likely one camera, which would pull away from the objectives of the thesis in a sense of providing a smaller dataset. Similarly, overhead drones require maintenance and expensively designed cameras which are lighter and do not inhibit its flight while capturing data, again, likely limiting collection to one unit.

2.8.7 Machine vision and machine learning with low-cost imagery vs Specular Reflection with a GPS robot, and Thermal Imaging vs Standard Imaging

The mirror-like reflection of waves, such as light, from a surface is known as specular reflection [58]. This is observed as a bright spot on the surface of a smooth and round fruit when light is shone on it (**Figure 2.11**). The problem is that Hass avocados have a bumpy skin texture, which may disrupt this method as it was designed for use on apples, which have a shinier, waxy skin. Adding to this, the camera would have to be mounted on a self-driven robot with a powerful light source and the operation would have to take place at night time for the best reflectance. These factors may hinder the objectives of this research as focus is drawn away from quantification of avocados to other non-scope methods like a GPS guidance system. Thermal imaging seems like a potential idea, however, the research available was too inadequate to pursue. To further investigate thermal imaging, some preliminary images of avocado were captured with a thermal imaging camera as part of this work and the results suggested that differentiation of avocados from thermal gradients would be challenging. Contrary to this, if a machine vision and machine learning model like YOLO was fed standard low-cost imagery from a smartphone, it will adapt the features of the desired object in any image and will learn to detect it in a variety of conditions. Though this could in theory be applied to thermal images too, this was considered a riskier option than standard imaging for this project.

2.8.8 Capabilities of Machine vision and machine learning with low-cost imagery

Machine vision and machine learning can be combined with drones, robots, mobile applications, and satellites, where the model could automatically process, in real time, the data it is collecting. The flexibility

and availability of a CNN make machine vision and machine learning a desirable option to continue with in this study for avocados, as an incredibly low cost, yet accurate method of counting and estimating [31, 52].

2.8.9 Machine vision and machine learning with low-cost imagery benefit through ground-based image collection

There are several studies relating to aerial or above ground data collection [21-23]. However, this type of data collection would not be able to quantify the size of the target fruit, or their location within the canopy. Applying machine vision and machine learning methods on images collected from within the canopy, would also benefit future research and investigations by other researchers in this area by providing a labelled dataset which could then be used to quantify the fruit and may also determine metrics like size, distance, colour, and average shape, which may help other studies.

Therefore, a machine vision-based approach involving the collection of a large, labelled dataset and the subsequent training of an existing state-of-the-art CNN for avocado detection and counting will be the focus of this thesis. As a result, the final sections of this review further examine the utility of such systems in the horticultural industry and some of the likely challenges around their implementation, especially those that relate to avocado.

2.9 Yield Decision Moving Forward

2.9.1 Machine vision and machine learning within the Horticulture Industry

Vision systems are increasing in popularity within horticultural industries [59]. They are useful in assisting a variety of practices like health and safety, post-harvest operations and more recently for identifying produce pre-harvest. The benefits of a machine vision system became more apparent as the COVID-19 pandemic prevented experienced workers from entering New Zealand [60]. Orchards became short staffed as a result, giving rise to pre-harvest problems like orchard management and estimation and also affecting post-harvest, as without experienced harvesters, many crops were left unpicked and thus dropped. By developing a machine vision model for certain tasks, it sets the framework for other assistive technologies to be developed, like picking robots and sorters which remove the need for repetitive tasks, resulting in a more productive orchard.

2.9.2 The problem of avocados

Avocado classification is complicated in object detection as their identifying features are very similar to their surrounding environment [31].

The ideal: A single green avocado selectively removed from its environment, placed against a white background in a well-lit room. The edges, skin texture, shape, size, and colour of the fruit can be clearly defined.

The actual: A green avocado within the dark canopy of a tree, surrounded by green oval shaped leaves all around, and half covered by another green avocado. The edges, skin texture, shape, size, and colour of the fruit are not well identified.

While occlusion is an ongoing issue for most applications utilising vision systems, avocados have other considerable disadvantages particularly when compared to ‘on tree’ kiwifruit, apples, oranges, and mangos, since similar colour and background features is a more significant factor in the avocado orchard. This is the current reality of avocados grown on commercial orchards, and it is bothersome for machine vision systems as it limits the features that are effective in identifying the fruit, as a result, there is a gap present in machine vision and machine learning applications regarding ‘on tree’ avocados.

CHAPTER 3

MATERIALS AND METHODOLOGY

This chapter reviews the data collection and also the justification of YOLOv3 and YOLOv5 as the convolutional neural network of choice. Appendix B contains a flow diagram to help explain the ordering withing this section.

3.1 Preliminary Data Collection, Equipment and Location:

Please note that this section only describes preliminary data collection upon first visit to the King avocado orchard. For more in depth detail about why this location was chosen, refer to 3.3.

- Zone 2, Block 49, King Avocado Orchard, Northland (Location)
- Canon EOS 550D (camera)
- YOLOv3 (CNN)
- Google Colab (platform)
- Python 3.8 (software)

Preliminary images were captured from the orchard using a Canon EOS 550D camera with the initial goal being to check if the avocados could be identified inside the canopy.

The YOLOv3 object detection neural network architecture was utilized in this part of the work. For training purposes, the network requires a series of images which have been manually labelled with 'bounding boxes' representing the location of avocados. During training, the network attempts to learn features and filters to identify avocados by altering network weights to minimize the error between the predicted and known avocado locations.

A preliminary set of 160 images were used, which were augmented horizontally flipping, giving a total of 320 images. 256 images were used for training, 60 for validation and 4 for testing. This is a typical dataset split in which approximately 80% of images are used for training, 20% for validation and a small number set aside for testing purposes. Since there was more than one avocado in each image, the total number of training instances (bounding boxes) was ~4000. See section 4.2 for preliminary results.

3.2 Plant Material

Persia Americana Hass are the target fruit of this study; hence Hass are also the subsequent target tree for this study. There are two variety of fruit trees on the orchard, Hass and Fuerte. Both have flowering cycles, where Hass is type 'A' and the Fuerte are type 'B', indicative of female and male flowers respectively. The timing of the Fuerte male flower compliments the timing of the Hass' female flowers and so the Fuerte are used to enhance the pollination and fruit bearing of the Hass, while the Fuerte itself will bear little to no fruit [61]. Section 3.3 goes more in detail on the timings for data collection as a result of this pollination process.

3.3 Time and Location

Data collection took place from August 3rd - 6th (2021) as avocados follow a specific pattern in their growth and maturing process, see (**Figure 3.1**). In southern hemisphere orchards, between April and August

is considered ideal time for counting and imaging [19]. This is after the previous harvest, after the removal of old fruit, after exponential fruit growth which increases the visibility of new fruit, and before the bloom of the next season's fruit.

Table 3.1 gives an overview of the location, trees and spacing as well as the number of trees sampled, and the amount of fruit counted. For quality data, the trees had to be similar, healthy, and resistant to irregular bearing [40, 41], hence access was obtained to a Northland orchard belonging to King Avocado. They are one of the largest avocado suppliers in New Zealand and maintain their orchard to a high standard. The dataset was focussed on zone 2, block 49 of the orchard (**Figure 4.1**), but zone 3, block 64 was also examined. The size and spacing of these trees (**Table 3.1**) was what the orchard manager described as the ideal tree for the future, which is a smaller sized tree allowing more trees per block, and harvestable by a single person without a cheery picker.

3.3.1 Environmental conditions during data collection

According to MetService [62], the weather conditions were heavily variant each day during data collection. Day 1 was an average of 11°C with occasional showers. Day 2 and Day 3 averaged 13°C, with overcast cloud clearing midday. Day 4 averaged 12°C with rain the whole day. Due to these unexpected weather changes, time became an issue and reduced the number of rows that were able to be eye counted as seen in section 3.4.1.

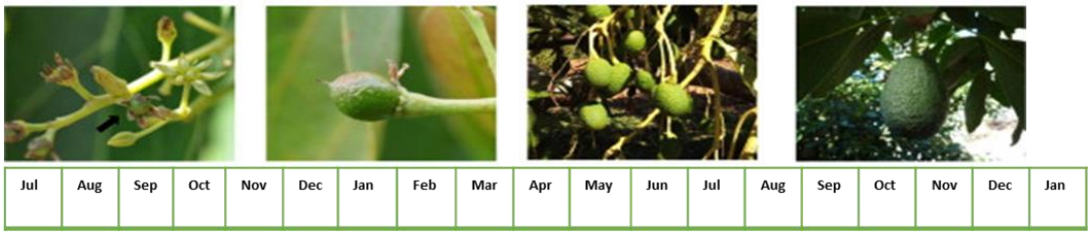

Figure 3.1. Showing the Southern Hemisphere avocado growth cycle, adapted from [63]

Table 3.1. Orchard Description used for Data Collection

LOCATION (LAT°, LONG°)	ZONE, BLOCK	ROW, TREE SPACING (M)	TREE HEIGHT (M)	ROWS IMAGED COUNT	HASS TREES IMAGED COUNT	ROWS SAMPLED COUNT	EYE COUNT TREES SAMPLED	HARVEST COUNT TREES SAMPLED
-34.987, 173.178	2, 49	5 x 2	3	12	433	5	132	87*
-34.983, 173.188	3, 64	3 x 3	4	3	71	1*	21*	0

* Collection was taken several months after photos and may be variant as a result, see section 3.4.4

3.4 Field Data Collection

3.4.1 Data Collection Design

Table 3.2 gives a description of the quantity of trees in each row of zone 2, block 49. There are 12 rows and 42 trees in each row, so 504 trees in total. Within this are Hass trees, Fuerte pollinators, new trees, diseased trees, and stakes indicating no current tree. Data collection hours were between 8am and 5pm, UTC +12. Data collection was shared between three researchers. To ensure that the data collected by each individual could be seamlessly integrated, each member followed the same image capturing process seen in section 3.4.3 and the same fruit counting process seen in section 3.4.4.

Table 3.2. Zone 2, Block 49 Cultivar and Fruit Count

ROW	HASS TREE COUNT	FUERTE TREE COUNT	NEW HASS TREE COUNT	NO TREE COUNT	DISEASED HASS TREE COUNT	EYE COUNT SAMPLE TREES	EYE COUNT FRUIT	HARVEST COUNT SAMPLE TREES	HARVEST FRUIT COUNT
1	40	0	2	0	0	0	0	0	0
2	35	7	0	0	0	0	0	0	0
3	37	0	4	1	0	0	0	0	0
4	36	0	1	4	1	0	0	0	0
5	32	9	1	0	0	32	1380	10	561*
6	38	0	3	1	0	0	0	0	0
7	37	1	4	0	0	38	1614	38	1783*
8	32	8	0	2	0	32	819	32	814*
9	40	0	2	0	0	10	374	3	190*
10	38	0	2	1	1	11	736	2	110*
11	31	9	0	2	0	9	162	2	85*
12	37	0	3	2	0	0	0	0	0

*Collection was taken several months after photos and may be variant as a result, see section 3.4.4

3.4.2 Field Data Collection Materials

To operate within budget restraints, a premium camera could not be used to capture images. However, images needed to be of sufficiently good quality to adequately allow for training. Given the capability of modern smartphones, an 8MP smartphone camera with 24 bit depth is capable of capturing 2^{24} colours [64] and is able to display it in 2K (2048x1080p), a greater definition than that of HD (1280x720p). The Vodafone P11 (aka the Alcatel 1B, TCL Communications, Hong Kong) costs under $100NZD and the rear camera specifications are as follows: 8 MP, f/2.0, wide, 1/4.0", 1.12µm, AF [65].

Alcatel 1B has a “wide” field of view [65]. When trialling the selected smartphone camera, a photo taken from a distance of 225 mm from a 225x300 mm object, revealed the angle of view (AOV) to be ~67° vertically and ~50° horizontally [66]. To assist in correctly orienting each picture, a handheld compass (Kathmandu, NZ) was used to orientate evenly around the circumference of the tree (see section 3.4.3). To

prevent battery discharge of the smartphone while capturing images in the orchard, a 15000 mAh portable charger and cable (Kmart, NZ) was used. To differentiate the trees from one another in the images, labels were attached to each tree, as seen in [52]. They indicated the row and tree number, for example, R4T7 is the equivalent of row 4 tree 7. Counts were recorded using a clicker counter and waterproof notebook (OfficeMax, NZ), see section 3.4.4.

A modification was made to the capturing hardware to speed it up. One member from MAFDL used an iPhone 11 (Apple, California) to capture images, which had a 90° horizontal AOV, and the following rear camera specifications: Lens 1: 12 MP, f/1.8, 26mm (wide), 1/2.55", 1.4µm, dual pixel PDAF, OIS; Lens 2: 12 MP, f/2.4, 120˚, 13mm (ultrawide), 1/3.6".

3.4.3 Image acquisition

The inside of the Hass canopy was used for the collection of the image dataset. This decision arose after personal communication with, the orchard manager, who advised that fruits are counted from within the canopy by counters, as this is the best way to view the avocados. There is clear contrast between the two perspectives (**Figure 3.2**), justifying the decision to capture inside the canopy.

Avocado trees have untrained 3-D canopies which creates more difficulty in capturing the harvest of a tree in any given image, unlike the trained 2-D like canopy used in modern apple orchards [67, 68], where images can more easily provide a quantification of fruit borne on the tree. To best represent the Hass canopy, several images need to be taken. To limit double counting of fruit, the pictures of any given Hass canopy should maximise the area of the canopy and minimise the overlap of each picture. Given that the horizontal AOV is 50°, and assuming the diameter average is 0.30 m for tree trunk and 3.00 m for Hass canopy in block 49, then 8 pictures would be needed to encompass the canopy to ensure the least overlap while capturing the most area (**Figure 3.3**). The smartphone camera was focused for each image and the exposure was manually adjusted to reduce the impact of back lighting (**Figure 4.3**). This study was made aware of the contrast adjustment method by machine learning expert, Dr Jaco Fourie of Lincoln Agritech. The method does result in some images having white backgrounds but allows for better visibility of the fruits inside the canopy and agrees with Kuznetsova [57] who did similar work with apples.

Alcatel 1B image capturing started by taking an image of the label of the current tree. Then the delegate positions at the base of the trunk, faces North, focusses the camera, adjusts the exposure, and takes the image. Then they rotate clockwise around the base of the tree to the next compass position (45° clockwise), repeating the capturing process until all compass directions are completed, then they move onto the next tree. The iPhone 11 images were taken in a similar fashion, but the starting direction was Northeast and clockwise rotation was 90° due to larger AOV (see **Figure 3.4**). All the captured images are sRGB colour representation and JPG format.

(a) (b)

Figure 3.2. Outside of the canopy (a), fruits are not easily visible, while inside (b) they are not

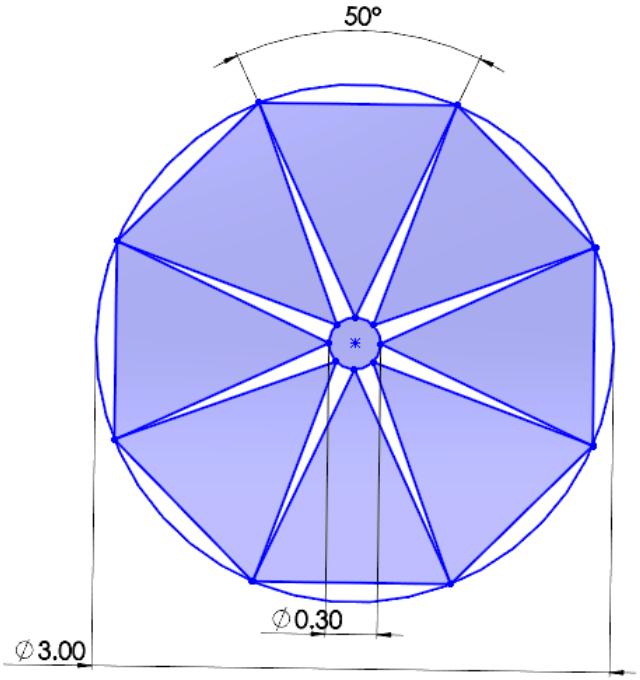

Figure 3.3. Sketch to visualise the canopy capture, where the blue triangles represent the AOV of the camera

(a) (b)

Figure 3.4. AOV of Alcatel 1B smartphone (a) vs the iPhone 11 (b)

3.4.4 Avocado Eye and Harvest Counting

Firstly, the eye count and harvest count were collected at different times of the year. The eye count took place in August, during the dates specified in section 3.3. The harvest count took place mid-November, the 13 - 17th.

Secondly, each delegate had a notebook to record, the zone, block, row, tree number, tree type, and the number of eye counted fruit. For example: zone 2, block 49, row 7, tree 6, Hass, 49.

Each delegate would split the Hass into quadrants to make the counting a more manageable process. The quadrants were North to East, East to South, South to West, and West to North. After counting the fruit in one quadrant, the delegate would record the count and move to the next quadrant. Before proper counting, each delegate used two trees to calibrate their counting ability, comparing with the other members post counting to check for error. **Table 3.3** shows reasonably consistent counting using this method. If at any point the delegate lost count, they reset the clicker counter and restarted the counting process for that quadrant.

Table 3.3. Practice eye counts

QUADRANT	N - E			E - S			S - W			W - N			TOTAL		
COUNTER	**A**	**B**	**C**	**A**	**B**	**C**	**A**	**B**	**C**	**A**	**B**	**C**	**A**	**B**	**C**
TREE 1	20	21	19	31	30	30	18	17	18	15	15	14	84	83	81
TREE 2	43	40	41	23	23	22	25	26	22	14	15	16	105	104	101

N: North, E: East, S: South, W: West

This study was not able to oversee the collection of the harvest count due to the effects of COVID-19 lockdown and had to rely on the orchard staff providing the results. This study provided King Avocado with a crop load protocol to follow when harvesting the avocados so that the counts would align with the specific trees from which data was collected by this study. Therefore, the purpose of this protocol was purely for consistency of data. Below is a simplified version of the protocol (See Appendix B for detailed protocol):

(i) Rows 5, 7, 8, 9, 10, and 11 were eye counted, hence only these rows shall be harvest counted
(ii) Pollinators and or young trees are not inclusive to the count
(iii) A single person will harvest each tree
(iv) Harvest personnel will count each avocado while harvesting into the fish box
(v) Harvest personnel write the number of fruits of each tree on the tag attached to each avocado tree
(vi) Fruit will be left in the fish box underneath the tree
(vii) If a tree has a zero count, write "0" on the tag
(viii) Personnel leading the harvest operation may double check the counts in the fish boxes to double check counts
(ix) Personnel leading the harvest operation may write down the count of each tree written in the tree tag in the attached table

3.5 Image analysis (Machine vision)

3.5.1 Image naming and labelling

Over 2000 images were taken to ensure an adequate number of images to train the model were collected [69], with the final augmentation dataset being ~4200 images, within which are over 40000 training instances (bounding boxes) of avocados.

This study made a script to rename the images using the Python programming language (Appendix B). Since the images were all taken consecutively, they were naturally in order according to date. The script passed through the images in a copied directory and renamed them to the following format, R#_T#_Direction. For example, once renamed, row 4, tree 7, Northeast Direction would be: R4_T7_NE. After the copied images were renamed, they were manually checked to ensure that they matched with the order of the original images. If the order was incorrect, the incorrect set of renamed images were deleted, and a new set of copied images were named, correcting for previous mistakes.

Following this was the application of bounding boxes (labels) to the avocados within the images of the dataset. This study used the YOLO format of labelling which normalises the bounding boxes as a fraction of the image width and height, therefore if the image size changes, it does not affect the bounding box. Documentation provided by Glenn Jocher [69], the creator of YOLOv5, explains best practices when labelling a custom dataset. For label consistency, every object within an image must be labelled, incomplete object labelling will lower the network's performance. For label accuracy, boxes should be as close to the object boundaries as possible, leaving no space in between. This study used an open source labelling program called LabelImg [70].

When using LabelImg, the user inputs a directory of images into the software, and for each image, boxes can be graphically drawn around objects using the cursor. In this case, avocados were the objects. Within (**Figure 3.5**), the first image shows an example of a graphically labelled avocado. This graphical box then gets transferred into a set of normalised coordinates, which YOLOv5 also predicts using equations (3.1) - (3.4) [11]. This is then saved into a YOLO format text file. The highlighted line of the text in (**Figure 3.5**) shows YOLO label format (c, b_x, b_y, b_h, b_w), where c is the class number (0 for only 1 class), b_x & b_y are the coordinates for the middle of the bounding box, and b_h & b_w are the dimensions of said box. For the example in **Figure 3.5** if the original image has dimensions 2048 x 2048 pixels, the middle of the bounding box will be at (1315, 1304.5) and its dimensions are 138.5 x 176.

$$b_x = \sigma(t_x) + c_x \quad (3.1)\ [11]$$

$$b_y = \sigma(t_y) + c_y \quad (3.2)\ [11]$$

$$b_h = p_w e^{t_w} \quad (3.3)\ [11]$$

$$b_w = p_h e^{t_h} \quad (3.4)\ [11]$$

Where,

(c_x, c_y): Offsets from the top left corner

(b_x, b_y): Coordinates for bounding box centre

b_h, b_w: Dimensions of bounding box

P_w P_h: Dimensions of the predicted box

$\sigma(t_{x\,,\,y})$: Cell wall offset to bounding box centre

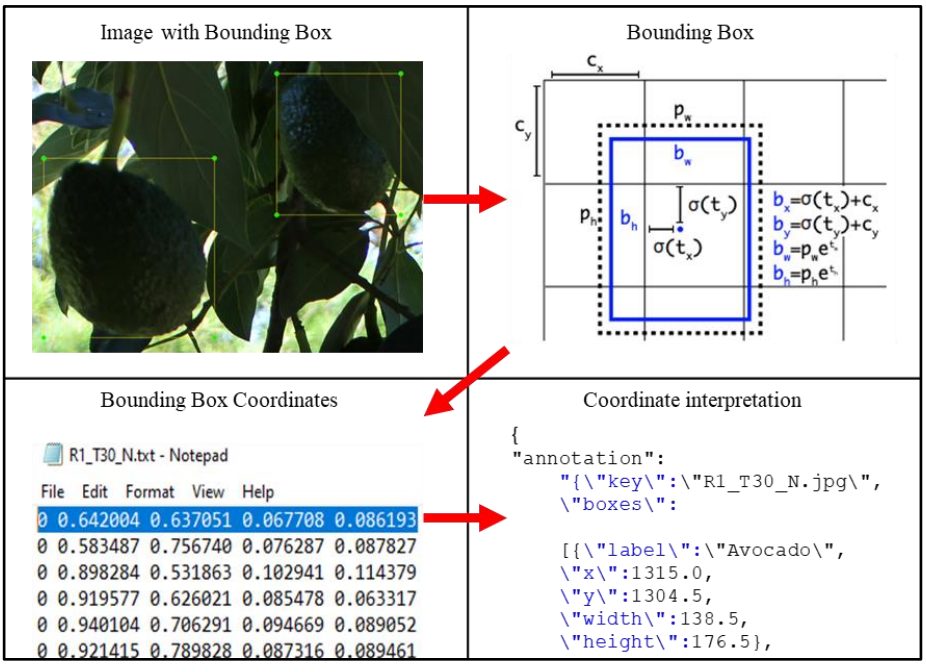

Figure 3.5. Flow diagram of YOLO labelling process

3.5.2 Image Augmentation

The next pre-processing step after labelling is the augmentation process, which entails editing certain features of the dataset to create a larger dataset which can result in better results from the training process. To apply augmentation to a dataset, there are a number of computer vision libraries available to choose

from, for example the OpenCV library is quite popular, compatible with python and C++, and is very well documented [71]. There are also two main ways to apply augmentations to an image, spatial augmentation, which changes the geometric characteristics like size, flip and rotation, and pixel augmentation, which changes the pixels within the image but leaves its geometric values the same [72]. The safe application of spatial and pixel augmentation depends on the test images. For example, when trying to identify a '6' or '9' the vertical and horizontal orientation matters, but less so the colour. In the case of green apples vs red apples, spatial augmentations will have no effect, but pixel augmentation like saturation can allow for more identifiable colour. This study used Roboflow, an online platform which helps with the pre-processing and augmentation of images for machine learning [73]. It is very simple, merely upload the dataset of images along with their respective label text file and Roboflow will create a project folder for this dataset. It is common in machine learning practices to split the dataset into three directories:

(i) training, images the network uses to train
(ii) validation, images the network uses to validate performance
(iii) testing, images the network has never seen and are used for inference.

A typical distribution is 70% training, 20% validation and 10% testing which Roboflow will do automatically (see

Figure 3.6). However, these ratios can be adjusted to suit desired performance needs. This study had 4280 images. 3524 in the training set, 504 in the validation set and 252 in the test set.

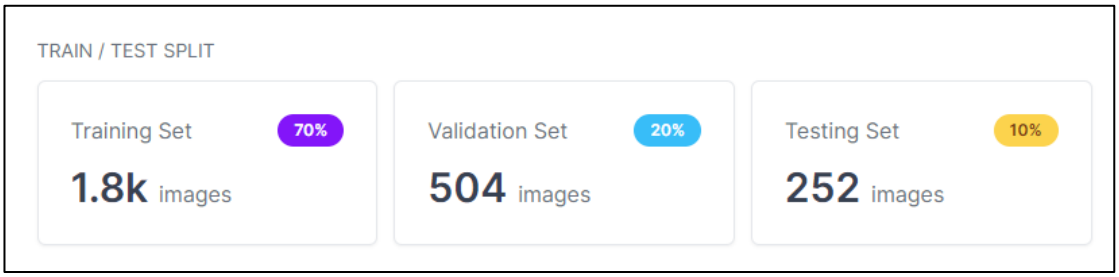

Figure 3.6. Example of uploaded images distributed into different sets using Roboflow.

Roboflow has 23 augmentations that can be applied after uploading [74]. A clever way to double a data set is to flip the images and bounding boxes horizontally or vertically. Following this, other augmentations like cropping, rotation, shear, grey scale, hue, noise, cut outs, mosaics, saturation, brightness, exposure, and blur can be applied. However, only the augmentations which were considered likely to be beneficial to the network were chosen. Assumptions made when deciding upon the above augmentations:

(i) Images will be used to identify avocados from an orchard in varying conditions and with cameras of varying quality, therefore brightness, saturation, blur and exposure augmentations will be useful
(ii) The avocado shape is similar between fruits; therefore, shears are not needed
(iii) The distance or relative size of avocados will vary naturally; therefore, cropping and cut outs are not needed
(iv) There will be multiple avocados in most images; therefore, mosaics are not needed

(v) Images will always be taken in colour and avocado colour is mostly green; therefore, grey scale and hue are not needed.

(vi) Since avocados are relatively uniform, this study opted also for horizontal and vertical rotations

When applying the chosen augmentations, judicious decisions need to be made, because the intensity of the changes will affect the outcome of the network. Since the desired outcome was to detect avocados amongst a complex environment, a decision was made to modify saturation, brightness and exposure by ±25%, which seemed to give enough variance but left visible, eye countable fruit. Gaussian blurring of the images by 1.5 pixels caused the neural network to place less emphasis on the texture of the fruit and more on the colour and shape which are the more defining characteristics at a distance (**Figure 3.7**).

Figure 3.7. Collage of augmented images with labels. The red boxes show the label of the avocado. The number '0' is the index of the class label, since avocados are the only class, the index is 0.

The reason for the emphasis on the minute details relating to this dataset, is because it is unique. There are currently no publicly available datasets for avocados in these volumes, or within an orchard environment. The final augmented dataset features over 4200 images and over 40 000 instances of avocados, captured with two different smartphone cameras and encompasses varying lighting and weather conditions. All images were labelled, this provided ample material to train a quality network [75, 76]. While there was the

potential to increase the dataset size more, doing so would have been unproductive due to the constraints of the hardware within this study.

3.5.3 Justification of YOLO

As discussed in section 1.2.2.2, YOLO is a real time convolutional neural network used to train image detection models. The benefit of using YOLO is the speed at which it can detect compared to other networks, whilst still maintaining accuracy comparable with other state-of-the-art object detection architectures. It should be noted that YOLOv3 was used initially for preliminary results shown in section 4.2 as its faster training time allowed for improvements to be implemented that would aid the objectives of this thesis. But YOLOv3 is known to have lower accuracies than that of newer YOLO versions (YOLOv4 and YOLOv5) [77]. Hence, the reader will see this jump from v3 to v5 in section 4.2.

Adding to this, [12, 57, 77-79] have used YOLO quite successfully to detect and quantify objects of interest, therefore in view of the objectives of this thesis, YOLO seems like an adequate convolutional neural network to choose for object detection and quantification, with YOLOv5 being the best choice for the objectives of this thesis.

3.5.4 Model Overview of YOLOv5

YOLOv5 is the fifth iteration of the YOLO framework and utilizes Cross Stage Partial Network (CSPNet) as its backbone and Path Aggregation Network (PANet) as its neck. As a result, its architecture is split into three: backbone, neck, and head, where each performs a different task. (**Figure 3.8**) shows a flow map of this process, and Appendix B contains example code of this process.

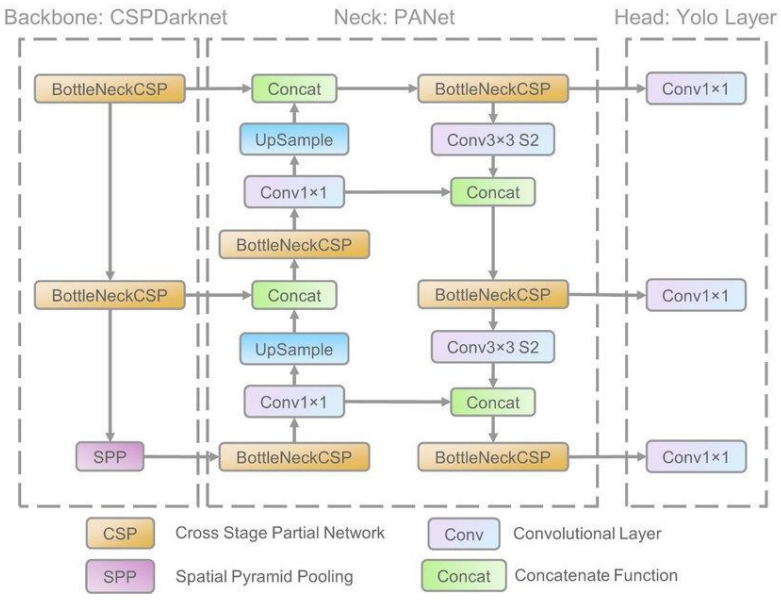

Figure 3.8. Model overview of YOLOv5 [80]

3.5.5 Metrics to assess performance of Machine learning networks

This study used precision, recall, mean average precision (mAP), and mAP (with 0.05 stepped IoU thresholds between 0.5-0.95) to assess the performance of the machine learning network. These metrics were used to assess various models in the 2010 Pascal VOC challenge, in which they clearly state a vision to provide 'standard evaluation procedures' for machine learning communities [81]. YOLOv5 also uses the

Jaccard or IoU (intersection over union) threshold to justify whether a prediction is a true positive or a false positive (3.5). Essentially, if both predicted and ground truth boxes are perfectly aligned, IoU = 1.

$$IoU = \frac{area(B_p \cap B_{gt})}{area(B_p \cup B_{gt})} \quad (3.5)[82]$$

Where, B_p = Predicted bounding box

B_{gt} = Ground Truth Bounding Box

The precision metric in (3.6) refers to the ratio of correctly predicted positives compared to the total number predicted. If there are no false positives, precision = 1.

$$Precision = \frac{TP}{TP + FP} \quad (3.6)[83]$$

Where, *TP*: True positive (Predicted as positive and was correct)

FP: False positive (Predicted as positive but was incorrect)

The recall metric in (3.7) refers to the ratio of correctly predicted objects compared to the labelled objects. If there are no false negatives, recall = 1.

$$Recall = \frac{TP}{TP + FN} \quad (3.7)[83]$$

Where, *FN:* False negative (Failed to predict an object that was there)

The AP (average precision) in (3.8)-(3.9) is a very popular metric used to evaluate detection models within scientific papers (note, for a singular class, mAP = AP). AP used to interpolate 11 evenly spaced points along the precision and recall curve, but now interpolates all the points on a precision and recall curve and determines the average across those points. This means $\overline{Precision} \neq AP$ and should not be confused as such. Essentially, the higher precision and recall, the higher the AP.

$$AP = \frac{1}{11} \sum_{r \in (0, 0.1 \cdots 1)} Pinterp(r) \quad (3.8)[81]$$

With,
$$Pinterp(\mathrm{r}) = \max_{\tilde{r}:\tilde{r} \geq r} p(\tilde{r}) \quad (3.9)[81]$$

Where, $\rho(\tilde{r})$: the measured precision at recall $\tilde{r}$

3.5.6 Training YOLOv5 model using Google Colab and Local Environment

In order to train a model on a custom dataset, a training environment needs to be set up. There are simple cloud based environments like Google Colab, Kaggle or Amazon AWS, but they often come with

disadvantages, like paying for GPU time or having limited training time. Alternatively, a local environment can be set up but that too has disadvantages, such as hardware and software dependencies resulting in a number of runtime errors. Early attempts to set up a local environment were thwarted due to dependency errors, so Google Colab was used as an interim measure because it enabled results to be gathered until the local environment could be established.

For the local environment, this study trained on an NVIDIA Quadro P2200. Apart from the YOLO requirements, the following software installs were required in order for the GPU to be used: cudatoolkit = 10.2.89, cudnn = 7.6.5 = cuda10.2_0, pytorch torch vision torchaudio cudatoolkit = 10.2.

The training was done with varying model types (**Figure 3.9**), varying image sizes, and a varying number of epochs, but batch size remained fixed at the default of 32.

Google Colab (GC) provides the user with free access to large amounts of hardware but restricts the time usage to a maximum of 12 hours. In practice, this tended to be half that, at around 6 hours. It made the training of datasets a tedious ordeal as an active session would timeout if the user was not active on the page. When GC times out, the session, along with any data, is lost. Multiple partly-trained models were lost due to page time outs which was an unfortunate loss of time as each took upwards of 4 hours. This study recommends a local training environment over GC, if possible, but local networks can be limited in hardware.

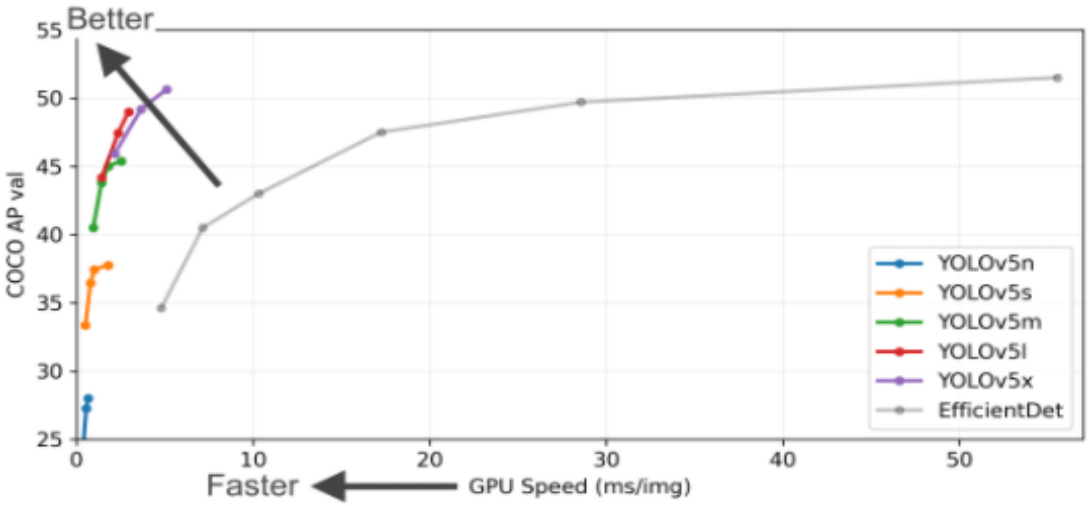

Figure 3.9. Showing examples of the different YOLOv5 results when trained on the COCO dataset using different model types [84]

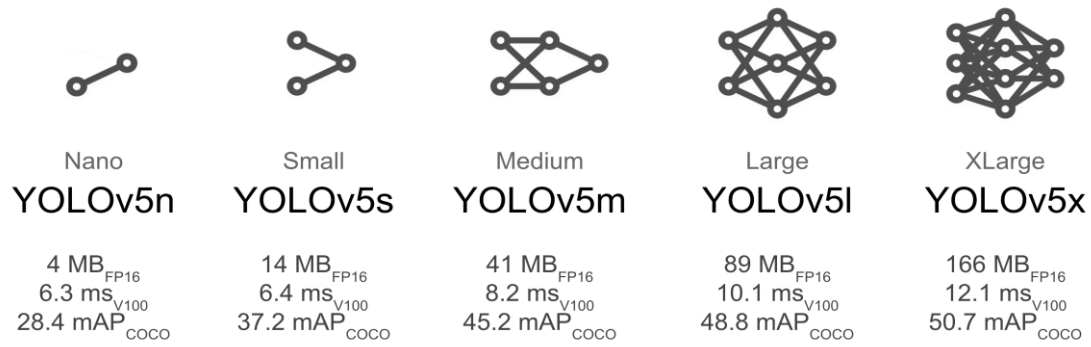

Figure 3.10. Showing the sizes of the different YOLOv5 structures, along with their speed and expected performance [69]

3.5.7 Statistical Data Analysis

Metrics from section 3.5.5 are used to assess the machine learning models. A model under 50% average precision (AP) is poor as a classifier, any model equal to 50% AP is a random classifier showing no discrimination between object and non-object, and any model above 50% AP is no longer a 'random classifier' [85], but more often than not will predict the correct object, with 100% AP being the ideal (note that mAP = AP for 1 class). Realistically a model will never reach the ideal, but a model approaching 100% (e.g., 85-99% AP) is considered a good-excellent model at classifying.

As for counts, the sample sizes were 87 orchard harvest count trees, and 133 eye count trees. YOLOv5 prediction counts were made for all sample trees. The pictures were input into YOLOv5 using a fixed confidence threshold of 0.6. The predicted fruit counts for the pictures were summed and stored as the total predicted count value for that tree. Linear regression was then performed for each machine learning model using Microsoft Excel [86]. The regression compared model performance for orchard harvest count, eye counts and the predicted counts.

3.6 Order of results in accordance with methodology

This section is for the benefit of the reader in preparation of the results section as there are many ideas working in conjunction. Hence, purely reading the results section may lead to a lack of understanding of the motivation behind each result.

Section 3.1 briefly discusses the methodology from an initial trip and its results link to section 4.1 and 4.2. From this initial trip, the orchard was mapped, and the viability of YOLO was checked, where version 3 was used for its inference speed.

Methods in sections 3.2, 3.3, and 3.4 were used to gather the most accurate image, eye, and harvest counts possible, these results can be seen in section 4.3.5.

Section 3.5, which goes into detail of image analysis and machine vision methods, links to the results shown in section 4.3. In which, section 3.5.6 which references model complexity and Colab links to section 4.3.2.

Image methods in section 3.5.2 were applied to all the training and validation datasets used to gather results shown in section 4.3.3, where differences were found between smaller and larger inference images. Section 4.3.2 and 4.3.3 were used as motivation to determine the best detection model for further results.

This chosen model would then:

I. Compare bounding box avocado counts (gathered with the method in section 3.5.1) against its predicted count using regression analysis, shown in section 4.3.4
II. Compare the orchard harvest count shown in section 3.4.1 against its predicted count using regression analysis, shown in section 4.3.5

The results seen in section 4.3.6 were to take advantage of the real time capabilities of YOLOv5, however, future work could look to develop this idea further.

CHAPTER 4

RESULTS & DISCUSSION

This chapter reviews the effectiveness of the trained model in its ability to detect, and quantify the avocado. For the benefit of the reader, the results section will flow in accordance

4.1 Orchard Data and Counting

To provide a relevant point of comparison for the results of this work, the precision, accuracy and time required by the current avocado counting methodology employed on orchard was investigated. The King Avocado orchard in Northland is broken down into different zones by orchard management, as shown in the layout depicted in (**Figure 4.1).** Also note the orchard's problem in section 1.1.

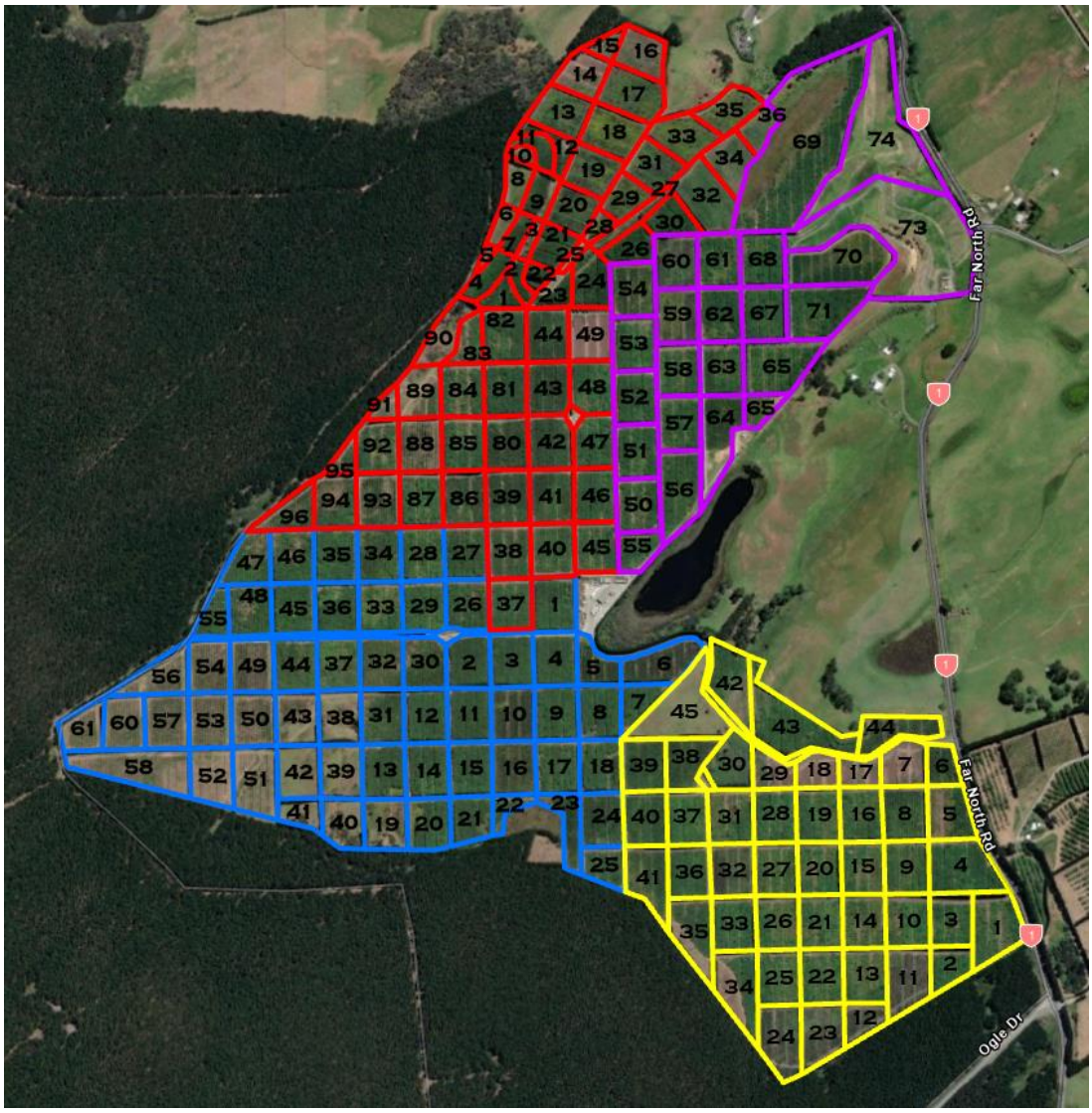

Figure 4.1. King avocado orchard broken down into the respective zones; yellow: zone 1, blue: zone 2, red: zone 3a, purple: zone 3b (note that zone 3a and zone 3b are combined to form zone 3). The numbers inside the zones represent the block numbers).

Currently, King Avocado estimates are made on a block-by-block basis. This is achieved by sending counters to certain rows of a given block, which are typically spaced about the centre of the block. The

counters have clickers and are instructed to only count trees which are not pollinators. The counts for all the trees in each counted row are then averaged to give a single tree basis which is subsequently multiplied by the number of trees in the block to provide an estimate of the count for that block. The summation of blocks in each zone is then used as an estimate of the count of avocado in that particular zone (**Figure 1.1**).

4.2 Avocado detection using "You Only Look Once" (YOLOv3) software

Figure 4.2 shows detection results from YOLOv3 under relatively ideal or best-case imaging conditions, which were created by using similar fruit sizing (uniform camera distance), contrasting background colour, well-lit, even lighting, visible texture and no occlusion.

With ideal conditions, all 16 avocados present in the photo are detected and the confidence levels are between 91-100%. This result showed promise that a CNN algorithm like YOLOv3 can detect avocado fruit and could be used to replace eye counting. Hence tests were then done using a small dataset of photos captured from the orchard.

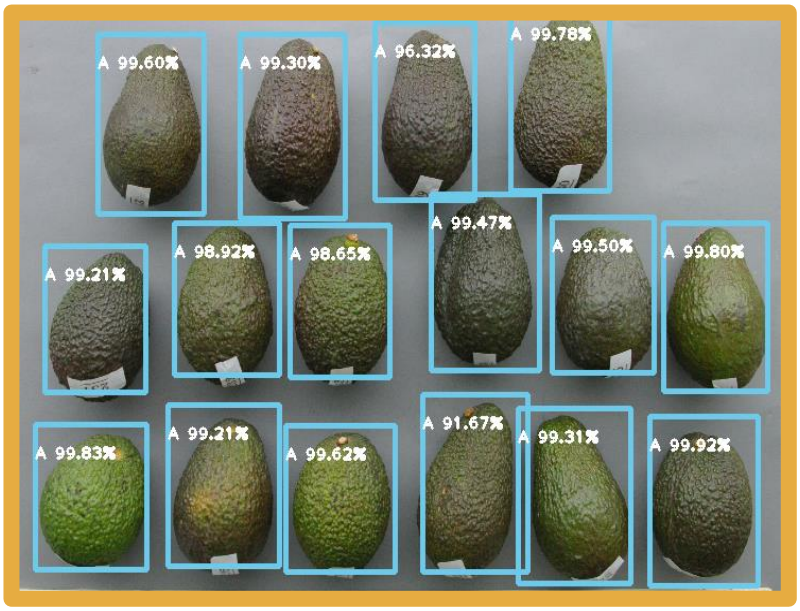

Figure 4.2. Detection results of YOLOv3, under relatively ideal detection conditions. 'A' is the label representing avocado, and the percentage value is the confidence level.

But these are not the real conditions of the avocados on the orchard, like that seen in **Figure 4.3**. The conditions present on the orchard contrast heavily with the ideal conditions; varying fruit sizing, canopy colour (which is similar to green fruit colour), leaf shape similar to avocado shape, uneven lighting, skin texture not as visible, and occlusion (covering of fruit) is present. Of the 31 visible avocados, 15 are occluded, with 6 of the 15 being heavily occluded.

Of the total 31, the model detected 15 avocados, 13 of which are the un-occluded avocados. Also, the confidences have dropped, where the new range is between 55-100%. The drop in detection and confidence is due to non-ideal conditions like similar canopy colour, occlusion, and the small dataset (small number of pictures). There was not enough variation in the training dataset for the model to detect 'difficult to spot' avocados.

Figure 4.3. (Top) Showing a detection result of the YOLOv3 model on a test photo from the preliminary orchard photo dataset, real detection conditions. (Bottom) A modified version with saturation and brightness increased by 100% respectively. 'A' is the label representing avocado, and the percentage value is the confidence level.

Table 4.1 summarises the results for **Figure 4.3** and shows the result when the same image was adjusted to increase the saturation and brightness by 100%. This result shows that augmented images can improve detection as it creates more ideal conditions for the model in detecting the fruit. The detection and confidence levels have both improved, with the total moving to 19 fruits being predicted and the confidence increasing, with a range of 65-100%. Also note the confidence levels of the previously detected fruits have increased, with some confidences improving from 50-60% to 90-100%. The predicted detection within this

photo is 19/31 avocado or ~60%. Therefore, improvements can be made to the training dataset to assist in identifying the fruits, this came through a larger dataset and through augmenting the photos to provide greater variability.

Table 4.1. YOLOv3 Detection Comparison

TYPE	NUMBER OF FRUITS DETECTED	NUMBER OF OCCLUDED FRUIT DETECTED	MEAN CONFIDENCE
IDEAL IMAGE	16/16	0/0	98%
ORCHARD IMAGE	15/31	2/15	80%
AUGMENTED ORCHARD IMAGE	19/31	5/15	88%

Confidence ideal is theoretically 100% but is unrealistic, hence 90-100% is ideal

4.3 Avocado detection using YOLOv5

4.3.1 Dataset size, Augmentation

This study trained on a dataset of 4280 images. 3524 in the training set, 504 in the validation set and 252 in the test set. The labelling of the dataset was exhaustive. There are usually multiple avocados in each photo, hence images took around 4-5 minutes to accurately label, where one bounding box represents one avocado instance. Around 160-200 hours of work were required to accurately label every visible avocado in each image, but this is common for machine learning practices and will not change until faster labelling methods are discovered [87].

Increasing the dataset size has shown to improve the performance of a model as it allows for greater training variance, producing better results when tested. There is improved mean average precision (mAP) for the augmentation model (**Figure 4.4(a)**) compared to no augmentation (see also section 4.3.2). The effect of dataset size shows to improve detection with increasing amounts of training images (**Figure 4.4(b)**).

This result is similar to another machine vision study (**Figure 4.5**) which showed the effect on average precision (AP) for increasing ‘training dataset’ size. The study involved detecting mangos using YOLOv3 [88]. Adding to this, another study involving tomatoes using a custom YOLO model, showed an increase in precision and recall (**Figure 4.6**) with increasing training images [78]. Hence the decision to use a large training dataset was justified.

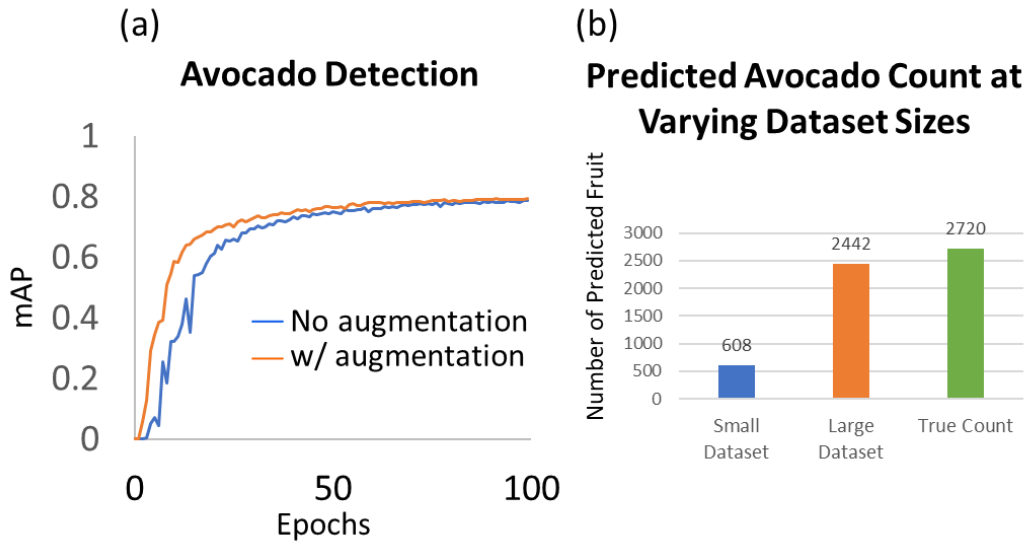

Figure 4.4. (a) Effect of augmentation on mAP from this study **(b)** Effect of increasing dataset size on detection, **Small Dataset**: 350 training images, **Large Dataset**: 3500 training images, **True Count**: Number of visibly counted avocados in the images.

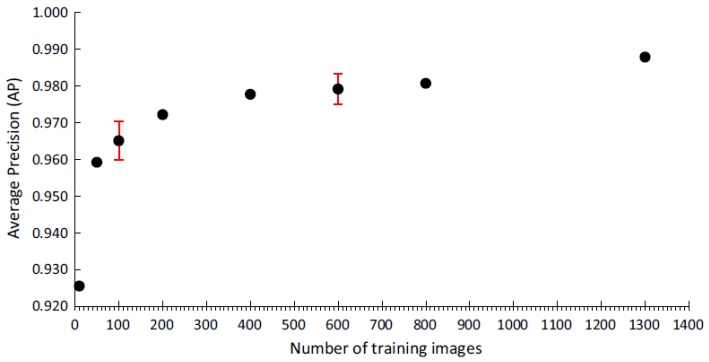

Figure 4.5. Effect of increasing dataset size on average precision (AP); from a similar machine vision study on mangos [88].

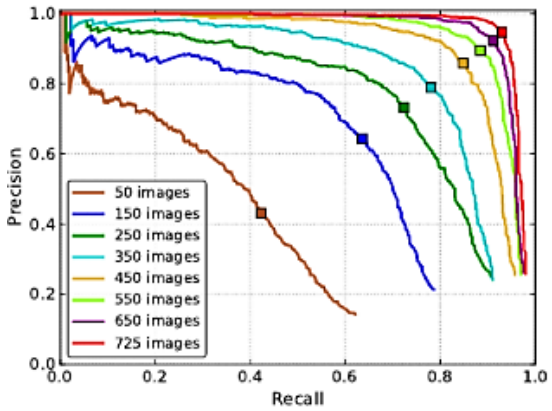

Figure 4.6. Effect on performance when dataset size increased; from a similar machine vision study on tomatoes [78].

4.3.2 Model Types

The purpose of this section serves as motivation for finding the best model to quantify the avocado fruit, i.e., which model type would be the best for detecting avocados.

Before the local (offline) training environment was established in this study (see section 3.5.6), results were captured using Google Colab (GC). GC's free environment was capable of training small ('s'), medium ('m') and large ('L') models before running out of memory or the web page timed out (12-hour time limit). Local environment training was restricted to 's' models (**Figure 3.9**). Note 'no_augments' (**Figure 4.7**) was trained using model 's' and refers to unaugmented training images. While 's', 'm' and 'L' (small, medium, and large respectively) refers to the size of the YOLOv5 model architecture and were trained with augmented images [69]. Note that metrics seen in 3.5.5 gauge the performance and these results are shown in (**Figure 4.7**) and compared in **Table 4.2**.

There is little difference between the precisions of the different model types in (**Figure 4.7** (a)), however the no augmentation model scored lowest with 87.15% and model 's' performed best with 88.09%. What can be noted is that the augmentation models pick up on avocados more quickly, but all models plateau at ~0.87. The closer precision approaches to 1, the better, however 0.88 shows promise when compared to similar studies [55, 78, 88]. The precision plateau is likely to be due to the small image size (e.g. < 1024x1024) and other objects different to an a avocado fruit in the image, such as leaves, branches, shadows, and overlapping fruit (occlusion).

Precision can often be improved by increasing the dataset size and clarity of the training/validation images. This study had the capacity to increase its training dataset size more but was limited in hardware and time to do such, but it presents an opportunity to improve this network in the future. Despite this, the dataset is already of substantial size (e.g. 3500+ training images), so this work examined image size as limiting factor, see section 4.3.3.

(**Figure 4.7** (b)) shows little difference in recall between the varying models, with the no augmentation model scoring lowest 69.85% and the 'L' model scoring best at 71.41%. All models seem to plateau at ~0.7. At this image size, and for all model types, 70% of the predicted avocados are within the labelled dataset. Ideally the recall should be reaching ~1, but as mentioned before, there are limiting factors and other objects in the image which are likely causing this plateau effect on recall.

The improvement of recall may lead to drops in precision, impacting other metrics. To improve recall means to change the way in which the dataset is labelled. Occlusion and bounding box size within the dataset images are likely the largest influencers to lower performance in recall. The original images, before they were scaled down, had resolutions of 8MP or more. As is practice, all avocado instances are labelled, even distant ones on other trees [69]. When the images are scaled down, distant fruits become very small causing the bounding boxes to become very small and this may have caused problems for the network detection. Likewise with occlusion, bounding boxes overlap, especially within bunched fruit, creating confusion for the network as to where the avocado is. The best way to deal with this is to train on larger image sizes and this is investigated in section 4.3.3. Hardware limitations meant training images could go no higher than 512x512 pixels. Future work could investigate higher resolution images.

Another possible solution would be changing the bounding box shape. This study used rectangular labels as YOLOv5 naturally uses this type to train on, but there are other networks that can handle splined labelling, like R-CNN, which would fit the avocado shape more precisely. However, splining would take significantly more time to label each avocado exactly, and the solution could not be implemented in real time, justifying this study's choice to continue with YOLOv5. Future work could examine circular labelling to improve performance, like in [78], which is similar to rectangle prediction, but instead predicts a circle.

Figure 4.7 (c) examines the mean average precision (mAP) of the models. There is noticeable difference in the values during training, however there is still little difference between them, where ~0.8 seems to be the plateau. The no augmentation model scored lowest with 78.81%, while model 'm' scored highest with 79.62%. This is a promising result, but compared to similar studies, the mAP of reliable networks may be closer to or above 90%. This study's best model was closer to this being 87% (see section 4.3.3). However, the model types in this section ('s', 'm', 'L') were to test if the structure impacted on performance, but it appears structure had little effect on performance. However, dataset size (number of images) and augmentation did have a noticeable impact on performance.

The mAP metric was calculated with an intersection over union (IoU) value of 0.5, meaning half the area of the predicted box must overlap with the ground truth box to be considered a true positive. The mAP could be improved by lowering the default YOLOv5 IoU threshold of 0.5 from to 0.4 or 0.3. However, doing so may be counterproductive as it would also impact the precision and recall metrics which make up mAP; and is therefore not recommended.

The difference between (c) and (d) in **Figure 4.7** is that the mAP values in (d) are calculated for changing IoU thresholds, i.e. every epoch, the mean of the mAP values with IoU values incremented in steps of 0.05 from 0.50 to 0.95 is given. Firstly, note the drop in values compared to (c), where the plateau has dropped from ~0.80 to ~0.52. In a deep learning study using this metric [89], they also report this drop in mAP 0.5:0.95. This is because the base IoU threshold in YOLOv5 is 0.5, but when that cut off point increases and becomes higher, it causes what used to be a TP (true positive) to become a FP (false positive). As a result, the precision will drop due to the reduced number of TP and the increased number of FP, leading to a drop in mAP.

There is also a noticeable difference between the unaugmented model, 416x416_no_augments, and the augmented models, 416x416_s, 416x416_m, 416x416_L, whereas this was not true before. The effect of the augmentations applied to the dataset have therefore improved the IoU predictions at higher values and are observed through this metric.

The best performing models are the models trained on augmentation datasets, while the architecture of model, 's', 'm' & 'L', had little difference at this image size. At this image size and number of epochs, ~0.52 seems to be the maximum.

The way to improve this metric is to decrease the number of false positives and or improve their IoU values. Generally, providing a larger dataset or increasing the image size are the two possible ways this can be achieved. By increasing the image size, naturally the labels will also increase in size, enabling the model to become more accurate in its predictions.

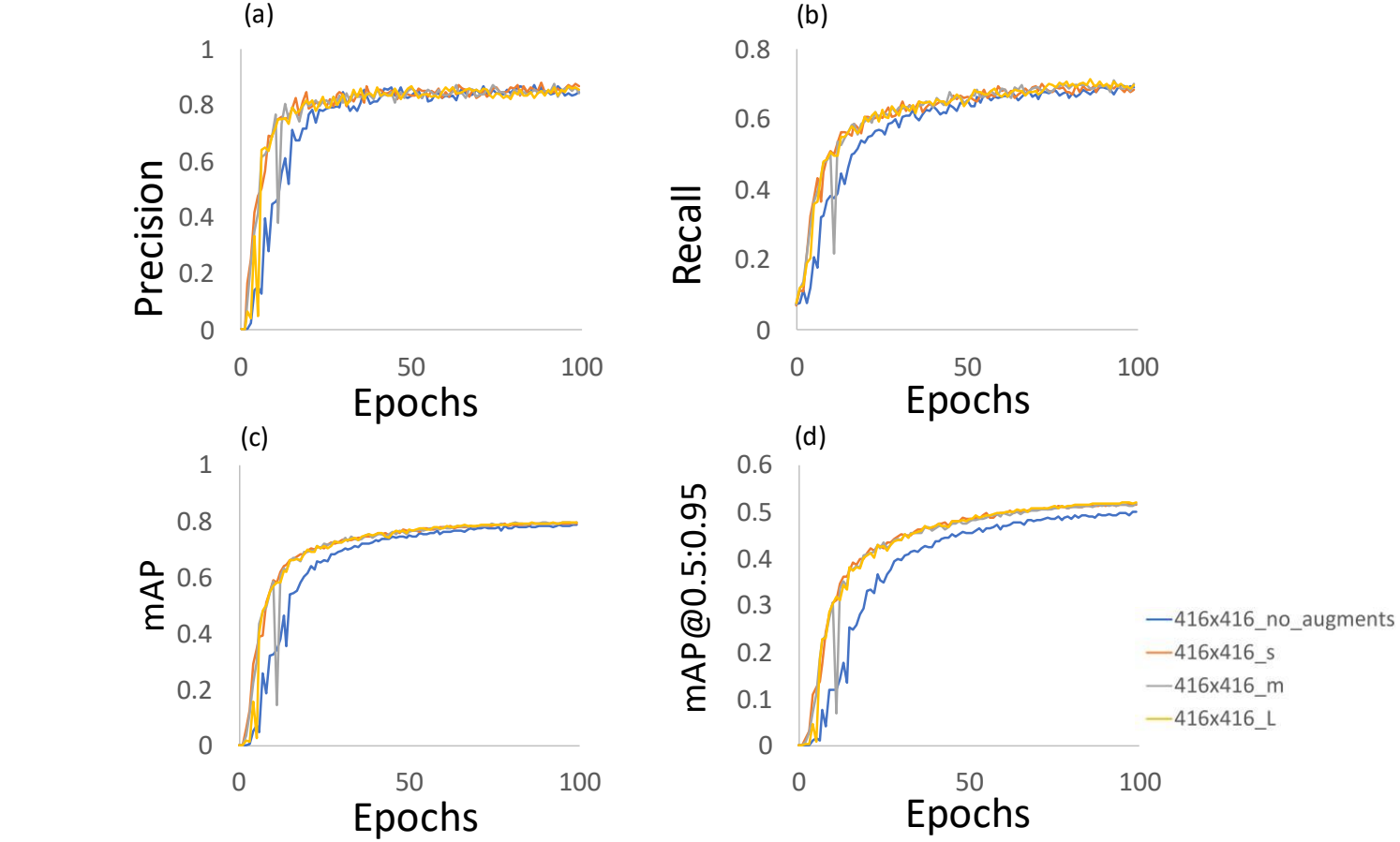

Figure 4.7. 416x416_no_augments(blue) 's' model with no augmentation, 416x416_s(orange) 's' model with augmentation, 416x416_m(grey) 'm' model with augmentation, 416x416_L(yellow). All models were at 416x416 pixel size.

Table 4.2. YOLOv5 model detection performance at fixed 416x416 image size

MODEL	PRECISION	RECALL	mAP	mAP 0.5:0.95
UNAUGMENTED (MODEL S)	87.15%	69.85%	78.81%	49.98%
AUGMENTED (MODEL S)	**88.09%**	70.21%	79.29%	51.78%
AUGMENTED (MODEL M)	87.82%	71.07%	**79.62%**	51.70%
AUGMENTED (MODEL L)	86.90%	**71.41%**	79.53%	**52.08%**

Bold font: BEST

Regarding the objectives of the thesis, the best performing model should be chosen to advance to the next stage of results as it should quantify the fruit the best. What can be concluded though is at the same image size and number of epochs, augmentations have more effect on the model than the type of model. Therefore, since model type 's' took the least amount of time to train, it was chosen to present further results. However, conclusions about image sizes needs to be investigated to further check the viability of YOLOv5 as a counting method for avocados and to determine if further improvements could be made.

4.3.3 Detection with Different Image size

Shown in (**Figure 4.8**), are model performance results for image sizes and epochs (number of training cycles) in a similar fashion to section 4.3.1 where precision, recall, mAP and mAP 0.5:0.95 are examined respectively.

(**Figure 4.8** (a)) showed the precision of smallest image size, 224x224, performing the worst with 82.53%, while the largest image size model 512x512_s performed best at 89.61% (note that 's' is the base model, and 's6' is a modified structure of model 's' designed to provide better and faster results from training).

448x448_s6 performed better than its base counterpart ,448x448_s, in all metrics, however, 512x512_s6 performed worse than its base counterpart, 512x512_s, in all metrics, which clash with the performance metrics Glenn Jocher (YOLOv5 creator) stated this model should achieve [84]. This study reran the training several times to attempt to correct the performance dips (seen at epoch 60 and 70) with no success. These performance dips did not happen with 448x448_s6, but based on documentation from PyTorch, it is possible that this study's local environment had unreproducible conditions at this image size [90]. This issue did not happen with any other model, so this study has confidence its methods were not the cause.

Another interest is the training length, of which three models were trained for 100 epochs and three were trained for 200 epochs. Notice that the number of epochs after 100 did not have a meaningful effect on the overall performance of the model, but rather the model is at risk of over training, having already reached its plateau after 100 epochs. This performance plateau was also seen for recall, mAP, and mAP 0.5:0.95. **(Figure 4.8** (b)) examines recall performance and shows the largest image size model 512x512_s performing best with 79.76%. While again the smallest image size model 224x224_s performed worst with 52.49%. This makes sense based on how the network detects the avocado fruits. The texture, shape and

colour of the avocados are defining features of the avocado, by scaling an image down, there are less pixels available to define those features to the network, hence it will predict more FN (false negatives) at lower image sizes. An increase in FN causes the recall value to drop, but a reduction in FN will cause the recall value to increase. Therefore, 512x512_s has the least FN, meaning that it performs the best in correctly detecting avocado fruits. This pattern looks like it will continue the more the image size increases, whereas the epochs beyond 100 had negligible effect on the recall (**Figure 4.8** (b)). Due to its performance drop, model 512x512_s6 is an outlier, shows a clear pattern that larger image sizes gave higher the mAP value, which agrees with [32]. This study was limited in the resources available to be able to train at image sizes greater than 512x512 pixels and this is therefore an opportunity for future work.

The mean average precision (mAP) metric for 512x512_s proves that the model is performing better in increasing image sizes, with model mAP scoring 86.9%. This is significant because it means approximately 87% of the time, the model is accurately predicting and detecting the avocado fruit. This result is similar to [31] in which 90% AP was achieved. (Average precision (AP) is calculated using (3.8), note, for 1 class, AP = mAP).

From this, it becomes clear that the best option for future development of this model would be to obtain greater hardware that is capable of storing larger images in temporary memory, so it is possible to train on images larger than 512x512 pixels. It also becomes clear that the 512x512_s model should be chosen for the counting of the avocado fruit within the orchard photos and is investigated in section 4.3.5.

Note also how there is negligible difference between the structure of the YOLOv5s training models. Type '448x448_s' and type '448x448_s6', where s is the base training structure and s6 is a newer training structure developed by YOLO. The 's6' structure is meant to reduce memory usage and improve training results, but in the case of avocados these were not that significant. Similar to (**Figure 4.8** (c)), the mAP over different thresholds shown in (**Figure 4.8** (d)) shows that the best performing model is the one trained with the largest images sizes, which is model 512x512_s. This is significant because it shows an almost doubled detection accuracy from 224x224_s being ~0.34 to 512x512_s being ~0.59 after 100 epochs.

From this, this study can deduce that the larger the image size, and thereby label size, the less FP there are, therefore the higher the mAP for each prediction.

The conclusions made in 4.3.3 are that the larger the image size is to its original ratio when trained, the better the performance of the model. Future work, in which powerful hardware can be attained, should look to explore this avenue more and find the plateau for image size. Also, it should be noted that the number of epochs (number of training cycles) past 100 tended to create over fitting and is therefore unnecessary. Adding to this, the structure of YOLOv5s training model ('s' and 's6') had little effect on the results.

From this point, the highest performing model, 512x512_s, will be used to quantify the number of avocado fruits within rows in block 49 of the orchard, but before this, test image sizes were scaled to see if model detection could be improved during inference (test images).

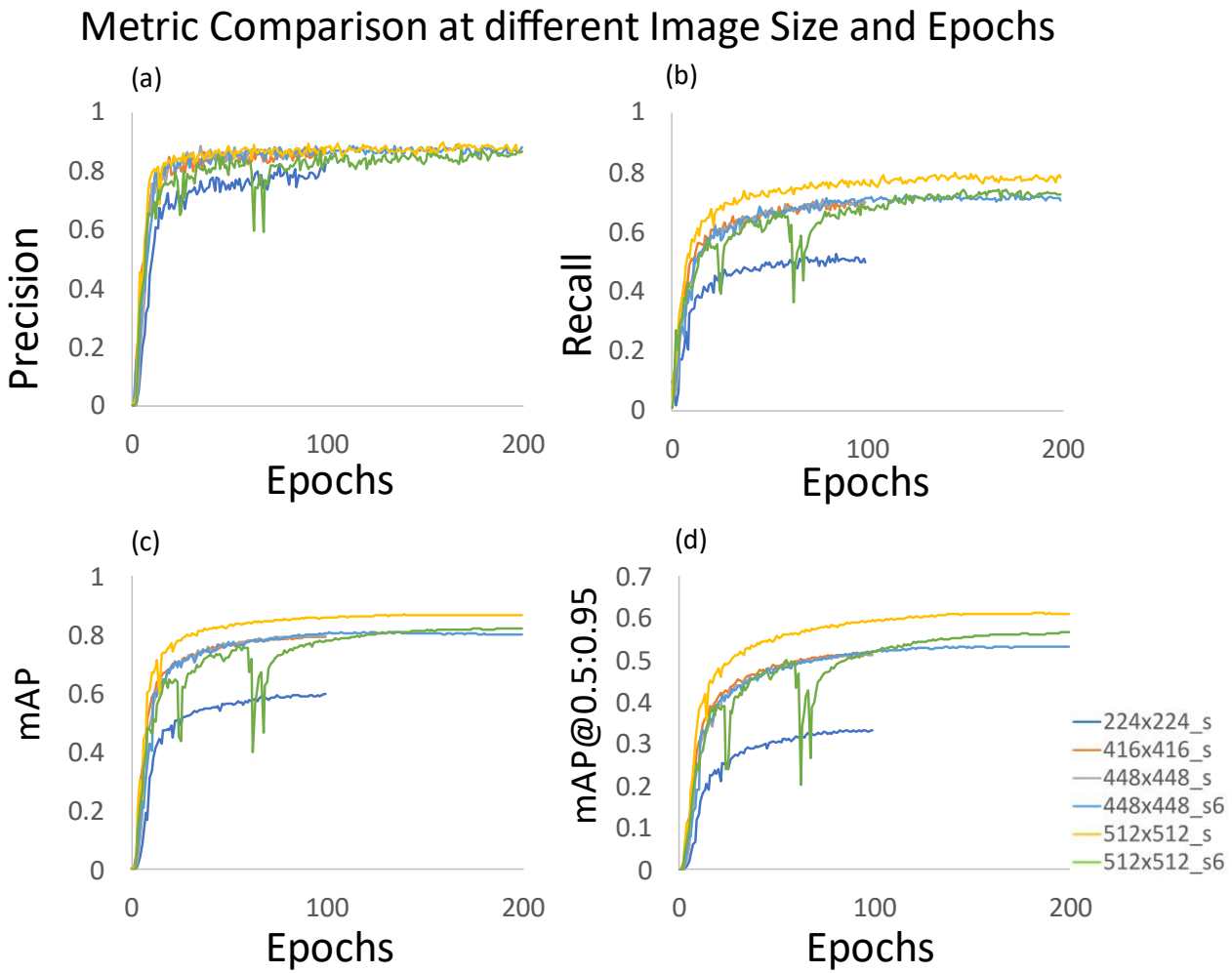

Figure 4.8 Image size in pixels, 224x224_s (Blue), 416x416_s (Orange), 448x448_s (Grey), 448x448_s (light blue), 512x512_s (yellow), 512x512_s6 (green).

Table 4.3. YOLOv5 model detection performance varying image sizes

MODEL	PRECISION	RECALL	mAP	mAP 0.5:0.95
224X224_S	82.53%	52.49%	59.86%	33.31%
416X416_S	88.09%	70.21%	79.29%	51.78%
448X448_S	88.67%	70.86%	79.74%	51.29%
448X448_S6	89.01%	72.50%	80.87%	53.22%
512X512_S	**89.61%**	**79.76%**	**86.93%**	**61.19%**
512X512_S6	87.37%	74.29%	82.19%	56.54%

Bold font: BEST

4.3.4 Predicted Count at Different Images Sizes

Within this section, the 252 labelled images that were within the test set were processed using the best performing model from section 4.3.3**Error! Reference source not found.** and the results are shown in **Figure 4.9**, with **Table 4.4** summarising the results and regression.

The test image sizes were scaled to various sizes from an initial size of 2048x2048 pixels, to elucidate which test image size performed the best (note that images sizes had to be multiples of 64x64 pixels). After the avocados were counted with each input image, the difference was calculated between the number of manually labelled bounding boxes and the number of predicted bounding boxes. The difference between the predicted counts and the ground truth counts was calculated, then the mean of difference (MD) was found using equation (4.1), where the ideal should be 0. Alongside this are 3 error metrics, where standard deviation (STD) which will show the range of the results, root mean square error (RMSE) is for regression analysis and mean absolute error (MAE) is also for regression analysis. Note that the goal is to accurately count avocados, hence the ideal errors will be those closest to 0 (see (4.2)- (4.4)).

$$MD = \frac{\sum(P_i - O_i)}{n} \tag{4.1}$$

Where,

MD: mean of difference

P_i : the estimated value,

O_i : observed value

n: total number of observations

$$STD = \sqrt{\frac{\sum(x - \bar{x})^2}{(n-1)}} \tag{4.2}[86]$$

Where,

STD: Standard deviation

n: total number of observations

x: each value from the sample

$\bar{x}$: the sample mean

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(P_i - O_i)^2}{n}} \quad (4.3)[91]$$

Where,

RMSE: Root mean square error

P_i : the estimated value

O_i : observed value

n: total number of observations

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - x_i| \quad (4.4)[92]$$

Where,

MAE: Mean absolute error

y_i: actual value for the ith observation

x_i: calculated value for the ith observation

n: total number of observations

While it is not conventional to pass across image sizes different to that which a network has trained on, this study examined whether it was possible to achieve greater detection doing so, with the thought that larger images contain a more defined texture, shape, and colour clarity of the fruit. The distance in which avocados were presented in the training images were naturally at different distances, meaning that the network naturally had different sized avocados to train on and detect. But, if the tested images sizes are too small, the distant avocado negatively affect the detection, so rather improve it.

Note the trend of the different image sizes, the coefficients tend to converge towards 1 as the images get larger, then veer away again. Likewise, the R^2 tends to move towards 1 as the image size increases, then moves away from 1 as the images get too big. This is due to the network not detecting avocados (false negatives) or picking up false positives amongst leaves, branches and or shadows caused by their relatively increased size compared to what the network is used to. The test image size with the coefficient closest to 1, and best R^2 was 832x832 pixels (**Figure 4.9**(d)).

From **Table 4.4**, the image size with the best MD, RMSE and MAE was size 640x640 pixels, but the image size with the best STD was 832x832. Note how the MD seems to approach 0 with increasing size, then moves away from it. This indicates that choosing the image sizes 640x640 would work the best for the counting of the avocados, but image size 832x832 will also be chosen for its better coefficient and y-intercept.

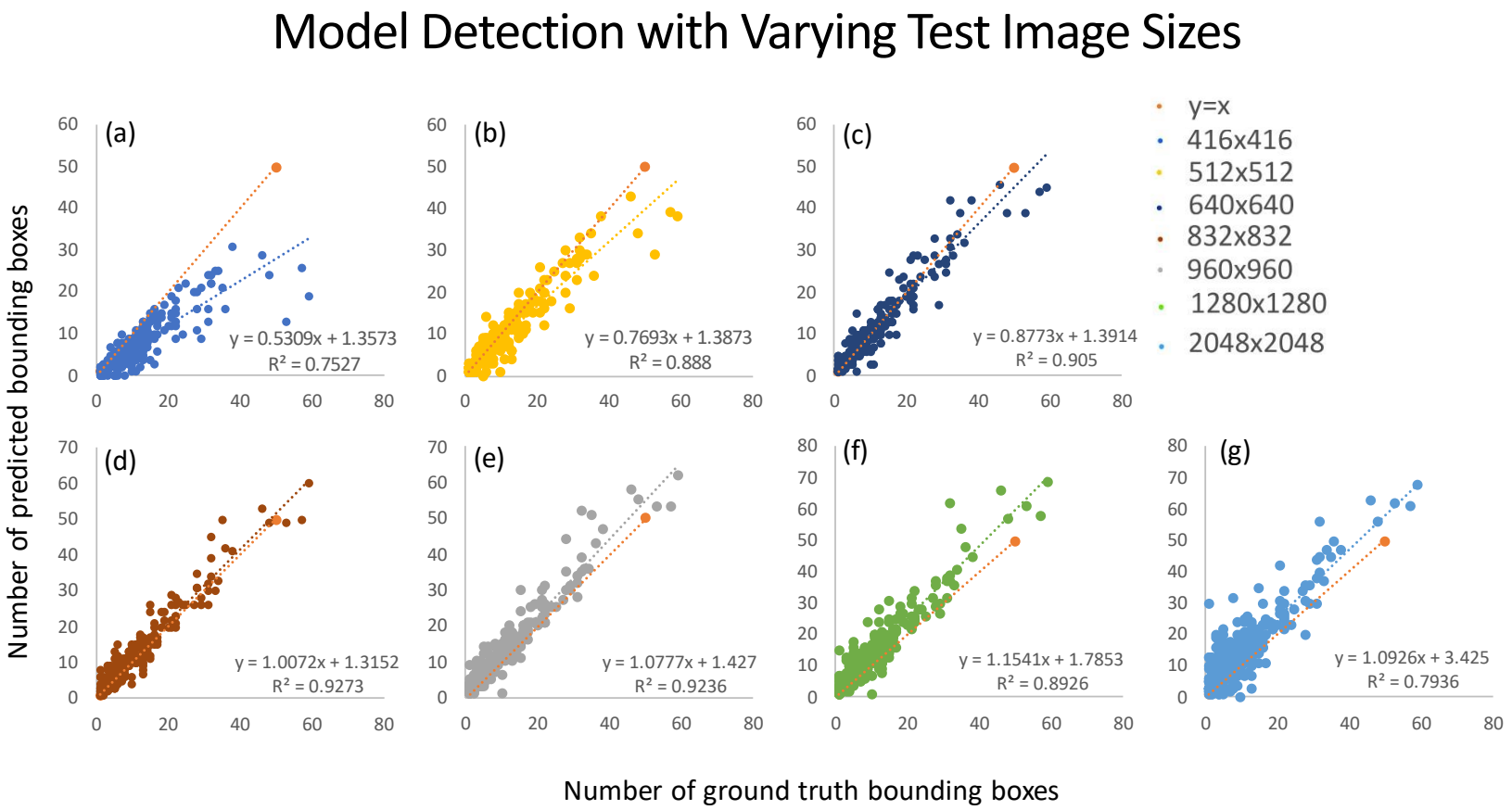

Figure 4.9 Predicted count from 252 test images of size (a) 416x416, (b) 512x512, (c) 640x640, (d) 832x832, (e) 960x960, (f) 1280x1280, & (g) 2048x2048 pixels using model trained on 512x512 pixel images.

Table 4.4. Avocado Predicted Count vs Ground Truth Bounding Boxes for varying image sizes

SIZE (PIXELS)	MD	STD	RMSE	MAE
416X416	-3.706	5.603	6.709	3.873
512X512	-1.103	3.583	3.917	2.671
640X640	**0.067**	3.102	**3.097**	**1.917**
832X832	1.393	**2.827**	3.147	2.187
960X960	2.265	3.201	3.743	2.079
1280X1280	3.448	4.297	5.504	3.861
2048X2048	4.424	5.660	7.176	5.385

Bold font: BEST. MD = Mean of Difference, STD = Standard Deviation, RMSE = Root Mean Square Error, MAE = Mean absolute error

4.3.5 Orchard Avocado Count

This section examines the performance of the model when used to count the orchard rows within zone 2, block 49 of the King Avocado Orchard. Note that only rows 5, 7, 8, 9, 10, and 11 were able to be counted as noted in section 3.4. Of the rows that are counted, they are separated into different groups: the 'orchard count' and the 'eye count'. The orchard count was collected by orchard personnel as this study was not able to travel due to COVID-19 border restrictions. However, the eye counts may be deemed more reliable due to larger sample size, also they were collected at the same time as the images, and they were overseen by this study. (**Figure 4.10**(a)) shows a coefficient of 0.47, (**Figure 4.10**(b)) shows a coefficient of 0.55. Both models had intercepts that were above 0, likely due to the model overcounting for values between 0-10. **Table 4.6** shows that the deep learning model performed the best with image size of 832x832, however this study's (**Figure 4.10**(c)) eye count was seen to have less error than the YOLOv5 models, and a coefficient of 0.78. There is a tapering off of predicted counts at higher fruit counts on both the 640x640 and 832x832 models. This is likely due to the occlusion naturally present within the environment and is not easily dealt with. Also, for both 640x640 and 832x832 models, there seems to be overcounting taking place between 0-10. The performance of the model is such that it can detect fruits on other trees in the distance explaining the over counting at lower fruit counts. Future work will look to address this

The trees in zone 2, block 49 are organised in North to South lines. The images being taken within one tree might overlap with other tree canopies. If the picture tree, in this case, Tree B, did not have any fruits and its canopy was thin enough, and if there were fruits on the North of South side of surrounding trees, then it could cause overcounting as the detection system developed will detect as many avocados in the image as it can and will not differentiate between the fruit within its tree and the tree next to it. It may be possible to limit this effect by changing the method in which the data is collected, perhaps vertically pointed photos instead of lateral.

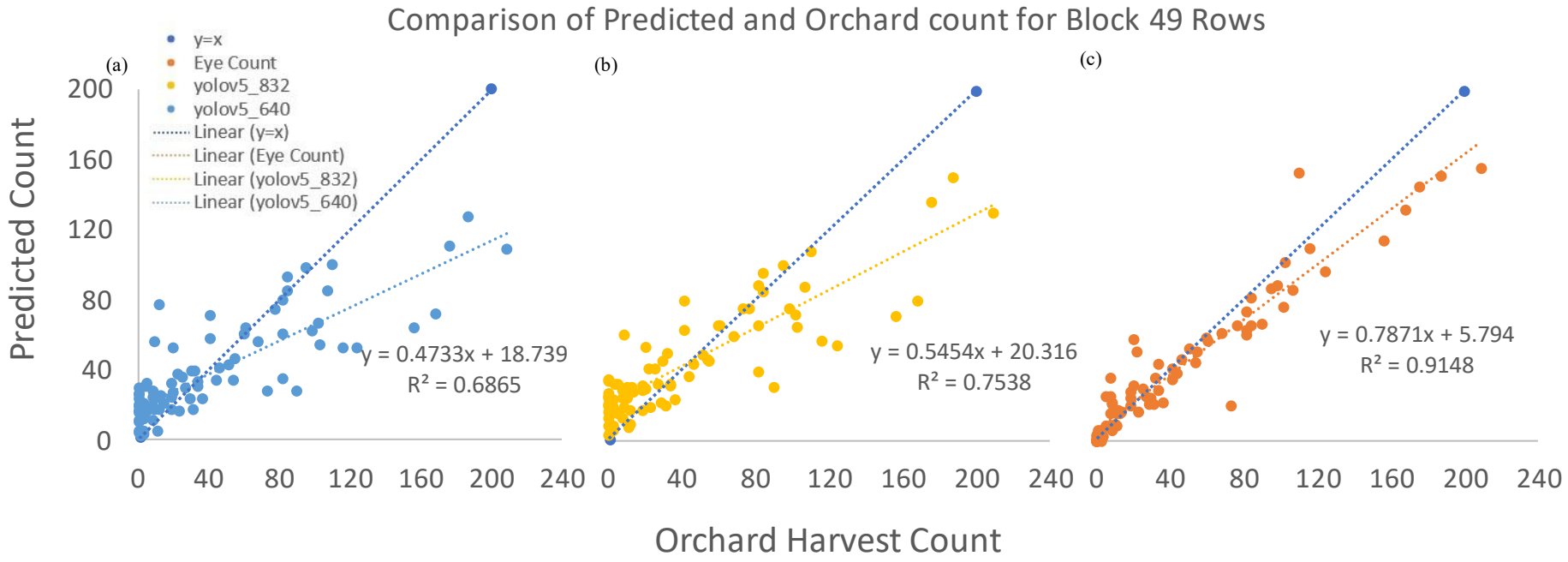

Figure 4.10. Shows the model's performance in counting avocados when compared to the orchard's harvest count of 87 sample trees, where: (a) yolov5_640 (sky blue) shows the YOLOv5 number of predicted fruit counts on test image size 640x640, (b) yolov5_832 (yellow) shows the YOLOv5 number of predicted fruit counts on test image size 832x832, (c) Eye count (orange) is the eye counted fruit collected by this study.

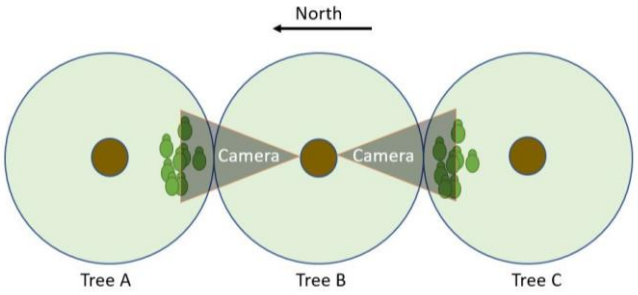

Figure 4.11. Showing how overcounting on a singular tree takes place where Tree B has no fruit, but surrounding Trees A and C have fruit

Figure 4.12 South facing image showing avocados seen in another tree

Another idea could be to analyse the North and South side count information from the trees to perhaps find out how much overcounting there is for these trees and find an experimental constant to adjust for it. Lastly, another potential idea, purely within the network, is to examine the size of the average bounding boxes and to remove detection from those under a certain size. The problem with the last potential solution is the fact that predicted bounding boxes are not always the same size as the original bounding boxes, and also that occluded avocados, which naturally have less surfaces area showing, will naturally have a smaller predicted bounding box, hence it would become quite finicky to deal with such issues.

In the rows 5, 7 & 8, the eye count performed the best. However, rows 9, 10 & 11 show that 832x832 performed the best. **Table 4.6** shows that 832x832 had the best mean of difference (MD) of 1.85, but the eye count had a lower standard deviation of 14.36 compared to the 26.94 of the 832x832. Based on the average, using pictures with YOLOv5 as a counting method is viable. Hence, fruit per tree averages were calculated for each method and compared to the harvest and the eye count value, see **Table 4.7**. As seen from the results in **Table 4.7**, it is clear that machine learning and vision methods will soon replace eye counting.

Table 4.5. Avocado Predicted and Eye count vs Orchard Count for individual rows

	Row 5		Row 7		Row 8		Row 9		Row 10		Row 11	
	SS = 10		SS = 38		SS = 32		SS = 3		SS = 2		SS = 2	
Method	**MD**	**STD**	**MD**	**STD**	**MD**	**STD**	**MD**	**STD**	**MD**	**STD**	**MD**	**STD**
640x640	-0.60	32.29	-13.24	33.68	8.84	14.51	-11.67	47.12	5.00	21.21	-4.50	0.71
832x832	3.70	32.83	-8.58	30.98	13.34	13.78	-0.667	34.12	12.00	19.80	0.50	3.54
Eye Count	-0.60	17.95	-4.45	17.35	-0.84	7.36	-8.00	19.98	-14.50	21.92	-5.00	9.90
Method	**RMSE**	**MAE**	**RMSE**	**MAE**	**RMSE**	**MAE**	**RMSE**	**MAE**	**RMSE**	**MAE**	**RMSE**	**MAE**
640x640	30.63	21.6	35.78	23.18	16.79	13.66	40.20	32.33	15.81	15.00	4.53	4.50
832x832	31.36	23.30	31.75	21.84	19.02	16.09	27.87	26.00	18.44	14.00	2.55	2.50
Eye Count	17.04	12.00	17.69	11.97	7.30	4.66	18.17	12.67	21.22	15.50	8.60	7.00

SS = sample size, MD = Mean of Difference, STD = Standard Deviation, RMSE = Root Mean Square Error, MAE = Mean absolute error

Table 4.6. Avocado Predicted Count vs Orchard Count

METHOD	MD	STD	RMSE	MAE
640X640	-2.99	28.87	28.85	19.20
832X832	1.85	26.94	26.85	19.41
EYE COUNT	-3.05	14.36	14.59	9.28

Sample size = 87, MD = Mean of Difference, STD = Standard Deviation, RMSE = Root Mean Square Error, MAE = Mean absolute error

Table 4.7. Avocado fruit per tree comparison

Method	**640x640**	**832x832**	**Eye Count**	**Harvest Count**	**640x640**	**832x832**	**Eye Count**
Total Fruit Count	3274	3695	3278	3543	5197	5912	5044
Trees	87	87	87	87	133	133	133
Fruit per tree	37.63	42.47	37.68	40.72	39.08	44.45	37.92

SS = sample size. Note that Harvest Count SS = 87, but this study's Eye Count SS = 133

4.3.6 Video Collection

The potential of video data collection could also assist in reducing the overall error, and future work could look to expand on this area and improve the network more in this regard. Instead of a snapshot in time where the network has one chance to detect the fruit, a real time method of detection could be used to track the fruit and count them across an orchard row. See https://youtu.be/hc82bEqUddc as evidence of this study's real time avocado detection, where **Table 4.8** summarises the detection of a few snapshots, see Appendix A for screenshots.

Table 4.8. Avocado Detection in real time

	FRAME 1	FRAME 2	FRAME 3	FRAME 4	FRAME 5	TOTAL
TIME STAMP	0:32	0:50	1:08	1:24	1:25	
NUMBER OF AVOCADOS EYE COUNTED	48	37	57	55	74	271
NUMBER OF PREDICTED AVOCADOS COUNTED*	43	36	55	39	66	239
ACCURACY	90%	97%	96%	71%	89%	88%

* Prediction done using YOLOv5 model at 0.6 confidence threshold

CHAPTER 5
CONCLUSION & RECOMMENDATIONS

5.1 Summary

The objective of this master's thesis was to assess the viability of avocado yield quantification using machine vision and machine learning. The results of this study indicate that a convolutional neural network like YOLOv5, trained on images from low cost smartphones, is a viable method for quantifying the yield of avocados.

After reviewing literature, gaps were present regarding 'on tree' avocado datasets, and machine vision and machine learning applications for avocados. However, machine vision studies on other fruits, like green apples and mangos, presented this study the opportunity to transfer the techniques learned from those studies to avocados. Therefore, this study developed and contributed a novel 'on tree' avocado dataset consisting of 4280 images, within which are 40000+ labelled avocados.

This study overcame the difficulties presented with green avocados clashing with green trees and leaves and, using its novel dataset, trained and contributed a YOLOv5 model capable of detecting and counting avocados from one of New Zealand's largest avocado orchards, King Avocado.

The limited computing hardware meant that this study was restricted in its ability to train a model on images larger than 512x512 pixels, but despite this, the performance of the best model reached 86.9% mAP at an IoU threshold of 0.6. Towards the end of this study, it was discovered that hardware could be rented online, i.e., a domain allowing access to a computer or server that has hardware suited to computationally demanding tasks like training a neural network. These domains act similar to Google Colab in that they are timed, but at a higher rent rate, private access can be gained. Renting hardware online could potentially solve this issue and should be investigated in future work.

This study demonstrated that image-based inference at 640x640 pixels showed that the model's predicted count deviated <5% from the ground truth bounding box count, with a mean absolute error of ~2. This indicates that for a given 'on tree' avocado image, there will be ±2 predicted avocados from the number of visibly seen avocados in the image. In terms of model accuracy, this could be improved further by tweaking augmentations and increasing image size at training and inference.

Orchard estimation was also promising, with fruit per tree of the best model being <5% compared to the harvest fruit per tree value. There is still an amount of error on tree to be accounted for, but I would attribute this as a benefit towards the model, as it is detecting avocados on other trees in the distance. Future work should look to limit this effect with an algorithm that passes the predicted bounding boxes produced, putting an area filter on the ones that are too small, and then recounting detections per trees.

Additionally, real time estimation results were also promising, showing a mean accuracy of 88% across 5 frames. There are challenges in this area, such as fruit tracking to prevent double counting or recounting. However, solutions exist to mitigate this, like stabilised mounting of the camera and use of the tracking algorithm known as DeepSORT [93].

Next steps in this work would probably look towards an automated collection of images, most likely using a robot or drone of some sort which could go around to capture data, and may present an idea to assist orchards struggling with labour shortages, not to mention the robot would be excellent at stabilising a camera.

There would be various challenges in constructing a robot capable of traversing the terrain within avocado orchards, i.e., pitfalls, uneven ground, debris. All the while, the robot needs to be capable of capturing high quality images to process. Future work needs to look into solutions to overcome this barrier.

5.2 Recommendations

Methodological improvements regarding single tree overcounting can be addressed in the future by incorporating machine vision distance technology, such as LiDAR, or through the implementing limits on bounding box area when training or inferencing a model.

Larger image sizes and datasets in the future may also allow for improved model performance. Due to GPU memory restrictions, renting hardware servers could be an option to address model training constraints. Alongside this, the experimentation of image augmentations for the avocado dataset was not explored to a great level of detail and could be a simple means of improving performance also.

5.3 Closing Sentiment

New Zealand Avocado is advancing into the future with high hopes of increasing sales and productivity significantly. In order to achieve industry goals, growers must make use of cutting-edge technology and innovative agricultural methods. Machine vision and machine learning has shown excellent potential for predicting and counting avocados, and this technology exists in the market for growers to readily adopt.

From the research presented in this thesis, the future of avocados within New Zealand is in the green and primed to grow robustly.

REFERENCES

[1] H. Chen, P. L. Morrell, V. E. Ashworth, M. De La Cruz, and M. T. Clegg, "Tracing the geographic origins of major avocado cultivars," *Journal of Heredity,* vol. 100, no. 1, pp. 56-65, 2009.

[2] R. W. Hodgson, "The California avocado industry," 1930.

[3] O. Khazan, "The selling of the avocado," *The Atlantic,* 2015. [Online]. Available: https://www.theatlantic.com/health/archive/2015/01/the-selling-of-the-avocado/385047/.

[4] N. Z. Horticulture, "FreshFacts," *Nutrition,* vol. 30, p. 31, 2017.

[5] N. T. Anderson, K. B. Walsh, and D. Wulfsohn, "Technologies for forecasting tree fruit load and harvest timing—from ground, sky and time," *Agronomy,* vol. 11, no. 7, p. 1409, 2021.

[6] A. N. Zealand, "Becoming an Avocado Grower," 2021. [Online]. Available: https://industry.nzavocado.co.nz/grow/becoming-an-avocado-grower/.

[7] (2019). *Avocado is the New Black*. [Online] Available: https://www.mpi.govt.nz/dmsdocument/43186-Opportunities-for-New-Zealands-Avacados

[8] N. Z. Avocado, "The different types of avocados," 2020. [Online]. Available: https://industry.nzavocado.co.nz/.

[9] M. Sonka, V. Hlavac, and R. Boyle, *Image processing, analysis, and machine vision*. Nelson Education, 2014.

[10] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779-788.

[11] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767,* 2018.

[12] U. Nepal and H. Eslamiat, "Comparing YOLOv3, YOLOv4 and YOLOv5 for autonomous landing spot detection in faulty UAVs," *Sensors,* vol. 22, no. 2, p. 464, 2022.

[13] C. Colantonio, "Innovative non-invasive techniques for integrated imaging diagnostics of paintings and data analysis," pp. 9-10, 2020.

[14] S. Moncada, "The Ultimate Guide to Convolutional Neural Networks (CNN)," ed, 2018.

[15] A. Stevenson, *Oxford dictionary of English*. Oxford University Press, USA, 2010.

[16] D. Wilimitis, "The kernel trick in support vector classification," *Online,* 2018.

[17] K. O'Shea and R. Nash, "An introduction to convolutional neural networks," *arXiv preprint arXiv:1511.08458,* 2015.

[18] A. McBratney, B. Whelan, T. Ancev, and J. Bouma, "Future directions of precision agriculture," *Precision agriculture,* vol. 6, no. 1, pp. 7-23, 2005.

[19] C. J. Lovatt, "Alternate bearing of 'Hass' avocado," *Calif. Avocado Soc. Yearb,* vol. 93, pp. 125-140, 2010.

[20] A. Robson, M. M. Rahman, J. Muir, A. Saint, C. Simpson, and C. Searle, "Evaluating satellite remote sensing as a method for measuring yield variability in Avocado and Macadamia tree crops," *Adv. Anim. Biosci,* vol. 8, pp. 498-504, 2017.

[21] A. Robson, M. M. Rahman, and J. Muir, "Using worldview satellite imagery to map yield in avocado (Persea americana): A case study in Bundaberg, Australia," *Remote Sensing,* vol. 9, no. 12, p. 1223, 2017.

[22] M. M. Rahman, A. Robson, and M. Bristow, "Exploring the potential of high resolution worldview-3 Imagery for estimating yield of mango," *Remote Sensing,* vol. 10, no. 12, p. 1866, 2018.

[23] J. Sarron, É. Malézieux, C. A. B. Sané, and É. Faye, "Mango yield mapping at the orchard scale based on tree structure and land cover assessed by UAV," *Remote Sensing,* vol. 10, no. 12, p. 1900, 2018.

[24] A. S. A. Salgadoe, A. J. Robson, D. W. Lamb, E. K. Dann, and C. Searle, "Quantifying the severity of phytophthora root rot disease in avocado trees using image analysis," *Remote Sensing,* vol. 10, no. 2, p. 226, 2018.

[25] S. L. Bangare, A. Dubal, P. S. Bangare, and S. Patil, "Reviewing otsu's method for image thresholding," *International Journal of Applied Engineering Research,* vol. 10, no. 9, pp. 21777-21783, 2015.

[26] S. Köhne, "Yield estimation based on measureable parameters," *South African Growers Association Yearbook,* vol. 8, p. 103, 1985.

[27] M. FOLDENAUER and S. KÖHNE, "A CROP ESTIMATION MODEL FOR AVOCADO."

[28] W. H. Maes and K. Steppe, "Perspectives for remote sensing with unmanned aerial vehicles in precision agriculture," *Trends in plant science,* vol. 24, no. 2, pp. 152-164, 2019.

[29] J. Das *et al.*, "Devices, systems, and methods for automated monitoring enabling precision agriculture," in *2015 IEEE International Conference on Automation Science and Engineering (CASE)*, 2015: IEEE, pp. 462-469.

[30] L. L. Detection, "Ranging—is a remote sensing method used to examine the surface of the Earth," *NOAA. Archived from the original on,* vol. 4, 2013.

[31] J. P. Vasconez, J. Delpiano, S. Vougioukas, and F. A. Cheein, "Comparison of convolutional neural networks in fruit detection and counting: A comprehensive evaluation," *Computers and Electronics in Agriculture,* vol. 173, p. 105348, 2020.

[32] A. Koirala, K. B. Walsh, Z. Wang, and C. McCarthy, "Deep learning–Method overview and review of use for fruit detection and yield estimation," *Computers and electronics in agriculture,* vol. 162, pp. 219-234, 2019.

[33] M. Woodson and C. Ross, "Automated Avocado Yield Forecasting Using Multi-Modal Imaging," 2016.

[34] J. Gené-Mola *et al.*, "Fruit detection, yield prediction and canopy geometric characterization using LiDAR with forced air flow," *Computers and Electronics in Agriculture,* vol. 168, p. 105121, 2020.

[35] B.-H. Cho, K. Koyama, E. O. Díaz, and S. Koseki, "Determination of "Hass" avocado ripeness during storage based on smartphone image and machine learning model," *Food and Bioprocess Technology,* vol. 13, no. 9, pp. 1579-1587, 2020.

[36] F. Westling, J. Underwood, and S. Örn, "Light interception modelling using unstructured LiDAR data in avocado orchards," *Computers and electronics in agriculture,* vol. 153, pp. 177-187, 2018.

[37] G. Moreno-Ortega, C. Pliego, D. Sarmiento, A. Barceló, and E. Martínez-Ferri, "Yield and fruit quality of avocado trees under different regimes of water supply in the subtropical coast of Spain," *Agricultural Water Management,* vol. 221, pp. 192-201, 2019.

[38] J. D. Oster, D. Stottlmyer, and M. Arpaia, "Salinity and water effects on 'Hass' avocado yields," *Journal of the American Society for Horticultural Science,* vol. 132, no. 2, pp. 253-261, 2007.

[39] S. Joseph *et al.*, "Biochar increases soil organic carbon, avocado yields and economic return over 4 years of cultivation," *Science of the Total Environment,* vol. 724, p. 138153, 2020.

[40] N. Z. Avocado, "Annual Reports," 2021. Accessed: 2021. [Online]. Available: https://industry.nzavocado.co.nz/about-us/annual-reports/

[41] ANZ, "AVOCADOS - HOLY GUACAMOLE!," *ANZ AGRI FOCUS,* 2018. [Online]. Available: https://www.anz.co.nz/content/dam/anzconz/documents/rural/agri-focus-march-2018.pdf?MOD=AJPERES.

[42] M. Islam, K. Wahid, and A. Dinh, "Assessment of ripening degree of avocado by electrical impedance spectroscopy and support vector machine," *Journal of Food Quality,* vol. 2018, 2018.
[43] F. A. A. ORGANIZATION, "Avocado Production in Asia and the Pacific," *RAP Publications,* 2000. [Online]. Available: http://www.fao.org/3/a-x6902e.pdf.
[44] N. Burns, "Avocado Price Forecasting," Visual 2020. [Online]. Available: https://www.flockconsulting.co.nz/post?id=2&datablog=avacado-price-forecasting.
[45] G. Trends, "Explore What the World is Searching," 2017. [Online]. Available: https://trends.google.com/trends/?geo=NZ.
[46] J. Rincon-Patino, E. Lasso, and J. C. Corrales, "Estimating avocado sales using machine learning algorithms and weather data," *Sustainability,* vol. 10, no. 10, p. 3498, 2018.
[47] S. Landahl and L. A. Terry, "Non-destructive discrimination of avocado fruit ripeness using laser Doppler vibrometry," *Biosystems Engineering,* vol. 194, pp. 251-260, 2020.
[48] C. GmbH, "Quantifying the Crop Load on Apple Trees using Hyperspectral Snapshot Imaging," 2020. [Online]. Available: https://cubert-gmbh.com/applications/qualification-of-crop-load-on-apple-trees/.
[49] Q. Wang, S. Nuske, M. Bergerman, and S. Singh, "Automated crop yield estimation for apple orchards," in *Experimental robotics*, 2013: Springer, pp. 745-758.
[50] W. C. Nut, "Almond Yield Forecasting by Drone," 2021. [Online]. Available: http://www.wcngg.com/2021/01/04/almond-yield-forecasting-by-drone/.
[51] T. T. Santos, L. L. de Souza, A. A. dos Santos, and S. Avila, "Grape detection, segmentation, and tracking using deep neural networks and three-dimensional association," *Computers and Electronics in Agriculture,* vol. 170, p. 105247, 2020.
[52] M. Stein, S. Bargoti, and J. Underwood, "Image based mango fruit detection, localisation and yield estimation using multiple view geometry," *Sensors,* vol. 16, no. 11, p. 1915, 2016.
[53] G. E. Dieter and L. C. Schmidt, *Engineering design*. McGraw-Hill Higher Education Boston, 2009.
[54] D. G. Ullman, "Toward the ideal mechanical engineering design support system," *Research in Engineering Design,* vol. 13, no. 2, pp. 55-64, 2002.
[55] J. Fourie, K. Pahalawatta, J. Hsiao, C. Bateman, and P. Carey, "Fusion of thermal and visible colour images for robust detection of people in forests," in *2019 International Conference on Image and Vision Computing New Zealand (IVCNZ)*, 2019: IEEE, pp. 1-6.
[56] J. R. Olatunji, G. Redding, C. Rowe, and A. East, "Reconstruction of kiwifruit fruit geometry using a CGAN trained on a synthetic dataset," *Computers and Electronics in Agriculture,* vol. 177, p. 105699, 2020.
[57] A. Kuznetsova, T. Maleva, and V. Soloviev, "YOLOv5 versus YOLOv3 for apple detection," in *Cyber-Physical Systems: Modelling and Intelligent Control*: Springer, 2021, pp. 349-358.
[58] K. Ikeuchi, *Computer vision: A reference guide*. Springer, 2021.
[59] A. Shikany, "Machine Vision Innovations," *Quality,* vol. 60, no. 13, pp. 41-41, 2021. [Online]. Available: https://www.qualitymag.com/articles/96774-machine-vision-innovations-result-in-big-growth-in-2021-and-beyond.
[60] A. Neef, "Legal and social protection for migrant farm workers: lessons from COVID-19," *Agriculture and Human Values,* vol. 37, no. 3, pp. 641-642, 2020.
[61] S. Gazit, "Pollination and fruit set of avocado," *Proc., First Int. Trop. Fruit Short Course: the Avocado, Univ. of Florida, Gainesville, FL,* pp. 88-92, 1977.
[62] MetService. "Meteorological Service New Zealand." https://www.metservice.com/ (accessed 2021).
[63] M. Alcaraz, T. G. Thorp, and J. Hormaza, "Phenological growth stages of avocado (Persea americana) according to the BBCH scale," *Scientia Horticulturae,* vol. 164, pp. 434-439, 2013.

[64] P. Wan, G. Cheung, D. Florencio, C. Zhang, and O. C. Au, "Image bit-depth enhancement via maximum a posteriori estimation of AC signal," *IEEE Transactions on Image Processing,* vol. 25, no. 6, pp. 2896-2909, 2016.
[65] GSMArena, "Phone Comparisons," 2022. [Online]. Available: https://www.gsmarena.com/.
[66] E. McCollough, "Industry: a monthly magazine devoted to science, engineering and mechanic arts," ed: Industrial Publishing Company, San Francisco, 1893, pp. 399-406.
[67] A. Lakso and L. Corelli Grappadelli, "Implications of pruning and training practices to carbon partitioning and fruit development in apple," in *I International Symposium on Training and Pruning of Fruit Trees 322*, 1991, pp. 231-240.
[68] J. N. Wünsche and A. N. Lakso, "Apple tree physiology: Implications for orchard and tree management," *Compact Fruit Tree,* vol. 33, no. 3, pp. 82-88, 2000.
[69] G. Jocher. "Tips for Best Training Results." https://github.com/ultralytics/yolov5/wiki/Tips-for-Best-Training-Results (accessed 2021).
[70] D. Tzutalin, "LabelImg. Git code," ed, 2015.
[71] G. Bradski and A. Kaehler, "OpenCV," *Dr. Dobb's journal of software tools,* vol. 3, 2000.
[72] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of big data,* vol. 6, no. 1, pp. 1-48, 2019.
[73] J. Solawetz and J. Nelson, "How to train yolov5 on a custom dataset," *Roboflow Blog,* 2020.
[74] J. Solawetz. "Image Augmentation." Roboflow Documentation. https://docs.roboflow.com/image-transformations/image-augmentation (accessed 2021).
[75] Ultralytics, "YOLOv5," 2021. [Online]. Available: https://github.com/ultralytics/yolov5.
[76] T. Mitsa, "How Do You Know You Have Enough Training Data," *Towards Data Science, https://towardsdatascience. com/how-do-you-know-you-have-enough-training-data-ad9blfd679ee Apr,* vol. 22, p. 15, 2019.
[77] V. G. P. Ramya. A, Dr. Amrutham Bhavya Vaishnavi, Sai Suraj Karra, "Comparison of YOLOv3, YOLOv4 and YOLOv5 Performance for Detection of Blood Cells," *International Research Journal of Engineering and Technology (IRJET),* vol. 8, no. 4, pp. 4225-4228, 2021. [Online]. Available: https://www.irjet.net/volume8-issue4.
[78] G. Liu, J. C. Nouaze, P. L. Touko Mbouembe, and J. H. Kim, "YOLO-tomato: A robust algorithm for tomato detection based on YOLOv3," *Sensors,* vol. 20, no. 7, p. 2145, 2020.
[79] Z. Wang, L. Jin, S. Wang, and H. Xu, "Apple stem/calyx real-time recognition using YOLO-v5 algorithm for fruit automatic loading system," *Postharvest Biology and Technology,* vol. 185, p. 111808, 2022.
[80] seekFire. "Overview of model structure about YOLOv5." https://github.com/ultralytics/yolov5/issues/280 (accessed 2021).
[81] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," *International journal of computer vision,* vol. 88, no. 2, pp. 303-338, 2010.
[82] P. Jaccard, "Étude comparative de la distribution florale dans une portion des Alpes et des Jura," *Bull Soc Vaudoise Sci Nat,* vol. 37, pp. 547-579, 1901.
[83] S. Shalev-Shwartz and S. Ben-David, *Understanding machine learning: From theory to algorithms*. Cambridge university press, 2014.
[84] G. Jocher *et al.*, "ultralytics/yolov5: v5. 0-YOLOv5-P6 1280 models AWS Supervise. ly and YouTube integrations," *Zenodo,* vol. 11, 2021.
[85] S. Narkhede, "Understanding auc-roc curve," *Towards Data Science,* vol. 26, no. 1, pp. 220-227, 2018.
[86] *Microsoft Excel*. (2019). [Online]. Available: https://office.microsoft.com/excel

[87] C.-W. Yu, Y.-L. Chen, K.-F. Lee, C.-H. Chen, and C.-Y. Hsiao, "Efficient Intelligent Automatic Image Annotation Method based on Machine Learning Techniques," in *2019 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW)*, 2019: IEEE, pp. 1-2.

[88] S. Bargoti and J. Underwood, "Deep fruit detection in orchards," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, 2017: IEEE, pp. 3626-3633.

[89] S. Kolawole, O. Osakuade, N. Saxena, and B. K. Olorisade, "Sign-to-Speech Model for Sign Language Understanding: A Case Study of Nigerian Sign Language," *arXiv preprint arXiv:2111.00995,* 2021.

[90] PyTorch. "REPRODUCIBILITY." PyTorch. https://pytorch.org/docs/stable/notes/randomness.html (accessed 2021).

[91] G. Geography, "How to Calculate Root Mean Square Error (RMSE) in Excel," *GIS Geography,* 2018. [Online]. Available: https://gisgeography.com/root-mean-square-error-rmse-gis/.

[92] G. Geography, "How to Calculate Mean Absolute Error (MAE) in Excel," *GIS Geography,* 2018. [Online]. Available: https://gisgeography.com/mean-absolute-error-mae-gis/.

[93] X. Hou, Y. Wang, and L.-P. Chau, "Vehicle tracking using deep sort with low confidence track filtering," in *2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 2019: IEEE, pp. 1-6.

[94] A. Kassim, T. Workneh, and C. Bezuidenhout, "A review on postharvest handling of avocado fruit," *African Journal of Agricultural Research,* vol. 8, no. 21, pp. 2385-2402, 2013.

[95] *YOLOv5s.yaml*. (2020). Git Hub. Accessed: 2021. [Online]. Available: https://github.com/ultralytics/yolov5/blob/4b5f4806bcd513b18171034c06364432ef2c19c2/models/yolov5s.yaml#L12-L25

APPENDIX A

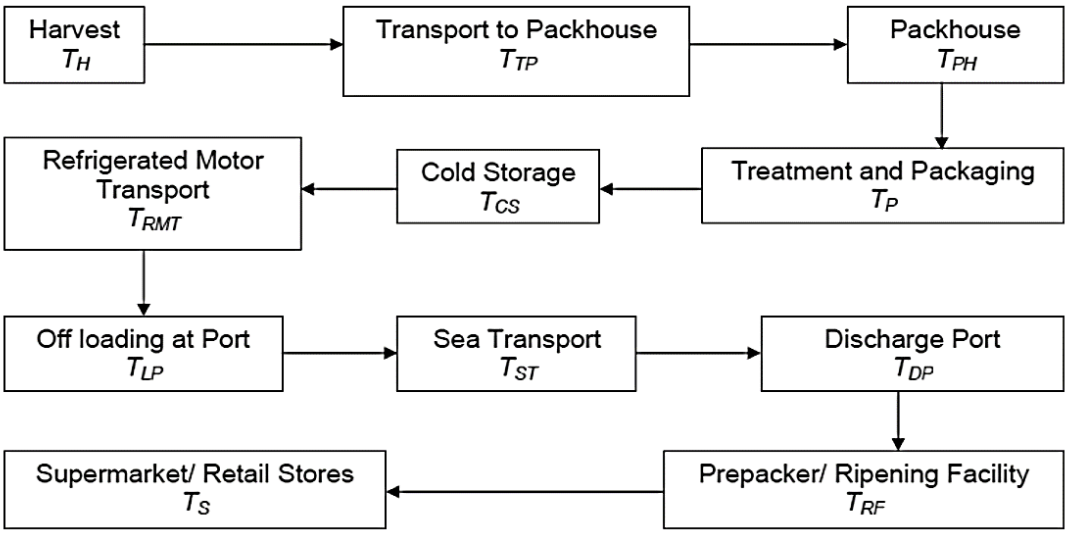

Figure A.1. Simple avocado cold chain [94].

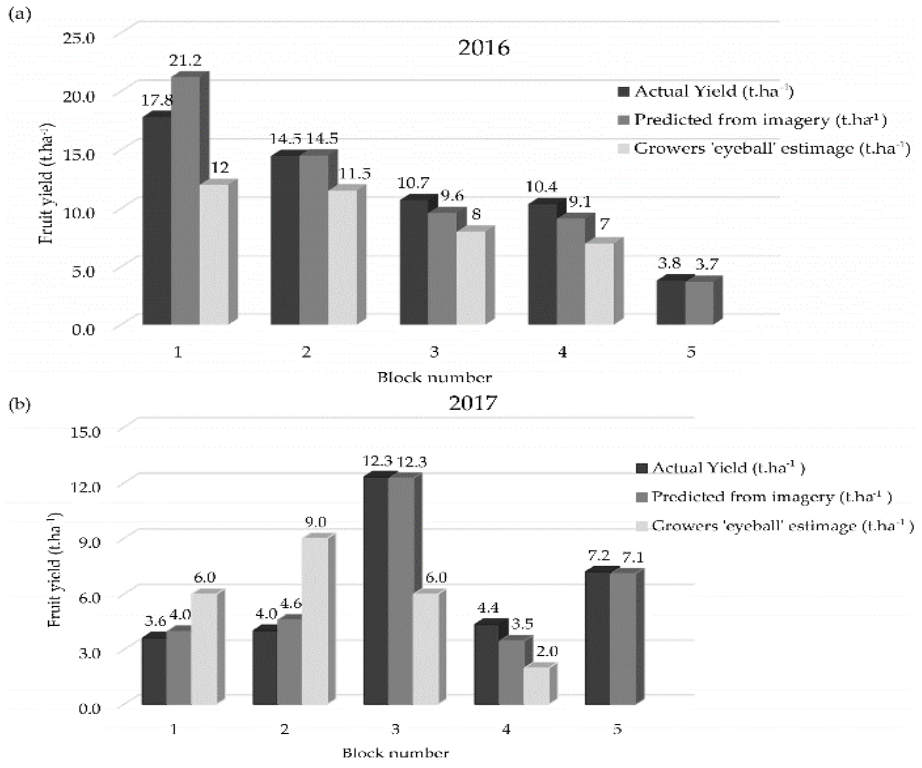

Figure A.2. Above shows a bar chart of the difference between the actual yield, the predicted yield from the satellite and that of visual assessment across years 2016 and 2017. Parameters compared are tonne per hectare. Note that there was no visual estimate for block 5 (Australia) [21].

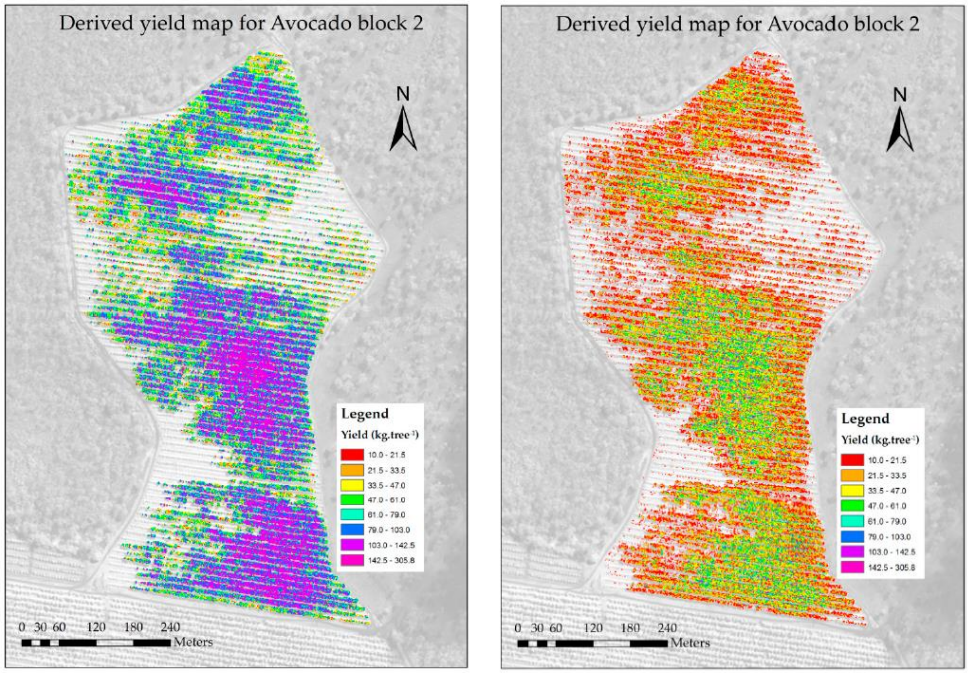

Figure A.3. Above shows the classified yield maps derived from the best fit model of the relationship between canopy reflectance and total fruit weight (kg/tree) for Block 2 in (a) 2016 and (b) 2017 [22].

Figure A.4. Snapshot of real time detection at 0:32

Figure A.5. Snapshot of real time detection at 0:50

Figure A.6. Snapshot of real time detection at 1:08

Figure A.7. Snapshot of real time detection at 1:24

Figure A.8. Snapshot of real time detection at 1:25

APPENDIX B

Labelling Script

```
import os


os.getcwd()
collection = "FILE PATH"
direction = ['','N', 'NE', 'E', 'SE', 'S', 'SW', 'W', 'NW']
count = 1
for i, filename in enumerate(os.listdir(collection)):
  if ((i%9)==0):
      count +=1
      os.rename(collection + "/" + filename, collection + "/" +
"_label" 'T' + str(count) + ".jpg")


  else:
      os.rename(collection + "/" + filename, collection + "/" +
"ROW NUMBER" 'T' + str(count) + '_' + direction[i%9] + ".jpg")
      print(filename)
```

Figure B.1. Example code of a Python script to label images

Avocado Annotation


```
{
"annotation":
    "{\"key\":\"R1_T30_N.jpg\",
    \"boxes\":

    [{\"label\":\"Avocado\",
    \"x\":1315.0,
    \"y\":1304.5,
    \"width\":138.5,
    \"height\":176.5},

    {\"label\":\"Avocado\",
    \"x\":1195.0,
    \"y\":1550.0,
    \"width\":156.0,
    \"height\":180.0},

    {\"label\":\"Avocado\",
    \"x\":1839.5,
    \"y\":1089.5,
    \"width\":211.0,
    \"height\":234.0},
}
```


Figure B.2. Showing 3 avocado bounding box annotations where x and y are the coordinates of the centre of the bounding box, and width and height are the dimensions of the bounding box

Default Code for Structure of YOLOv5 model 's'

```
# parameters
nc: 1  # number of classes
depth_multiple: 0.33  # model depth multiple
width_multiple: 0.50  # layer channel multiple

# anchors
anchors:
- [10,13, 16,30, 33,23]  # P3/8
- [30,61, 62,45, 59,119]  # P4/16
- [116,90, 156,198, 373,326]  # P5/32

# YOLOv5 backbone
backbone:
# [from, number, module, args]
[[-1, 1, Focus, [64, 3]],  # 0-P1/2
[-1, 1, Conv, [128, 3, 2]],  # 1-P2/4
[-1, 3, BottleneckCSP, [128]],
[-1, 1, Conv, [256, 3, 2]],  # 3-P3/8
[-1, 9, BottleneckCSP, [256]],
[-1, 1, Conv, [512, 3, 2]],  # 5-P4/16
[-1, 9, BottleneckCSP, [512]],
[-1, 1, Conv, [1024, 3, 2]],  # 7-P5/32
[-1, 1, SPP, [1024, [5, 9, 13]]],
[-1, 3, BottleneckCSP, [1024, False]],  # 9
]

# YOLOv5 head
head:
[[-1, 1, Conv, [512, 1, 1]],
[-1, 1, nn.Upsample, [None, 2, 'nearest']],
[[-1, 6], 1, Concat, [1]],  # cat backbone P4
[-1, 3, BottleneckCSP, [512, False]],  # 13

[-1, 1, Conv, [256, 1, 1]],
[-1, 1, nn.Upsample, [None, 2, 'nearest']],
[[-1, 4], 1, Concat, [1]],  # cat backbone P3
[-1, 3, BottleneckCSP, [256, False]],  # 17 (P3/8-small)

[-1, 1, Conv, [256, 3, 2]],
[[-1, 14], 1, Concat, [1]],  # cat head P4
[-1, 3, BottleneckCSP, [512, False]],  # 20 (P4/16-medium)

[-1, 1, Conv, [512, 3, 2]],
[[-1, 10], 1, Concat, [1]],  # cat head P5
[-1, 3, BottleneckCSP, [1024, False]],  # 23 (P5/32-large)

[[17, 20, 23], 1, Detect, [nc, anchors]],  # Detect(P3, P4, P5)
]
```

Figure B.3. Showing the YOLOv5s structure [95]

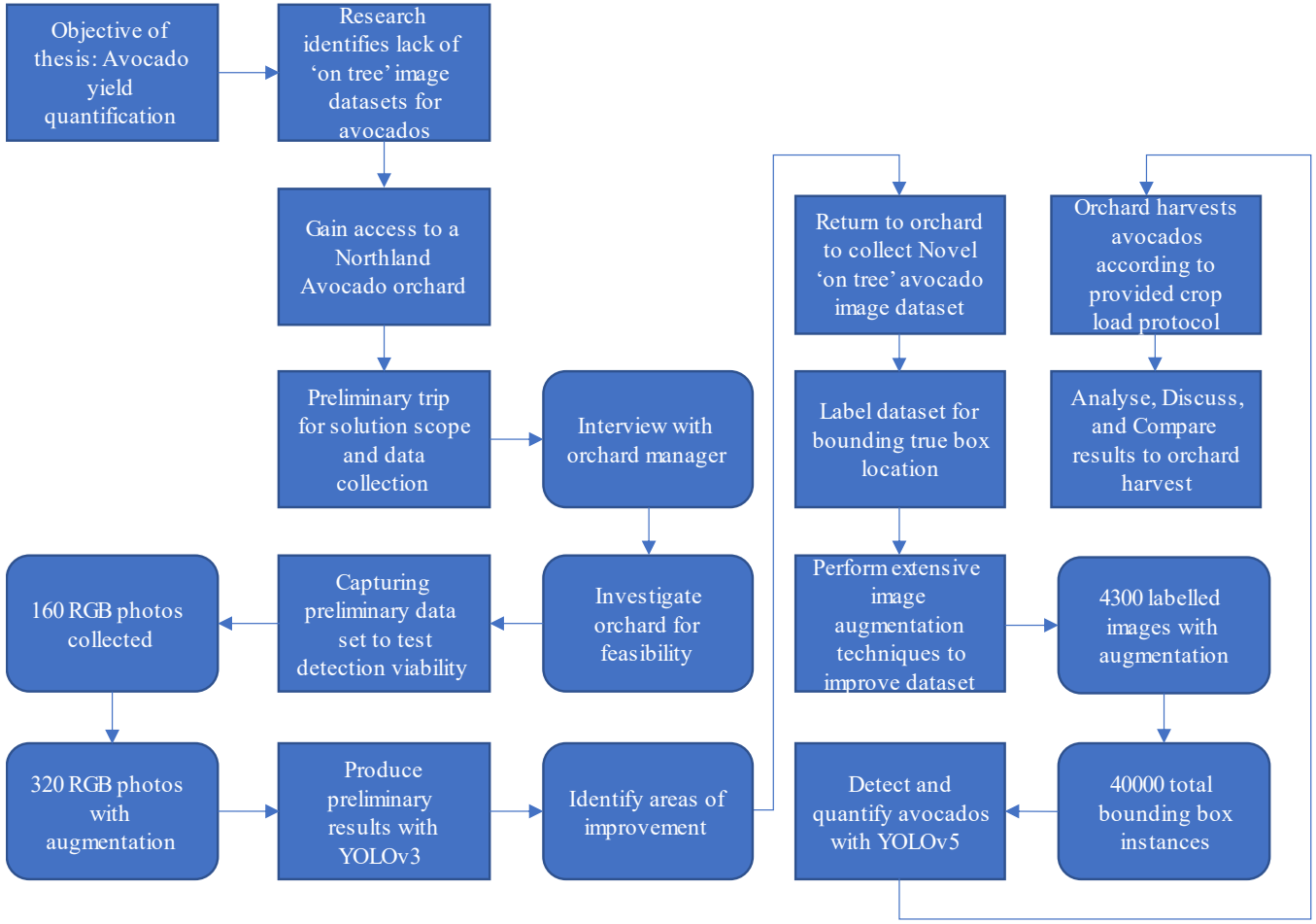

Figure B.4. Flow chart relating to the order in Chapter 3 as determined by the objectives of the thesis

Crop load protocol for zone 2, block 49

A. Avocado trees to be used on the study

The protocol applies to the following row/trees. Each tree has a tag mark indicating the row and the tree number

- **Row 5,** trees 1-42 (Except for the trees, 2, 7, 12, 17, 22, 27, 32, 37, which are pollinators or young trees)
- **Row 7**, trees 1-42 (Except for the trees 5, 6, 9, 22, 41, which are pollinators or young trees)
- **Row 8**, trees 1-41 (Except for the trees, 5, 10, 14, 19, 24, 29, 34, 39, which are pollinators or young trees)

- **Row 9,** trees 1-11
- **Row 10**, trees 1-11
- **Row 11**, trees 1-11 (Except for the trees 3, 8, which are pollinators or young trees)

B. Procedures to be conducted

Before the harvesting start (Instructions for harvest personnel leader)

1. Locate a fishbox under the following avocado trees

Row 5. Trees: 1, 5, 9, 13, 16, 20, 24, 29, 35, 39, 42

Row 7. Trees: 2, 7, 12, 15, 19, 24, 28, 33, 35, 38, 40

Row 8. Trees: 1, 4, 7, 9, 13, 16, 22, 28, 32, 37, 41

Row 9. Trees: 2, 7,10

Row 10. Trees: 1, 9

Row 11. **Trees:** 4, 10

***During harvest operation (Instructions for harvest personnel*)**

2. A single person should harvest each tree
3. Ask the harvest personnel to count each avocado while harvesting
4. Ask harvest personnel to write the number of fruits of each tree in the tag attached to each avocado tree.
5. For the trees with a fish box underneath, the fruit should be left on the corresponding box. While, for the trees without fish boxes, fruit can be handled directly to the bin.
6. If the trees have "0 (zero)" fruit write down in the table a number "0"

After harvesting operations (Instructions for harvest personnel leader)

7. Personnel leading the harvest operation may count in each fish box the number of fruit and write it down in the attached tables (**in the tables bold number are the trees with a Fish box**)*
8. Personnel leading the harvest operation may write down the count of each tree written in the tree tag in the attached table.

*If step 6 can be performed at the same time the harvest personnel is counting the avocadoes, the information can be used to identify potential counting errors and correct them simultaneously.