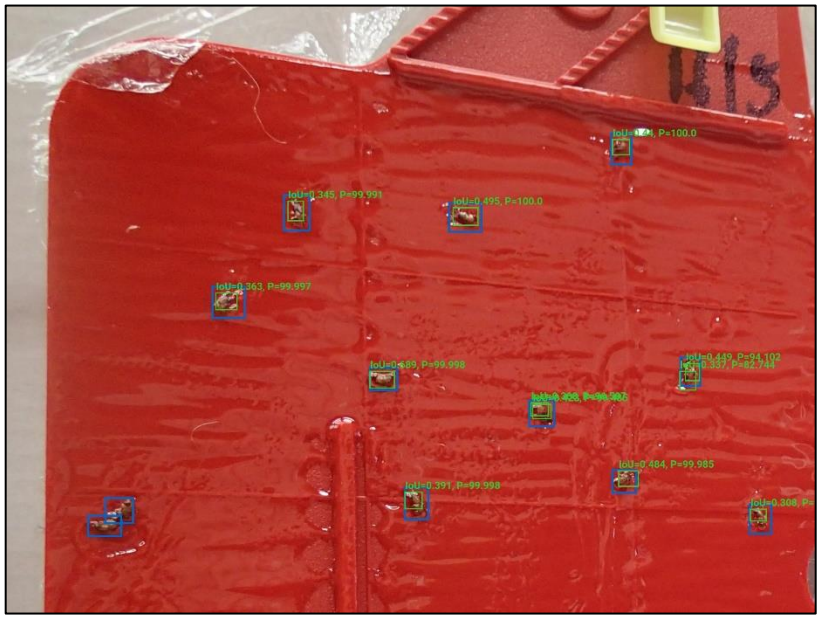
Geo-information Science and Remote Sensing

Thesis Report GIRS-2020-02

Object Detection for Automated Airborne Pest Monitoring of *Drosophila suzukii*

The effect of view angle and distance on the accuracy of an object detection model

Daniel Lamont



31-01-2020

WAGENINGEN
UNIVERSITY & RESEARCH

Object Detection for Automated Airborne Pest Monitoring of *Drosophila suzukii*

The effect of view angle and distance on the accuracy of an object detection model

Daniel Lamont

Supervisors:

dr. Peter P.J. Roosjen
dr.ir. Lammert Kooistra

A thesis submitted in partial fulfilment of the degree of Master of Science
at Wageningen University and Research Centre,
The Netherlands.

31-01-2020
Wageningen, The Netherlands

Thesis code number: *GRS-80436*
Thesis Report: *GIRS-2020-02*
Wageningen University and Research Centre
Laboratory of Geo-Information Science and Remote Sensing

Contents

Abstract

Integrated pest management (IPM) is a method of crop protection that combines different strategies and practices with the goal to reduce the negative impacts of current harmful pest management strategies. A critical part of IPM is the monitoring of crops to obtain data that is needed to make management decisions. A major pest in agriculture focused on fruits is the fruit fly species the "spotted wing drosophila" (*Drosophila suzukii*), which damages soft-skinned fruits such as raspberries, peaches, cherries etc. Unlike other fruit fly species that use damaged or overripe fruits for reproduction, *Drosophila suzukii* damages fruits before the ripening stage, causing huge economic loss to farmers. Therefore, an early detection system is needed to detect *Drosophila suzukii* as soon as possible. One way to do this is by using a combination of unmanned aerial vehicles (UAV), Deep learning (DL) object detection models (ODM's) and sticky traps to automatically monitor and detect the presence of these flies. The use of UAVs in such systems needs to be examined as there are many different aspects that could cause problems. In this thesis, the goal is to determine how the pose of a UAV towards the sticky-traps affects the accuracy of the ODM. Therefore, an experimental set-up was constructed where images were made of multiple traps containing *Drosophila suzukii* flies at different view angles and distances; 0°, 10°, 20°, 30° and 40°, and 14 cm, 24 cm and 34 cm. Furthermore, an ODM was trained to detect the flies and used to determine the effect of the mentioned variables. The accuracy of the ODM was quantified with the use of three measurement units; precision, recall and the harmonic average between the two: the F_1 score. Afterwards a statistical analysis was conducted, using a combination of the Mann & Whitney U test, ANOVA-test and a polynomial linear regression. The position of 24 cm was optimal and resulted in the highest recall, precision and F_1 score. The position of 14 cm resulted in a higher accuracy than 34 cm. The effect of the view angle was most pronounced at the distance of 24 cm and 34 cm and, while the negative effect of the view angle was very limited for precision, it was strong for recall. The position of 24 cm was optimal and resulted in the highest recall, precision and F_1 score. However, a position of 14 cm resulted in a higher accuracy compared to a distance of 34 cm. Furthermore, the larger the view angle, the larger the negative effect on the precision, recall and F_1. Additionally, the negative effect of the view angle on the accuracy was three times larger than that of the distance (within the parameters of this experiment). Thus, the results of this thesis indicate that positioning of the UAV is an important aspect to research as it has clear effects on the ability of an ODM to correctly detect objects.

Chapter 1 Introduction

1.1 Introduction

One of the major issues facing agriculture since its emergence roughly 10,000 years ago, is that of agricultural pests. These can consist of animals, plants or fungi that cause damage to agriculture by either physically destroying the crop, competing for resources or by acting as a vector for diseases that negatively affect the crop. Therefore, pest control, or the act of suppressing agricultural pests, is of major importance to minimize the economic loss of agriculture (Metcalf & Luckmann, 1994). Chemical pesticides have resulted in improved food safety, resulting in a general increase in human health and quality of life due to the increased yield of crops (Dobson & Cooper, 2007). Although, in contrast, pesticides can also have direct negative effects on humans who come in contact with them (Fantke, Friedrich & Jolliet, 2012). Furthermore, they also negatively affect agriculturally beneficial organisms and cause environmental damage (Miller & Spoolman, 2014). Therefore, it is important to minimize the use of pesticides.

Integrated pest management (IPM) was introduced around the 1950s. IPM is a framework of crop protection that combines different strategies and practices with an emphasis on environmental health, human health and natural pest control mechanisms (FAO, 2019). A critical part of IPM is the monitoring of the crops, as data is needed on e.g. the crops, environmental parameters and pest amount, to make informed management decisions. An important part of this is to determine when and where a population threshold is reached to be able to implement certain crop protection methods. The overall goal of an IPM is to maintain a low usage of pesticides by reduced and partial application (Ehler, 2006).

A pest that is causing major damage to soft fruits such as raspberries, peaches, cherries and other soft-skinned fruits is a fruit fly; *Drosophila suzukii* (figure 1.1). Originating from Japan and first reported in 2008 in North America and 2009 in Europe, it has become a major pest (Mortelmans, Casteels, & Beliën, 2012). Unlike other fruit fly species that use damaged or overripe fruits, the female *Drosophila suzukii* have a serrated ovipositor that lets them use fruits that are still in the ripening stage to oviposit eggs in. This results in damage to the fruit, reproduction of *Drosophila suzukii* in the crop and increased possibility of secondary infection by pathogens (Lee, et al., 2011).

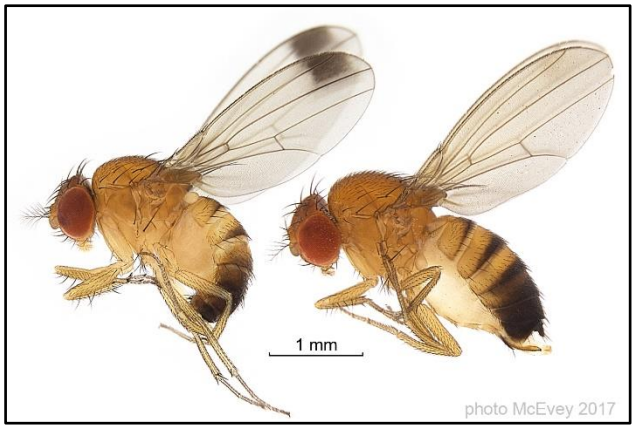


Figure 1.1 Male and female *Drosophila suzukii* flies. The fly on the left is the male and the fly on the right is female. Notice the black spots on the wings of the male (McEvery, 2017).

Traditional methods of obtaining quantitative and qualitative data on crop and pest characteristics consist of trained laborers that have to go into the field to manually inspect traps or crops and document data. This is an expensive, time consuming, labour intensive activity and also prone to human error (Bennet, 2010). A relatively new form of obtaining field data is via the combination of images obtained with the use of a camera mounted on a UAV, object detection/object recognition software using machine learning (ML) software and deep learning (DL) algorithms.

ML is a type of artificial intelligence where the model is provided with input and the requested output. With the use of statistical analyses, the model is able to "learn" how to interpret the input to obtain a specific output. A subset of ML is called deep learning. In DL, the model interprets data features and its relationships very much like in ML. The main difference between ML and DL is that the latter can automatically discover specific visual patterns to be used for classification, whereas with ML the specific features need to be provided manually. Due to this, DL models can be trained using images, while ML models need to be programmed. The major advantage of using DL to automatically detect specific objects (for example insects) in images is that data can be obtained faster, more reliable and on a more regular basis, reducing the needs of human workers for monitoring, thus reducing the labour costs. Using artificial intelligence to detect objects is not limited to detecting pests (insects, weeds, rodents etc.), but can also be adapted to detect nutrient deficiencies, droughts and many other issues in agriculture (Kamilaris & Prenafeta-Boldu, 2018).

A DL algorithm consists of multiple layers that are connected, with each layer having a certain number of nodes (a deep neural network). The input could be a set of images that contain a certain object such as a car. The input is passed on to the first layer of nodes, where each node analyses a pattern between the pixels; edges, curves, corners etc. This output is then passed to the next layers of nodes that find more abstract relationships between the patterns, to be passed on to the next layers and so on. The network is trained by providing images containing specific objects. Each node is given a certain value that depends on the importance of that specific feature. The value of each node is updated during the training phase via a feedback loop. This feedback loop is how the DL model "learns", improving the detection with each convolution. (Krizhevsky, Sutskever & Hinton,2012).

Object detection and object recognition are both technologies in the field of computer vision. With object detection, the goals are to detect instances and the locations of objects in an image and recognize what the objects are. Depending on the purpose of the object detection model (ODM), objects can be people, cars, traffic lights, animals etc. With object detection, the algorithm is capable of determining what class the object is. For example, if an image shows a cat and a dog, the output would be a bounding box around both the cat and the dog, with a probability given that the object in the image is of a specific class (Ren, He, Girshick, & Sun, 2015). In the field of image analysis, deep neural networks are used that are called convolutional neural networks (CNN). A CNN is very similar to a deep neural network, but contains one or more "convolutional layer(s)", which use filters to process small pieces of the image that are important for classification into a new output (feature map). This output is then used as input further down the network. CNNs have been used in many different image classification tasks, ranging from everyday objects (Chen, Xiang, Liu, & Pan, 2014), animals (Kellenberger, Marcos, & Tuia, 2018) and agricultural pests (Faria,

et al., 2014; Ding & Taylor, 2016; Cheng, Zhang, Chen, Wu, & Yue, 2017; Chen et al., 2018) among others. Therefore, CNNs can be potentially useful for DS monitoring.

Images used as input for an ODM need to be of a certain quality to be useful; low amount of noise, low amount of blur and certain levels of brightness (Grm, Struc, Artiges, Caron, & Ekenel, 2017). It is unknown how far the images can deviate from the optimal conditions before the accuracy of the object detection starts to drop. Furthermore, the image quality also depends on the pose of the UAV towards the trap. Many consumer-positioning systems provide inaccurate positioning and orientation data (Colomina & Molina, 2014; Fangning, Zhou, Xiong, Hasheminnasab, & Habib, 2018). Additionally, a positioning bias can occur due to vibrations from hovering and shaking due to wind. Overall, many different deviations in pose accuracy are presented in literature, ranging from a deviation of 3-5 cm to 20-30 cm in horizontal and vertical positioning, respectively (Benassi, et al., 2017). This large variation in accuracy will likely result in large qualitative diversity of images. Therefore, it is important to determine how the pose of the UAV affects the accuracy of an ODM to determine the maximum deviation a UAV may have in its pose.

Due to the (relatively) low quality of the camera of the UAV for close-up images, the flies were only visible as mere grey blotches and were unable to be detected as actual flies. Therefore, the decision was made to use a hand held camera which produced better quality images at close distances. The implications of this will be further discussed in chapter 2 and chapter 4. Monitoring the presence of DS with the use of a UAV-based camera can be used as input for a decision support system which could help farmers in their decision-making.

1.2 Research objectives

The objective of this research is to make a set of recommendations that describe the maximum view angle, the sub-optimal view angle and the distance a camera can be placed from the trap, while still maintaining a predefined accuracy threshold. In the field, a UAV that is programmed to obtain images will experience conditions that could negatively influence the quality of the image, resulting in sub-optimal data input for the detection algorithm. These conditions could consist of; positioning bias, orientation bias, wind and illumination differences. Therefore, it is important to determine how sub-optimal orientation affects the quality of the data and the classification accuracy of the classification model.

In this thesis, the parameters of these conditions will consist of; 1) the distance between the trap and the camera, 2) the view angle between the trap and the camera, and 3) a combination of the distance and the view angle between the trap and the camera.

To determine how the model will perform in practice with imperfect conditions, the following research questions will be answered;

Main research questions:

1. Can *Drosophila suzukii* be detected in images using a CNN based object detection model?
    1.1 Which object detection model will be suited for detecting *Drosophila suzukii?*

2. How does the positioning of the camera affect the accuracy of the object detection model?
    2.1 How does the distance between the camera and the trap affect the accuracy of the object detection model?
    2.2 How does the view angle between the centre of the camera and the trap affect the accuracy of the object detection model?

Hypotheses:

1. In literature, multiple small agricultural pests have successfully been classified with the use of multiple detection models (Kamilaris & Prenafeta-Boldu, 2018). These pests were often of a similar size as *Drosophila suzukii* flies. Thus, the hypothesis is that *Drosophila suzukii* flies can be detected using a CNN based ODM.
    1.1 A preliminary literature study was conducted and it was determined that the Faster R-CNN will be a good candidate for the detection of DS.
2. The hypothesis is that there is an optimal view angle and distance where the classification accuracy is the highest. At a certain view angle and distance, the accuracy will start to drop until it becomes too unreliable.
    2.1 The hypothesis is that deviating from the optimal distance, will negatively affect the accuracy of the ODM. Furthermore, if the distance increases from the optimal distance, the accuracy will most likely be negatively more affected than when the distance decreases. The reason for this is that with an increase in distance, the relative size of the flies will be smaller in the image than with a decrease in distance, making it harder for the ODM to accurately detect the flies.
    2.2 The hypothesis is that with an increasing view angle, the accuracy of the ODM will have an increasingly more difficult task at detecting flies on the target. The reason for this is thought to be due to the smaller surface area of a fly on the image due to the view angle, making it more difficult for the ODM to detect the fly. Also, the different perspective of the flies due to the increasing view angle will negatively affect the detection.

The methods used to answer these research questions can be found in chapter 2. In chapter 3, the results of the experiment are presented. Chapter 4 discusses the implications of each result and in chapter 5 the conclusions that were deducted from the results are presented.

Chapter 2 Methods and materials

2.1 Introduction

In this chapter the experiment will be explained; how it is built and what materials were used. First a general summary will explain the overall set-up and experiment. This will then be followed by a detailed description of each separate section.

The goal of this thesis is to determine how an ODM for the detection of *Drosophila suzukii* flies is affected by the pose of the camera towards the trap containing the flies. To determine this, an experiment was designed that would answer the research questions. This resulted in several components that had to be designed and conducted. First, an experimental set-up was devised that would simulate the deviations in position and orientation a UAV could encounter in a real-life scenario. Second, an ODM that could detect very small objects in an image was selected. To determine how well the model is at detecting *Drosophila suzukii* flies on a trap, a method for quantifying the accuracy was chosen from literature. The ODM also had to be trained and therefor a training set had to be created. This was done by manually adding *Drosophila suzukii* flies to multiple traps. The traps were then used in the experiment to create a dataset of images made at different view angles and distances. To be able to test the model, each image was manually labelled, resulting in 600 labelled images. From these images, a selection was made of images made at a view angle of 0° and a distance of 24 cm, which were then used for the training of the ODM. After the training, the remaining images were used to test the ODM on its accuracy at different view angles and distances. The results were then manually inspected and documented to be analysed.

2.2 Design of the experiment

The experimental set-up (figure 2.1) was designed to be able to make images of a trap containing *Drosophila suzukii* flies from different view angles and distances. The first step was to determine what the optimal distance of the camera from the trap was. The criteria for the "optimal distance" are: the distance at which the image is the sharpest, the clearest and at which the flies were clearly visible. Furthermore, the assumption was made that the object-detection model will always perform better or the same at a view angle of 0°. Therefore, the optimal distance was determined at a view angle of 0°. The optimal distance was determined by positioning the camera as close to the trap as possible while still having the entire trap in the image.

The camera was then positioned further away from the trap with steps of ± 4 cm and an image was taken. The images were then manually inspected to determine if the *Drosophila suzukii* flies were clearly visible and if the image was sharp and focused. Around a distance of 24 cm, the images became sharp, and each fly became more visible and detailed. Eventually the sharpest images were from a distance of 24 cm (measured from the lens to the front of the trap) and this was chosen as the optimal distance.

The distance of the two sub-optimal distances was determined based on the optimal distance and how much the image visually changed when increasing/decreasing the distance. At a distance of around 10-12 cm the visual change in an image was such that it became significant (visually determined). Therefore, a distance between the three positions of 10 cm was chosen. This resulted in a sub-optimal distance of 14

cm and 34 cm, and an optimal distance of 24 cm. The difference between the view angles was decided to be 10°, ranging from 0° till 40°. This resulted in five separate view angles (including 0°) at which three distances were tested. The maximum view angle of 40° was chosen because at that view angle the front of the trap was still visible, while at 50° parts of the trap were no longer visible. Furthermore, increasing the maximum view angle of the experiment, while still maintain a 10° difference, also increased the size of the data-set to such an extent that it would become impractical to manually label and manually analyse that much images. The positions in the experimental set-up were only conducted on the left side of the 0° line. This was done because the effect of the view angle was assumed to be the same on both sides of the 0° line.

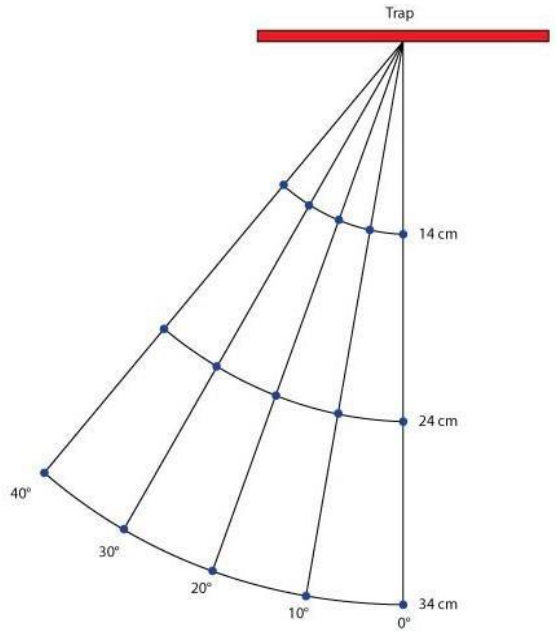


Figure 2.1 Experimental set-up. This image shows the experimental set-up from the top. Each dot represents the location of the camera where an image was taken for each trap.

The experiment was conducted at the Wageningen University in a room which was illuminated with artificial light and partly sunlight. The camera positions of the experimental set-up (figure 2.1) were drawn on a tabletop with a pencil. The camera was elevated with a block of wood so as to have the lens at a similar height as the centre of the trap.

2.3 Object-detection model

A literature study was conducted to determine what ODMs could be used for. It was important that publications were used in which open-source ODMs were used with emphasis on pest detection. This was because open-source was desired due to the free availability of these models, the availability of support information and the availability of additional software widely found on the internet. The focus on pest detection is due to the similar goal of this thesis. The most important similarity is that the objects of detection are very small and usually numerous in each image. Thus, it was important to find the best suitable ODM for this purpose. A number of papers used ODMs for the use of pest detection, as described in the review by Kamilaris & Prenafeta-Boldu (2018), but the most applicable of these publications was

Fuentes, Yoon, Kim, & Park, (2017). In this paper, the goal was to find the most suitable deep-learning architecture for the task of a faster and accurate detection of diseases and pests in plants. They looked at three main families of deep-learning architecture that are commonly used; *Faster Region-based Convolutional Neural Network* (Faster-RCNN) (Ren, He, Girshick, & Sun, 2015), *Region-based Fully Convolutional Network* (R-FCN) (Dai, Li, He, & Sun, 2016) and *Single Shot Multibox Detector* (SSD) (Liu, et al., 2016).

Region proposal network (RPN) is a model in object detection, where bounding boxes are proposed in an image where there is a likelihood that an object is present. Then in these proposal regions, a convolutional neural network preforms the classification. Faster R-CNN and R-FCN are able to perform the region proposals on the feature map (processed original image using filters), instead of on the original image. Faster R-CNN and R-FCN are able to 1) RPN classify object/not object, 2) RPN regress box coordinates, 3) Final classification score (object classes) and 4) Final box coordinates.

Single Shot Multibox Detector does not work with a proposal network, instead a fixed set of default object positions/scales and aspect ratios is defined. The network then generates scores for the presence of each object category in each default box and produces adjustments to the default box to better match the object shape. Faster R-CNN and R-FCN are both a "two-stage" ODM, due to the first stage being the RPN, and the second stage being the classification. SSD is a "single-stage" ODM, as there is no RPN needed (Fuentes, Yoon, Kim, & Park, 2017).

Of these three, Faster R-CNN had a better performance than both R-FCN and SSD. In Huang et al. (2017), the three mentioned meta-architectures were also compared to each other on basis of prediction accuracy and computation time. Faster R-CNN had the overall highest accuracy, but also had the longest GPU computation time (Huang, et al. 2017). Therefore, Faster R-CNN is the likely candidate, as accuracy is preferred above minimal computation time in the case of DS, due to the perceived difficulty of classifying a specific species of fruit fly.

Instead of using images of tomato plants containing diseases or pests, Fuentes, Yoon, Kim, & Park, (2017) used images obtained from the *Microsoft Common Objects in Context* database (COCO) to test the three different ODMs. COCO consists of over 2.5 million images of complex everyday scenes of common objects in their natural context (Lin, et al., 2014). The model with the highest accuracy (*mean Average Precision* was used, thus the mean AP*)* was Faster R-CNN, although it did also have the longest computation time.

Another option was *You-Only-Look-Once* (YOLO) (Redmon, Divvala, Girshick, & Farhadi, 2016). YOLO is an object-detection architecture where the input image is divided into grids (similar to SSD). Each square in the grid gets a confidence score, whether there is an object in it or not. The advantage of YOLO is that it is one of the fastest ODMs available. However, YOLO version 2 can have difficulties in identifying small objects in images, which makes in less practical in the scope of this thesis (Redmon, Divvala, Girshick, & Farhadi, 2016).

For this research Faster R-CNN was chosen, due to the high accuracy (especially with smaller objects), large amount of support information found on the internet, and relative ease of use.

2.4 Creating the image dataset

The image dataset was created for two purposes; the first is to be able to train the Faster R-CNN ODM to detect *Drosophila suzukii* flies, and the second is to test the ODM on images made at different view angles and distances. A total of eight traps were used, with a total of fifteen images made per trap. These fifteen images correspond to the fifteen positions of the experimental set-up (figure 2.1). This adds up to 120 images. Each image was then processed according to the image processing pipeline (figure 2.2). First dataset 1 was created. Due to the large size of each image (±2.5 mb and 3968 x 2976 pixels), each image was segmented using a custom python script (Split.py). This segmentation resulted in four equal sized images (±1.1 mb and 1984x 1488 pixels). The segmentation was done because the training phase of the model would be hampered due to the very large image size, as would the actual detection by the model. Furthermore, it also made the manual labelling and manual analysis of the images more manageable. The segmentation resulted in a total of 480 images. Each image was then manually labelled with "LabellMg", a graphical image annotation software package that can be used to label images for ODMs (Tzutalin, 2015). Afterwards, all the images made at a view angle of 0° and a distance of 24 cm were used for the construction of data set 2 and dataset 3.

The reason for only using images made at 0° for the training, is that this thesis is about the effect of view angle and distance on the accuracy of the ODM. Therefore, the training was only done with images made at the theorized optimal distance and optimal view angle. If images made at other view angles and distances were also used in the training, the results of the accuracy of the ODM would be very difficult to interpret correctly.

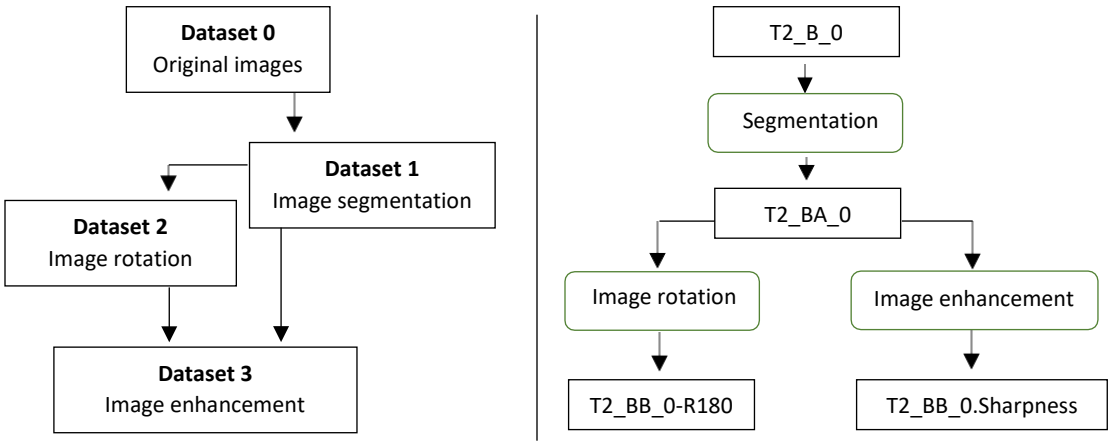


Figure 2.2 Image processing pipeline. Each image in the dataset is segmented and augmented according to this pipeline. For example, the image of trap 2 at a distance of 24 cm (B) and a view angle of 0° is segmented. Segment A is then both rotated and becomes T2_BB_0-R180, and enhanced and becomes T2_BB_0-Sharpness. This results in three images that can be used for the training phase.

Dataset 2 and dataset 3 are the augmented images. Augmentation is done to increase the number of images in the dataset by altering the images in such a way that for the training of the ODM, they seem to be different images altogether (Perez & Wang, 2017). In this thesis, multiple types of augmentation were done: rotation, changes in the colour balance, contrast, brightness and sharpness of an image. In dataset

2, the images were rotated with a custom python script (Rotate.py), which would randomly rotate the image by either 90°, 180° or 270°. The coordinates of the manually labelled boxes were also translated by the python script to fit the rotation. Each image of dataset 1 was then used to make dataset 3. Dataset 3 consists of images that have been visually altered with a python script (augmentation.py) by either altering the colour balance, the contrast, the image brightness or the sharpness (figure 2.3 image). The degree of alteration is the "enhancement factor" which is a number between 0.0 and 1.0. For colour balance, an enhancement factor of 0.0 will give a black and white image. With contrast, 0.0 will give a solid grey image, and for brightness a black image. An enhancement factor of 1.0 will give the original image. For sharpness the enhancement factor is a number between 0.0 and 2.0 where 1.0 is the original image, 0.0 is a blurred image and 2.0 produces a sharpened image. The enhancement factor is therefore randomly chosen between a number of 0.15 and 1 (0.15-2 for sharpness). A minimum enhancement factor of 0.15 was chosen as to have an actual visible change in the altered images. The type of augmentation done by the script is also randomly chosen.

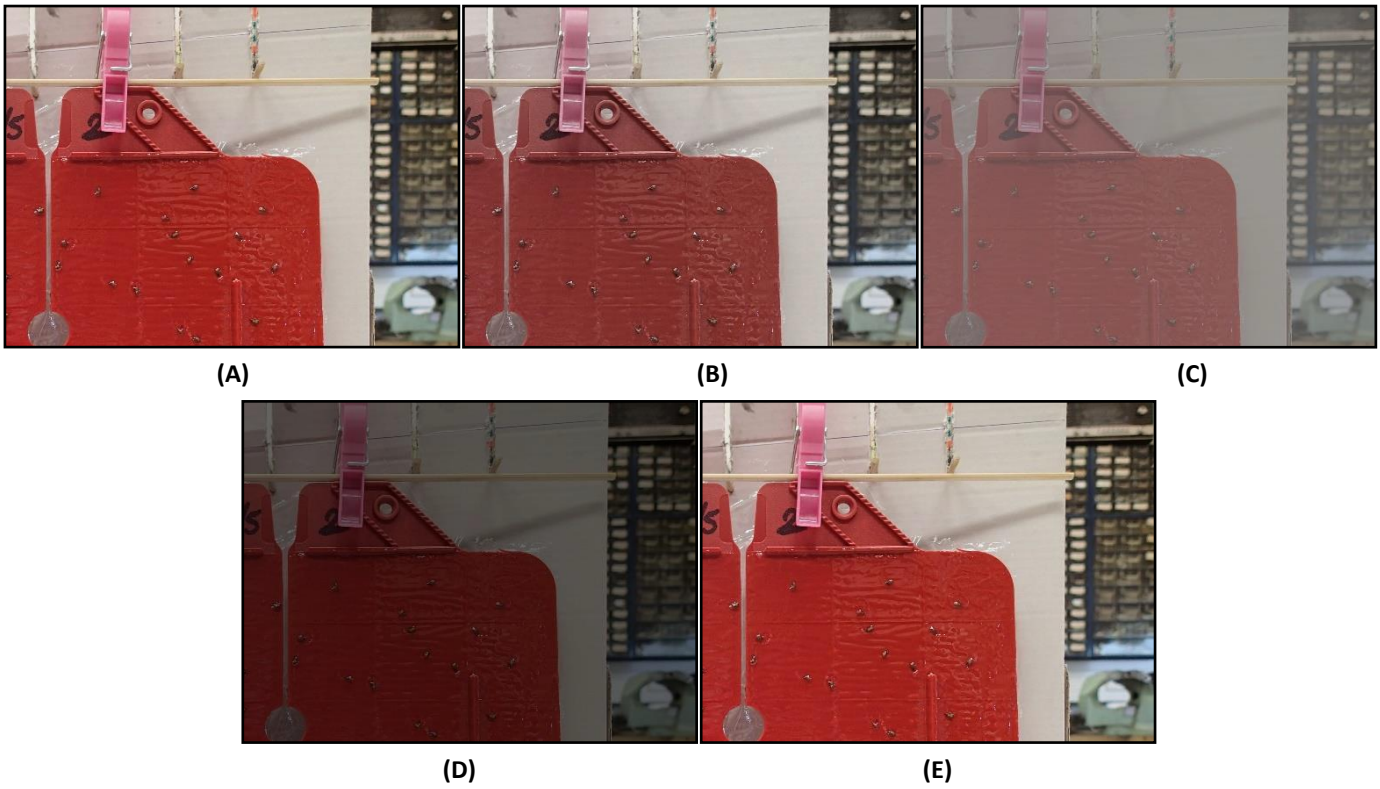


Figure 2.3 Image enhancement example. The first image is the original (segmented) image (A). The following images are altered in their; colour balance (B), contrast (C), brightness (D) and sharpness (E) respectively.

2.5 Training object detection model

The Faster R-CNN model was trained using only images that were made at a distance of 24 cm and a view angle of 0° and the enhanced and augmented images made from these images. To train the ODM, two datasets were created from the three datasets shown in figure 2.2; the training dataset and the test dataset. The training dataset is used by the ODM to train the neural network while the test dataset was used by the ODM to evaluate the neural network. Both datasets were made by first creating a file consisting of 96

images with a part of the segmented images of trap 2, and trap 3, the augmented and enhanced images of trap 2 and trap 3 and enhanced and augmented images from a selection of the remaining traps. The file was then randomly split up in the training dataset consisting of 80 images and the test dataset consisting of the remaining 16 images. The training was conducted on a desktop pc. The training ran until the loss consistently dropped below 0.05. The "loss" is a quantification of how well the ODM models the given data. When the loss value is small, it means that the prediction of the ODM has an insignificant small deviation from the actual results. After the training was conducted the ODM was used to analyse the images of trap 2 and trap 3 that were made at a view angle of ≥10° and the remaining traps at all view angles and distances. This resulted in a total of 456 images that were used for the statistical analyses. All the steps taken to install and prepare the ODM were taken from the github page of "EdjeElectronics" The specific Faster R-CNN model that was used from EdjeElectronics (2018) is the "faster R-CNN Inception Resnet V2 (COCO)" version. The default configuration setting was used with only one class. The configuration file ("faster_rcnn_inception_v2_pets.config) can be found at the github page of EdjeElectronics (EdjeElectronics, 2018).

2.6 Quantifying accuracy

To be able to determine how well the ODM was in detecting *Drosophila suzukii* flies on an image, a method of quantifying the accuracy was selected form literature. A widespread method used in the field of object detection is that of *precision* and *recall*, which together can be used to calculate the F_1 score. In general, the F_1 score is solely used as a measure of accuracy, and precision and recall are only used to calculate F_1. In this thesis recall and precision scores will also be used separately as both provide valuable information that otherwise is lost in the transformation to the F_1 score (Hand & Christen, 2018).

Intersect over union

To determine how well each individual detection is made, the intersect of union is used. When the ODM identifies a fly, it will place a *bounding box* over the object that the ODM predicted as a *Drosophila suzukii* fly. The coordinates of these bounding boxes combined with the manually labelled bounding boxes were then used to calculate the intersect of union (IoU). If the IoU was 0.25 or more, the detection is considered successful. In literature (Hui, 2018) the value of 0.5 is usually used, however due to the very small size of the targets in the images, it was chosen to reduce this number to 0.25.

Precision

The precision explains what portion of the positive identifications were actually correct. A positive identification is an identification made by the ODM of an object that it perceives to be of a specific class it was trained to detect, in this case a *Drosophila suzukii* fly. A positive identification is split between true positive (TP) and false positives (FP). The TP are actual objects of a specific class, while FP are objects or parts of an image that are identified as a specific class, yet in reality, are wrongly classified.

$$Precision = \frac{TP}{TP + FP} \tag{2.1}$$

Recall

Recall looks at the fraction of objects that are correctly detected compared to the objects that were incorrectly miss detected. For example, if an image contains ten *Drosophila suzukii* flies and the ODM only detected four of them, the recall value would be 0.4, as only 40% of the flies on the image were correctly detected.

$$Recall = \frac{TP}{TP + FN} \quad (2.2)$$

F_1- score

The F_1 score is calculated using the recall and precision value to create an average score. An F_1 value of 1.0 means that the image has a perfect precision and recall score. The lowest value is 0.0, which means that there has not been a single correct identification made in the image. The F_1 score is used as an average score of the object detection of an image. However, both the recall and precision will also be separately analysed in this thesis to have a more in depth analyses of the ODM.

$$\mathbf{F_1} = 2 \cdot \frac{precision \cdot recall}{precision + recall} \quad (2.3)$$

2.7 Statistical analyses

Multiple statistical analyses have been done on the result data. First, for the comparison of the precision, recall and F_1 scores (dependent variables), a Mann & Whitney U test was conducted on the mean value of each distance at a view angle of 0° to determine if there was a significant difference between the distances (paragraph 3.2). Second, another Mann & Whitney U test was performed to determine if the mean values at each view angle were significantly different from each other (paragraph 3.3). The Mann & Whitney U test is similar to the traditional t-test, yet the assumption of normal distribution is not required. This is important as it is not known if classification accuracy is based on a normal distribution. Next, the precision value was plotted versus the recall value (paragraph 3.4) to determine how the difference in view angle changes both dependent variables. This was done three times, once for each distance. Lastly, a polynomial linear regression was conducted for each of the dependent variables (paragraph 3.5). This was done to determine how much the variables distance and view angle change each of the dependent variables. First, the data was plotted for distance (figure 3.12) and then for view angle (3.13). Then the trendline was drawn to determine which type of trendline would explain the pattern of the data the best, which was the polynomial trendline. To justify the use of a linear regression on the data, certain assumptions have to be met, which can be found in Annex I. The results (table 3.7, 3.8, 3.9, 3.10, 3.11 and 3.12) of the model used were then further analysed using an ANOVA test. The ANOVA test results show how much effect each model-variable has on the end result compared to each other. The entire polynomial linear regression was done using the R programming language in R-studio.

2.8 Polynomial linear regression

A polynomial linear regression model was made for precision, recall and the F_1 score to determine how both the view angle and the distance affect the three dependent variables. The model used in the analysis is shown in figure 2.4. The model takes the separate effect of distance, the separate effect of the view angle, the separate effect of both the distance and the view angle squared and the interaction between the distance and the view angle into account. The model is run three times, once for each dependent variable separately.

$$y = \beta_0 + \beta_1 x_1 + \beta_1 x_1^2 + \beta_2 x_2 + \beta_2 x_2^2 + \beta x_1 x_2$$

Figure 2.4 Polynomial regression model. The model is written in the R programming language and run in RStudio.

The variable “y” stands for either “precision”, “recall” or “F_1”. “ß0”, stands for the intercept, which is the value of the x-axis, if the “y” value is zero. The variable “ß1x1” stands for the amount the “y” value changes if “x1” would change with a unit of 1. In this case it stands for the effect of distance, thus if distance would change with 1 cm, the “y” value would change with the value of “ß1”. The variable “ß2x2” represents the change of “y” if the view angle would change with a value of 1°. The variables “β x 2” and “β x 2” are the polynomial factors of distance and view angle respectively. The variable “ßx1x2”, represents the interaction of distance and view angle. It is the amount the “y” value would change if the distance and the view angle change with a value of 1 cm and 1°. A polynomial regression is seen as a special form of a multiple linear regression. This is because the variables x1 and x2 are transformed to be able to capture a curvilinear relationship between the variables and “y”, yet still represents a linear correlation.

Chapter 3 Results

3.1 Introduction

In this chapter the results of the experiment will be presented. The dependent variables (precision, recall and F_1 score) are used as a grading system to show the effect of the predictor variables (distance and view angle). Even though the precision and recall make up the F_1 score, they are still used separately in the result section and further in the discussion and conclusion. The reason for this is because the precision and the recall provide valuable information about the accuracy of the ODM. The recall shows how well the model is in the identification of an object in an image as a specific class. In figure 3.1 the total amount of flies is fourteen, but the model only detected nine flies, and one object incorrectly as a fly. For this image, the recall was 0.64, meaning that 64% of the flies that were actually on the image were correctly identified as a *Drosophila suzukii*.

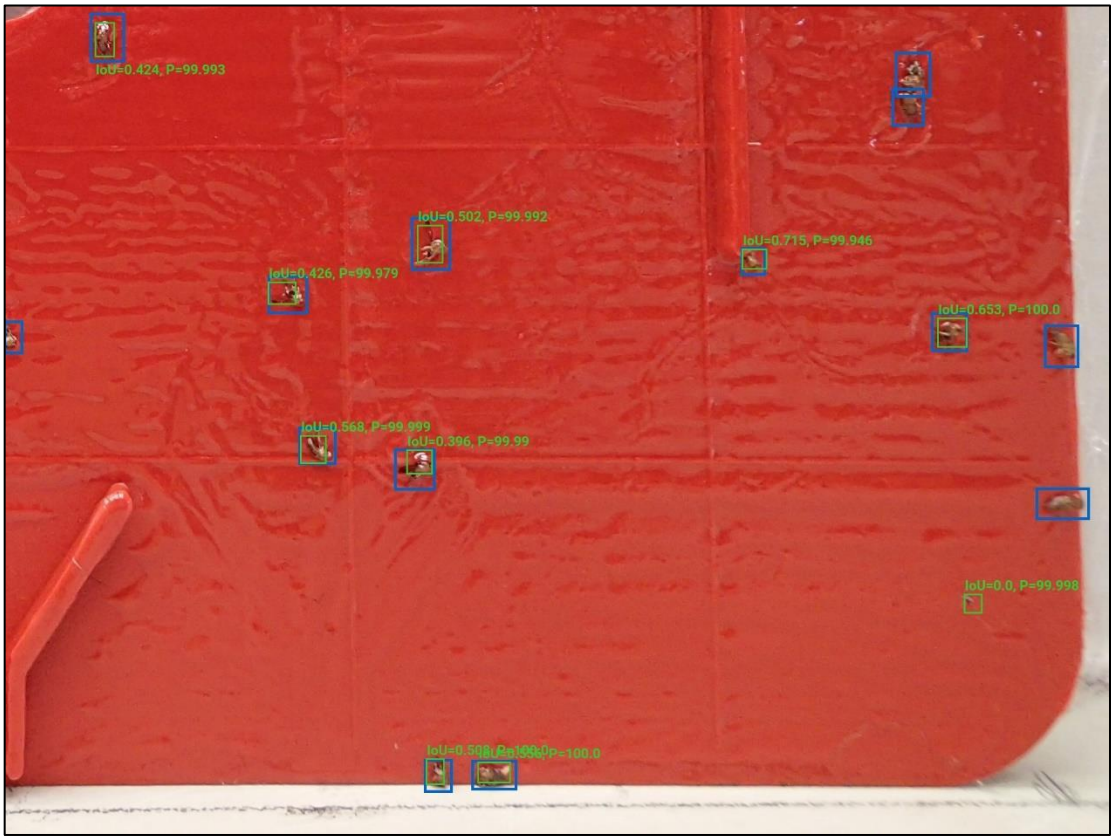


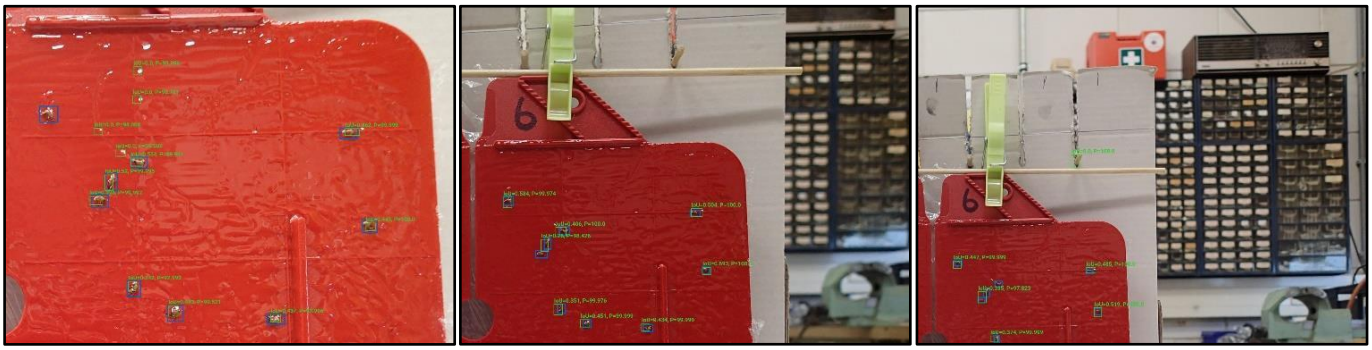
Figure 3.1 Precision and recall example. The blue squares are the manually labelled flies and the green squares are the predictions made by the model. The model generated rectangles also have a intersect of union score (IoU) and a probability score (P).

Precision on the other hand explains more about the actual identification instead of the detection. In figure 3.1, the total amount of objects identified as *Drosophila suzukii* fly is ten, of these only nine were actual *Drosophila suzukii* flies. The calculation for precision is; precision = true positives / (true positives + false positives) (9/10). The precision value of this image is 0.9. This means that of the positive

identifications made by the ODM 90% is correct.
The experiment was conducted at three different distances and five different viewing view angles (15 unique positions). First, the results of the distances are shown. Then the results of the view angles are presented, followed by a comparison of the precision and recall scores. This is followed by the results of the linear regression model. Lastly the effect the view angles and distances have on the sizes of the flies in an image is shown. This will be further discussed in detail in the discussion chapter (chapter 4).

3.2 Distance comparison results

In figure 3.2 an example is given of how distance effects the accuracy of the ODM.

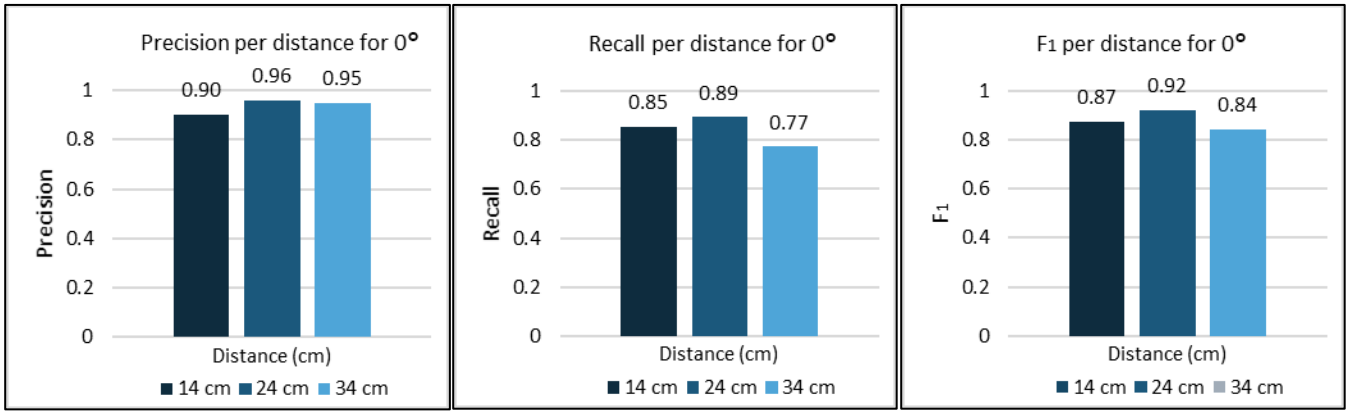


(A) **(B)** **(C)**

Figure 3.2 Model results at different distances. Images A, B and C are taken at an angle of 0°, with A taken at a distance of 14 cm, B at a distance of 24 cm and C at a distance of 34 cm. The scores are; A: precision (0.67), recall (0.89), F_1 (0.76), B: precision (1.00), recall (0.89), F_1 (0.94), C: precision (0.83), recall (0.71), F_1

These three images show some of the visual effects that distance has on the images. In image 3.2A, the background is almost non-existent while in image 3.2B this is partly visible in the right part of the image. Image C exists of more background than actual trap. Furthermore, the fraction that the size of a fly compared to the size of the entire image is, is greatly reduced the further the camera is placed from the trap. In image A, it is possible to distinguish a fly with the naked eye, while in image C this is hardly possible. The results on the fly size will be presented later in this chapter, and both visual effects presented here will be discussed further in the discussion chapter.

In figure 3.3, the mean precision, mean recall and mean F_1 score for each image (3.3A, 3.3B and 3.3C) are shown. In table 3.1, the p-value is given for the comparison shown in figure 3.3.



(A) **(B)** **(C)**

Figure 3.3 Mean distance results at 0°. The mean recall, mean precision and mean F_1 score are shown here for the distances of 14 cm, 24 cm and 34 cm. All the results are from the view angle of 0°.

For figure 3.3A, the mean precision value is lowest with a value of 0.90 at a distance of 14 cm and highest with a value of 0.96 at a distance of 24 cm. The mean precision at a distance of 34 cm is 0.95, which is insignificantly different with a p-value of 0.697 from the score at 24 cm as shown in table 3.1.

Table 3.1 Mann & Whitney U test results for distance comparison.

Distance	Precision (p-value)	Recall (p-value)	F_1-score (p-value)
14/24	0.006	0.330	0.053
24/34	0.697	0.002	0.003
14/34	0.026	0.045	0.302

The optimal distance of 24 cm has the highest mean precision value, but it seems that the change in mean precision due to distance is minimal when increasing the distance (34 cm), and is strong when decreasing (14 cm) the distance (figure 3.3A). For the comparison between a distance of 14 cm and a distance of 34 cm, the result is also shown to be significant with a p-value of 0.026. This reinforces the idea that precision is more negatively affected when the distance is decreased from the optimal position. In figure 3.3B, the mean recall value for the optimal distance of 24 cm is the highest score with a value of 0.89. The mean recall value at a distance of 14 cm, 0.85, is not significantly different from that at 24 cm, while the mean recall score at a distance of 34 cm is significantly different from 24 cm. For recall, it seems that the reverse of what we see at precision is true. When the distance increases from the optimal position, the recall value is negatively affected, while when decreasing the distance seems to have no real negative effect. When comparing the results of the recall value with a distance of 14 cm and 34 cm, we can see that the difference is significant with a p-value of 0.045. For F_1, (figure 3.3C), when the distance increases, it has a larger negative effect on the overall accuracy of the object detection then when the distance is decreased. The p-value for the comparison of 24 cm with 34 cm shows that the difference is significant, but for the comparison of 14 cm with 24 cm, the difference is almost significant with the p-value of 0.053.

For the comparison of F_1 score between 14 cm and 34 cm, the p-value is 0.302, which indicates that the difference is not significant. Thus, overall the results indicate that decreasing the distance will negatively affect the precision value more than the recall, while increasing the distance will negatively affect the

recall more than the precision. Overall it also looks like the negative effect of distance on recall is larger than the negative effect of distance on precision. This is emphasized by the Mann & Whitney U test result of the F_1 score of 0.003 for the comparison of 24 cm and 34 cm, which shows that the difference between these distances is significant while that of 14 cm and 24 cm is less significant.

3.3 View angle comparison results

For a general visual idea how the view angle of a camera towards an object changes the image, see figure 3.4. Just as with figure 3.2, the amount of total background visible in the image changes with the view angle. This is somewhat unclear between image 3.4B and 3.4C, but for image 3.4A, 3.4D and 3.4E it is clearly visible that with the increase of the view angle, the amount of background increases. Furthermore, the size of the target and the size of the individual flies in the image also decreases with an increase of view angle. The scores clearly show how view angle has an effect on the accuracy of the detection, with image 3.4A and 3.4B having maximum scores of 1.00 for each dependent variable, but this decreases in 3.4C slightly and in 3.4D more and also has a very strong drop in image 3.4E.

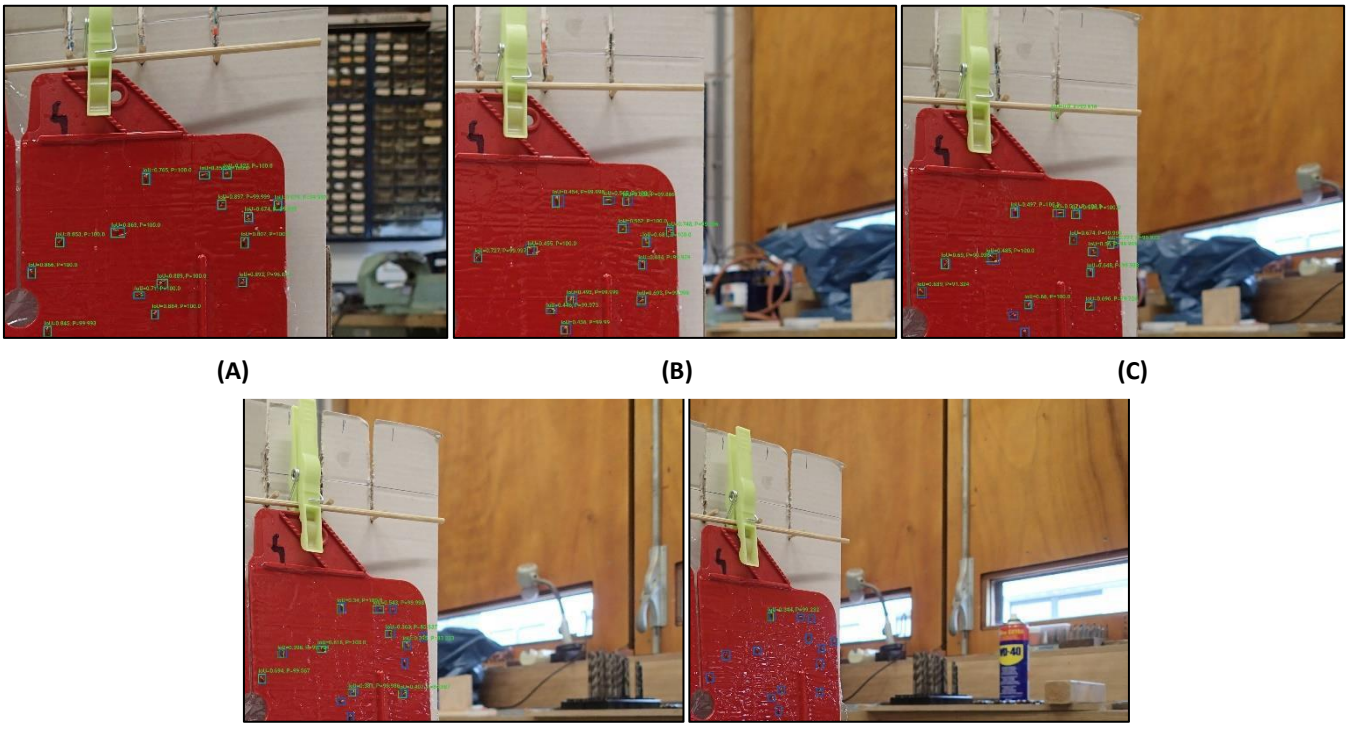


Figure 3.4 Model results at different view angles. Images A, B, C, D and E are taken at a distance of 24 cm and at an angle of 0°, 10°, 20°, 30°, and 40° respectively. The scores are; A: precision (1.00), recall (1.00), F_1 (1.00), B: precision (1.00), recall (1.00), F_1 (1.00), C: precision (0.92), recall (0.86), F_1 (0.89), D: precision (1.00), recall (0.64), F_1 (0.78), E: precision (1.00), recall (0.07) and F_1 (0.13).

Furthermore, it also shows how the precision can have a high value even though the ODM misses a lot of *Drosophila suzukii* flies in the image. For example, in image 3.4E the model is only able to detect one

Drosophila suzukii fly, missing 13 *Drosophila suzukii* flies, yet the score is 1.00 due to how precision is calculated. Therefore, it is important to include recall and the F_1 score.

Another feature that is visible is the reflection of the light on the glue of the trap. In image 3.4A, the reflection is minimal yet in image 3.4D and especially image 3.4E this reflection is very strong. This can influence the accuracy of the ODM as shown in figure 3.5.

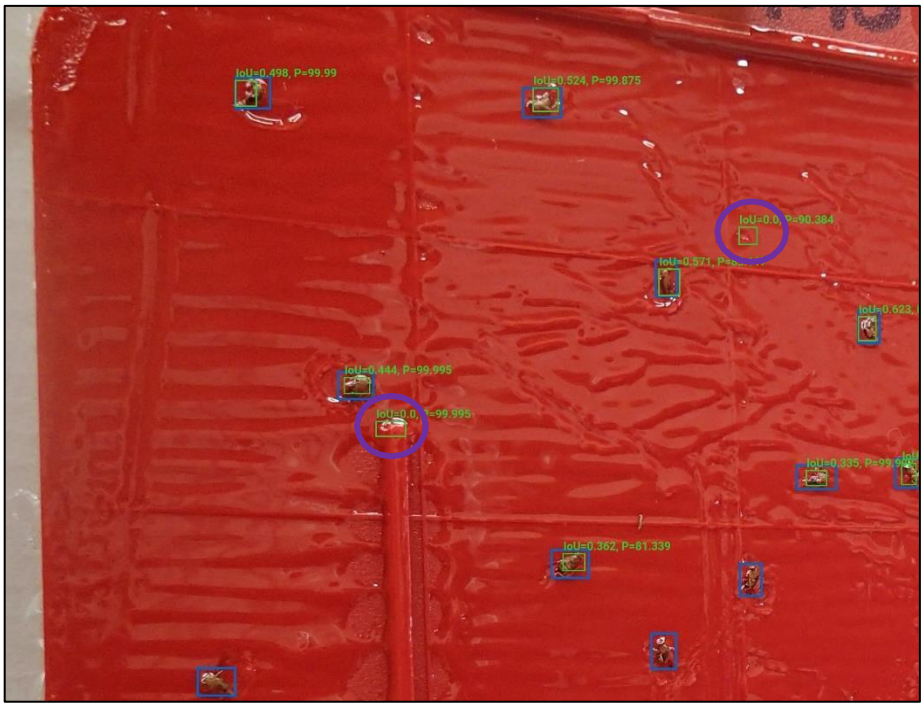
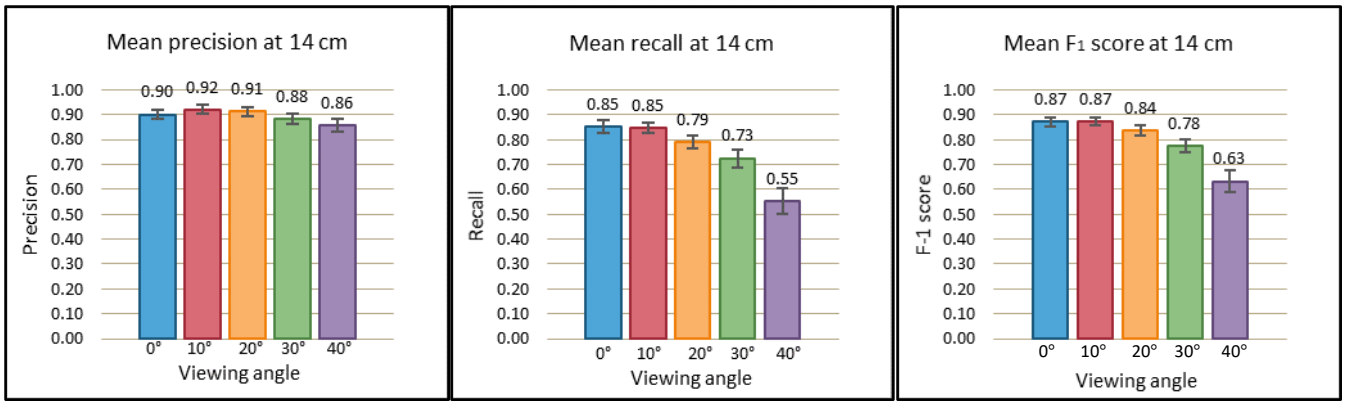


Figure 3.5 Reflection example. The image is taken at a distance of 14 cm and a view angle of 30°. The purple circles indicate the reflections that were detected as a fly by the ODM.

The detected reflection on the left is caused due to the glue and a plastic ridge on the trap itself. The detected reflection on the right is less noticeable but overall the image shows that the ODM can be mistaken due to various types and sizes of reflections.

The mean results of the dependent variables are divided on distance and are shown in figure 3.6, figure 3.7 and figure 3.8. The figures 3.6, 3.7 and 3.8 can also be found in Annex III. Annex III was made to have a complete view of the results of paragraph 3.3. The view angles are colour coded and are the same for all figures.



(A) **(B)** **(C)**

Figure 3.6 Mean precision, mean recall and mean F_1 score at 14 cm. figure 3.6A shows the results of precision, 3.6B shows the results of the recall and 3.6C shows the results of F_1 score. The first (blue) bar represents 0°, the second (red) bar represents 10°, the third (yellow) bar represents 20°, the fourth (green) bar represents 30° and the final (purple) represents 40°.

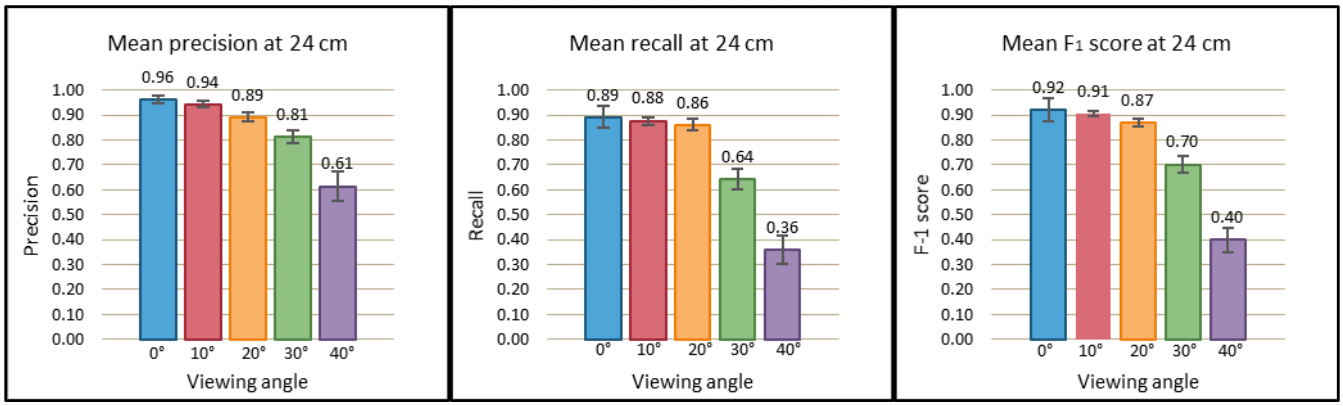
In figure 3.6A, the results of precision at a distance of 14 cm are shown. What is striking is that the difference between the view angles is only a maximum of 0.06, which is between 10° and 40°. Overall, the precision does not really change. Table 3.2 shows the results of the Mann & Whitney U test which indicates if there is a significant difference between the precision values of the different view angles.

Table 3.2 Mann & Whitney U test results for view angle comparison at a distance of 14 cm.

View angle comparison	Precision (p-value)	Recall (p-value)	F_1-score (p-value)
0° - 10°	0.366	0.846	0.796
0° - 20°	0.332	0.071	0.245
0° - 30°	0.366	0.002	0.007
0° - 40°	0.196	0.002	0.002

The comparison was made between the precision scores at 0° and the precision scores at either 10°, 20°, 30° or 40°. Table 3.2 shows that the difference between the precision values shown in figure 3.6A were not significant. This indicates that the increase in the view angle has either an insignificant effect or no effect on the precision in this experiment. Figure 3.6B shows the recall value of each view angle at a distance of 14 cm. There is a strong pattern between the size of the view angle and the recall. The largest difference is between 0° (or 10°) and 40° with a difference of 0.30. Table 3.2 also shows that the difference between 0° and 20° is insignificant, although it is close to being so. For 30° and 40° the recall value is significant and indicates that an increase in view angle does have a significant negative effect on the accuracy of the ODM. The F_1 score is shown in figure 3.6C. The largest difference in F_1 score is 0.24 between 0° (or 10°) and 40°.

Figure 3.6 indicates that the recall value is negatively affected by the view angle more than the precision is affected by the view angle.



(A) **(B)** **(C)**

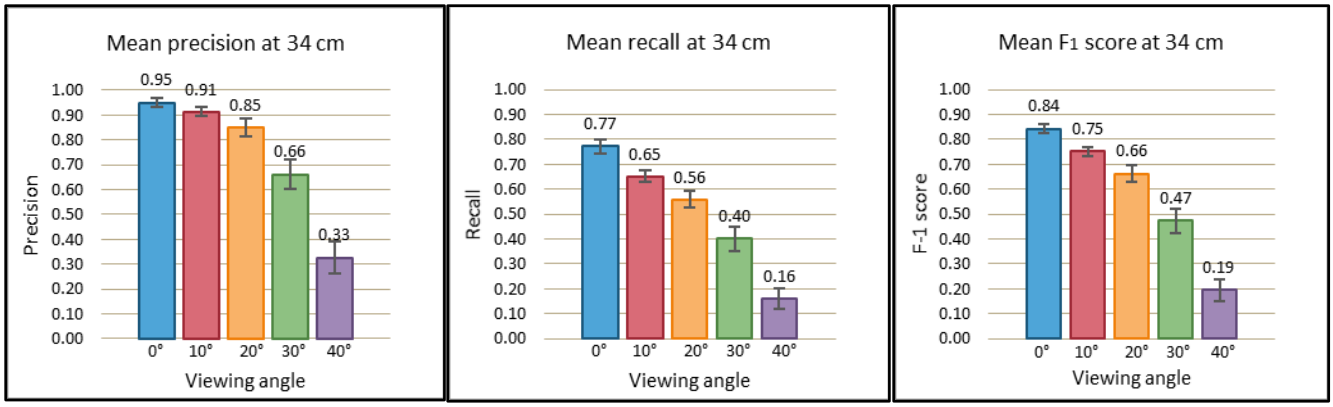
Figure 3.7 Mean precision, mean recall and mean F_1 score at 24 cm.

The results of the dependent variables at 24 cm are shown in figure 3.7. It is shown that the dependent variables are a lot more affected by the view angle at 24 cm then they are at 14 cm. Figure 3.6A shows the mean precision. The maximum mean score is at 0° with a value of 0.96. The minimum mean precision is at 40° with a score of 0.61. The effect of the view angle seems to have a larger effect on the precision at 24 cm compared to at 14 cm.

Table 3.3 Mann & Whitney U test results for view angle comparison at a distance of 24 cm.

View angle comparison	Precision (p-value)	Recall (p-value)	F_1-score (p-value)
0° - 10°	0.33	0.365	0.332
0° - 20°	0.02	0.155	0.061
0° - 30°	0.004	0.002	0.002
0° - 40°	0.003	0.002	0.002

In table 3.3, the p-values are shown for the comparison with 0° and the four view angles. The difference between 0° and 10° was insignificant. The difference between 0° and 20° on the other hand was significant with a p-value of 0.02. The difference increases and was significant for 30°, with a precision value of 0.81. It reaches the maximum difference at 40° with a difference of 0.35. For recall (figure 3.7B), the pattern is similar to that of precision but has a more pronounced drop in recall after 20°, with 30° and 40° having much lower values than at 14 cm. The difference between 0° and 30° and 0° and 40° are both highly significant, which implies that the ODM has difficulty detecting flies on the trap as an object of interest at wide view angles. The mean F_1 score shows the general pattern of precision and recall, with 10° and 20° being rather unaffected and a significant effect at 30° and 40° (figure 3.7C).



(A) **(B)** **(C)**

Figure 3.8 Mean precision, mean recall and mean F_1 score at 34 cm.

The results of the dependent variables at a distance of 34 cm are shown in figure 3.8 and the significance score is shown in table 3.4. For mean precision, there is a large difference between the maximum and minimum score, with 0° having a value of 0.95 and 40° having a value of 0.33 (figure 3.8A). Table 3.4 shows that the difference between 0° and 10° is insignificant. At 0° and 20° the difference does become significant. When comparing figure 3.8A and 3.7A we can see that there is a strong effect of distance on the mean precision, which is also visible in figure 3.8B and figure 3.7B for recall. The recall values are all a lot lower at a distance of 34 cm, also having the lowest mean recall score of the three distances (figure 3.8B). What is also noticeable is that, unlike with the precision, the recall of 0° does have a relatively low score (paragraph 3.2). In table 3.4 we can see that the differences between 0° and the other view angles are all significant. This suggests that the relationship between view angle and recall at a distance of 34 cm is more pronounced.

Table 3.4 Mann & Whitney U test results for view angle comparison at a distance of 34 cm

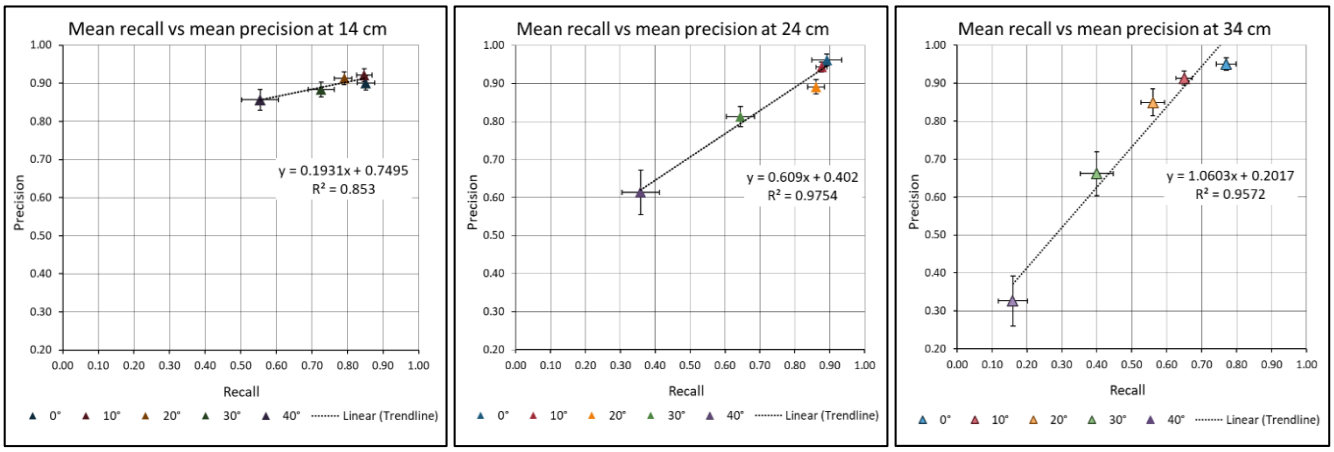
View angle comparison	Precision (p-value)	Recall (p-value)	F_1-score (p-value)
0° - 10°	0.093	0.028	0.014
0° - 20°	0.007	0.003	0.002
0° - 30°	0.004	0.003	0.002
0° - 40°	0.002	0.002	0.002

In figure 3.8C the F_1 score is shown. The pattern of reduced accuracy is clearly visible, with significant differences between 0° and the other view angles. Also, for 20°, 30° and 40° the steps between the F_1 scores become larger with the increased view angle. This pattern is also visible in figure 3.6C and 3.7C. However, the significant in difference between 0° and 20° is not large enough for both the F_1 score at 14 cm and 24 cm. The results shown in this chapter suggests that the increase in distance also enlarges the negative effect of the view angle on precision, recall and thus F_1 score. This is quite evident when comparing the figure 3.6 with figure 3.7 and figure 3.8, which shows that the minimum value of either precision, recall and F_1 score are a lot lower at a distance of 24 cm and 34 cm than they are at a distance of 14 cm. On the other hand, the highest mean values are all shown in figure 3.7 which is at a distance of

24 cm, which indicates that this is the optimal position for the ODM in this experimental set-up and in combination with the model of camera used.

3.4 Precision vs recall

Here, the mean precision and the mean recall are compared to see how these two variables respond to the change in view angle. These mean values are the same as used in paragraph 3.2 and paragraph 3.3, but in this comparison the focus is on comparing how the two respond variables respond to the different view angles. The results are shown in figure 3.9.



(A) **(B)** **(C)**

Figure 3.9 Mean precision vs mean recall. The recall is on the x-axis and the precision is on the y-axis. The trendline is based on the mean values of recall and precision.

Figure 3.9A, shows that the precision (on the y-axis) has a minimal response to the increase of view angle at a distance of 14 cm. The recall has a stronger response to the increase in view angle. This response is minimal for the view angles of 0°, 10° and 20°, and becomes more pronounced at 30°. At 40° the effect is a lot stronger, with a mean recall value of 0.55. For a distance of 14 cm, recall seems to be a lot more affected compared to precision. This is also visible in the equation of the trendline where the coefficient is 0.1913. This means that the precision will increase or decrease with 0.1931 for each value of 1.0 for recall. The R^2 value is 0.853, which indicates that the variation in the data is still quite large and that the equation is unable to describe ± 15% of the data. In graph 3.9B the results are shown for a distance of 24 cm. There is a strong trend visible, with a stronger effect of the view angle on the precision score. In the right upper corner, the intersect of precision and recall are shown for 0°, 10° and 20°. They are quite close to each other, with 0° and 10° overlapping. The equation shows that recall also responds stronger to the view angle than precision, with the constant of the x-value being 0.609. The R^2 of 0.9754 indicates that the equation explains the pattern very well. In graph 3.9C the results for a distance of 34 cm are shown. The relationship seen in figure 3.9A and 3.9B becomes a lot more pronounced, with recall and precision affected almost the same by the view angle. Unlike in figure 3.9A and 3.9B, the precision seems to be slightly more affected by the view angle then recall is. Overall, it seems that precision responds strong to the increase of the view angle when the distance increases.

Another pattern visible in the three graphs is that the standard deviation at each data point becomes larger (in general) with the increase of the view angle. This is most likely due to the increase in variation that occurs on the precision and recall score with the increase of view angle. The graph showing all the values for each dependent variable and distance can be found in annex II.

3.5 Linear regression

To determine how distance impacts the accuracy of the model compared to the view angle, a linear regression model has been made and tested with the dataset. In the previous results presented in this chapter, the mean value was used for all the calculations. For the linear regression the mean score of the three dependent variables was taken for each whole trap instead of a $1/4^{th}$ segment of the image (figure 3.11). This was done to minimize the amount of noise in the data. The noise in the data is partly due to the difference in distance between the left segments and right segments of the trap when it is at a view angle (Figure 3.10).

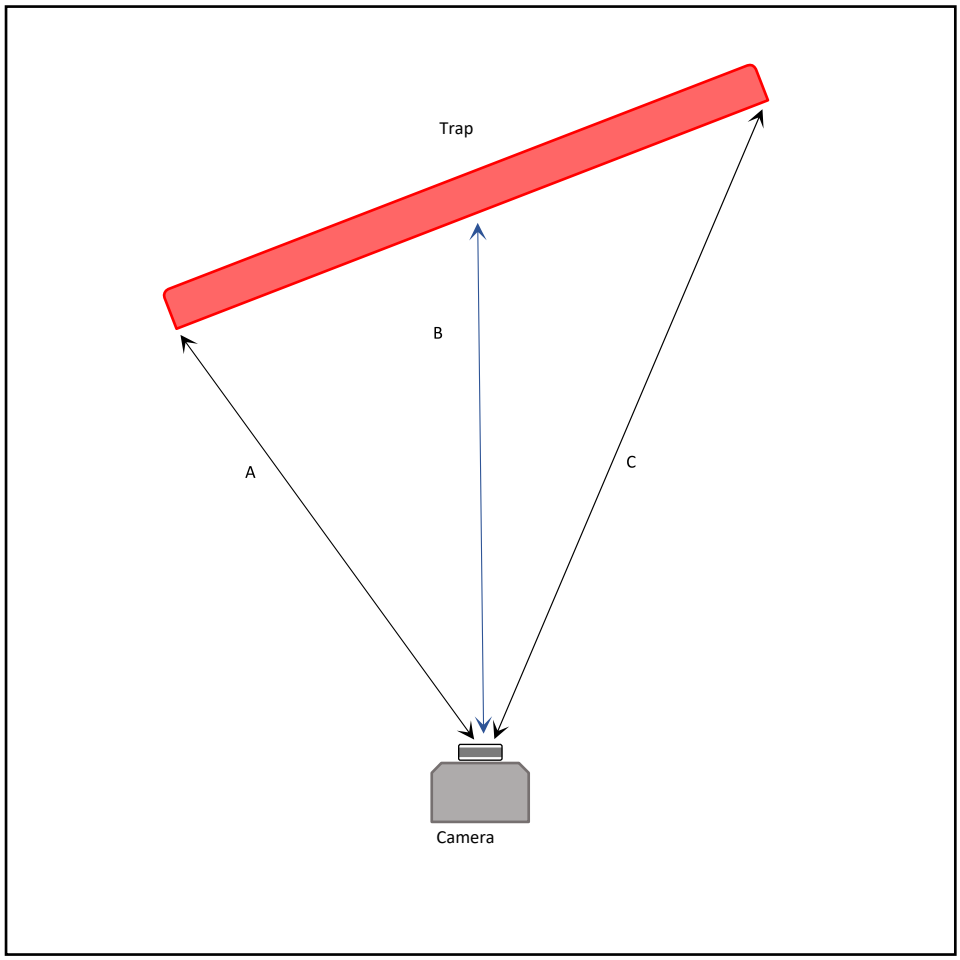


Figure 3.10 Corner distance example. Line B is the distance that is used in the dataset and line A and B are the actual distances of the two sides of the view angle. The larger the view angle, the more these sides differ in distance to the camera.

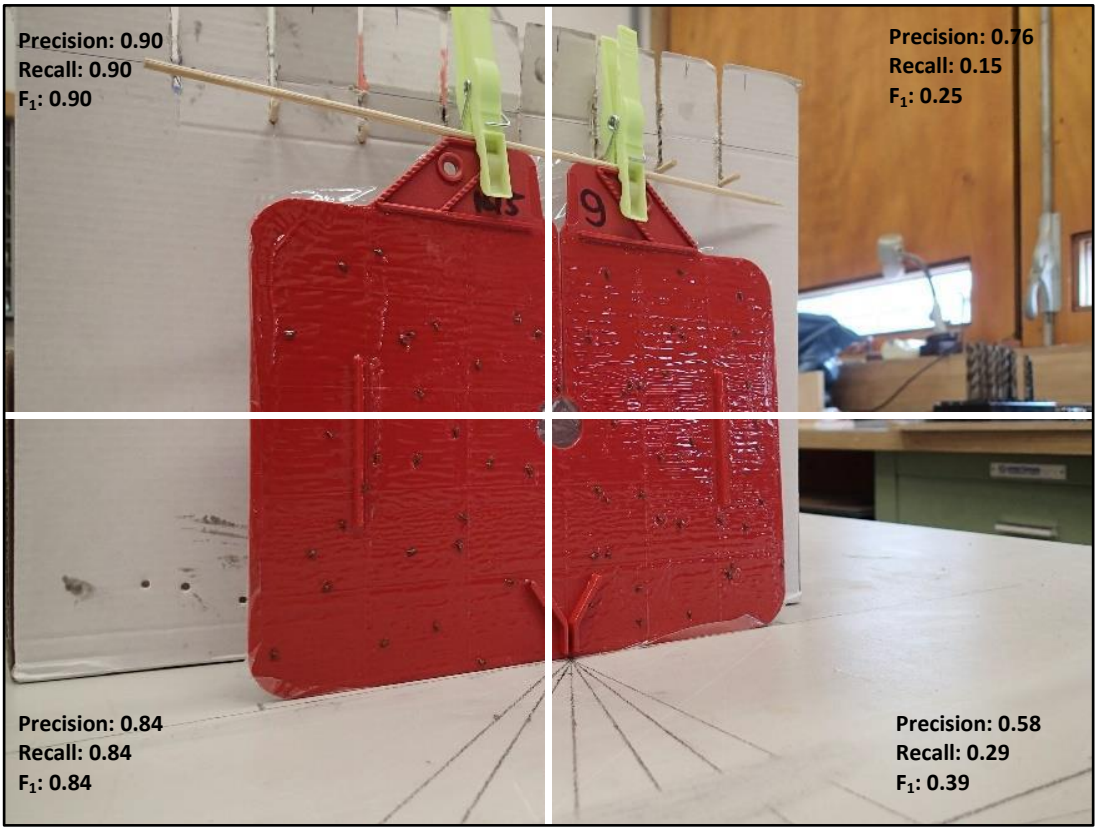


Figure 3.11 Linear regression image example. The right segments have a far lower score for each dependent variable compared the ones on the left. This is most likely due to the difference in distance. The average value of the entire image (all four segments together) is, precision: 0.77, recall: 0.55 and F_1: 0.60. The angle of this image is 30°.

In figure 3.10, the difference in distance between the three arrows is clearly visible. This effect was not of importance for the analysis in the previous part of this chapter because the actual scores of each segment were used. For example, in figure 3.11 each segment has a score for each dependent variable. For each image that was made at the same distance and same view angle, the values were averaged. These are the average values used in paragraph 3.1 to 3.4. Thus, it does not matter if the average scores of a whole image (four segments together) or each separate image is used as these will all be added to each other to create an average score for each dependent variable. Using the scores of each separate segment in the linear regression though, would result in a lot of noise in the data, negatively affecting the reliability and accuracy of the model. This noise is due to each segment of the same image being labelled as the same distance (14 cm, 24cm or 34 cm), while in reality this distance is different as shown in figure 3.10 and 3.11. To prevent this, the average score of a whole image is used instead of the four separate scores. Thus, for the image in figure 3.11, the scores shown in the description were used for the linear regression model.

Linear relationship in the data

To determine if the relationship between the dependent variables and the independent variables of distance and view angle are linear, a scatterplot is constructed for each dependent variable with distance

and with view angle (figure 3.12). In figure 3.12, the relationship is shown for the independent variables with distance. The first pattern that is clearly visible, is that the values of the y-axis deviate more the further the distance becomes. The R^2 value of each linear trendline is shown in the graph. Because of the large deviation in the data, it is difficult to produce a trendline that explains a large part of the pattern in the data. Therefore, the R^2 values of different types of trendlines are presented in table 3.5.

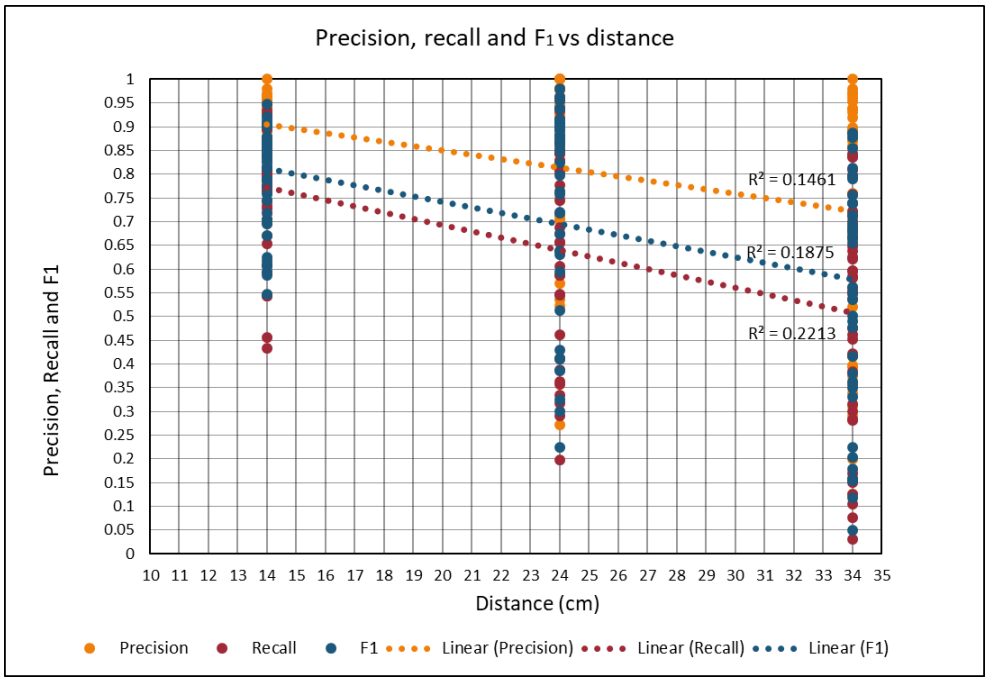


Figure 3.12 Scatterplot for linear regression with distance. The x-axis shows the distance in cm and the y-axis shows the score of precision, recall and F_1. The precision is shown in yellow, the recall in red and the F_1 score is shown in blue. The trendlines have the same colour as the corresponding independent variable.

When looking at the R^2 value of each type of trendline in table 3.5, we can see that the polynomial trendline has the highest value. The differences are not very large but it seems that all three dependent variables have somewhat of a polynomial relationship with distance instead of a straight linear one.

Table 3.5 R^2 values for the different types of trendlines in figure 3.12.

Trendlines	Precision (R^2)	Recall (R^2)	F_1 (R^2)
Linear	0.1461	0.1875	0.2213
Exponential	0.1367	0.1712	0.1974
Logarithmic	0.1380	0.1711	0.1979
Polynomial	0.1502	0.2044	0.2531
Power	0.1279	0.1564	0.1722

In figure 3.13, the data of each independent variable is shown plotted against view angle. Here, it seems that the relationship between each dependent variable and the view angle is stronger compared to distance.

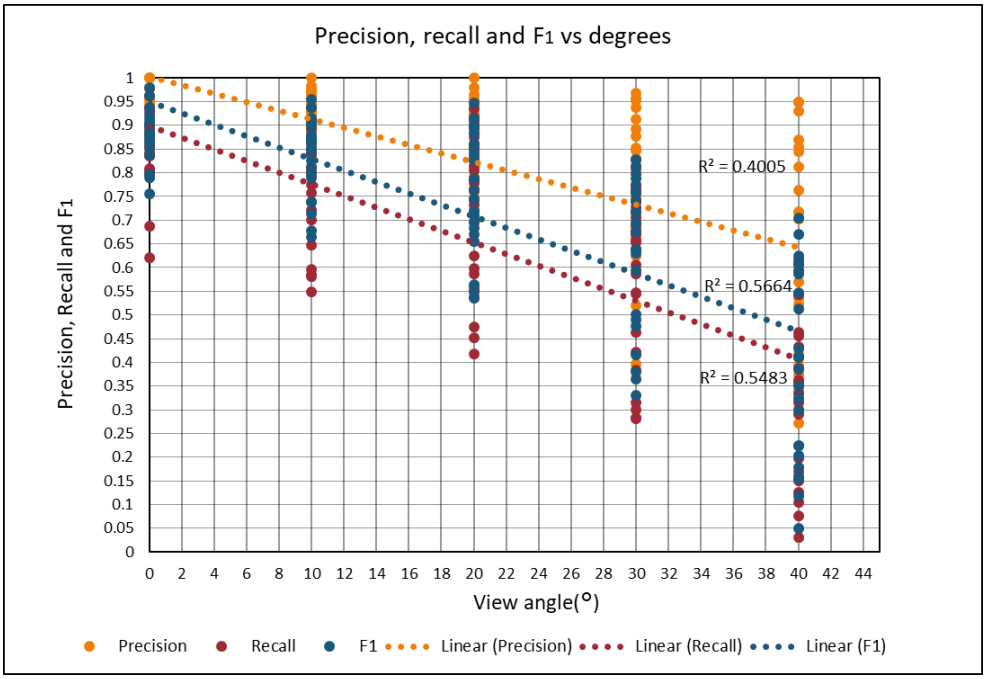


Figure 3.13 Scatterplot for linear regression with view angle. The x-axis shows the view angle and the y-axis shows the score of precision, recall and F_1. The precision is shown in yellow, the recall in red and the F_1 score is shown in blue. The trendlines have the same colour as the corresponding dependent variable.

The R^2 values in the graph are of a linear trendline and indicate that the view angle at which an image is taken, is a stronger predictor of the trend in the data when compared to distance. Furthermore, the R^2 values shown in table 3.6, show that the relationship in figure 3.13 is better explained with a polynomial trendline.

Table 3.6 R^2 values for the different types of trendlines in figure 3.13.

Trendlines	Precision (R^2)	Recall (R^2)	F_1 (R^2)
Linear	0.4005	0.5664	0.5483
Exponential	0.3247	0.4249	0.4111
Logarithmic	0.0978	0.1629	0.1750
Polynomial	0.4704	0.6337	0.5957
Power	0.0695	0.1012	0.1054

For Logarithmic and Power, the 0 values in the data were changes to 1e-8. This is done because both cannot be used with values of 0 in the dataset.

This polynomial relationship for both explanatory variables is also visible in figure 3.3. The curve is caused because the optimal distance is not at a position closest to the image but at a distance of 24 cm. In figures 3.6, 3.7 and 3.8, where the dependent variables are plotted against the view angle, this curvature is also visible. Therefore, for both the view angle and the distance a polynomial factor is added in the model of the linear regression.

Before the linear regression model is run, four assumptions need to be checked:

1. A linear relationship in the data
2. Independence of the dependent variables
3. Normality of errors
4. Constant error variance

The results of the four assumptions are somewhat outside the scope of this thesis and are therefore placed in annex I. Furthermore, in this chapter the most important results are shown for each dependent variable.

Precision

The results of the model are shown here in table 3.7. The results of the ANOVA test are shown in table 3.8. Table 3.7 exists of two parts. The first part are the results of each variable in the model and the second part are the R^2 and the p-value of the model overall. The first column in table 3.7 shows the variables used in the model.

Table 3.7 Linear regression model results for precision

Precision	**Estimate**	**Std. Error**	**t-value**	**Pr(>t)**
Intercept	0.6585	0.1063	6.194	1.09e-8
Distance	0.01839	0.00902	2.039	0.0439
View angle	0.02088	0.00292	7.150	1.08e-10
Distance^2	-0.000264	0.000185	-1.442	0.1523
View angle^2	-0.000314	5.264e-5	-5.961	3.21e-8
Distance:View angle	-0.000706	7.718e-5	-9.147	4.04e-15

Adjusted R^2	0.7774
p-value	<2.2e-16

The second column shows the coefficient of each variable (the relative effect of each variable), the third column shows the standard error of each coefficient, the fourth column gives the t-value which explains the size of the difference relative to the variation in the data. The larger the t-value is, the more the relative difference is. In the fifth column the p-value of the t-value is given. The adjusted R^2 shows how much of the patterns in the data can be explained using the model. The R^2 is adjusted for the number of variables in the model. The p-value shows how significant the model is.

For the model of precision, we see that all the variables except distance^2, preform a reasonable job at predicting the precision score. The distance^2 has a small t-value and a high p-value. This means that this variable is not significant (>0.05) for this model. Also, the standard error of distance^2 is very large with an error of 70%. Overall though, this model explains ± 77% (adjusted R^2) of the data with a p- value of <2.2e-16.

The ANOVA test results show the proportion of the variation (“mean sq”) that each variable predicts. For example, the view angle has a mean sum of square of 1.73, while the distance has a sum square of 0.64. This means that in this model the view angle has a larger impact on the prediction of the precision than distance has.

Table 3.8 ANOVA test results for precision

Precision	Mean sq	F value	Pr(>F)
Distance	0.63992	74.963	5.125e-14
View angle	1.73350	203.292	<2.2e-16
Distance^2	0.01772	2.078	3.205e-8
View angle^2	0.30298	35.532	3.205e-8
Distance:View angle	0.71346	83.669	4.039e-15
Residuals	0.00853		

According to the results, the view angle is the most important predictor followed by the interaction of distance and the view angle (distance:view angle) with a sum of squares value of 0.71. The distance^2 has the lowest mean sum of squares value, meaning that it has a small impact on the precision. The variable residuals show the proportion of the variation that is not explained by one of the model variables. The mean of squares of the residuals is very low with a value of 0.00853. The F-value shows the significance of each variable, and the Pr(>F) shows the p-value of the F-value.

Recall

Table 3.9 shows the results of the linear regression for recall. Firstly, unlike with precision, all of the model variables are significant. Furthermore, the R^2 is higher with a value of 0.8787 than for the regression of precision.

Table 3.9 Linear regression model results for recall

Recall	Estimate	Std. Error	t value	Pr(>t)
Intercept	0.5099	0.09199	5.543	2.12e-7
Distance	0.03667	0.007802	4.700	7.73e-6
View angle	0.009278	0.002526	3.673	0.000375
Distance^2	-0.000872	0.0001587	-5.495	2.62e-7
View angle ^{2}	-0.000304	4.554e-5	-6.671	1.12e-9
Distance: View angle	-0.000378	6.678e-5	-5.666	1.23e-7

Adjusted R^2	0.8787
p-value	<2.2e-16

In table 3.10, the ANOVA test results are shown for recall. Both the distance and the view angle explain a large proportion of the variation with the view angle explaining the largest part of the recall score.

Table 3.10 ANOVA test results for recall

Recall	Mean sq	F value	Pr(>F)
Distance	1.3184	206.525	<2.2e-16
View angle	3.2552	509.918	<2.2e-16
Distance^2	0.1928	30.200	2.623e-7
View angle ^{2}	0.2841	44.502	1.124e-9
Distance: View angle	0.2049	32.100	1.227e-7
Residuals	0.0064		

The polynomial variables (distance^2 and view angle^2) have a relatively small explanatory part in the model, but are highly significant. The interaction of distance and view angle have a similar limited impact on the recall score but are also very significant.

F_1

Table 3.11 shows the results of the linear regression for the F_1 score. Here it seems that all the model variables are significant in explaining the pattern predicted by the model. Also, the standard error of all the model variables are very low. Furthermore, the R^2 value is very high with a value of 0.8991, meaning that the model explains almost 90% of the data.

Table 3.11 Linear regression model results for F_1

Precision	Estimate	Std. Error	t value	Pr(>t)
Intercept	0.5844	0.08121	7.197	8.59-11
Distance	0.02819	0.006888	4.092	8.26-5
View angle	0.01445	0.002230	6.477	2.84-9
Distance^2	-0.000611	0.0001401	-4.357	3.02e-5
View angle ^{2}	-0.000349	4.021e-5	-8.681	4.56e-14
Distance: View angle	-0.000507	5.895e-5	-8.599	6.97e-14
Adjusted R^2	0.8991			
p-value	<2.2e-16			

Moving further to the ANOVA test results shown in table 3.12, we can see that the view angle explains the largest part of the model's prediction, almost three times more than distance.

Table 3.12 ANOVA test results for F_1

Precision	Mean sq	F value	Pr(>F)
Distance	1.05634	212.324	<2.2e-16
View angle	3.13925	630.990	<2.2e-16
Distance^2	0.09443	18.980	3.020e-5
View angle ^{2}	0.37491	75.358	4.555e-14
Distance: View angle	0.36786	73.939	6.966e-14
Residuals	0.00498		

When we compare the mean sum of squares of the distance and distance^2 with that of view angle and view angle^2" for each model, we see that proportion that is explained by the view angle is always larger. This suggests that the view angle at which an image is made has a larger effect on the accuracy of the ODM than that the distance has. It is important to emphasize that the regression for these models was done on data that has a minimum and maximum distance of 14 and 34 cm respectively and a minimum and maximum view angle of 0° and 40° respectively.

Chapter 4 Discussion

4.1 Introduction

In this chapter all the results will be discussed and presented in the same order as in the result chapter. Furthermore, the results will be compared to other similar research that was found in the literature. Lastly, a reflection on the experimental set-up is given.

4.2 The effect of distance on the accuracy of the object detection model

For precision the difference between a distance of 24 cm (0.96) and 34 cm (0.95) is not significant, however, the difference between 14 cm (0.90) and 24 cm is significant (figure 3.3A). It is unexpected that the precision score is lower at a distance of 14 cm, as it would be expected that a camera positioned closer to the target would result in a higher score than when it is placed further than the optimal distance of 24 cm from the target. The reason for this could be due to the size of the flies. The size of the flies is relatively larger in images made at a distance of 14 cm compared to images made at a distance of 34 cm. Due to the small size at a distance of 34 cm, the ODM has less information (amount of pixels) to work with, which decreases the chance of detection and classification. Furthermore, as stated in the introduction segment (3.1) of chapter 3, precision only provides information about objects that were actually detected by the ODM, ignoring the undetected objects. It could be that at a distance of 14 cm the ODM detects more objects and labels these as a *Drosophilia suzukii* fly compared to images made at a distance of 34 cm, but of these multiple objects detected, some could be misidentified. Meanwhile, at a distance of 34 cm there could be only a few of the flies actually identified but with a higher probability of actually being *Drosophilia suzukii* flies. This theory is further supported by the observations made with the independent variable recall.

Recall is calculated by looking at the fraction of objects correctly identified compared to the total amount of objects that are actually in an image. In figure 3.3B, the mean recall score is shown for the three distances. The difference in mean recall for 14 cm and 24 cm with a mean recall score of 0.85 and 0.89 respectively is statically not significant ($p = 0.330$). But the difference between 24 cm and 34 cm (0.77) is highly significant ($p = 0.002$). This indicates that at a distance of 34 cm the number of flies that are actually detected is significantly less compared to images made at 14 cm and 24 cm. Although the precision indicates that of these flies that are actually detected, a high number are correctly identified. Also, when comparing the mean recall results between the distance of 14 cm and that of 34 cm we can see that there is a significant difference between the two ($p = 0.045$). Thus, for recall, the results suggest that a position closer to the trap will result in a higher score than if the camera is placed further from the trap than the optimal distance.

Moving further to the mean F_1 score (figure 3.3C), the difference between 14 cm (0.87) and 24 cm (0.92) is almost significant ($p = 0.053$). It does indicate that the optimal position of 24 cm is most likely better compared to the distance of 14 cm. The difference between 24 cm and 34 cm (0.84) is significant ($p = 0.003$). This indicates that the optimal position of 24 cm is also better compared to the distance of 34 cm. When comparing the mean F_1 score of 14 cm with the mean F_1 score of 34 cm, it shows that there is no

significant difference between the two distances (p = 0.302). This is however the case when comparing both distances for the precision with a p-value of 0.026 and the recall score with a p-value of 0.045.

4.3 The effect of the view angle on the accuracy of the object detection model

The relation between the view angle at which the image is made and the dependent variables are shown in paragraph 3.3. It is important to remember how the view angle affects the actual image. This is shown in figure 3.4. Not only does the relative size of the trap change in the image with the change in view angle, the distance between the camera and different parts of the trap also changes, as shown in figure 3.10. Also, the view angle changes how the reflection of light on the trap acts, which will be further discussed in this chapter. However, the results of the mean precision, mean recall and mean F_1 score will be discussed first.

In figure 3.6 the results are shown for 14 cm. The mean precision scores are shown in figure 3.6A and the Mann & Whitney U test results in table 3.2. The hypothesis (paragraph 1.2) is that a view angle of 0° is the most optimal and that the larger the view angle, the lower the precision, recall and F_1 scores will become. The mean precision scores are all similar to each other with no significant difference. One reason for this could be that the effect of the view angle is very minimal on the precision at such a close distance. In figure 2.1 in chapter 2, the experimental set-up is shown. The figure shows that the relative difference in position is much larger at a larger distance (e.g. 24 cm or 34 cm) than compared to a closer distance (14 cm). This is visible when comparing the mean precision at 14 cm with that of the mean precision at 24 cm (figure 3.7A) and 34 cm (figure 3.8A) where the precision significantly drops with the increase of the view angle. The mean precision score at 24 cm is significantly different between 0° and 20° (p = 0.02), 30 (p = 0.004) and 40° (p = 0.003), but not with 10° (p = 0.33). The mean precision score at 34 cm is similar for 0°, 10° and 20° compared to the mean precision score at 24 cm, which suggests that the precision is similarly affected by these view angles at both the distance of 24 cm and 34 cm. There is however a significant difference between both distances for the view angles of 30° and 40°. This suggests that for precision, a deviation in the view angle larger than 20° has a very strong negative effect the further the camera is placed away from the optimal position. The other way around though, it seems that in case of a large deviation in view angle, the camera should be placed closer to the target than the hypothesized optimal distance of 24 cm, as the mean precision scores at 14 cm are significantly similar to each other. A reason for this could be that the size of an object on the image becomes smaller the larger the distance and the view angle become, making it more difficult for the ODM to detect the objects. In case of the distance of 14 cm, this effect is so small that it goes unnoticed in the mean precision values. For the distance of both 24 cm and 34 cm however, this effect is noticeable as the larger the distance becomes, the more pronounced the effect of the view angle also becomes. This is also shortly discussed in the beginning of paragraph 3.3 and illustrated in figure 3.4. In summary, the negative effect of the view angle on the precision seems to be strongest when the distance increases and is minimal if the camera is placed close to the target. For recall, the negative effect of the view angle at a distance of 14 cm (figure 3.7B) is stronger than for the precision. Unlike with precision, the distance of 14 cm has significant differences between 0° and 30° (0.002) and 40° (0.002). At a distance of 24 cm (figure 3.7B) the mean recall scores at the view angles of 0° and 10° are similar to that of the mean recall scores at 14 cm. The mean recall scores at the view angles of 20°, 30° and 40° on the other hand are much lower at a distance of 24 cm compared to 14 cm. This could be again due to the

enlarging effect that the view angle has on the distance. The mean recall scores at a distance of 34 cm (figure 3.8B) are the lowest of the three distances, with a pattern of decreasing mean recall score with each increasing view angle. All the mean recall values are significantly different from the mean recall value at 0°, showing that the larger the view angle, the lower the recall becomes. Thus, it seems here that the ability of the ODM to correctly identify the number of objects in an image is severely hampered if the image is made at the optimal distance in combination with a view angle larger than 20°. If the camera is positioned at a further distance than the optimal distance, the negative effect of the view angle becomes larger. If the camera is positioned closer to the target than the optimal position, the effect of the view angle becomes a lot less pronounced compared to the optimal distance and the larger then optimal distance.

4.4 Precision vs recall discussion

It is interesting to visualize how precision and recall respond to the change in view angle at each distance (figure 3.9). Depending on what the goal is of an ODM, one could prefer to focus on increasing the precision over recall or vice versa.

The mean values presented here (figure 3.9) show a somewhat misleading representation as they suggest that there is a very clear distribution, in reality however, there are many different values spread over the different view angles (annex II). For example, there are images made at a view angle of 40° where the precision and/or recall value are between 0.9 and 1.0. There are also images made at a view angle of 0° that have a very low precision and/or recall value. Nonetheless, the comparison of the mean values of precision and recall show an obvious pattern.

In figure 3.9A the mean precision value for each view angle is very similar, the mean recall value on the other hand differs a lot for each view angle. In figure 3.9B and 3.9C the mean precision value starts to deviate further, with a lower mean score for each increasing view angle. There is almost no difference in mean precision and mean recall at a view angle of 0° at the three different distances. But when the view angle becomes larger than 0° the mean scores start to drop, with precision being the most profound. Thus, if the view angle at which the image is made can be correctly controlled, the distance would have only a minor negative effect on the precision and the recall. If the orientation of the camera is difficult to control and the precision is preferred, a position close to the target would be preferred as it would likely result in the highest mean precision scores due to the buffering effect the close distance has on the negative effect of the view angle. If recall would be the focus and the orientation of the camera is difficult to control, a close position to the target would also be preferred. However, the buffering effect of distance is less compared to the precision. In summary, it seems that the distance has minor effect (to a certain limit) on the precision and the recall score as long as the view angle is as close as possible to 0°. If the view angle becomes large, the distance should be reduced to reduce the negative effect of the view angle.

As mentioned in paragraph 3.1, the precision is dependent on the correct identification of the total amount of objects detected by the ODM. The recall on the other hand is dependent on the correct identification of the total amount of relevant objects in the image. There are multiple cases where the recall is reduced due to the ODM detecting dirt on the trap as a fly, but also random objects in the background or the wooden sticks holding the trap vertically. This occurs in images made at all different view angles and distances, although with a pattern of more errors at larger view angles and distances. This could be part of

the reason why the pattern of reduced mean recall is visible when the distance and view angle increase. For precision the same pattern is visible. This is likely partly caused due to the relative size of the fly decreasing with the increase of view angle and distance, making it more difficult for the ODM to detect them.

4.5 Linear regression discussion

The linear regression shows that for precision, recall and F_1 score, the view angle is responsible for most of the change. Distance also has a significant effect on the dependent variables. Overall it seems that the two variables do a decent job at explaining the patterns in the data.

The ability of the model to correctly classify the detected objects in an image (precision) is clearly also dependent on other variables that were not included in the regression model. The R^2 value for the model of precision is the lowest of the three with a value of 0.77. The ability of the model to correctly identify and count the total amount of objects in an image (recall) is more dependent on the used model variables, and has a high R^2 value of 0.899.

The linear regression clearly shows that the view angle at which an image is made has a much stronger impact on the outcome compared to the distance. This is most likely due to a combination of the relative fly size discussed in paragraph 3.2 and paragraph 3.3, but also partly due to the increased light reflection of the glue on the trap. Also, the increased distance of one part of the trap (figure 3.10) would likely increase the difficulty for the ODM to correctly detect *Drosophila suzukii* flies. Furthermore, the view angle would likely cause problems for the camera to correctly focus the lens. This would result in parts of the images being unfocused and therefore reducing the likelihood of the ODM to correctly detect *Drosophila suzukii* flies. For distance, the problem with the relative size of the flies would be similar but the reflection of light, difference in distance of the parts of the trap and the focus of the camera would be limited. This would explain why the view angle has significantly more impact on the accuracy of the ODM than distance does.

4.6 Literature and points of consideration

Varying kinds of research has been done in the field of agriculture, ecology and pest management in combination with object detection. Many papers discussing object detection using DL methods in combination with monitoring activities are done on large animals from which images have been obtained by camera traps (Norouzzadeh, et al., 2018; Schneider, Taylor, & Kremer, 2018) and there are even approaches that identify specific individual animals that have been captured on camera before (Schneider, Taylor, Linquist, & Kremer, 2018). The use of UAV's for animal detection has also been researched, for example in Gray, et al. (2019), UAV images were used to detect sea turtles and in Rey, Volpi, Joost & Tuia (2017) animals in the African savanna were detected with the use of UAVs. However, these animals were all relatively large making it a lot easier for an ODM to correctly identify them. For camera traps this is even more evident, as the effective detection distance at which a camera trap can detect an animal is usually dependent on the body mass of the animal, with larger animals being detected at larger distances (Hofmeester, Rowcliffe, & Jansen, 2017). The size of the animals in the images are often much larger than is the case for animal pests in agriculture. In pest management in agriculture there has already been

extensive use of ODMs. For example, plant diseases have been detected with the use of DL and object detection (Wu, Zhang, & Meng, 2019; Selvaraj, et al., 2019; Xing, Lee & Lee, 2019; Zhao, He, & Xu, 2012). But many of the images used in these researches were obtained by hand, and often very close to the object of interest. In Babu & Rao (2007), leaves were scanned using a computer scanner to obtain images. Furthermore, many plant diseases that have been detected with the use of ODM's have been diseases that show clear symptoms, such as discoloration, spots and leaf damage. Difficulties arise with the identification of plant diseases when these symptoms become less visible (Kamilaris & Prenafeta-Boldu, 2018). For weed detection, there have been multiple studies done where UAV's have been used to obtain the images (Hung, Xu, & Sukkarieh, 2014; Bah, Dericquebourg & Hafiane, 2018; Bah, Hafiane, & Canals, 2018; Gao, et al., 2018). But also, for weed detection many images contain large patches of the weeds or close up imagery is used. Unlike with diseases or weeds, invertebrate pests are more difficult to detect. This is due to the small size of the animals such as caterpillars, aphids or beetles. In many publications that have used DL ODMs, the small size of the animal is compensated with close-up images. For example, in research by Ding & Taylor (2016), a neural network-based moth detection pipeline was constructed. The images used were all close-up images where there was often little background. A similar strategy of manually obtaining images was used by Liu, Gao, Yang, Zhang, & He (2016) and Asefpour Vakilian & Massah (2013). Unlike in the previously mentioned papers, in Shen, Zhou, Li, Jian, & Jayas (2018) images were made of insects that are pests for grain farmers and in these images multiple insects were detected per image. An interesting side note is that the neural network used in this paper was based on Faster R- CNN, the same that is used in this thesis. But in Shen, Zhou, Li, Jian, & Jayas (2018) the insects were placed in a small container and the images were made manually. A paper that has done something similar as was done in this thesis is that of Nguyen & Hung (2018). In this research traps with glue were used to capture six pest types. The images were then used to train and validate an ODM. Another similar research is that of Sun, et al. (2018), where pheromone traps were used to capture and monitor the red turpentine beetle (*Dendroctonus valens).* The detection was done using a DL object detection. In this research, the camera was facing downwards and the trap itself was illuminated with LED lighting. Thus, each image was objected to the same controlled environment. In this thesis, the experimental set-up was in such a way that images were also manually made, but a key difference is that the position and the view angle of the camera was not based on optimal position (except for one position) but on a semi-random selection of the positions and orientations. In summary, many researches in literature have used DL and object detection to detect organisms in either their natural habitat or in laboratory environments. But in many, the size of the animal was either not an issue, pre-filtered or addressed in another way. In this thesis, the small size of the flies was often much smaller than the invertebrate animals mentioned. Furthermore, no research in literature was found where the effect of the view angle and/or distance at which the image was made was researched for the automatic detection using DL object detection methods for agricultural pests or similar objects. Also, the method to determine if a fly was detected or not (IoU) is not very suitable for small targets, as the common threshold of 0.5 is often too high. A possible method instead of IoU for determining a correct identification could be to see if the model generated bounding box intersects the centre point of the manually drawn bounding box.

Although the size of the *Drosophila suzukii* flies is a problem for having a high object detection accuracy, there are other issues that are worth mentioning. The first is that the camera used in this thesis was not a

UAV camera. The goal was first to use UAV obtained images to train and test the ODM. This was however not possible because of the low quality of the images which caused the flies to be unclear and blurry. It would have been more appropriate for this research to be able to use a camera attached to a UAV and having the UAV obtain images while flying to simulate a real world scenario more realistically.

This was however not possible as a license was needed to fly the UAV. Another important issue is that the traps were prepared with only *Drosophila suzukii* flies. In a real-world scenario there would have been multiple species of arthropods on the trap. Due to this, a complete analysis of the accuracy was impossible to achieve. It is unknown if the view angle and the distance would have influenced the accuracy the same if there were multiple types of arthropods. Furthermore, the lack of multiple classes would have likely also affected how the accuracy of the ODM would respond to the view angle and the distance. A dataset was provided at the beginning of this thesis with images made of the trap in a natural environment containing many different species of arthropods. However, many of the animals on the traps were mislabelled or not labelled at all (including *Drosophila suzukii* flies). The choice was made to construct an own image dataset due to this, because there was an assumption that the already existing dataset would have to many errors, negatively affecting the training phase of the ODM. This resulted in images containing only *Drosophila suzukii* flies. Therefore, it is unknown if the ODM is actually detecting *Drosophila suzukii* flies based on characteristics unique to that species or that it is detecting brown/black spots on a red background. Possible recommendations for improvements in the experimental design will be discussed in the conclusion; chapter 5.

Chapter 5 Conclusion and recommendations

5.1 Introduction

In this chapter a summary for each segment of the discussion chapter will be given together with the answers to the research questions. Furthermore, a summary of the recommendations will be given for potential improvements of the used experimental set-up and other experiments that could perhaps follow this one.

5.2 Discussion summary

The effect of distance on the accuracy of the object detection model

Precision and recall respond differently to the change in distance. When the camera is placed close to the target (< 24 cm) recall will drop and precision will be less affected. If the camera is placed further away from the trap (> 24 cm) precision will drop and recall is somewhat less affected. In general, the best distance would be the optimal position. In this research this distance is 24 cm, but this is dependent on the camera used and the size of the target.

The effect of the view angle on the accuracy of the object detection model

The negative effect that distance has on the accuracy of the ODM is enlarged when the view angle increases. This effect becomes stronger at view angles larger than 20°. The recall is more negatively affected by this than the precision, although the effect of the view angle is evident for both. The negative effect of the view angle can be partly compensated by decreasing the distance at which the image is made.

Precision vs Recall

At a distance of 14 cm, recall has a larger drop with each increasing view angle than precision. When the distance also increases, precision becomes relatively more affected. If a high precision is wanted, a distance close to the target will compensate the effect of the view angle. The same is true for recall but less so than for precision.

Linear regression

The negative effect of the view angle on the accuracy of the ODM is ±3 times stronger than the negative effect of the distance. Thus, a correct pose would be more important than a correct distance for the accuracy.

5.3 Answering the research questions

Here a short summary is given of the answers to the research questions.

1. Can *Drosophila suzukii* be detected in images using a CNN based object detection model?

   Drosophila suzukii can be detected using a CNN based ODM. But due to the design of the experiment, where only one class was used and no other possible objects were placed on the trap, it is unclear if the ODM was able to distinguish between multiple classes. However, in literature there were multiple experiments where a similar CNN based ODM was used to detect multiple classes of arthropods.

1.1 Which object detection model will be suited for detecting *Drosophila suzukii*?

Out of the four ODM's (Faster R-CNN, R-FCN, SSD and YOLO) Faster R-CNN was chosen on the basis of literature research. Faster R-CNN has a high accuracy for small objects which YOLO was reported to have problems with. Furthermore, Faster R-CNN has a large amount of support information available and is relatively easy to use compared to R-FCN and SSD. Faster R-CNN and R-FCN are also two-stage ODM's, which makes the use simpler as it is not necessary to use a separate RPN.

2. How does the positioning of the camera affect the accuracy of the object detection model?

If the camera is placed at a different position than the optimal position, the accuracy of the ODM will be negatively affected. The strength of this effect is however different for each position.

2.1 How does the distance between the camera and the trap affect the accuracy of the object detection model?

Diverting from the optimal distance has a negative effect on the accuracy of the ODM. If the camera is placed closer to the target this negative effect will be smaller than when the camera is placed further away from the optimal position.

2.2 How does the view angle between the centre of the camera and the trap affect the accuracy of the object detection model?

The view angle has a negative effect on the accuracy of the ODM, with a stronger negative effect when the view angle increases. The negative effect of the view angle is also almost three times stronger than the negative effect of the distance (in the experimental set-up used). Thus, in a real-world scenario, the focus should be on using a UAV that can correctly position itself in front of the trap/target and orientate correctly. The distance can vary somewhat without having significant effect. However, the view angle should always be as low as possible, as this has a much stronger negative effect on the accuracy of the ODM.

5.4 Recommendations

Most of the suggestions for improvements have been discussed in detail in paragraph 4.6, but here a summary will be made. Also, other possible experiments will be mentioned here.

1. *Use of a UAV camera that can be used at close distances.*
   The problem with the camera of the UAV (DJI Mavic Pro 2) that was going to be used for this thesis was that it was unable to focus on the target at a close distance. While on the other hand the distance at which the camera was able to have a clear focus, the *Drosophila suzukii* flies were too small to make out. It would be interesting to obtain images of the target with a hovering UAV to simulate real world situations. Issues concerning the use of a hovering drone could then be researched.

2. *Multiple classes for the ODM.*
   Having the ODM trained on multiple classes of animals would simulate real world scenarios more accurately than in this thesis. Not only the detection of these classes would be interesting, but also how the ODM handles this in terms of accuracy.

3. *Different method than IoU to determine if a fly is correctly detected or not.*
   In this thesis the IoU threshold was 0.25, which was decided due to the small size of both the manually labelled bounding boxes and the model generated bounding boxes. This however could result in edge cases in which it is hard to determine if a target is correctly identified or not. A better method of calculating the IoU would prevent these edge cases.

4. *Larger image dataset to train the ODM*
   A larger image dataset with more diverse images (different background and classes) to train the ODM would most likely result in a higher detection accuracy.

5. *For testing the ODM, images should be obtained in the field to simulate similar environmental factors such as illumination and shadows.*

6. *Use of multiple types of cameras and/or filters too determine if the negative effect of the view angle and/or distance can be neutralized.*

7. *Comparing different ODM's to determine if the negative effect of the view angle and/or distance can be addressed with the type of model.*
   Also, it would be interesting to quantify the accuracy of different types of ODM's (such as the four ODM's considered in this thesis) with this data set.

8. *Determine the size of each object in the image (Drosophila suzukii fly or other objects) using the size of the label boxes.*
   This information in combination with the view angle and distance at which the image can be used to determine what kind of correlations exists between the size of the target and the accuracy of the ODM.

Overall, the focus of most of the recommendations is on trying to simulate real world scenarios more and see how this effect the accuracy of the ODM. However, this would most likely cost a significantly larger amount of manual labour for the labelling activities, which will have to be done very diligently as part of the images will be used for the training of the ODM.

References

Asefpour Vakilian, K., & Massah, J. (2013). Performance evaluation of a machine vision system for insect pests identification of field crops using artificial neural networks. *Archives of phytopathology and plant protection*, *46*(11), 1262-1269.

Babu, M. P., & Rao, B. S. (2007). Leaves recognition using back propagation neural network-advice for pest & disease control on crops. *IndiaKisan. Net: Expert Advisory System*.

Bah, M. D., Hafiane, A., & Canals, R. (2018). Deep learning with unsupervised data labelling for weed detection in line crops in UAV images. *Remote sensing*, *10*(11), 1690.

Bah, M. D., Dericquebourg, E., Hafiane, A., & Canals, R. (2018). Deep learning-based classification system for identifying weeds using high-resolution UAV imagery. In *Science and Information Conference* (pp. 176-187). Springer, Cham.

Benassi, F., Dall'Asta, E., Diotri, F., Forlani, G., Morra di Cella, U., Roncella, R., & Santise, M. (2017). Testing accuracy and repeatability of UAV blocks oriented with GNSS-supported aerial triangulation. *Remote Sensing*, *9*(2), 172.

Bennet, G. W. (2010). *Truman's Scientific Guide to Pest Management Operations 7th Edition* (7 ed.). Questex Media Group LLC & Purd.

Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica: Journal of the Econometric Society*, 1287-1294.

Chen, J., Fan, Y., Wang, T., Zhang, C., Qiu, Z., & He, Y. (2018). Automatic segmentation and counting of aphid nymphs on leaves using convolutional neural networks. *Agronomy*, *8*(8), 129.

Chen, X., Xiang, S., Liu, C. L., & Pan, C. H. (2014). Vehicle detection in satellite images by hybrid deep convolutional neural networks. *IEEE Geoscience and remote sensing letters*, *11*(10), 1797-1801.

Cheng, X., Zhang, Y., Chen, Y., Wu, Y., & Yue, Y. (2017). Pest identification via deep residual learning in complex background. *Computers and Electronics in Agriculture* (141), 351-356.

Colomina, I., & Molina, P. (2014). Unmanned aerial systems for photogrammetry and remote sensing: A review. *ISPRS Journal of photogrammetry and remote sensing*, *92*, 79-97.

Dai, J., Li, Y., He, K., & Sun, J. (2016). R-FCN: Object detection via region-based fully convolutional networks. In *Advances in neural information processing systems* (pp. 379-387).

Ding, W., & Taylor, G. (2016). Automatic moth detection from trap images for pest management. *Computers and Electronics in Agriculture*, *123*, 17-28.

Cooper, J., & Dobson, H. (2007). The benefits of pesticides to mankind and the environment. *Crop Protection*, *26*(9), 1337-1348.

EdjeElectronic. (2018). *Tensorflow Object Detection API tutorial train multiple objects windows 10*. Retrieved from github.com: https://github.com/EdjeElectronics/TensorFlow- Object-Detection-API-Tutorial-Train-Multiple-Objects-Windows-10

Ehler, L. E. (2006). Integrated pest management (IPM): definition, historical development and implementation, and the other IPM. *Pest management science*, *62*(9), 787-789.

He, F., Zhou, T., Xiong, W., Hasheminnasab, S. M., & Habib, A. (2018). Automated aerial triangulation for UAV-based mapping. *Remote Sensing*, *10*(12), 1952.

Fantke, P., Friedrich, R., & Jolliet, O. (2012). Health impact and damage cost assessment of pesticides in Europe. *Environment International, 49*, 9-17.

FAO. (2019). *AGP - Integrated Pest Management.* Retrieved February 17, 2019, from http://www.fao.org: http://www.fao.org/agriculture/crops/core-themes/theme/pests/ipm/en/

Faria, F. A., Perre, P., Zucchi, R. A., Jorge, L. R., Lewinsohn, T. M., Rocha, A., & Torres, R. D. S. (2014). Automatic identification of fruit flies (Diptera: Tephritidae). *Journal of Visual Communication and Image Representation*, *25*(7), 1516-1527.

Fuentes, A., Yoon, S., Kim, S. C., & Park, D. S. (2017). A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition. *Sensors*, *17*(9), 2022.

Gao, J., Liao, W., Nuyttens, D., Lootens, P., Vangeyte, J., Pižurica, A., ... & Pieters, J. G. (2018). Fusion of pixel and object-based features for weed mapping using unmanned aerial vehicle imagery. *International journal of applied earth observation and geoinformation*, *67*, 43-53.

Gray, P. C., Fleishman, A. B., Klein, D. J., McKown, M. W., Bézy, V. S., Lohmann, K. J., & Johnston, D. W. (2019). A convolutional neural network for detecting sea turtles in drone imagery. *Methods in Ecology and Evolution*, *10*(3), 345-355.

Grm, K., Štruc, V., Artiges, A., Caron, M., & Ekenel, H. K. (2017). Strengths and weaknesses of deep learning models for face recognition against image degradations. *Iet Biometrics*, *7*(1), 81-89.

Hand, D., & Christen, P. (2018). A note on using the F-measure for evaluating record linkage algorithms. *Statistics and Computing*, *28*(3), 539-547.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097-1105).

Hofmeester, T. R., Rowcliffe, J. M., & Jansen, P. A. (2017). A simple method for estimating the effective detection distance of camera traps. *Remote Sensing in Ecology and Conservation*, *3*(2), 81-89.

Huang, J., Rathod, V., Sun, C., Zhu, M., Korattikara, A., Fathi, A., ... & Murphy, K. (2017). Speed/accuracy trade-offs for modern convolutional object detectors. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7310-7311).

Hui, J. (2018). *mAP (mean Average Precision) for Object Detection.* Retrieved from www.medium.com: https://medium.com/@jonathan_hui/map-mean-average-precision-for- object-detection-45c121a31173

Hung, C., Xu, Z., & Sukkarieh, S. (2014). Feature learning based approach for weed classification using high resolution aerial images from a digital camera mounted on a UAV. *Remote Sensing*, *6*(12), 12037-12054.

Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Computers and electronics in agriculture*, *147*, 70-90.

Kellenberger, B., Marcos, D., & Tuia, D. (2018). Detecting mammals in UAV images: Best practices to address a substantially imbalanced dataset with deep learning. *Remote sensing of environment*, *216*, 139-153.

Lee, J. C., Bruck, D. J., Dreves, A. J., Ioriatti, C., Vogt, H., & Baufeld, P. (2011). In focus: spotted wing drosophila, Drosophila suzukii, across perspectives. *Pest management science*, *67*(11), 1349-1351.

Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., ... & Zitnick, C. L. (2014, September). Microsoft coco: Common objects in context. In *European conference on computer vision* (pp. 740-755). Springer, Cham.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016, October). Ssd: Single shot multibox detector. In *European conference on computer vision* (pp. 21-37). Springer, Cham.

Liu, Z., Gao, J., Yang, G., Zhang, H., & He, Y. (2016). Localization and classification of paddy field pests using a saliency map and deep convolutional neural network. *Scientific reports*, *6*, 20410.

Perez, L., & Wang, J. (2017). The effectiveness of data augmentation in image classification using deep learning. *arXiv preprint arXiv:1712.04621*

McEvery, F. S. (2017). *High resolution diagnostic images of Drosophila suzukii (Diptera: Drosophilidae).* Retrieved February 21, 2019, from Figshare.com: https://doi.org/10.6084/m9.figshare.5363431.v1

Metcalf, R. L., & Luckmann, W. H. (Eds.). (1994). *Introduction to Insect Pest Management* (Vol. 3). New York: Wiley-Interscience.

Miller, G. T., & Spoolman, S. (2014). Biodiversity. In G. T. Miller, *Sustaining the Earth* (pp. 211-216). Cengage Learning.

Mortelmans, J., Casteels, H., & Beliën, T. (2012). Drosophila suzukii (Diptera: Drosophilidae): A pest species new to Belgium. *Belg. J. Zool*, *142*(2), 143-146.

Nam, N. T., & Hung, P. D. (2018, June). Pest detection on traps using deep convolutional neural networks. In *Proceedings of the 2018 International Conference on Control and Computer Vision* (pp. 33-38).

Norouzzadeh, M. S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M. S., Packer, C., & Clune, J. (2018). Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *Proceedings of the National Academy of Sciences*, *115*(25), E5716-E5725.

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

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems* (pp. 91-99).

Rey, N., Volpi, M., Joost, S., & Tuia, D. (2017). Detecting animals in African Savanna with UAVs and the crowds. *Remote Sensing of Environment*, *200*, 341-351.

Schneider, S., Taylor, G. W., & Kremer, S. (2018\). Deep learning object detection methods for ecological camera trap data. In *2018 15th Conference on Computer and Robot Vision (CRV)* (pp. 321-328). IEEE.

Schneider, S., Taylor, G. W., Linquist, S., & Kremer, S. C. (2019). Past, present and future approaches using computer vision for animal re-identification from camera trap data. *Methods in Ecology and Evolution*, *10*(4), 461-470.

Selvaraj, M. G., Vergara, A., Ruiz, H., Safari, N., Elayabalan, S., Ocimati, W., & Blomme, G. (2019). AI-powered banana diseases and pest detection. *Plant Methods*, *15*(1), 92.

Shen, Y., Zhou, H., Li, J., Jian, F., & Jayas, D. S. (2018). Detection of stored-grain insects using deep learning. *Computers and Electronics in Agriculture*, *145*, 319-325.

Sun, Y., Liu, X., Yuan, M., Ren, L., Wang, J., & Chen, Z. (2018). Automatic in-trap pest detection using deep learning for pheromone-based Dendroctonus valens monitoring. *Biosystems engineering*, *176*, 140-150.

Tzutalin. (2015). https://github.com/tzutalin/labelImg. *LabelImg*. Retrieved from https://github.com.

Wu, Q., Zhang, K., & Meng, J. (2019). Identification of Soybean Leaf Diseases via Deep Learning. *Journal of The Institution of Engineers (India): Series A*, *100*(4), 659-666.

Xing, S., Lee, M., & Lee, K. K. (2019). Citrus Pests and Diseases Recognition Model Using Weakly Dense Connected Convolution Network. *Sensors*, *19*(14), 3195.

Zhao, Y., He, Y., & Xu, X. (2012). A novel algorithm for damage recognition on pest-infested oilseed rape leaves. *Computers and electronics in agriculture*, *89*, 41-50.

Zhong, Y., Gao, J., Lei, Q., & Zhou, Y. (2018). A vision-based counting and recognition system for flying insects in intelligent agriculture. *Sensors*, *18*(5), 1489.

Annex I Linear regression assumptions

A linear relationship in the data

In paragraph 3.5, figure 3.12 we can see that the data has very large deviations, with for example F_1 scores ranging from ±0.15 to 1.00. Furthermore, the small amount of different distances (only three different distances) makes it more difficult to have a clear linear relationship. For figure 3.13 in paragraph 3.5, the deviation is also large, although somewhat smaller than for figure 3.12. Also, the larger number of different view angles makes the relationship easier to see. Even though the large deviation in the data and relatively small amount of different distances and view angles, there seems to be a linear relationship, with on average the "y" value decreasing with an increasing "x" value. Although there needs to be more research done to determine this.

Independence of the dependent variables

The question here is if the observations of the 'y' variable are dependent or independent of each other. An example of a situation where the 'y' variable is dependent is in the case of the growth of a plant, where the total biomass increase reached at day 10 influences the total biomass increase reached at day 20. In the case of this experiment, the score of precision (or F_1 or recall) at a view angle of 20° and a distance of 14 cm does not influence the precision score at a view angle of 30° at a distance of 14 cm, for example. This means that the independent variable(s) are not dependent.

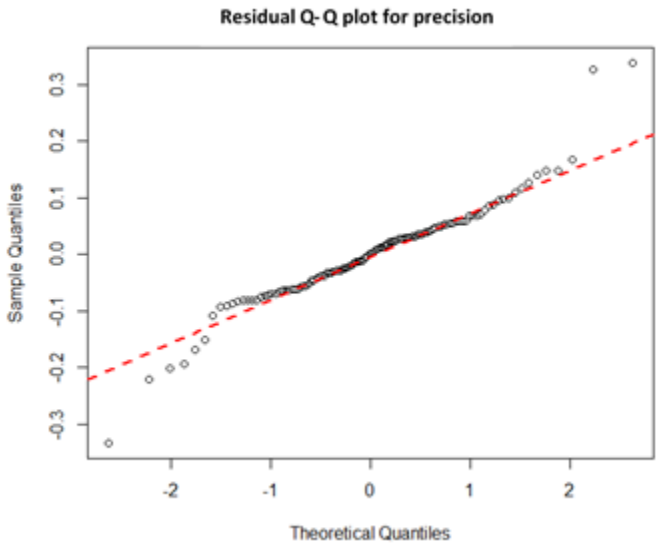


Figure A.1 Residual quantile-quantile plot for precision. The graph shows the distribution of the residual data for precision. The hollow circles represent each residual data point and the dotted red line represents the normal distribution line. On the x-axis the theoretical quantiles are shown, which are the model generated residuals. On the y-axis the sample quantiles are shown, which are the residuals of the experiment data

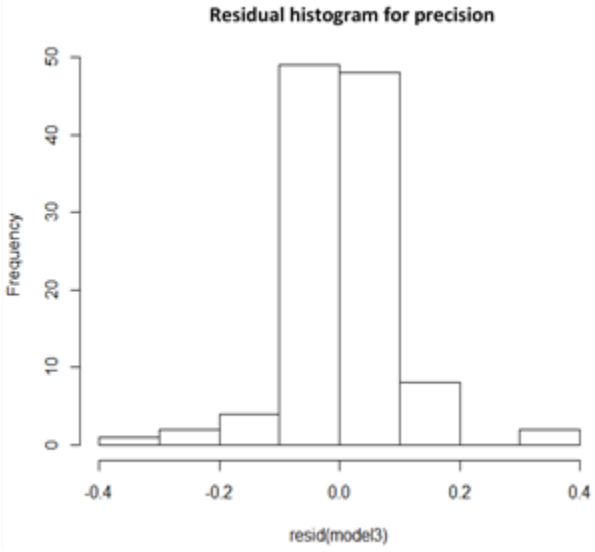


Figure A.2 Residual histogram for precision. The graph shows the distribution of the residuals. The x-axis shows the values of the residuals (how large the error is) and the y-axis shows the frequency an error occurs (within a specific value range).

In figure A.1, the residual Q-Q plot is shown for precision. First thing that is visible is that there are data points on both the left (theoretical quantiles of -3 and -1.5) and the right part (theoretical quantiles of 2 and 3) that are strongly deviated from the normal distribution line. Due to the low amount of data points that show this pattern, these points seem to be outliers. Furthermore, there is a large amount of data points that are centred in the middle of the graph. This is also shown in figure A.2, where most of the residual data points occur in the middle of the graph. Overall it seems that the model does a decent job at predicting the precision value based on the distance and the view angle, although there are some strong outliers, indicating that the regression model has problems with modelling the data collected in the experiment.

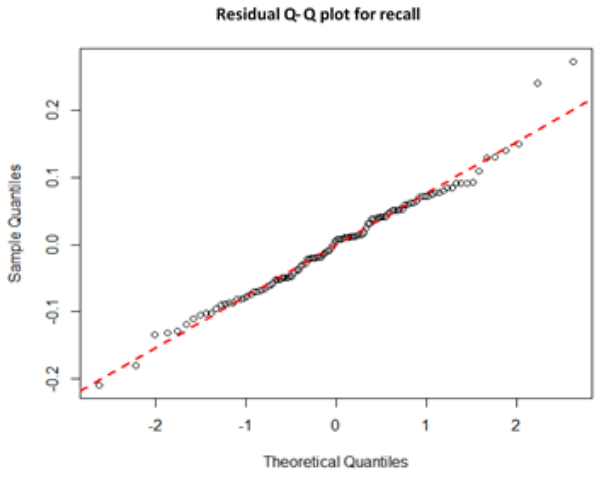


Figure A.3 Residual quantile-quantile plot for recall

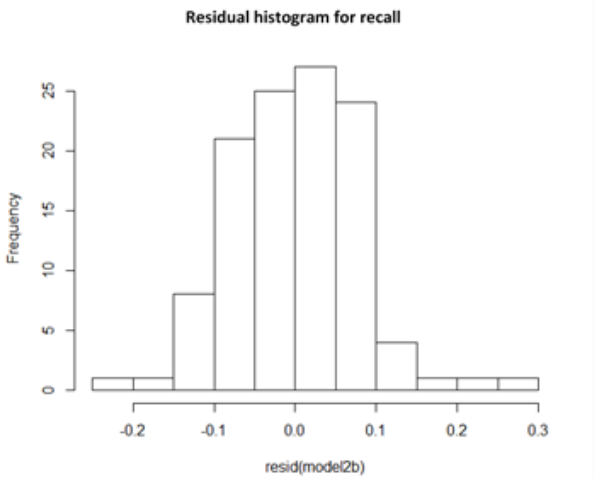


Figure A.4 Residual histogram for recall

In figure A.3, the residual Q-Q plot for recall is shown. Unlike in the Q-Q plot for precision, the Q-Q plot for recall shows far less deviation from the normal distribution line. Similar to precision though, there are also two quantile data points in the upper right corner. Another aspect that is visible is that most of the quantiles data points are oscillating around the normal distribution line. It is unclear why this pattern is occurring. Overall, the graph shows that the quantiles are nicely placed on, or very close to the normal distribution line. In figure A.4, the residual histogram is shown for recall. Here we see that there is a decent skewness to the left side of the graph. Although this is not clearly visible in the Q-Q graph.

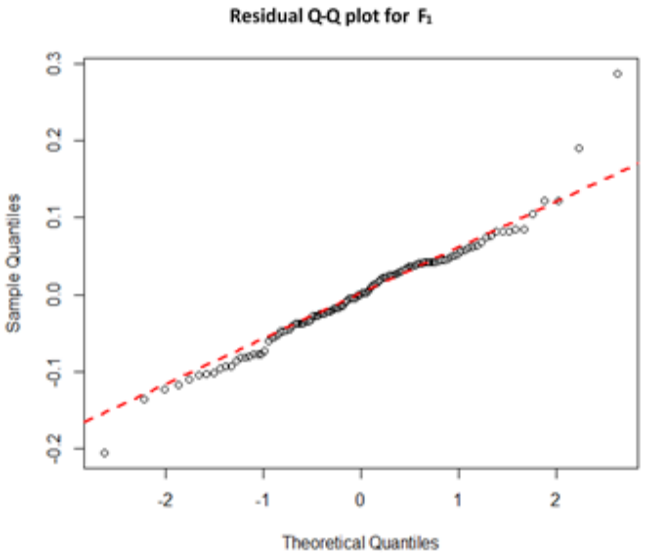


Figure A.5 Residual quantile-quantile plot for F_1

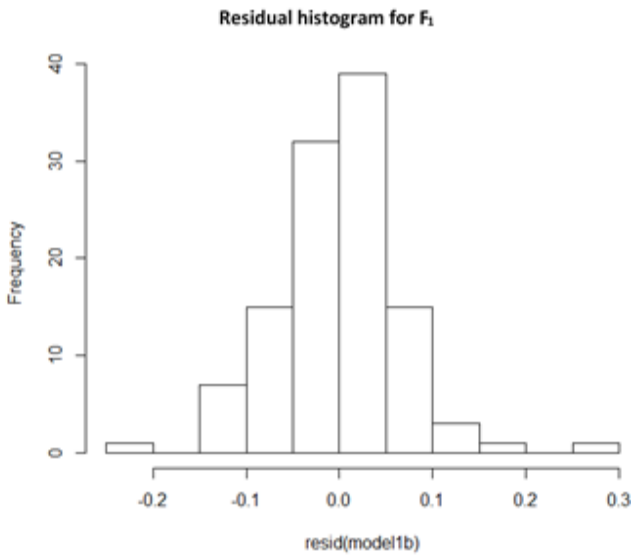


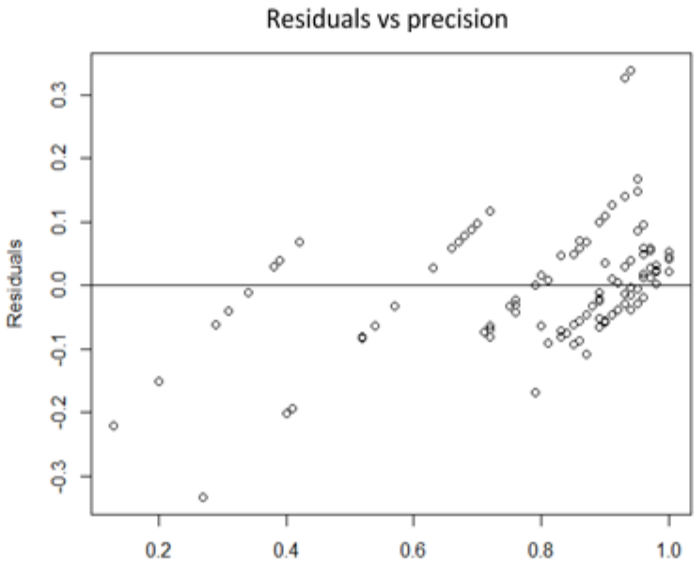
Figure A.6 Residual histogram for F_1

In figure A.5, the residual quantile-quantile plot is shown for F_1. At first glance there seems to be a good fit, with almost all the points positioned on the normal distribution line. Again, there are two points in the upper right corner that deviate very strongly. What is also visible is that in the left corner there are a decent amount of points that are positioned a bit under the normal distribution line. These points clearly follow the normal distribution line in parallel fashion, although very close to the normal distribution line. Overall there is a strong linear relationship between the explanatory variables and the independent variables.

Constant error variance

To check if there is a constant error variance in the data produced by the model, the Breusch-Pagan test (BP test) will be conducted on each model. (Breusch & Pagan, 1979). The Breusch-Pagan test, tests for heteroskedasticity. If the p-value is ≤ 0.05, then the there is a pattern in the error variance, meaning that there is no constant error variance. If the p-value is ≥ 0.05, then the error variance is constant. Furthermore, for each independent variable, a graph is shown where the residuals are plotted versusthe independent variable values.

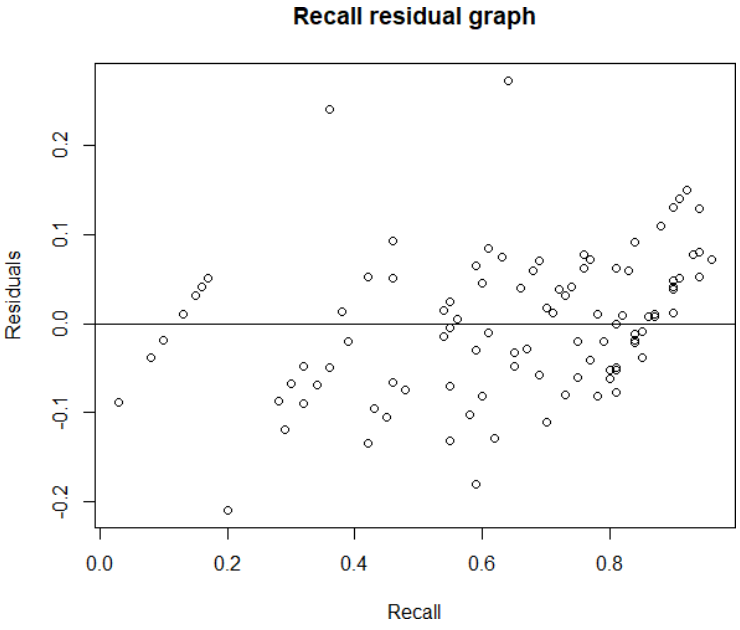
In figure A.7 we can see the residuals plotted versus the precision values predicted by the model. Also, in the figure, the results of the BP test are shown. The p-value is smaller than 0.05, indicating that there is no constant error variance for precision. In the left part of the graph we can see that there are residuals mostly in the negative part of the y-axis while there are almost no residuals in the positive part of the y-axis between the precision values of 0 and 0.6. This result could indicate that the relationship between precision and the variables used in the model is not linear.



BP = 15.219	df = 5	p-value = 0.009476

Figure A.7 Residuals vs precision. In this graph the residuals of the model are plotted versus the precision values predicted by the model. The precision values are placed on the x-axis and the residual values are placed on the y-axis. Right from the graph, the results are shown for the Breusch-Pagan test. The BP value is the actual result of the calculation made with the test, and is used together with the number of variables in the model (df) to calculate the p-value

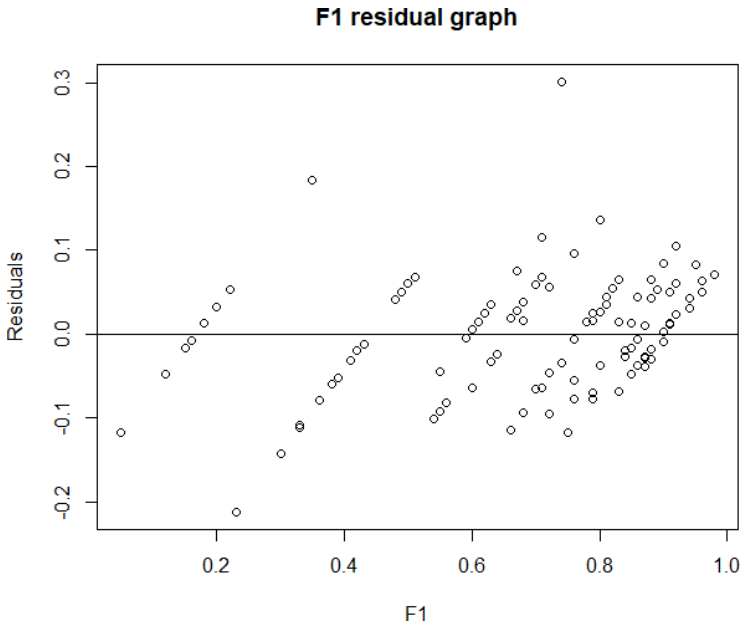
In figure A.8, the graph of the recall residuals plotted versus the recall values is shown. According to the results of the BP-test, the distribution of the error variance is random. This suggests that the relationship between recall and the variables used in the model is linear.



BP = 6.9682	df = 5	p-value = 0.223

Figure A.8 Residuals vs recall

In figure A.9 the residuals of the predicted F_1 score is plotted against the predicted F_1 score. According to the p-value, which is larger than 0.05, we can reject the null-hypothesis. This result indicates that the data shows a linear relationship for F_1 and the five variables used in the model.



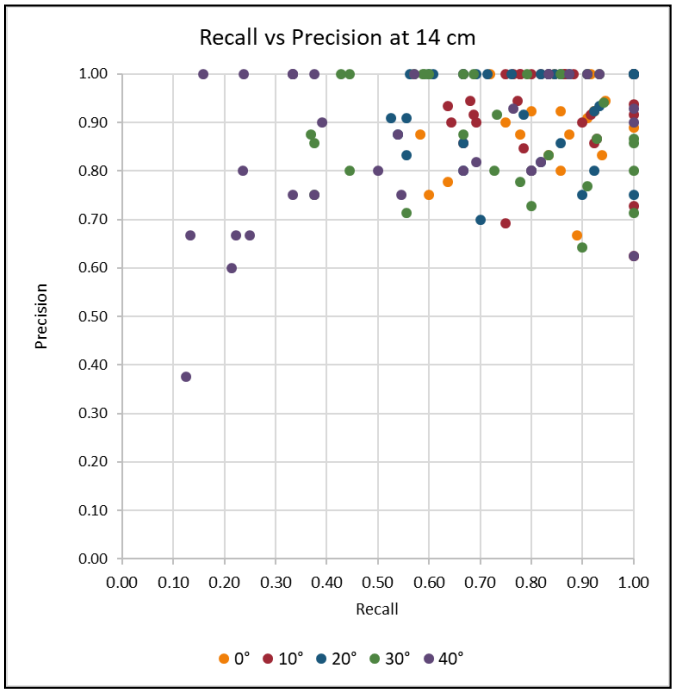
BP = 9.0045	df = 5	p-value = 0.1089

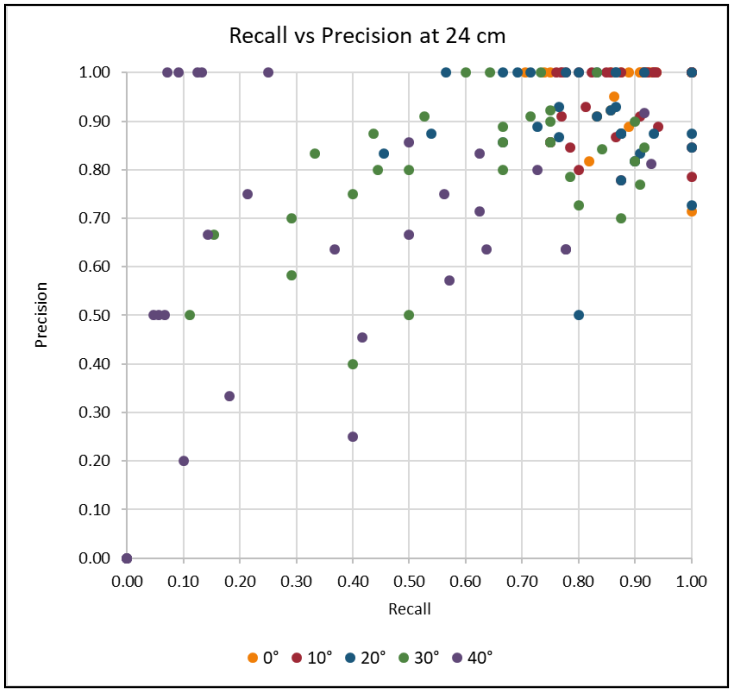
Figure A.9 Residuals vs F_1

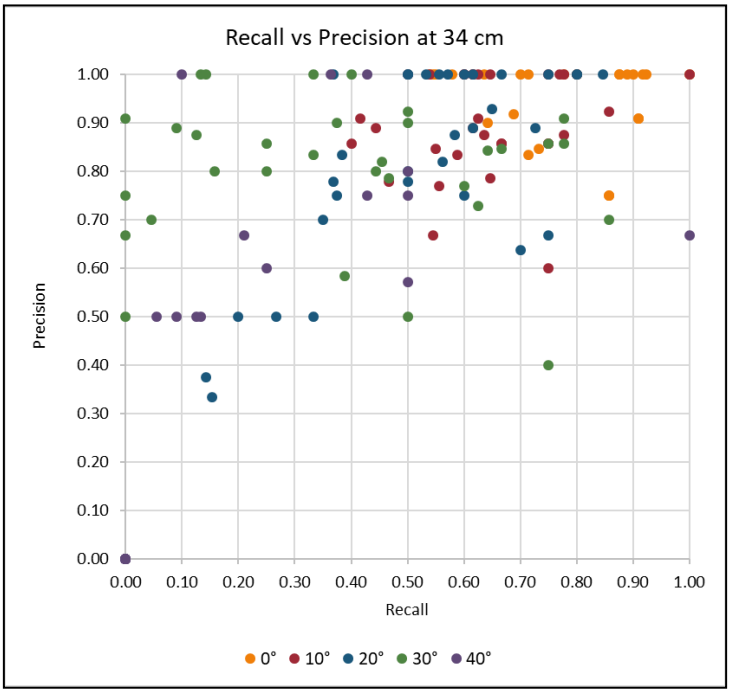
Overall, the assumptions seem to be met for recall and F_1. For precision it seems that assumption four (constant error variance) is not met. Even though precision has not met all the assumption the model is still used further in this thesis, although it should be handled with some scepticism.

Annex II Precision vs Recall

In this annex, the figures 3.8A, 3.8B and 3.8C are shown but instead of the mean scores, the total data is shown. Each view angle is represented with a separate colour and with the recall placed on the x-axis and the precision placed on the y-axis.







Annex III Precision, recall and F_1 scores for each view angle at each distance

In this annex, the figures 3.6, 3.7 and 3.8 are placed on the same page. The reason for this is that it makes it easier to see the relationship between the dependent variables and the view angle for each distance. See chapter 3 for a detailed description of each figure.