Norwegian University
of Life Sciences

Master's Thesis 2020 30 ECTS
Faculty of Science and Technology

Length Measurement System to Assist Automation in the Slaughtering Process

Lengdemålesystem for å tilpasse automasjon i slakterprosessen

Ruben Cornelius Stokland
Mechanical Engineering, Process Technology and Product Development

Length measurement system to assist automation in the slaughtering process

By

Ruben Cornelius Stokland

Norwegian University of Life Sciences

Faculty of Science and Technology

Master's Thesis

Spring 2020

Preface

This is my final project to complete my master's degree in Process and Energy Technology at the Norwegian University of Life Sciences. The choice of thesis was influenced by my love for food, my interest in automation and the infinite limits of computer programming. I feel lucky that I have been able to work with such a relevant and important topic. The knowledge I have acquired has been a part of getting me a permanent job in the food industry.

The master's thesis was an exciting process from finding a project to complete the final paper. Through this period, I had both challenging and encouraging days. I have learned to work independently, to search for and in academic literature and to use computational tools. Along the way of completion, I have discussed a lot of ideas and become actively involved in interesting conversations. People I have talked to have been crucial for the result. Before I thank the contributors, I want to say I am satisfied with the thesis and can proudly present it.

Ås, June 1, 2020

Ruben Cornelius Stokland

Acknowledgements

I want to express my sincere gratitude to people who have supported me throughout six years of education. The academic community at Ås is an enlightening, encouraging and caring society. I have had a wonderful time as a student thanks to professors and fellow students. Those individuals and organizations who have contributed to this thesis which I want to give a special appreciation are:

Research Professor Alex Mason for presenting project ideas, embracing me to be a part of his project and guidance throughout the thesis. (assistant supervisor)

Associate Professor Odd Ivar Lekang, my principal supervisor for his guidance throughout the thesis.

Associate Professor Jan Kåre Bøe for his help and support throughout the thesis. (assistant supervisor)

PhD Candidate Ian de Medeiros Esper for helping me with the camera setup and software implementation on Supervisely.

RoBUTCHER for accessing the lab and use of equipment's. MeaTable for also funding the overall project.

Fatland Oslo AS and the helping staff, particularly Slobodan Momić, for letting me record at the production line. Their quick response before the restrictions because of SARS-CoV-2 were imposed made a significant part in getting the results.

The Library and Writing Centre at NMBU for courses, lectures, assistance, retrieving articles and English support.

Classmates, student organizations and clubs to make me shift focus and relax.

And finally, my family and close friends for their love and support.

Thank you!

Abstract

The meat industry sees a development towards more local, sustainable and self-sufficient food. This means the whole process of livestock farming, butchering and distributing. If a slaughtering process were automated, it would replace the hardest work of a butcher and make it easier to produce local meat products. This is the aim of a larger project which this thesis is a part of; the primary task is to develop a system to measure the length of pigs so that a machine can adapt to the biological variation inherent in the raw material. Hence, the goal with this thesis is to make a precise, robust and contactless measurement system.

This thesis begins with addressing the properties of sensors and cameras for measurement purposes. A depth camera had the best potential and was chosen for further usage. Furthermore, the thesis discloses the applications for recognition types, and distinguishes between different detection and segmentation models. Three models were selected for testing and aligned next to each other for comparison. When a pig is located in a picture, it is possible to reduce the amount of data in the depth image and processing it to a 3D point cloud. The point cloud needs filtration due to a rough and noisy surface. At this point, the nose and hind feet can be pinpointed and the distance between them calculated to obtain a carcass length estimation.

To get the results, a depth camera was installed in a pork slaughter-line (Fatland, Oslo). The placement of the camera was adjusted during recording, which led to a consequential measurement error. Unfortunately, it was not possible to get more data as the industry no longer allowed outsiders to visit the plant during the COVID-19 outbreak. As a consequence of the measurement error, the system's accuracy was unobtainable. However, the precision could be found by adjusting the manual measurement using a calibration ratio. A few comparisons were made of the length differences, and the total error was ±4 cm. The accuracy and precision can be improved by a variety of ways. If the pigs in the production line would stop at the exact place, the error is likely to decline. Additionally, some of the aforementioned operations may be easier to implement as a result of this.

Sammendrag

Kjøttindustrien har en utvikling som går i retning av mer lokale, bærekraftige og selvforsynte matvarer. Det vil si i hele prosessen fra husdyrhold til slakting til distribusjon. Hvis slakterprosessen var automatisert, ville det kunne erstattet det tyngste arbeidet til en slakter, og gjort det lettere å produsere lokale kjøttprodukter. Dette er målet til det overordnede prosjektet som denne masteravhandlingen er en del av; hovedoppgaven er å utvikle et system som måler lengden på griser, slik at en maskin kan tilpasse seg de biologiske variasjonene i råmaterialet. Målet for denne avhandlingen er derfor å lage et presist, robust og kontaktløst målesystem.

Denne avhandlingen begynner med å adressere egenskapene til sensorer og kameraer som målegrunnlag. Et dybdekamera hadde størst potensial og ble valgt for videre bruk. Oppgaven beskriver deretter bruksområdene for gjenkjenningstyper, og skjelner forskjellen mellom ulike deteksjons- og segmenteringsmodeller. Tre modeller ble valgt ut for testing, og satt opp mot hverandre for sammenligning. Når en gris er lokalisert i et bilde, er det mulig å redusere datamengden i dybdebildet og prosessere det til en 3D-punktsky. Punktskyen trenger filtrering på grunn av en ru og støyet overflate. På dette tidspunktet kan nesen og bakbena bli nøye angitt, og lengden mellom dem beregnes for å estimere kadaverlengden.

For å få resultatene, ble et dybdekamera installert i en svineslakterline (Fatland, Oslo). Plasseringen av kamera ble justert mens det tok bilder, noe som førte til en følgefeil på målingene. Uheldigvis var det ikke mulig å få tilgang på mer data, ettersom industrien ikke tillot utenforstående å besøke anlegget under COVID-19 utbruddet. Som en konsekvens av målefeilen var det umulig å finne systemets nøyaktighet. Imidlertid kunne presisjonen bli funnet ved å justere den manuelle målingen ved å bruke et kalibreringsforhold. Noen få sammenligninger ble gjort av lengdeforskjellene, og den totale feilen var på ±4 cm. Nøyaktigheten og presisjonen kan forbedres på mange måter. Hvis grisene i produksjonslinjen kunne stoppe på nøyaktig det samme stedet, ville feilen sannsynligvis avta. I tillegg kan operasjonene nevnt ovenfor være lettere å iverksette som et resultat av dette.

Terms and Abbreviations

3D	Three dimensional
AI	Artificial intelligence
BCC	Beef Classification Center
CNN	Convolutional neural network
CT	Computer tomography
DAT	Data file
DEXA	Dual energy x-ray absorptiometry
DMRI	Danish Technological Institute
DTL	Data transformation language
FPS	Frames per second
Fast R-CNN	Fast region-based convolutional neural network
Faster R-CNN	Faster region-based convolutional neural network
IMRaD	Introduction, (theory), method, results and discussion
IoU	Intersection of union
JPG/JPEG	Joint Photographic Experts Group file
LED	Light-emitting diode
LiDAR	Light detection and ranging
mAP	Mean Average Precision
Mask R-CNN	Mask region-based convolutional neural network
MFC	Meat Factory Cell
MLA	Meat and Livestock Australia
NIR	Near infrared
NIRS	Near infrared spectroscopy
PCD	Kodak Photo CD Image file
PV	Photovoltaic
PZT	Lead zirconate titaneate
R-CNN	Region-based convolutional neural network
RGB	Red, green, blue
RGBD	Red, green, blue, depth
RMS	Root mean square
ROI	Region of interest
RPN	Reginal proposed network
SEXA	Single energy x-ray absorptiometry
SI	The international system of units
SONAR	Sound navigation ranging
SPECT	Single-photon emission computed tomography
SSD	Single shot detector
TOF	Time of flight
UK	United Kingdom
UN	United Nations
VGG	Visual Geometry Group
YOLO	You only look once

List of Contents

1 Introduction

1.1 Background

The supply of food is essential for human survival. Meat and vegetables are vital for the nourishment of the body. They are the main source of fat, protein, carbohydrates and vitamins. The access for food is the most fundamental part of living. Any other activity or chore becomes meaningless, due to lack of energy and bodily functions. In other words, the food industry is one of the cornerstones in the society. In 2018 the annual meat production reached 370 million tons, and continues to increase [9]. The push for development in the meat industry is a result of higher demand. Customers seek variety of meat in virtue of taste, texture, juiciness and proportion of proteins. Pork and poultry is the largest share of meat production at approximately 35-40% worldwide [10]. Since the demand for food is equal to the increase of human growth, the efficiency of food production must constantly improve.

Nations are compelled to be more self-sufficient because of the food security. In the spring of 2020 during the writing of this thesis, the world was in the midst of the coronavirus outbreak. Some of the repercussions in this context were travel restrictions and border shutdowns, which also affected the transportation of goods. To be able to resist similar situations, nations cannot rely solely on imported goods, but have their own product manufacturing.

The Norwegian University of Life Sciences (NMBU) was originally based on agriculture. In recent years, the professional diversity has grown along with the emerging field of research. The overall commitment is sustainability, both at campus and as a contributor worldwide. Hence, the main target areas for further development is interdisciplinarity, digitalization, educational environment and research. The University are also fundamentally rooted in UN Sustainable Development Goals. Out of 17 goals, four of the most relevant topics are addressed in this paper. The following goals are 2, 3, 8 and 9, see figure 1. The master's thesis will reflect the University's core values mentioned above [11].

Figure 1. Four focus areas that are reflected throughout the thesis. [1]

The traditional way to produce meat is by hand: A butcher kills, slaughter and dissects skin, flesh and bone. Entrails are being removed while other body parts are carved for cooking. Nowadays the technical improvement is based on automatization. For instance, conveyor belts and hooks are helping the butcher with the transportation and lifting of heavy carcasses. Moreover, machines and robotic

arms do simple repetitive task and have replaced some of the work of a butcher. The push for further automatization is due, not only to effectiveness, but also as a result of a dwindling labor supply [12]. A reason for this is that the profession is associated with hard physical work and strain injuries.

To maintain the supply of pork a more modern and technological production process is needed. Factories with linear production lines are fast but risk a total stop in production when an element is defect. A way of preventing interruptions, is to have a vertical production structure. A vertical structure means that a robot can do multiple tasks and sophisticatedly navigate the carcass to desired sections. New inventions like this in the pork supply chain are necessary for the growth in the meat industry. [13]

A Meat Factory Cell (MFC) is the main concept of the EU-funded RoBUTCHER project [14] [15]. The cell is a station where a robot, or a collaboration between a robot and a butcher, is doing most of the primary task. The idea builds on a vertical structure where the parallel cells increase the capacity. Benefits like a flexible production and space efficiency are highly valued. If the robot is unable to precisely divide muscle tissue and fat, a butcher can take charge; meanwhile, the robot holds and stretches the body parts. In this case the operation is easier to accomplish and does not demand heavy lifting.

Hygiene in the cell is another important part of the slaughtering process. The report of Alvseike O. et al. [16] contains an outside-in approach for the order of operation; in contrary to standard procedures. Since the early withdrawal of the intestines constitutes a greater risk of contamination, an outside-in operation is conducted. The report emphasizes the importance of having a safe environment for the product and employees, which will be reflected upon in this thesis.

To make a meat production process fully autonomous, advanced and complex systems are needed. Artificial intelligence, machine learning and neural networks are examples of tools to streamline disciplines. The meat factory cell requires knowledge about the pig in the early stage of the automatic process. This thesis focuses on development of a pre-slaughter measuring tool to obtain information about the length of a pig.

Inside the meat factory cell there are scalable robots which require the length input before the carcass enters. The length of a pig is determined by the outstretched height from top to bottom. Every pig is being transported by railings in the ceiling and are hanging by their hind leg. Somewhere along this transport path, the length measurement system should be mounted. Cameras and sensors can be used to scan the pig to obtain the required information.

1.2 Literature Review

In the meat industry, the technological progress has been slow compared to other industries. Due to the challenges of handling the soft tissues, the automated tasks attached to slaughter have been overdue. Now the robotics and software technology are established to process big data. This gives a possibility of finding a solution to the challenges. The following paragraph contains a selection of concepts that has similarities with the automatic length measurement system.

Automated systems for livestock often focus on weight measurement. One early study was done with living pigs at the Research Centre Foulum in Denmark [17]. The purpose of the experiment was to determine the life weight from image analysis. A total of 416 pigs from the range 25 kg – 100 kg was weighed and filmed. The area of the body was compared to the real weight, and the relationship between them was described in a diagram. This diagram showed that the weight increased exponentially with the area. An equivalent study was done with sheep [18]. Furthermore, a digital body measurement, including length and height has been done on cows [19]. Photos from different angles were captured for the purpose of image analysis.

There has also been lots of studies on pig detection systems. One example is automatic tracking with video recording to monitor their health [20]. The basis of the article was to detect patterns of action to reveal their internal state. In general, the goal was to improve the pig's standard of living. This article is taken from the journal Scientific Reports which was published by Nature Research. The following article is also about the pig's welfare [21]. It addresses the impact of aggressive behavior and focus on how to recognize actions like chasing and head knocking. A Kinect Depth Sensor was installed above the stall as a standalone system.

Companies and articles with the state-of-the-art appliances are the most relevant for this thesis. SCOTT, Frontmatec, Meat and Livestock Australia (MLA), Danish Technological Institute (DMRI), etc. are examples of companies that are leading the cutting-edge technology for food and meat automation. SCOTT has a system that sends X-rays for generating 3D models [22]. The model is further applied for length measurement and cutting operations. Frontmatec has Amongst other developed a third generation Beef Classification Center (BCC-3^{TM})[23]. It is quantifying the meat's content for product sorting and prize estimation. The BCC consists of a multi-view stereo system for a complete rendered 3D image. MLA made a report in 2015 about a various of technologies used for meat processing [24]. This report will be referred to later in the thesis.

An article based on a thesis with similar approach was done in the University of Nebraska-Lincoln [25]. The article is about image detection and separation of dense pig herds. A big dataset was trained to get the detection model. To filter background noise, a depth camera with a maximum threshold was

used. Along with an image segmentation, the researchers managed to make a good model within the testing conditions. The last supplement is an article by G. Purnell from Grimsby Institute of Further & Higher Education, UK [26]. It addresses the challenges and positive outcomes of automation within meat production. Overall, the article expresses the urge for further development. As an encouragement, the work to succeed with a length measurement system is strengthened.

1.3 Aim and Objectives

The rationale for the work is to modernize parts of the meat industry. For instance, to improve the slaughtering process, working conditions, reduce waste and contribute to the RoBUTCHER project with a measurement system. Some of the parts mentioned will hopefully be impacted by the completion of this thesis.

The master thesis is focusing on how to automatically measure the length of every pig, so that a robot can be scaled to customize each carcass. The measurement happens when the pig is hanging upside-down on a production line. One of the main criteria is to develop a non-contact measurement system. At the MFC, the pig is attached with two hooks trough the feet and around the muzzle. This means that the desirable length is from the hind feet to the nose. Moreover, the system should be flexible, adaptable, and cost-effective as the slaughtering process are intended for small to medium sized factories.

Aim

- Develop a precise, robust and contactless measurement system to estimate the length of pigs in a production line.

Objectives

- Perform a comprehensive literature search and review to evaluate the current available approaches for carcass length measurement.
- Develop a selection of models to detect pigs at the production line.
- Test and compare the results from different models.
- Make a script to convert depth images into a point cloud.
- Filter the noise of a point cloud to resemble the pig.
- Estimate the length of a pig to prepare the machine at the MFC.
- Establish the accuracy, precision and uncertainty of the system.

Bottlenecks

- The economy associated with the project and the estimation of effectiveness.
- The scaling of the robot in the Meat Factory Cell is done later by someone else.

- The dataset for segmentation and training is limited to time.
- The measurement system is fitted to a specific line but can easily be adapted to other lines.
- Number of test samples to get statistical significance.
- Environmental aspects (food waste, energy consumption, resources, recycling, environmental footprint)
- COVID-19; no access to meat industry after 12.03.2020 – all slaughter lines closed to outside visitors.

2 Theory and Technology

2.1 Measurements and Methodology

Measurement, scaling and customization is part of every daily activity. For instance, tailored clothing, cooking, cleaning, transportation, and so forth. These activities are bound by different physical and natural conditions which limit the area of application. Consequently, the focus on accuracy within materials are essential. Different parts of the world have varying measurement units, but the most common is the SI units. This international system of units is implemented in this report.

Over the last few centuries, automatic sensors have increasingly replaced manual measurements. The implementation of sensors has been responsible for better accuracy, faster production and in general a more reliable measuring tool. Sensor technologies differ in size, appearance and application. Some of their wide usage areas are temperature, pressure, force, flux, lux, sound, radiation, humidity, voltage, movement, etc. However, the main focus of this thesis is length measurement.

In the last decades camera- and sensor technologies have been used for meat automation. Due to sanitary reasons, the measurement system must not be in contact with the carcasses. The following subsections contain technologies that can be applicable. Technologies that do not satisfies the acceptance criteria will not be included in the theory chapter. The absolute criteria for a length measurement technology is:

- Non-contact
- Not harmful or toxic to the meat
- Not contaminated from the production
- Consistent and reliable in a long term
- Inexpensive

The science behind this text is based on books and scientific articles. Secondary sources are used as support and for illustrations. For this reason, the sources are split into two sections in the references. Moreover, every figure that has not been provided with a source is either the author's own work or made in collaboration with coworkers.

Throughout the thesis, a IMRaD model is structuring the framework [27]. In addition to the model, the structure follows a logical progress. The reason for this is to gradually introduce the reader as problems arise. Along with the intention of making the text easy to read, the hope is to get the reader interested, fascinated and engaged.

2.2 Sensor Technologies

Electromagnetic radiation is the most used physical property for sensing. The difference in wavelength is categorized by the following types of radiation: Radio wave, microwave, infrared, visible light, ultraviolet, x-ray and gamma ray. This list is sorted by decreasing wavelength but has a reverse proportional energy level. Most of the radiation types are tested in the meat industry for measurement purposes.

2.2.1 Optical Sensors

Optical units are commonly applied in different sensors. A term often used is LiDAR (light detection and ranging). LiDARs do mainly consist of a transmitter and a receiver. They are entirely dependent on the reflection of the target. The transmitter can be a light source that sends out visible light, but it is also very common to send near infrared radiation (NIR). NIR is not affected by other light sources because the light frequency does not intersect. That means the sensor get completely surveillance in total darkness.

Light-emitting diodes (LED) are often used as transmitters. However, the receiver must detect the light or infrared radiation. The most common receivers are photoconductive, photodiodes and photovoltaic (PV) devices. All of these devices are used in open air. Photoconductive converts incident light into a resistance in an electric circuit. In other words when the light signal increases, the resistance will also increase. PV devices generate voltage when the irradiation is above a certain threshold. In contrast photodiodes uses the current to describe the amount of irradiation. [3]

Another medium for the optical sensor is fiber, which operates inside a cable. The fiber-optic cable can use the same technology as in open air, but do not necessarily go in a straight line. At bigger distances, the usage of infrared and radio waves are more common. Infrared radiation and radio waves have a lower frequency than visible light. Some of their applications are astronomy and navigation. [3]

2.2.2 Spectroscopy Technologies

Techniques of comparing the intersection between radiation and matter are valuable. Because atoms absorb and emit electromagnetic radiation, a method of scanning a material's content and texture is possible. The name of this method is called spectroscopy. Raman spectroscopy is an example of where a laser is pointed at the meat. Returning photons give information about the state of the molecules due to a frequency change. Near infrared spectroscopy (NIRS) sends infrared radiation that is being absorbed and scattered both on the surface and within the meat.

Devices using ionizing gamma rays, like SPECT (Single-photon emission computed tomography), is considered unsustainable. Gamma rays has the highest energy level and can be harmful for the meat. Apart from this, spectroscopy has a low impact on the surroundings. Spectroscopy technologies are

fitted for classifying the content of fat, protein and moisture. Furthermore, to measure the pH and get an understanding of the flavors. [24]

2.2.3 Ultrasonic Sensors

A SONAR (sound navigation ranging) is a general term of all sensors that uses sound waves. An ultrasonic device uses sound frequencies above 20 kHz, around the limit of what a human ear can hear. The main difference between optical and ultrasonic sensors is that the waves are respectively of light and sound. While the speed of light is approximately 10^8 m/s (in vacuum) the sound travels at 343 m/s (in air at 20^{o}C).

Despite the difference both technologies include a transmitter and a receiver. Ultrasonic sensors measure the length to an object either by calculating the time of flight (TOF) or by the phase shift in the sound wave. In addition to the two methods the angle between the transmitter and the receiver can also determine the distance. For instance, a combine technique between more than one type of measurement system gives a more reliable result. A piezoelectric crystal is often used in ultrasonic sensors because it can be both the transmitter and the receiver. [2, 3]

The medium where the light, sound or laser are traveling, is essential for the speed and range. A denser medium, like liquids or solids, conserves the energy and preserves the speed of sound. Furthermore, the medium is affected by temperature. Due to environmental changes, a precise sensor system should include a complex algorithm that comprehend individual factors. Other factors can be humidity, turbulence, foreign particles or pressure. [2, 3]

Another way of excluding differences in the air is to have one extra set of transmitter and receiver. On the contrary to the sensor, these parts must face each other, with a constant distance. The unknown speed will then be found and used as an offset value for other sensors. Compared to optical sensors, the ultrasonic sensors are not restricted to air transportation only. The sound can penetrate a variety of solid substances. Because of a higher density, the speed of sound is increased. The sound wave behaves in a similar way as the frequent pressure difference, especially in solids. [2, 3]

Devices usually emits radiation in every direction, hence a spherical direction. However, when the radiation returns, the beginning of a sound wave is small before it ramps up to full amplitude, see figure 2. The reaction time of a receiver determines the error. Furthermore, if the sensor is close to a wall or other objects, the risk of getting a wrong measurement result is increased. The focusing part of the sensor is still able to detect, because

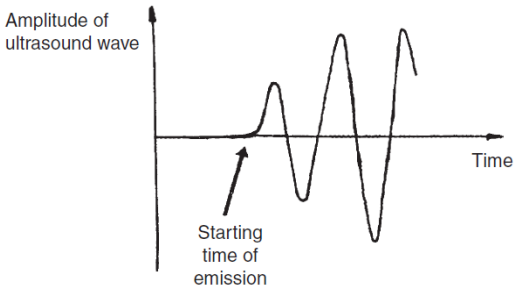


Figure 2. The curve of a sound wave when it hits the receiver. A sensitive receiver gives a smaller margin of error because it detects an incoming wave before it reaches a full amplitude. [3]

the peak energy is centered. When this energy returns, the sensor can find the length to the object. [2, 3]

If implementing the doppler effect, a sensor can also detect the speed of the object by sensing the frequency shift. This speed is proportional to z-axis which is closer or further away from the sensor. When the sensor also detects angular movement, the absolute speed can be calculated. [2, 3]

The sound energy escalates over distance; hence the waves attenuates. As a result, the ultrasonic sensor has a maximum range of detection. By the sound intensity formula, the frequency is essential for this limit. Because the ultrasonic waves have a relatively high frequency, the ranges are usually not a problem. The receiver has a detection threshold which means that when an incoming signal begins to affect, some of the first waves might not be detected. This causes a small but consistently error. Another problem appears if machines nearby also produce a high frequently sound. The sensor should therefore be placed in an area where other ultrasonic waves are absent. [2, 3]

2.2.4 Piezoelectric Transducers

Piezoelectric transducers are often applied in ultrasonic sensors to detect ultrasonic frequencies. Incoming sound wave hits the sensitive surface of a transducer. The pressure is converted to a corresponding amount of voltage, which eventually are turned into an electrical signal. Piezoelectric transducers consist of a thin film that vibrates synchronous with the ultrasound. A typical film can be a oxide lead zirconate titanate (PZT)[28]. The vibration is sinusoidal fluctuations. Also, it is possible to find the minimum- and maximum values of these pressure differences. The electric charges are furthermore reorienting inside the PZT, and the surface of the film is being induced. As a result, it causes a constant flow of electrons which can be measured in voltage, see equation 1. To summarize, a piezoelectric transducer cannot do the mapping itself, but can be applied in other technologies. [2, 3]

$$V = \frac{kFd}{A}$$

Equation 1 [3]

where k is the piezoelectric constant, F is the force [g], d is the thickness [mm] and A the area of a selected section [mm^2].

2.2.5 X-Ray Technologies

There are a great variety of x-ray technologies. In general, the high radiational energy is penetrating the pig's body. At a specific energy level, the retardation of electrons remains the same in a homogenous substance. Only the skeleton provides a resistance, which consequently becomes the operational region. The information is often used for detecting and separating body parts. A robot can

cut along bones and ribs without piercing trough them. Slaughtering and meat processing then become easier to automate. [24]

Single Energy X-Ray Absorptiometry (SEXA) has been popular in the meat industry. There is however a shift towards more extensive methods. What separates a Single- from a Dual Energy X-Ray (DEXA) is a one versus two levels of intensity. The DEXA can discriminate between two unique resistances, e.g. bones and soft tissue. Upcoming methods as 3D X-Ray and CT scanning (Computer Tomography) has a fully viewpoint from every angle of the body. In addition to the skeleton, the two methods can estimate the location and weight of muscles and fat with high accuracy. [24]

2.3 Depth Camera

In contrast to sensor technologies, cameras process the image by themselves. The image is usually a replica of the real time situation within the visual spectra. Depth cameras on the other hand gets an additional dimension. In depth perception and imaging, advanced technology is put together. Most of the technologies has their origin from how human eyes can spot depth. A few examples are motion, familiar sizes, shading, parallax, stereopsis and convergence. The last two examples can only be attained with a binocular vision, while the remaining need a deeper understanding of an object or a physical property. Humans automatically respond to depth, but a camera will on the other hand need a complex algorithm. Machine learning, deep learning and AI (artificial intelligence) is often used for interpreting the output data from the camera. [29]

While a regular camera uses pixels with a code for the coloring RGB (red green blue) a depth camera includes an additional value D (depth). One method of getting the depth information is to project a stripy stricture. When an object or a background is curved, the projected pattern will deform. The amount of deformation result in how far away parts of an object is.

In the example on figure 3, a ball spreads the stripes. By this a camera can know that the center of the ball is closest to the lens of the camera. When the ball moves towards the lens, the thickness of every stripe increases and vice versa. If both the emitter and receiver are at one place, the pattern stays the same. However,

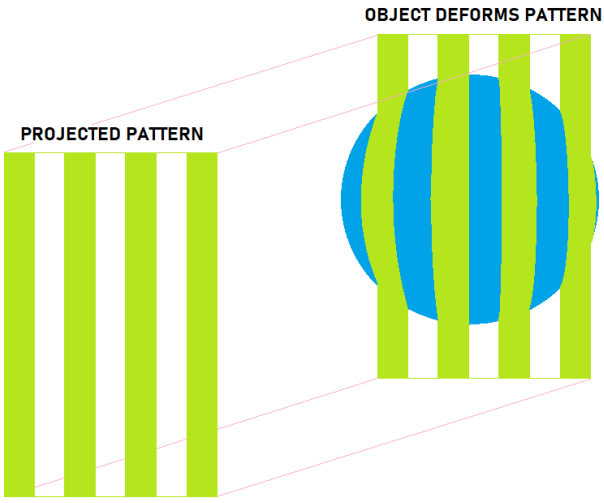


Figure 3. A simple illustration of how a 3D object deforms the projected pattern when the emitter and reciever are apart. In this case, a projector emits green stripes and the reader receives the deformated pattern. The projector is located at the bottom left of the reader.

when they are placed next to each other with a little distance, the object deformation pattern appears. This is a stereo vision technique which create the depth perception. [29]

Intel RealSense D415 is a depth camera with three lenses, se figure 4. Two of them are used for stereo image sensing and third is an RGB lens, same as regular cameras. The middle part, consisting of two white rectangles, project infrared light to the surroundings. Figure 5 is a magnified fragment of a picture taken with the depth camera. It illustrates the combine pattern projected on a rough surface. In sunlight pictures become even more clear due to more infrared radiation. As the matter of fact, every other source of infrared increases the quality of depth measuring. That is because the camera includes two stereo depth sensors that can compare both of the images. [30]

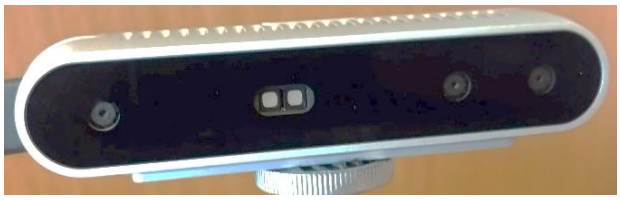


Figure 5. An Intel RealSense D415 depth camera with a projector in the middle, two stereo senses and one RGB lens at the right side. The distance between the stereo lenses are 7.0 cm.

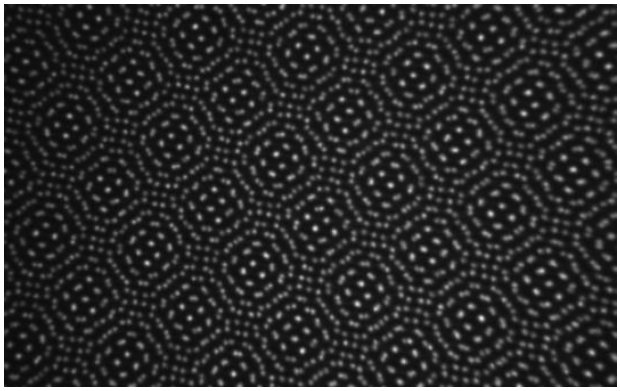


Figure 4. The projected pattern from the D415 camera on a smooth surface.

The depth can be estimated by a combination of different methods. In this case structured light and stereo depth are used. When the same point is located by both stereo lenses, triangulation is used to calculate the distance from camera to this point, see equation 2. For this operation the distance between the lenses are essensial to determine the depth. On the D415 camera, this distance is 7.0 cm. The triangulation metod is elaborated in subsection 2.6.1.

The root mean square (RMS) deviation on a tuned D415 camera is illustrated by the green graph in figure 6. As the distance increases, the error increases exponentially. A distance of 1 meter gives an aproximation of 1 mm error while a 2 meter distance gives 6 mm error. The orange graph is for compairson. It is testet on a Intel 435 camera which has a wider field of view and is meant for faster moving scenarios. Those factors are likely the reason for the worse outcome.

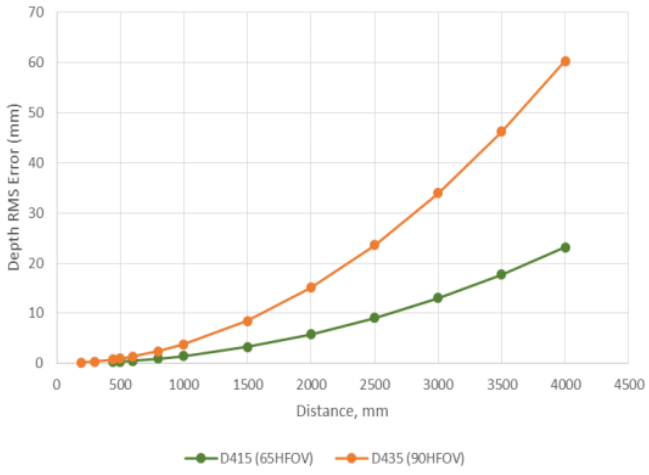


Figure 6. The RMS error of a D415 camera (green) and a D435 camera (orange) with respect to the depth estimation. [7]

2.4 Evaluation of Criteria

The procedure of finding an optimal solution to the objective is by comparison. The layout is inspired by Pugh's Methodology [31] where a few selected criteria are ranked, see table 1. Collecting data and further estimation are two of the most central points of application. Because the assessment is not particularly accurate, a rough scale of comparison is made. The scale is designed with three values. Pluss corresponds to a positive trait while the negative sign corresponds to a negative or unsuited trait. The equal sign can either have a neutral influence or an uncertainty associated with the ranking.

The table consists of four camera or sensor technologies. Lidar (A), Sonar (B), Depth Camera (C), Piezoelectric transducer (D), Spectroscopy (E) and X-Ray (F). These technologies are rated within seven criteria which are described as follows. Accuracy – the properties of the device can achieve a length measurement within a margin of 2 % and a confidence level of 95 %. In other words, the estimated length should be consistently ±4 cm of a real length of 200 cm. Noise cancelling – a system that can comprehend and reduce unwanted noise. The camera or sensor does not need to do the cancelling itself but get enough information so that another device can do the job. Complexity – possibilities for adapting the system to different environments. It should be robust and multi-optional for a high reliability. Cost – the price of a device that fulfills the basic requirements. It is intended for long-term industrial which means the cheapest or best device is irrelevant. Environmentally friendly – which in simple terms means a sustainable product with no toxic effects nor use of scarce materials. A plus means a direct positive effect. Other appliances – if the device can serve another purpose and be useful in the slaughtering process. Overall impression – a subjective opinion of whether it is customized for length measurement within a factory's environment.

Table 1. A comparison of 6 different technologies for the selection of the best length measurement system. Lidar (A), Sonar (B), Depth Camera (C), Piezoelectric transducer (D), Spectroscopy (E) and X-Ray (F). The plus, minus and equal sign correspond to a positive-, negative- or an uncertain/neutral trait, respectively.

Row	Criteria		A	B	C	D	E	F
1	Accuracy		+	+	+	=	+	=
2	Noise cancelling		+	+	+	+	=	=
3	Complexity		-	-	+	=	=	=
4	Cost		+	+	+	+	+	-
5	Environmentally friendly		=	-	=	-	=	-
6	Other appliances		=	=	+	=	+	+
7	Overall impression		+	=	+	-	-	-
		SUM	+3	+1	+6	0	+2	-2

Depth cameras has the highest value and lidars secondary. Both can be a useful tool for length measuring. The remaining technologies in the table are without a doubt very useful, but for other applications. Depth camera is nevertheless assumed to be the best option. Thus, the next section will elaborate some of its appliances.

2.5 Depth Camera Elaboration

A snapshot of a depth camera gives a picture that consists of one depth layer and one colored layer. To be able to estimate a specific length of a pig, a way of detecting it in the picture is important. When the pig is found, everything else on the picture should be filtered out, so that the body is the only thing left. The next operation is to turn the two-dimensional body into a three-dimensional point cloud, by using the depth information.

2.5.1 Recognition Types

There are many ways of finding an object in a picture. Table 2 show five different processes which are further elaborated in this subsection.

Table 2. Classification of different ways of identifying, locate and segregate objects.

Types of recognition	**Specialization**	**Models**
Image Classification	Lists what is in the picture	
Object Localization	Locates where an object is	
Object Detection	Combination of the types above	Faster R-CNN, YOLO, SSD
Semantic Segmentation	Highlights the specific pixels, edges and lines of an object	U-Net
Instance Segmentation	Divides every individual object	Mask R-CNN

Object detection

One way of finding a pig in a picture is to use an object detection system. Most of the existing detection systems use the information from a color image. The separation process of distinguishing objects from each other is based on changes in color, lightness, structure, etc. The result of analyzing a pixel to the surrounding pixels is to detect patterns. An example of pattern can be edges, lines, shapes, corners, circles, squares, textures or objects. These patterns are merged into a complex algorithm. After the processing of data, the system should be able to detect objects. Figure 7 shows a possible outcome from an object detection system.

Furthermore, a depth camera records overlapping color- and depth images. If the object is detected, it is possible to convert this area into a 3D image. The conversion is done by turning all pixels inside the frame into cartesian coordinates. As a result, a point cloud can be made. With this point cloud it is easier to estimate sizes and have a look at different angles. The advantage of having a depth camera is not only to measure depth and length, but also to remove background noise. Consequently, the contour of an object then becomes more visible.

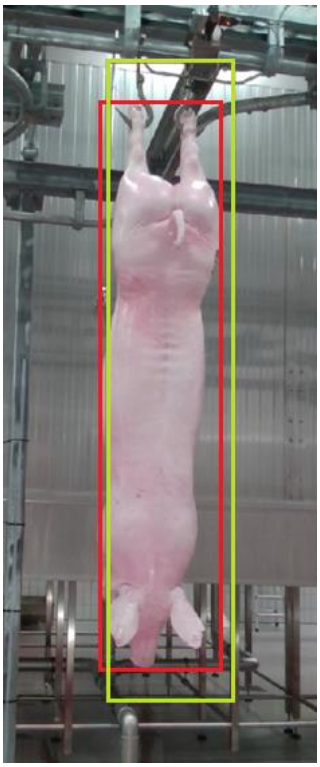


Figure 7. An illustration of the object detection, where the green frame is the proposal and the red frame is the benchmark.

Segmentation

A segmentation process is described as a division of multiple areas. These areas are separated by coloring and are classified as one type of object. An example may be a picture of a cat. The picture has 50 sections of which 20 makes the whole cat. All the cat-like-sections are merged into one section. If the sections are divided correctly, it is possible to only see the cat. As a result of high consistency and accuracy, other filtering processes are superfluous. The distance and the properties can more easily be found. Xie's report [32] from the John Hopkins University gives various examples of segmentation tasks.

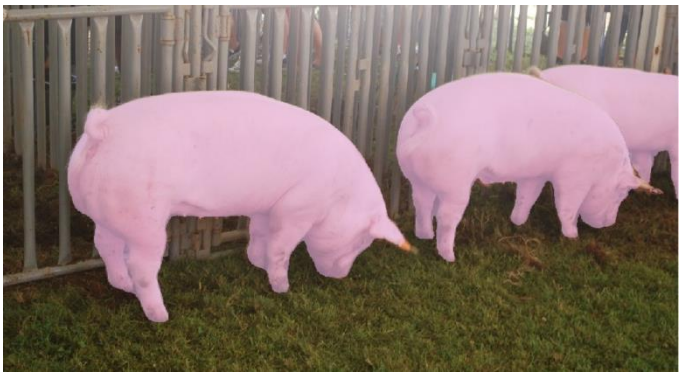


Figure 8. An image of three pigs where the outline is manually segmented. The segmented layer is represented as a pink color and is barely visible due to a precise work. [4]

Semantic and instance segmentation is a precise method of determine whether a pixel is inside an object of not. The only difference of the two types, is that instance segmentation distinguishes between individual objects of the same kind. A segmentation process is in comparison to object detection much more time consuming. Additionally, the procedure of pinpointing every pixel within an object is difficult. Lines and edges around the object are manually carved from the background, see figure 8. The purpose is to make sure that the neural network comprehends the properties of a desired object. A large dataset of many segmented objects improves the skill level to separate objects from the surroundings. Furthermore, the higher quality of annotations, the more accurate is the result. Advantages with a segmentation model is to separate overlapping objects. In comparison to the object detection, the segmentation process is filtering everything except the desired object.

2.5.2 Anchor- and Bounding Boxes

The detection technologies mentioned above use anchor boxes. An anchor box is a suggested bounding box, which is a pinpointed parallelogram, see figure 9. To be able to detect one or more objects within an image, it should contain multiple of boxes. These boxes can be strategically and randomly spread around a picture or be a part of a complete grid. It is important to have overlapping boxes so that parts of an object are not secluded.

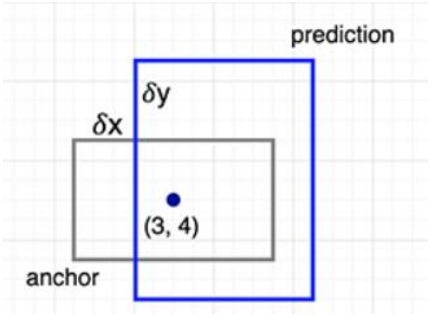


Figure 9. A model has an impression that an object is located at a point and surrounds it with an anchor box. [5]

To detect different shaped objects, the anchor boxes are also constructed to have different rectangular sizes. High and wide boxes are important to for example detect humans and cars, respectively. Every box has an algorithm to search for objects. The algorithm is based on probability with values between 0 and 1. A high value assures the algorithm to be more confident of its prediction. Especially when nearby anchor boxes have similar values.

IoU (intersection over union) is a method of finding how much an object is intersecting with the given anchor box. One anchor box can contain multiple objects. For every assumed object there are a corresponding probability factor or confidence score. When two or more objects have a higher probability factor than a threshold, the anchor box splits into the number of objects that were detected. Due to the recognized object, the box is centered on that object. This process is repeated and hence detect big and small objects in including intersecting objects.

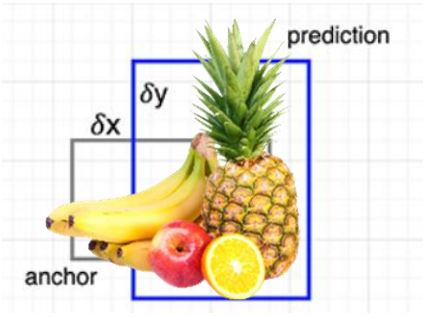


Figure 10. The anchor box gives a confidence score of the fruits, and then predicts the location of a pineapple. [5]

Figure 10 shows an example of the IoU method. The anchor box is giving a confidence score of a banana (0.94), a pineapple (0.89) and an apple (0.74). Since the threshold is 0.80 the apple is secluded. The banana and pineapple on the contrary are given a prediction of the location. In this example, only the pineapple is predicted. However, the prediction does not match perfectly. That is the reason why the process is repeated. [33, 34]

2.5.3 Models

A model is in this context the program that operates the object detection or segmentation process. This is an overview of the most applied models:

- CNN (Convolutional Neural Network)
- R-CNN (Region-Based Convolutional Neural Network)
- Fast R-CNN
- Faster R-CNN
- You Only Look Once (YOLO)
- Single Shot Detector (SSD)
- U-Net

CNN programs in general consist of three layers: Convolutional layer, pooling layer and a complete connected layer. The convolutional layer makes up the base layer which is the pre-processing par of object detection. Instead of converting RGB(D) code to a long string, this code is being processed via a matrix at the region of interest (ROI), see figure 11. Furthermore, the 3x3 grid is downsampled by interpolation, Kernel ([[101][010][101]]), max- or average pooling. With the calculated numbers the new picture usually becomes blurred. In contrary to the original picture, it is easier to detect edges from left, right, up and down. The grid near the edge is represented by a high number. At the final layer a program search for nearby numbers, and selects routinely every pixel above a given threshold. This defines the outline of an object or area in a picture. [35]

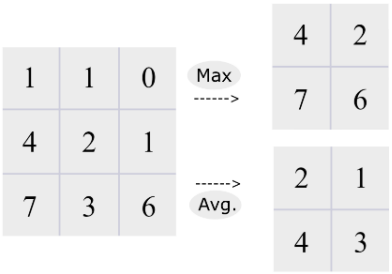


Figure 11. A simple illustration of how a picture can be processed. The grid with numbers represents a pixel with a color code. This grid is downsampled by max pooling and average pooling. Kernel = [3.2]

R-CNN divides the picture to over 2000 proposed regions, see figure 12. Every individual region is selected on the premise of the picture's distinctiveness. An object can therefore be a part of many regions, but it is also possible to have multiple objects within one region. R-CNN has a way of sorting this picture; every region independent of each other is examined to see if it looks like and fits into one or more categories. Furthermore, it connects regions within the same category. It also split regions with more than one category, as the region is in between objects. If many regions in a small area detect for example a car, the chance of this to be true is higher. This means also that the model comprehends overlapping objects. In later updates like Fast- and Faster R-CNN, the usage of RPN (Regional Proposed Network) increases the speed of finding appropriated regions. The time of finding an object is much

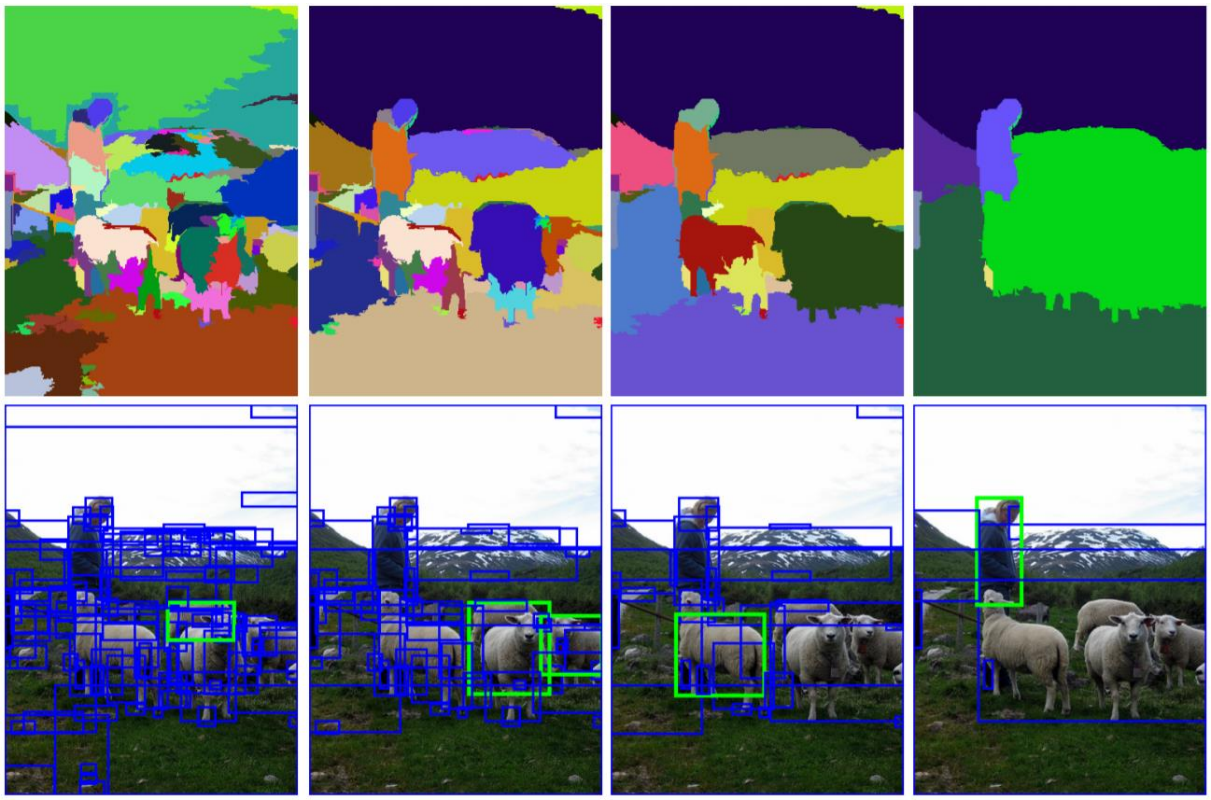


Figure 12. An example of proposed regions which are categorized and reduced in number. The R-CNN model distinguishes a person from sheep, the sky and the ground. [8]

less because the information from a pooled picture has less resolution. Hence, the useful information is not necessarily less. Mask R-CNN is an extended branch of Faster R-CNN which generates segmented masks for every object. [35]

YOLO. As CNN programs go back to pixels multiple times and separates regions after processing the picture, YOLO only look once. In a similar way as R-CNN, YOLO classify regions. While every region is selected, the regression of every box is done simultaneously. As of today, YOLO has three versions. Version 1 uses the information in a grid to determine where the bounding boxes are going to be placed. The consecutive version finds bounding boxes with help of anchor boxes. Real time appliances use Fast YOLO, which is developed from version 2.

YOLO9000 is also a separate version. Instead of using anchor boxes it operates a more complex dimension clusters (cluster centroids). The ultimate version does the prediction by logistic regression solved in a pyramid structure by downsizing the picture to more layers. If the score is in the upper half (0.5 – 1.0), the box prior is in front of another object. Two of the boxes must as a result somehow intersect with each other. [34, 35]

SSD. The Single Shot Detector are explanatory, and only takes one shot at the picture. The bounding boxes are at a fixed location and in size. Automatic placements of these boxes take less time than searching for good placements, hence RPN. On the other side, boxes placed in the background are unnecessary. Another case is when a highly dense area of the picture only has a few boxes. The detection will in that case not be successful. [36]

U-Net. U-Net is an intricate neural network that was developed for biomedical image segmentation. It has its name from the U-structure seen on figure 13. A simple way of explaining its application is to divide the U-turn in half. The left half consists of four operations. At every operation, the image is copied and sent to the opposing side of the "U". Afterwards, the convolutional layers are pooled, which means that the image deforms and get additional channels. As the resolution is halved the channels are doubled, but at the right half of the "U" this is vice versa. [6]

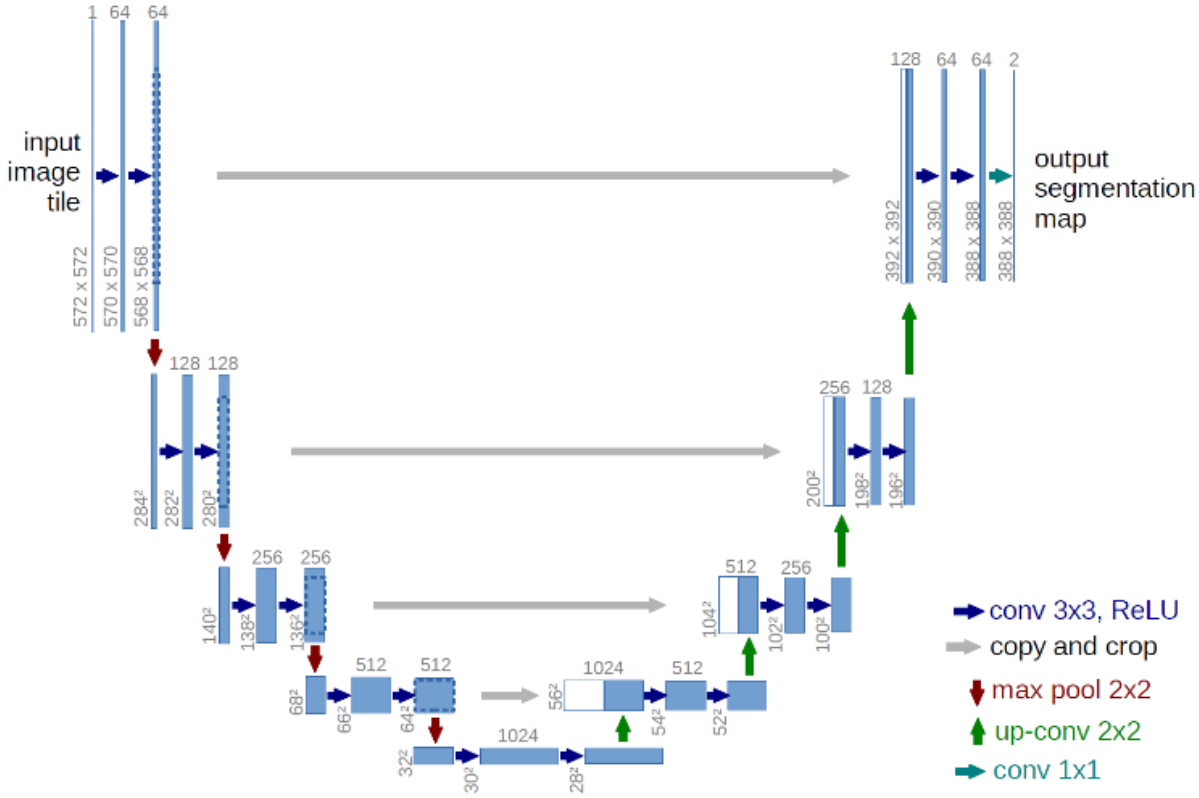


Figure 13. The architecture of a U-Net model, which illustrates the image processing from the input to the output. The model can find and segment a desired object in a picture. [6]

Comparison of performance

A fast detection algorithm is compromising the accuracy of detection. Models as Fast- and Faster R-CNN are two-stage detectors, while YOLO and SSD are one-stage detectors. One-stage detectors are therefore more applicable to real-time applications. See table 3 for the comparison. The resolution and the FPS indicates the processing capacity of the models. The mean average precision (mAP) is a score of their accuracy.

Table 3. A performance comparison of three models. The models were tested on the same premise, using the PASCAL VOC 2007 and 2012 with a batch size of one. [36]

	PASCAL VOC 2007		PASCAL VOC 2012		
Model	**mAP** [07+12]	**FPS**	**mAP** [07++12]	**Boxes**	**Resolution**
Faster R-CNN VGG-16 [37]	73.2 %	7	70.4 %	6000	1000 x 600
YOLO [38]	63.4 %	45	57.9 %	98	448 x 448
SSD300 [36]	74.3 %	46	72.4 %	8732	300 x 300

2.6 Length and Depth Calculations

In this thesis different methods are used for determining spatial proportions. For the depth perception, a depth camera uses triangulation and most sensors use TOF. Length estimation from a 3D point cloud uses the Pythagorean theorem. Without conversion via a point cloud, a poorer substitute can be a self-made equation.

2.6.1 Triangulation

Triangulation is a method for distance measuring, see the illustration on figure 14. If one of the three sides are known and two of the angles in a triangle, all the distances can be calculated. This principle is used in astronomy to find the distance to remote stars and galaxies, but also to find short distances (centimeters and meters). The formula is written as

$$D = \frac{L \sin\alpha \sin\beta}{\sin(\beta - \alpha)}$$ *Equation 2 [2]*

where D is the distance, L the length between A and B and the angle α (ΔCAB) and β (ΔABC).

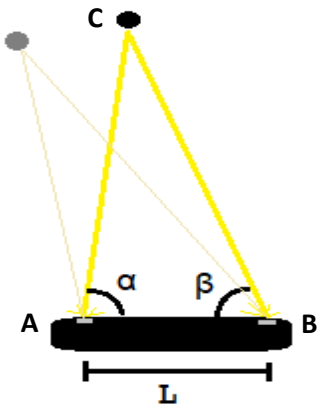


Figure 14. The black oval represents the depth camera, the black circle the object and the yellow lines the light trajectory of the two lenses. The distance from the camera to the object can be calculated.

2.6.2 Time of Flight

TOF is refered as a way of measuring distances when the speed of an energy source is known. The system needs a transmitter, reciever and a timer. A simple test is to measure how long time a sound takes before the echo returns. For example, you stand in front of a rock wall and make a short scream. After 2 seconds the scream reappairs. The travel distance is 686 m, because the speed of sound travel at 343 m/s. Since the sound waves must travel back and forth, the distance is halved, hence you stand 343 meters from the wall.

A comparison between the uncertainty of triangulation and TOF is showed in figure 15. TOF has a constant error while triangulation has an exponential error as the distance to the target increases. On the contrary, a field based approach is to find distance based on the energyoutput from a source. The intensity decrease with respect to distance is usually constant, and so calculation are quite simple. [2]

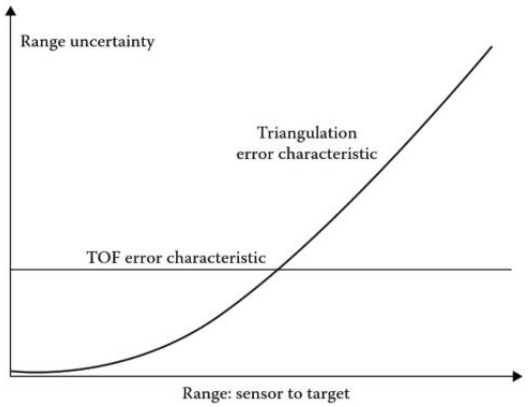


Figure 15. The uncertainty level of triangulation and TOF with respect to the distance to a target. [2]

2.6.3 Pythagorean Theorem

The Pythagoras theorem is commonly used to determine one the sides of a right triangle. Hence, the hypotenuse or length L is defined by

$$x^2 + y^2 = L^2$$ *Equation 3 [39]*

where the x and y represent the legs. In a spatial dimension, however, the distance between two points can be rewritten as

$$(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2 = L^2$$ *Equation 4*

and the location of point 1 and 2 is respectively (x_1, y_1, z_1) and (x_2, y_2, z_2).

2.6.4 Customized Height Formula

Equation 5 describes the height conversion between an electronic image and a real-life object. It is based on a test with one camera and one known length which is recorded at multiple distances. By comparing the number of pixels and the real height with respect to distance from the camera, the relationship can be defined. Furthermore, the correlation is expressed with a constant k. If the

correlation is exponential it should be k^D. This equation is limited to one specific camera and image setting and is not recommended for other instances. The height is as follows

$$H = \frac{D * n}{k}$$ *Equation 5*

where D is the depth, n is the number of vertical pixels and the constant k = 9 for this system.

2.7 Concept of Application

The starting process of finding a suitable object detection and measurement system is described in this section. However, an accurate and consistent segmentation process can substitute most of the applications. Some of the following concepts on the contrary, might be of interest if the segmentation does not work. A D415 depth camera was used for showing concepts within object detection. The python script was inspired by Dorodnicov [40]. In the following pictures, the color illustrates the distance to the camera objective. Black color is put to default at 0 meters, blue is the closest measurable distance while dark red is furthest away. The heading above each picture is part of the object detection and include the distance to the camera. See the explanation in the figure text.

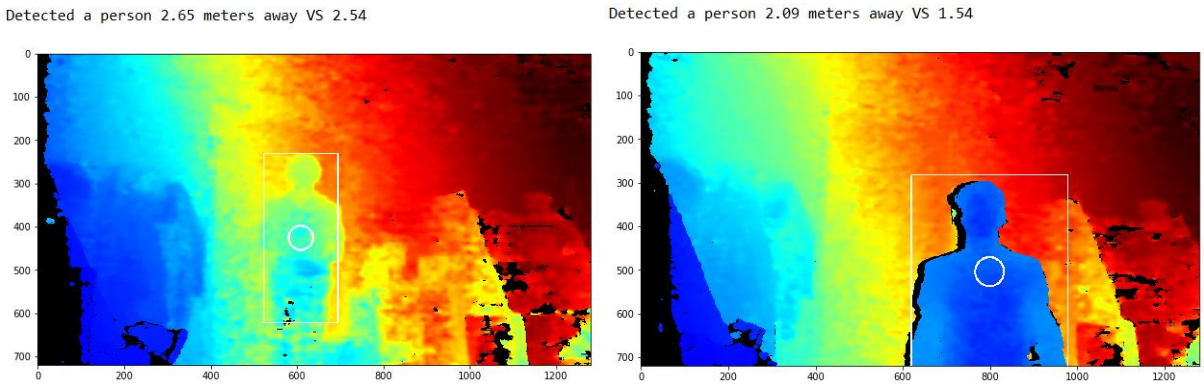


Figure 16. Two pictures taken with a depth camera and processed by an object detection algorithm. The distance from the person to the camera is calculated in the heading. First distance is an average from every pixel inside the rectangle versus the distance in the circle, both in meters. (A) The person is further away, (B) the person is closer. The axis show the resolution (1280 x 720).

Figure 16 include two pictures with a person at different distances from the camera. In the first picture there are not a big difference in the depth measurement because the person stands close to the door (A). The mean distance in the large rectangle is a little less than the mean in the circle, which is more accurate. In contrary, if a person is standing way in front of a background, the result is otherwise (B). The biggest differences in depth are when a person or object is much less than the big rectangle e.g. when a person is stretching out his arms (C), see figure 17. Exactly half of the area in the large frame is non-person, and are also much more behind. That is why the difference is bigger than the two

pictures above. In the last picture the person is holding both of his arms to the left (D), and this results in a not-centered frame. The little circle does not contain any pixels from the person and can be very devastating in the length measurement. It is therefore important to know the type of object to be detected. Another thing is what kind of surroundings are being captured by the camera; multiple objects, overlapping objects and other noise (E).

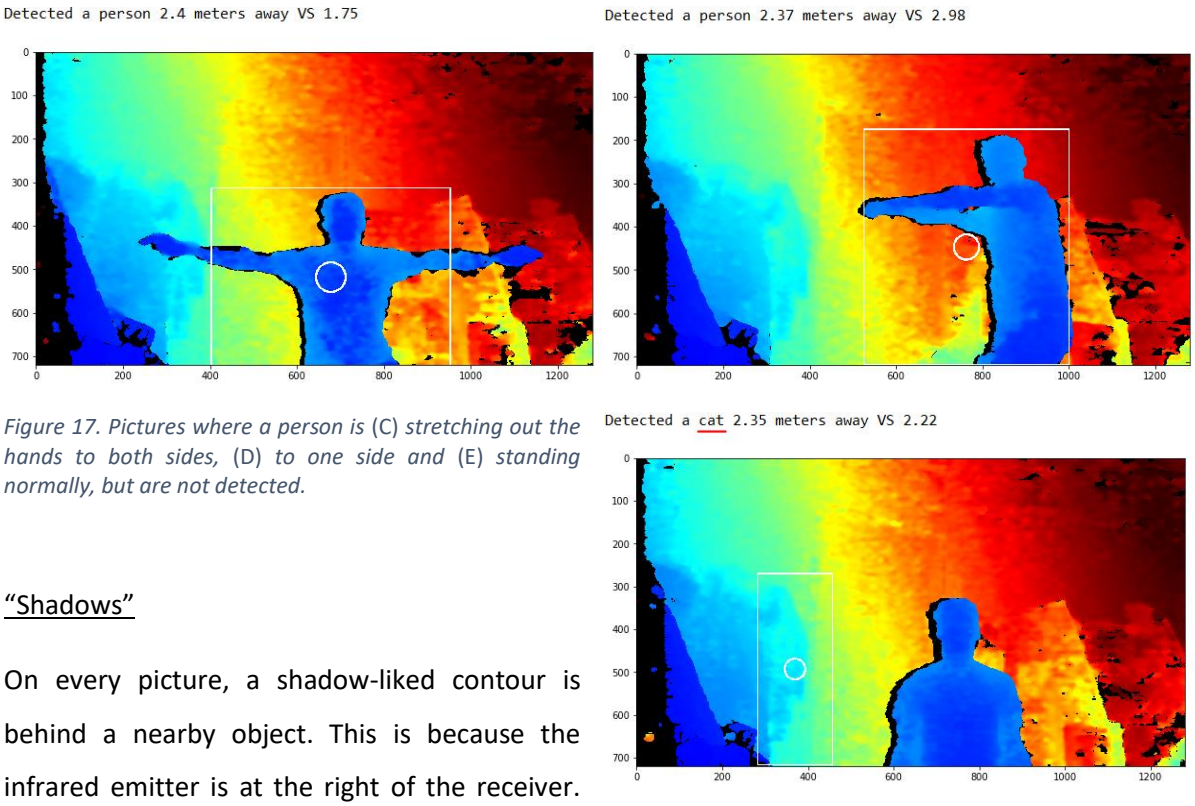


Figure 17. Pictures where a person is (C) *stretching out the hands to both sides,* (D) *to one side and* (E) *standing normally, but are not detected.*

"Shadows"

On every picture, a shadow-liked contour is behind a nearby object. This is because the infrared emitter is at the right of the receiver. The black shadow is represented by a length of 0 meters, which results in a systematic error. As the background influences the algorithm to get further behind, the shadow causes a closer result. A way of removing shadows is to set a threshold. 0.16 meter is the minimum depth of the camera. To remove the background a maximum threshold can be set. Another possibility is to make a histogram with all the distances in the bounding box.

The length measurement requires that the camera is perpendicular to the center of the object. If the camera sees the object at an angle, the amount of pixels in the picture is not taken into account. There are two ways of solving this problem. The first is to take an object's outline and make a line segment from one side to another, in a cartesian coordinate. Afterwards it is possible to calculate the length by using the Pythagoras theorem, see equation 4. The second solution is when the camera is stationary, and the angle of the object is known. By using this angle, it is possible to do a conversion from the apparent length to the real length.

Histogram analysis

To be more certain that the circle is in fact at the stomach of a person, a histogram of the pixel-depth inside the parallelogram is made. The average depth inside the circle should be placed right before a peak in the histogram, because the stomach is a little closer than the rest of the body. Figure 18 is a good example of the accordance between a person and the histogram. Inside the white frame the body is closest to the camera, which is indicated with a peak. The background on the other hand is more of a smooth noise. The y-axis is the amount of pixels.

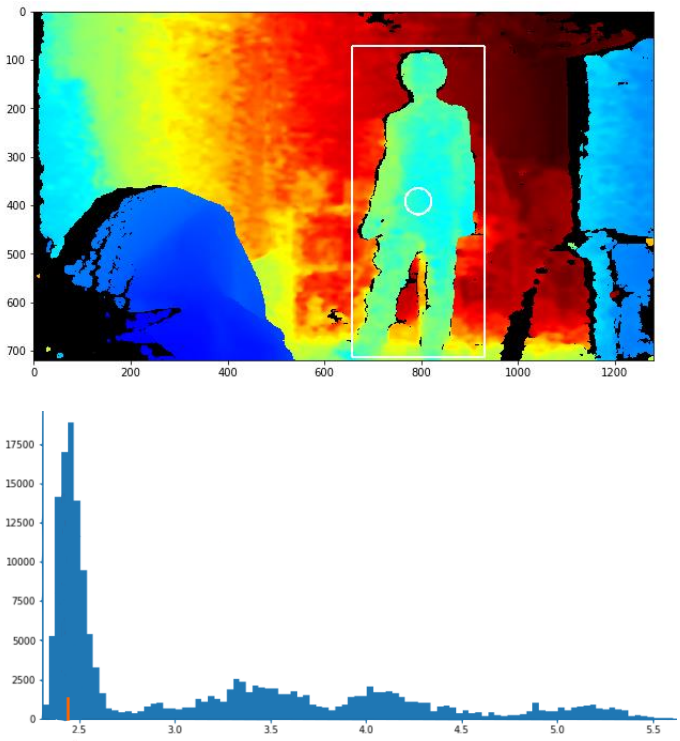


Figure 18. Object detection of a person with a corresponding histogram. The depth within the white frame makes up the histogram. X-axis equals the depth in meters and y-axis the number of pixels. The orange pillar is the average depth inside the white circle at 2.4 meters.

Multiple peaks

Other pictures might include objects and walls that also make the histogram peak. The inner circle would still be right before or at one of the peaks. If that is the case, there is usually no problem. Still, when objects overlap, has the similar depth, or is partly visible, the differentiation is much harder.

The shape of a histogram

A table or a square-shaped item (A, B) will be more visible at the histogram-subdivision than a tiny item (C), see figure 19. Objects with a flat surface will likewise be visible. The visibility is indicated by a clearly defined peak (A). If the camera is not directed perpendicular to the flat object or background wall, the so-called peak is more distributed along the x-axis (B). A chair has a lot of background noise, so it is harder to detect the acual chair (C). The first peak is the chair and the floor is the middle part and the wall/door is the biggest two peaks. The circle-distance is represented by the orange part of the histogram.

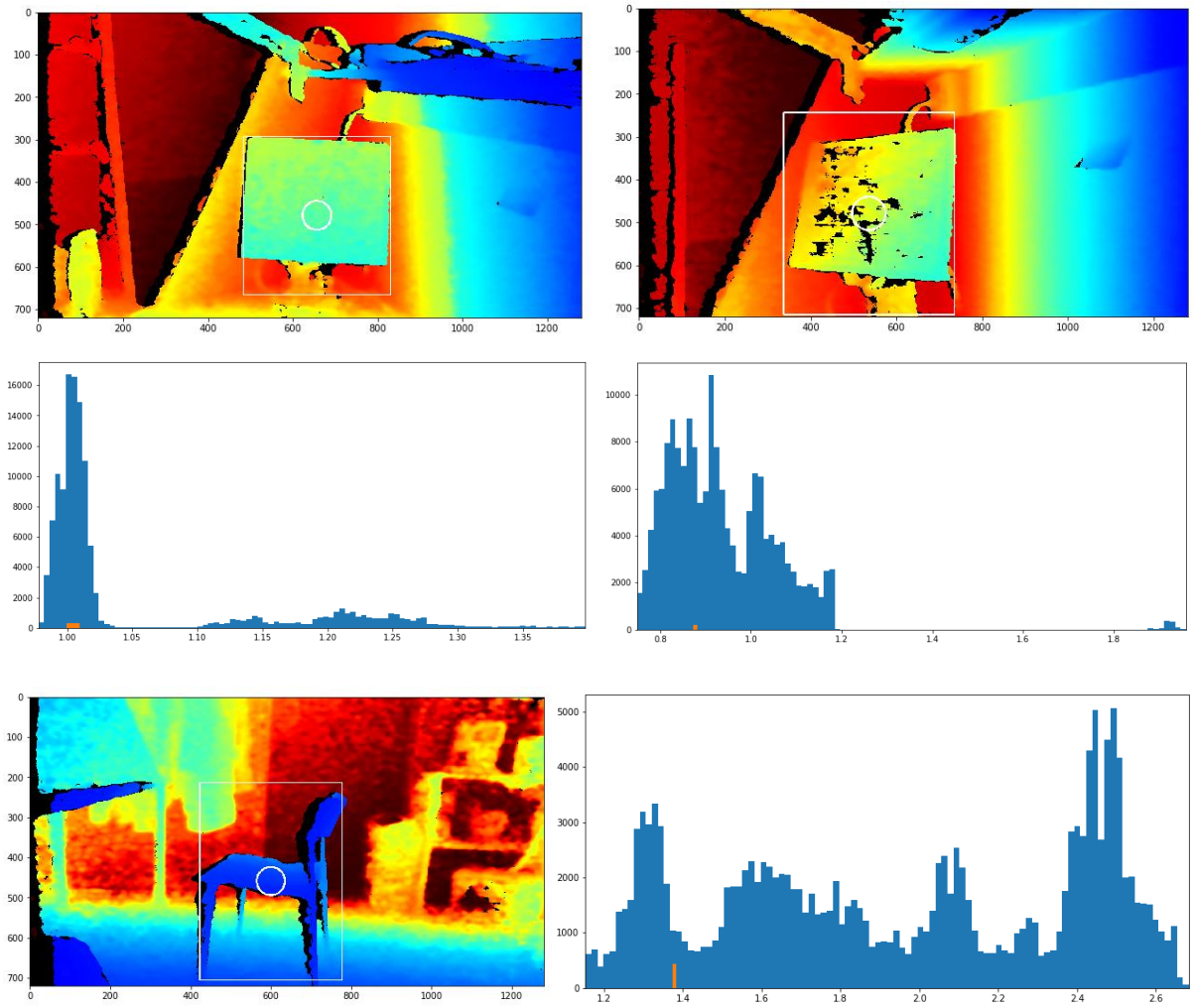


Figure 19. The first two pictures of a computer screen (A and B) with different angles has the consecutive histogram beneath. The last picture of a chair (C) has the histogram to the right.

Partially visible objects

If the purpose of object detection is to find the lengh of an object, the object should be fully inside the picture frame. When the bounding box is either near the upper or lower camera frame, the posibility of missing one part of the object is present. The camera can solve the problem by increasing the distance to the object, adjust more up/down or make the script approximate the length. The approximation is done by comparing the object with other similar objects. When other objects are in front of the detected object, the same principle can be applied. This method is however unaccurate due to the uncertainty of the object's missing part.

Minimum value detection

Minimum value detection is based on a condition of which the pig is closest to the camera. As the pigs are transported by a conveyor rail, the front legs are at a fixed point. The nose of the pig is the only unknown placement to calculate the length of the whole pig. A sensor can therefore be placed on the

floor underneath the pig. If the length between the sensor and the closest body part of the pig can be measured, then the length of the pig is determined. To avoid spilled blood on the sensor, it can be placed a little to one side, and above the ground due to the dripping and splashing. The length estimation can be corrected by including the angle from zenith to the nose. However, the uncertainty increases in proportion to this angle. Moreover, a camera on the floor is in general not very practical, especially not at a slaughtering line. It can affect the cleaning, contamination and space of which workers need to walk.

Placement of the camera

The location and aim of the camera are essential to the quality of the color and depth image. A pig can be a little more than two meters from top to bottom, in a hanging state. The applications for the D415 depth camera can be determined by testing or looking at the technical specifications. To detect a 2-meter object in height, the depth should be at least 2.5 meters away. On the contrary, a 1-meter object should not be further away than 8 meters. The margin of error and the distance are proportional because of higher resolution. Another source of error is when the distance to an object is varying, e.g. the stomach is closer to the lens than the head and feet. In this case the error and the distance are inversely proportional. For the most accurate measurement, the object should fill as much of the frame as possible. Furthermore, the camera and the middle part of the object should be in the same height.

If the camera is held in a portrait format, a tall object can be placed even closer. Comparison between the angle of width and height of the camera are respectively $72^{o} \pm 2^{o}$ and $40^{o} \pm 1^{o}$. In other words, the sufficient distance will be reduced to approximately 1.5 m. Because of narrow corridors, the turning of the camera might be a better solution. More precise properties of the D415 camera can be found in Carfagni's article [41] which supports the tested distances.

Adapting to a slaughtering line

The aim for the thesis is to have a length measurement system. Consequently, the depth information is used to determine the height and length. In figure 20, a person is standing in a semi-similar position a pig would hang. By using the self-made formula, see equation 5, the estimated height is a few centimeters beneath the real length. The height corresponding calculation is H = 3.56 m * (602 – 81) pixels / 9. However, the object detection frame is often too high or low. For example, the frame cuts the fingers. While the estimated distance is higher, the framed height is shorter. This makes the estimated height closer to the real height than it should.

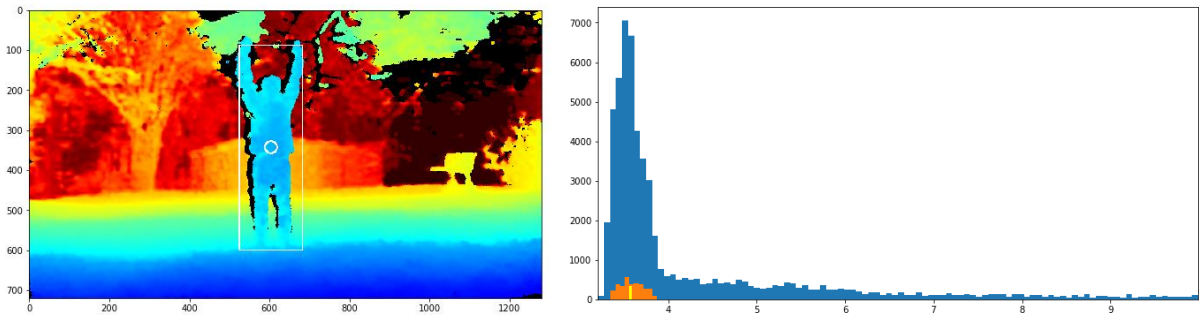


Figure 20. A test of comparing estimated depth (in the circle) versus the real depth. The results showed respectively 356 cm and 350 cm. By converting the depth to a height estimation, the result is 206 versus the real height of 210 cm.

3 Method

This is a length measurement system based on the recordings from a depth camera. A detection or segmentation model was needed to locate the pigs. Three models were chosen due to different characteristics: U-Net, Mask R-CNN and YOLO. They were compared to find the best solution. Furthermore, a process of converting the depth image into a 3D representation was needed. For this, the Computer Vision Toolbox was used in MATLAB. This program was also used to filter the background, highlight the pig and make an estimate of the length.

Figure 21. A setup for recording pigs at a slaughtering line with RealSense D415 Depth camera. This picture was taken after one hour of production, when the floor was completely covered with blood. The measuring tape was attached to the beam behind the pig.

3.1 Setup

For the setup, the surroundings had to be similar as possible to a meat factory cell. This was found in a slaughtering line for pigs. An Intel RealSense D415 Depth camera was set to record at Fatland Oslo AS, see figure 21. It was done on the day prior to the coronavirus restrictions. The recording should ideally be done when the system is complete. For the sake of safety and the need of data, the recording was done in advance of the finished length measurement system. In total, the duration of recording was around three hours. During that time 250 carcasses passed through the production line. To ensure

that the measuring system stays within the safety margin, a measuring tape was installed in the background. It was attached to a beam with cable ties for every 50 cm to make it more visible. A picture of the measuring tape was also taken at the same exact location as where the pigs went through for comparison. This was done in order to avoid contact with the carcasses. To be able to calculate the length of the pig, a few parameters are required: The distance from the camera to the line (225 cm), from the line to the measuring tape (225 cm) and the height of the camera (170 cm) was listed. However, the standings were adjusted a few times. It is worth noticing that the camera was turned sideways, since a portrait format was best suited for the object.

A box was connected to the depth camera. The box contained a recording program, storage card and an on/off switch. Along with the box, a computer was connected to verify that the recording system was running properly. A floodlight aimed at the pigs was placed below the camera. The shutter speed of a camera determines the blurriness which depends on the speed of the pig and the brightness reflected by it. The speed was very inconsistent as the pig was transitioning from one rail to another. In between the rails, at the intersection, the pigs had a short stop of different durations. Furthermore, the intersection was about 3 meters in length, resulting in a stoppage on various places.

The settings on the camera was set to 6 FPS (Frames per second) and two files were intended as output. A JPG file provided a standard image format with RGB colors. A DAT file on the contrary is a generic data file which contained the depth information. This file was not directly used but converted to a more readable PCD file. The PCD file was processed by a point cloud toolbox in MATLAB to make a visual 3D image. An overview of the D415 depth camera is given in table 4.

Table 4. The technological specifications of an Intel RealSense D415 Depth camera [42]. The range is dependent on calibration, scene and light conditions. Even if the depth- and RGB properties are different, they can be done equal.

Features	**Traits**	
Technology	Active IR Stereo	
Range	0.16 m – 10 m	
Rolling shutter	1.4 µm × 1.4 µm pixel size	
Dimensions	99 mm × 20 mm × 23 mm	
	Depth properties	RGB properties
Resolution	1280 × 720	1920 × 1080
Field of view	(65° ± 2°) × (40° ± 1°) × (72° ± 2°)	69.4 ° × 42.5 ° × 77.0 ° (±3°)
Frame rate	90 fps	30 fps
Open source software	RealSense SDK 2.0	
Processor	Intel RealSense Vision Processor D4	

3.2 Training a Model

A model to process raw data from a picture is needed for object detection and segmentation. There are essentially three steps to make a complete model: Data labeling, data preparation and neural network training. Supervisely is a webpage and a machine learning tool to easily train a dataset. The dataset should include an adequate number of pictures with pigs. A total of 202 pictures with 324 pigs were collected. 134 pictures were taken inside the factory in Fatland while 68 pictures were collected from internet. The first set of pictures were supposed to specialize in the exact surroundings including the pose of the pig. The other set were given a more general knowledge of what separates the pig from other objects. This set consisted of a variety of motives, both living and dead pigs and often multiple of pigs in one picture.

When the dataset was implemented to Supervisely, a manually and time-consuming task awaited. For object detection a rectangular frame was drawn around the pig, while a segmentation model requires carved outlines of the pigs, see figure 22 (A). The whole pig was highlighted with a pink color when finished (B). After this process a DTL (Data Transformation Language) was needed. DTL is an overview of the transformation from raw data to an organized model.

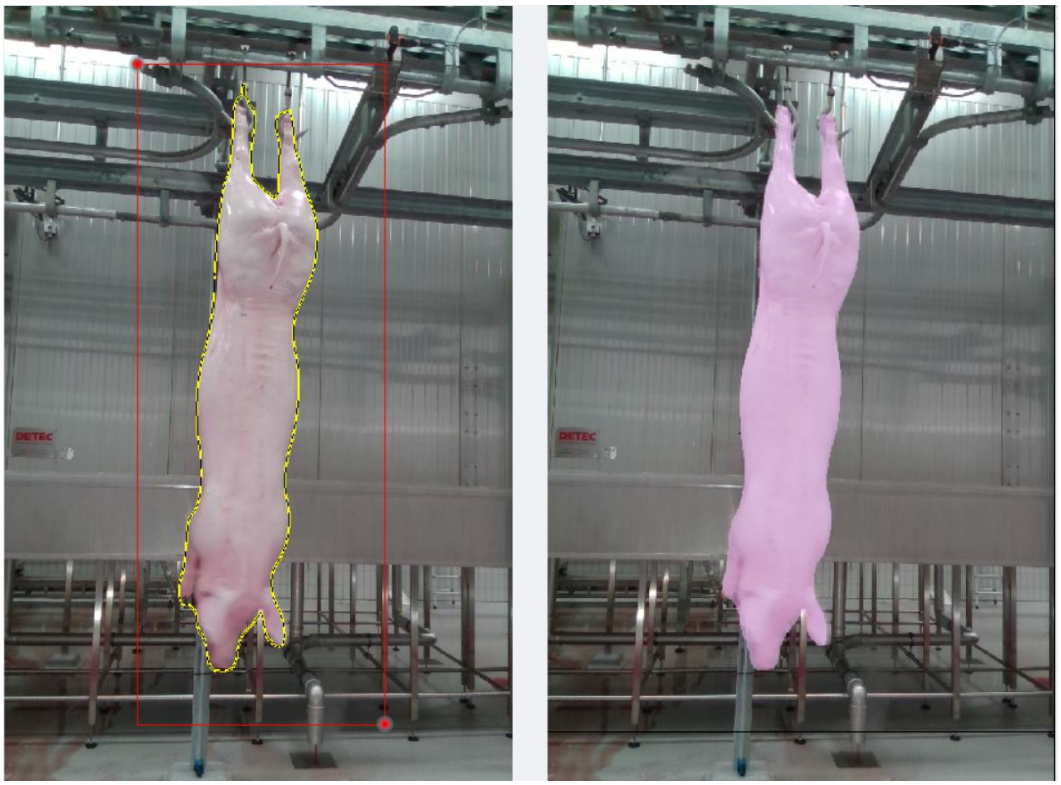


Figure 22. (A) Labeling the pig with respect to a type of model. Red rectangle is indicated for object detection and yellow outline for segmentation. (B) A complete version of a labeled pig.

The semantic segmentation was first tested. A prototype model was made with a separate set of pictures. It was supposed to find all the edges of the carcass to create the mask on the pig. After the

first test, the mask had a few minor flaws. To improve the model, the dataset was increased. A method for getting a larger dataset was implemented, without having to add and draw new pictures of pigs. The already preloaded pictures were replicated with distinct changes. A picture could be both rotated and zoomed in to a random area. Every picture was replicated 44 times. The final dataset had 8'888 pictures with 13'892 pigs.

In the last part of training a neural network was set to process the data. A U-Net model was chosen for segmentation purposes. After the final part of training the model was ready for use. Two models were done prior to the final model. Model version 1 used only the dataset with the internet pictures, while version 2 was the same as the final model but with less epochs. The full layout of version 3 is illustrated in figure 23 and consists of 8 steps.

1. Two sets of pictures were imported to Supervisely.
2. Manual segmentation of the pigs.
3. A DTL was inserted.
4. One dataset was made up by the pictures and the DTL script. A U-Net model with VGG weights was chosen (Visual Geometry Group, developed by the University of Oxford).
5. The dataset was trained by the U-Net model. Two additionally sets of unseen pictures were imported.
6. A validation set called "test-set" was made up by the pictures. The finished model was ready.
7. The test-set with new pictures was processed in the model.
8. The final result, where a segmented layer was more or less overlapping the pig.

This process was repeated two times, by using a YOLO V3 model and a Mask R-CNN model.

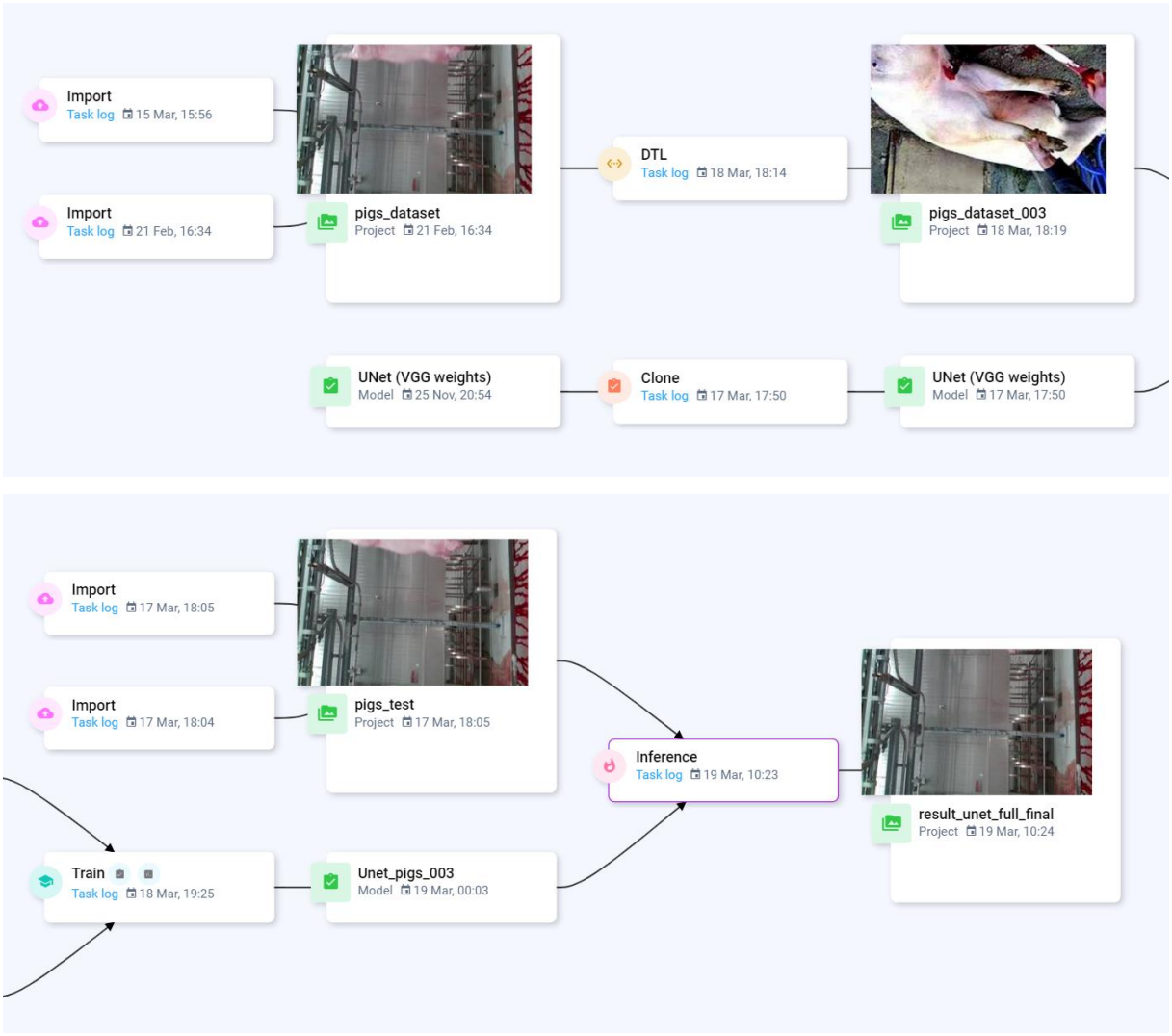


Figure 23. A layout of the full process at Supervisely. There are 8 steps in total, where the first four are at the top and connects with the bottom four.

3.3 Measurements

There are two methods for calculating the length of a pig. The first is to use the trained model to detect and outline the pig. Every pixel index value is pulled out of the outline from the JPG file and converted to the PCD. Only the depth pixels of the pig are included to render a new 3D image of the pig. Since the camera has one viewpoint, only one side of the pig is visible. When the highest and lowest point on the vertical axis are pinpointed, the length can be calculated. The Pythagorean theorem uses the x, y and z-values to find the length. This method was intended for use, but due to time constraints the second method was used instead.

The second method is to only use the point cloud. To be able to filter out the background and other noise, a rectangular prism surrounding the region of interest is set as boundary, see figure 24. The roof and rails will be inside the upper part of the region, due to a connection to the feet. To be able to set a reference point at the hind feet, a straight trajectory is set and illustrated in figure 25. The trajectory line is locked to the x- and z-axis (horizontal and depth) where the hooks are moving.

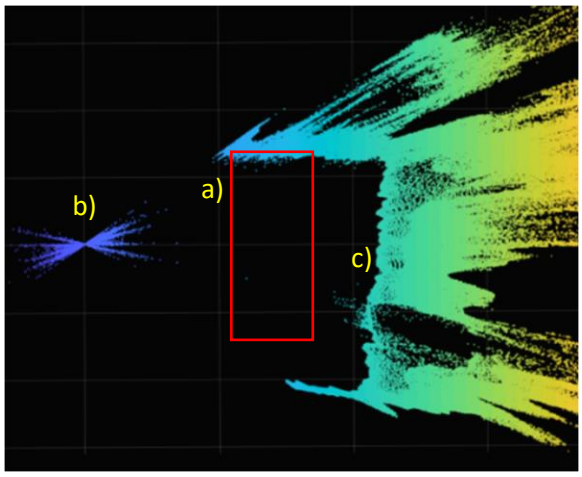


Figure 24. A side view of a point cloud image. a) the red rectangle is where the pigs are moving, b) depth camera, c) opposing wall.

One way of locating the feet along the line, is to take the average of all the pixels in a plane. To prevent any disturbances from the rails, the plane should be a few centimeters beneath the line. This way of pinpointing the feet is accurate even if the pig is swinging from one side to the other. An improvement can be to have multiple planes down the body. The average of each plane is inserted to a polynomial regression. At the intersection between the regression curve and the trajectory line are the feet. The nose on the contrary is easier to find due to an open space around the pig's head. It is found in the same way as the first method, by searching for the lowest vertical pixel. To prevent a measurement error from a remote pixel, the average of the lowest 30 pixels is used instead.

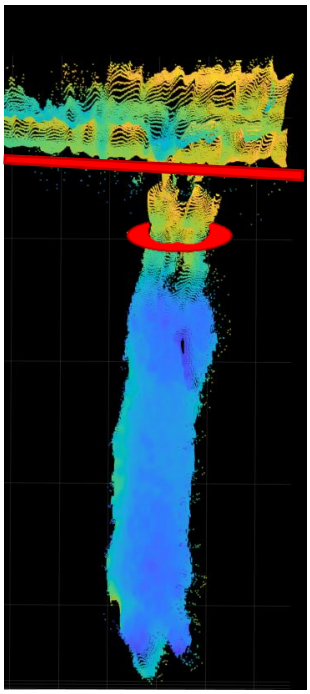


Figure 25. An illustration of how the trajectory and the plane visually looks on a point cloud.

The point cloud needs further processing. This includes to denoise the wrongly distributed pixels outside the surface of the pig. Two factors are used for this purpose. A threshold is set to eliminate a pixel if it has a long distance to the closest cluster of pixels. Furthermore, a number neighbor value (NumNeighbor) is the number of pixels inside the cluster or other nearby pixels. Higher values are synonymous with higher criteria, which eventually can remove important parts of the pig. Nevertheless, the process will both improve the image visually and the viability of the length measurement. The smooth leveling of pixels are done prior to the selection of the two extremes. Another way of improvement is to have a faster shutter speed or having the pigs to stop at the same place. This will result a much better image quality without a blurry pig.

To make a credible measurement, comparing the real length with the estimated length is necessary. Every parameter should be precisely measured in order to not cause any errors. The real length is calculated from the parameters measured at the factory. Admittedly, the parameters were measured prior to the camera adjustments and caused consequential errors. As a consequence, the system's accuracy was impossible to find. However, by adjusting the manual measurements, it was possible to neutralize this error. The precision could therefore be found instead of the accuracy.

The manual measurements were adjusted for every new camera location. For example, at the first location, the distance from the camera to the production line was the same as the distance from the production line to the measuring tape. From the camera's perspective, the pig's length would been equal to half of the tape's length, see figure 26. In this example the ratio was 0.5, but as the camera was moved closer to the production line, this ratio increased. Despite the inconvenience, the adjustment is a natural way of fine-tuning the relationship between the real and estimated length. Different calibration methods are also very common for measuring tools.

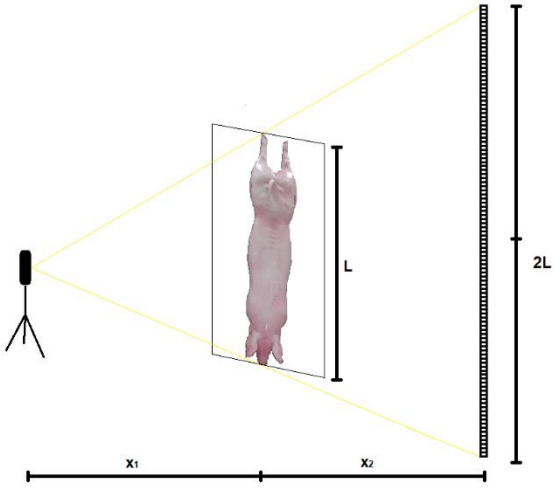


Figure 26. An illustration of how the real length is calculated. The ratio used to adjust this length with respect to the measuring tape is determined by the two distances x_1 and x_2.

4 Results

The result section is divided in two parts. The first section consists of a comparison of models which have been tested. Test results are illustrated as a contour in a picture. The second section show the measurements which uses data from the point cloud.

4.1 Compairson of Models

Three different types of models were selected: U-Net with semantic segmentation, Mask R-CNN with instance segmentation and YOLO version 3 with object detection. Every model was tested multiple times, see table 5 for the full overview. The different versions were built up by a dataset consisting of pictures from Fatland and a diverse mix from internet. As the number of pictures and epochs in each dataset vary, the result also differs. The best version from every model is included in this section while the rest is attached in Appendix - A. They are included on the basis of comparison and to get a deeper insight in how the neural network works. The validation/test-set is the same for all the versions, for comparison reasons. However, none of the pictures in the test-set are the same as the remaining datasets.

Table 5. An overview of the datasets, pictures and epochs used to make a model. The test-set differ from the rest of the datasets as it contains unique pictures. It is used for testing the models' performance.

Dataset	Fatland	Other	Total	Epochs	Best
Test-set	34	17	51		
U-Net $_{V0}$	134	68	202		
U-Net $_{V1}$	0	2'992	2'992		
U-Net $_{V2}$	5'896	2'992	8'888	2	✓
U-Net $_{V3}$	5'896	2'992	8'888	5	
Mask R-CNN $_{V1}$	1'608	816	2'424	2	
Mask R-CNN $_{V2}$	2'278	1'130	3'408	5	✓
YOLO V3 $_{V1}$	1'608	816	2'424	2	
YOLO V3 $_{V2}$	1'608	816	2'424	6	✓

Figure 27. A result from U-Net V1. It contains errors that are to be addressed in the discussion.

Four U-Net versions were made altogether. The first version was a prototype which was made to test the whole layout (neural network, cluster, DTL). Version 2 (V2) gave the best result and chosen to represent the U-Net model. Figure 27 is taken at the factory (Fatland). It contains errors which are found in all versions. For example the segmented area of blood on the floor and human skin.

Mask R-CNN version 1 uses the same segmented dataset as U-Net, but with only 24 replicas. The second version was added a thousand pictures with lighter and darker surroundings. YOLO detects a pig by framing it. Consequently, the pigs within the dataset had to be labeled with boxes and not segmentation. A few selected pictures illustrate the result which is shown in the figures below.

Figure 28 is a composed image of four identical pictures of a pig. The first image is the original while the other tree are the best version of each model.

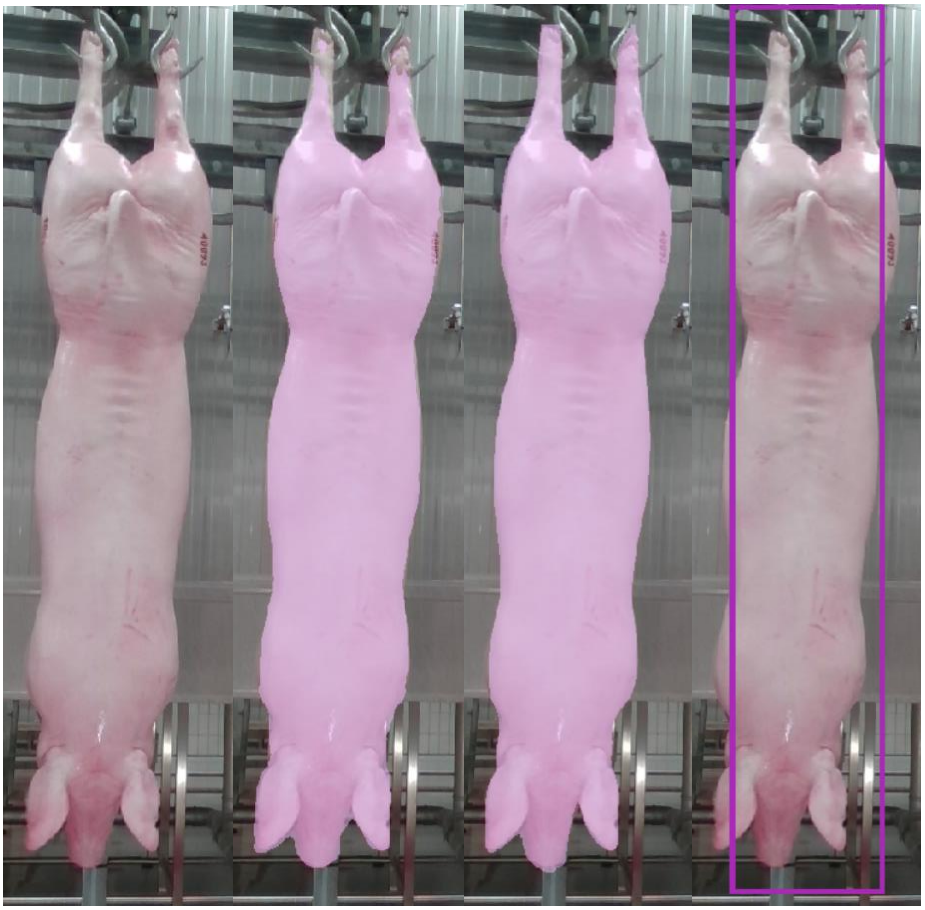


Figure 28. A collection of images put together. From left: the original image, U-Net, Mask R-CNN and YOLO V3.

The next set of image has the same layout as the previous but with a more blurry reference picture.

Figure 29. A collection of blurred images put together. From left: the original image, U-Net, Mask R-CNN and YOLO V3.

Last set of image is taken at NMBU inside the lab. The pig is laying at the operating table, which already is customized to its length.

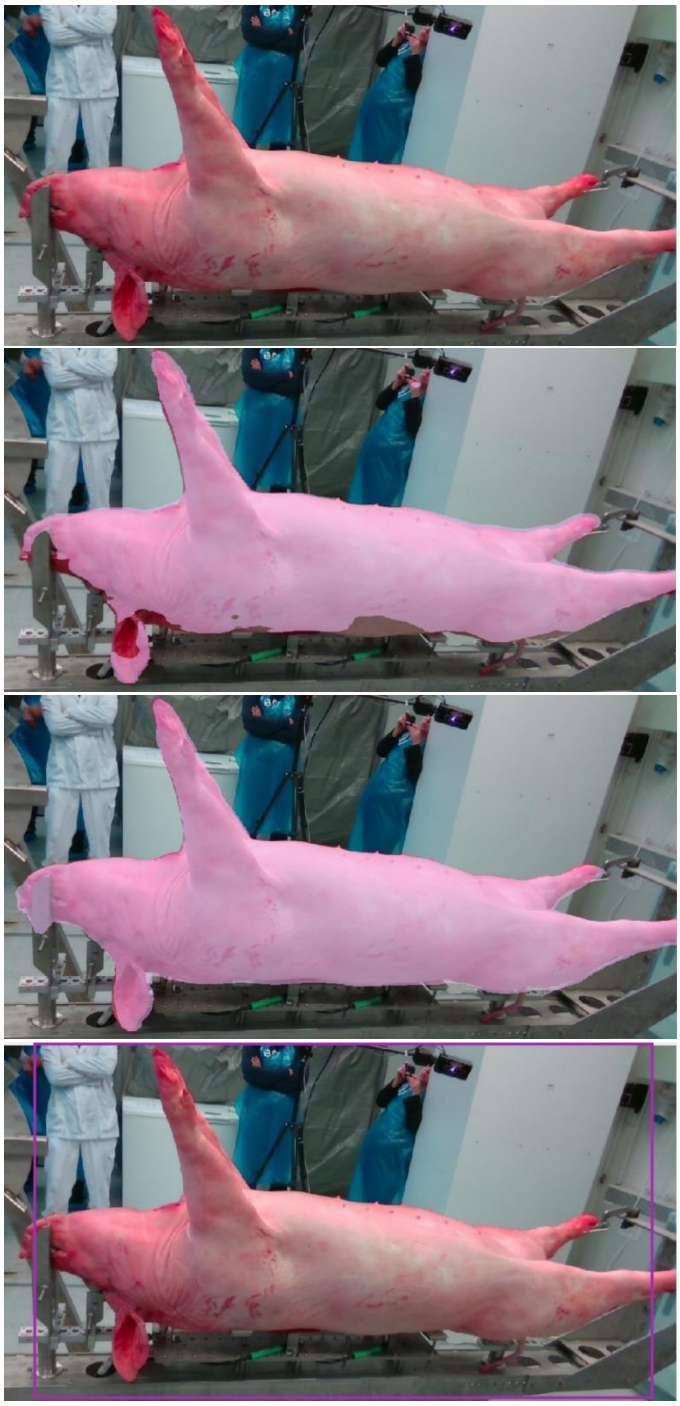


Figure 30. A collection of images from the lab. From top: the original image, U-Net, Mask R-CNN and YOLO V3.

4.2 Measurements Using Point Cloud

A full point cloud image from the factory is illustrated in figure 31. This image is filtered down to the ROI seen in figure 32. Only the points within ROI is used for measurements.

Point clouds made from one depth camera will solely cover one side of a pig. Consequently, the backside is unmapped. The perspective on the following figures is from the camera's point of view.

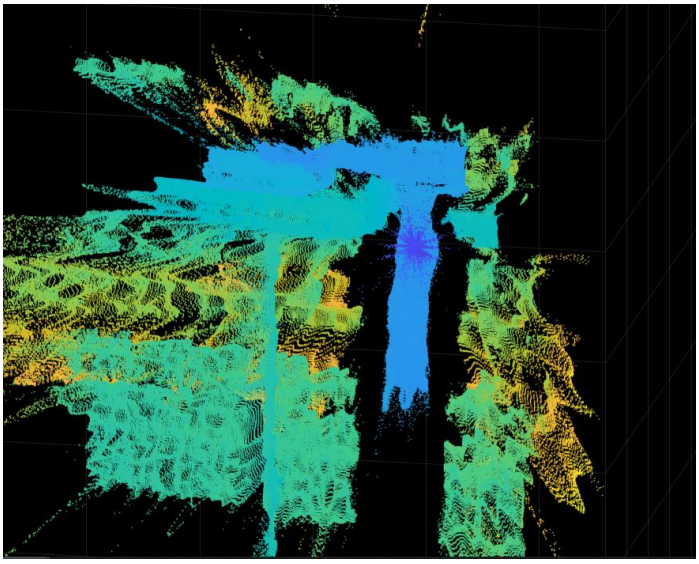


Figure 31. An unfiltered point cloud of the production line with the pig in the center.

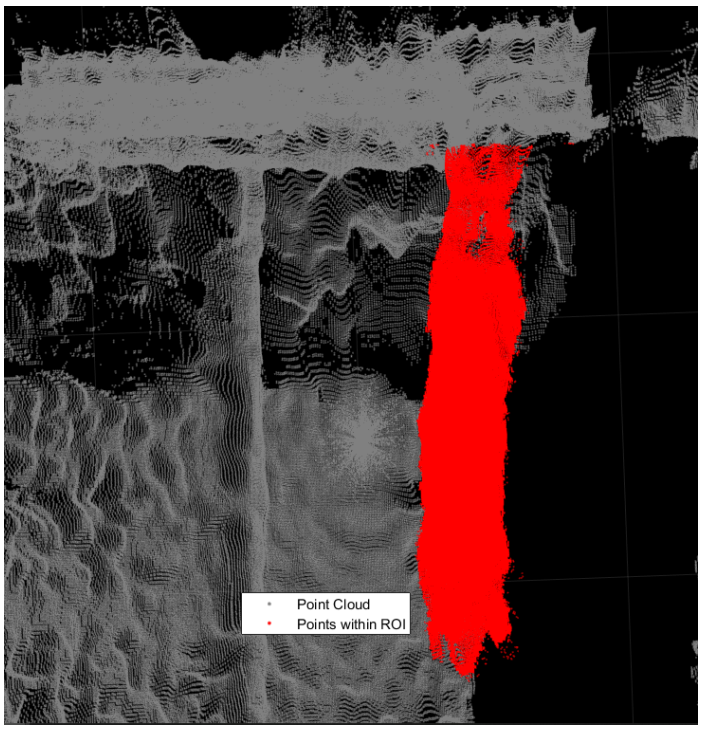


Figure 32. A highlight of the region of interest. The gray area is irrelevant and secluded from the point cloud.

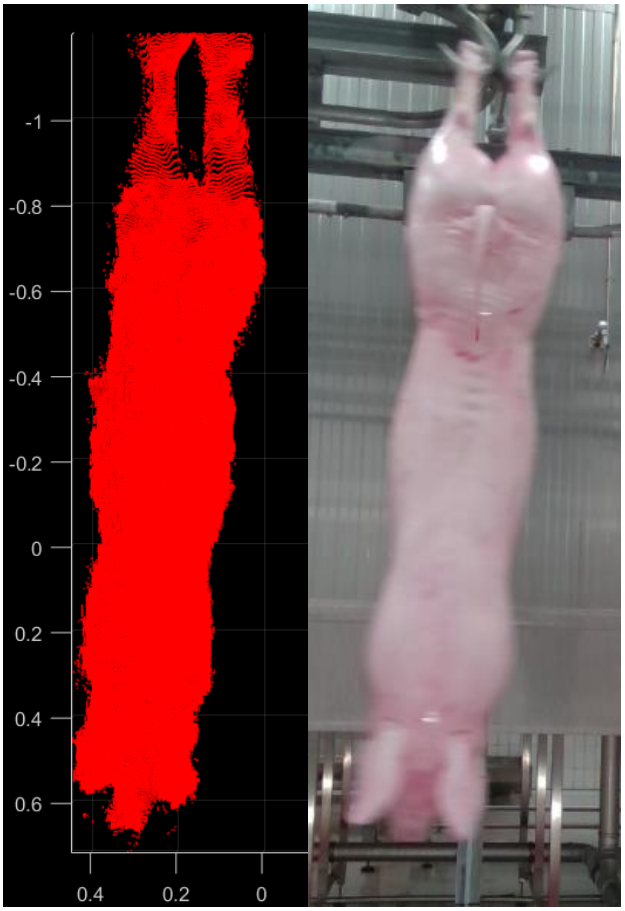


Figure 33. A point cloud and a picture lined next to each other. The axes are in meters, and the view section is from the x-axis and y-axis.

Figure 33 and 35 show the ROI and a matching picture. The length estimation is based on this final point cloud, and the axes are in meters. Moreover, the origin of the coordinate system is not in the bottom left corner, because the ROI is only a segment from the whole point cloud.

The overview of the real length and the consecutive estimation is referred in the Appendix - B. It is of less interest because the manual measurement had consequential errors. That is also why it was done so few manual measurements. Nonetheless, the length differences gives a clue of the precision, as the box plot in figure 34 shows.

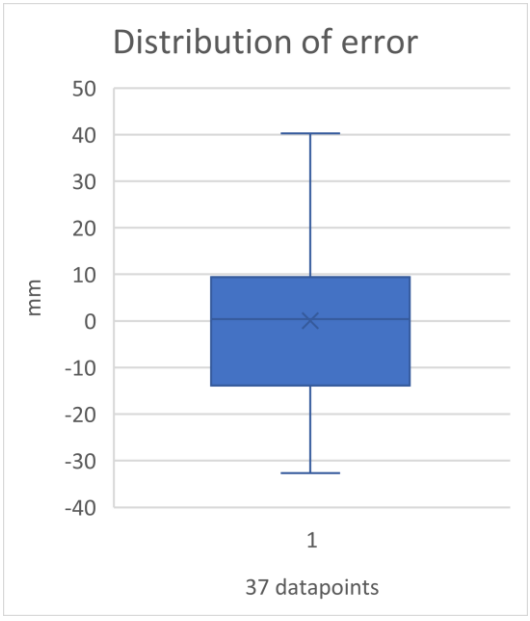


Figure 34. A box plot of the length deviation. In other words, the difference of the real and estimated values. Every value was adjusted with respect to the camera location. The values are taken from the "Difference" column in Appendix - B.

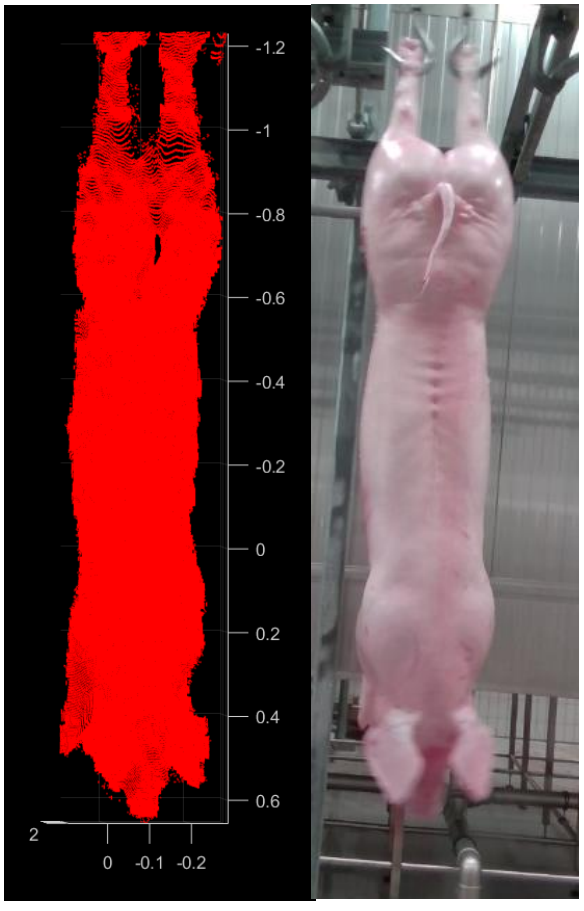


Figure 35. Another comparison of a point cloud and a picture.

5 Discussion

The first step to find a reliable measurement system was to address different technologies. A handful of sensors were considered, but a depth camera was decided as the best solution. Furthermore, a type of recognition system would allow a detection, location or segmentation of the pig. Three types of models were compared in order to get the best result. The benefit of this system is to be able to refine the image to only include the pig and convert the depth image to a point cloud.

Due to lack of time, the segmentation was not used to narrow the point cloud. Instead, a cuboid where the pigs were moving was constructed. A cuboid is an elongated box; because nothing else was inside this cuboid, the ROI was not influenced by anything. For this instance, the use of setting manual boundaries made no difference from using a segmentation model. After a filter process, the point cloud can be used for length estimation. It is worth mentioning that the measurement system is supposed to be installed at the production line and do a live estimation. Due to the corona situation, this was not possible. Nevertheless, the system performs in the same way regardless of whether the pictures are taken live or not.

5.1 Models

The reason for choosing two segmentation processes are simple. The advantages of using segmentation versus object detection, is the customization around the pig. When other objects are nearby, the point cloud largely contains dots at the surface of the body. On the contrary, when the frame of the object detection consists of more than the pig, the length measurement system needs further processing and filtration. At most factories, workers are the main source of disturbance. It is highly recommended to set restrictions to a small section, to reduce the risk of error. The difference between semantic and instance segmentation is absent when the pigs are passing by. However, the problem with semantic segmentation occurs when more than one pig is within the camera range. This means that the U-Net model cannot differentiate between two pigs. A likely outcome in such a situation will be a wrongly estimated length.

The first blurry picture at the result section include some errors, see page 34. It is not representative for the results as a whole but reflects how the neural network is operating. On the floor, beneath the carcass is a blood puddle. The puddle has a similar color as the carcass and is the reason for the wrongly segmented area. Human skin is another misinterpret area due to the coloring. Conversely, at the lab on page 37, the U-Net model did not segment areas with dark color, e.g. inside the ear. By implementing an even wider range of pigs, the algorithm can amplify the significance of shape and texture. For instance, pictures with gray, brown and black pigs, including a diversity lightning condition will improve the model.

5.1.1 U-Net Model

The U-Net model was very smooth around the mid-section of the pig at the factory. Places where the body and segmentation layer did not overlap, was at the feet, head and number tag. The feet were definitely the hardest area because of the interaction of the hooks, the variety of color and the thickness of the feet. At the head section the segmentation layer was slightly thicker. Nonetheless, to ensure a point cloud with every pixel on the pig, this layer should have a tiny margin. With the MATLAB script used in this thesis, the purpose of the segmentation process is to determine the location of the tip of the nose. The rest of the body is also important to pinpoint the uppermost part of the hind feet. Due to the issues associated with segmenting the feet, a trajectory was set to locate them more accurately. As a consequence, the need for a highly fitted layer became less important.

At the lab, the segmented layer was almost the opposite of the factory images. On one hand, the back and head have big chunks of flesh cut away from the layer. On the other hand, all the legs and feet are completely covered. Even if the approximation is good, the risk of estimating wrong is rather large because of the big flaws. As mentioned, the result could have been improved by increasing the size and variation of the trained dataset.

5.1.2 Mask R-CNN

Mask R-CNN was better than the U-Net at almost every level. The outline of the head is tighter and both legs are completely covered. However, blurry pictures have a tendency to enlarge the thickness of the legs and feet. Another central part of differentiating the two models is looking at the objects interacting with the pig. The interceding objects are hooks and muzzle located in the factory and the lab, respectively. The Mask R-CNN model is ignoring these objects in a greater extent, while U-Net is excluding them. This might be a reason for the difficulty in the segmentation process.

5.1.3 YOLO

YOLO was only detecting the pig and framing it. This model should therefore be applied in a different way than previous models. The result in the first collection of pictures on page 35 show a narrower width and a taller frame than the pig. Even though it is displaced, it does not cause a problem. As long the nose and the hind feet are within the frame. In other words, intersection at the sides is irrelevant for the length measurement. The information from the frame is compatible with the MATLAB script where the depth and height are already set. When the width also is determined, a cuboid around the pig is defined. For instance, an object detection model can detect multiple pigs. If the pigs are separated, and the cuboids are not overlapping, it is possible to estimate several pigs simultaneously. Other objects within the bottom part of the cuboid, however, makes it difficult to pinpoint the tip of the nose.

5.1.4 Comparison

All the models had significantly better results on the stagnant pig. As a picture becomes blurrier with moving objects, the same thing happens to the roughness of the point cloud. For this reason, it is important to have regular stops in the production line. It is also preferred to make the pigs stop at the exact same location for visual inspection and comparison. Besides, if this is possible, there is no need to locate the hind feet because they are at the same place. Now the only unknown point for estimating the length is the nose. The depth camera can therefore be placed much closer, focusing on the bottom part. Another way of improving the image quality, is to have depth camera with higher resolution, faster shutter speed and increase the brightness. To be more precise, it is the brightness reflected from the pig that improves the quality.

5.2 Point Cloud

A point cloud will to a certain extent have noise. It is however a desire to reduse the noise and assemble the dots alike the real object. Filtration methods and management systems can be used for further improvement. In the thesis, the filtration done on the point cloud was thresholding and use of NumNeighbor value. There are many possibilities, but the customization should be adapted to every environment. If a D415 camera is decided for further use, it might improve the performance by tuning the camera settings. A walkthrough of this is done by Grunnet-Jepsen [7]. It might also be other cameras that are better equiped to fulfull the requirements.

The majoraty of the point clouds had only minor errors. The dots around the nose constitute the biggest error for the length estimation. By a visual inspection, the biggest error caused by a single dot is approximatly 5 cm. Because of displaced pixels, a method of averaging the bottom 30 pixels was made. As a result, the 5 cm error changed to less than 1 cm.

The point cloud covers only one half of the pig. If two or more depth cameras had been installed at different angles, the whole body would be charted. The measurement uncertainty would be increased a bit; however, other applications could benefit more from a complete point cloud. It is for example possible to find the volume, mass and body shape.

5.2.1 Real Length Versus Estimated Length

The camera took pictures as the carcasses went across the screen. By manually measure the length of a carcass while knowing the dimensions of the setup, it is possible to calculate the real length. The real length is furthermore compared to the length estimation done by the system. Since the camera was adjusted while running, the real length was adjusted afterwards. In retrospect, the measuring should have been done in between and after the adjustments. Due to this, it was not worth increasing the number of samples. A total of 37 datapoints were used for comparison. The only indication of the

system's skillset, is the difference of the two lengths. A handful of comparisons were made, and the distribution is showed in the box plot on page 39. The extrema displayed in the box plot was ±4 cm. However, 25 out of 37 measurements were within the range of ±1.5 cm, hence the system's precision was 68%. To increase the reliability of the system, the distances at the setup should be measured more precisely. An even better method is to measure the length of the pigs manually, by contact. This is of course only for testing and requires the right permissions.

By comparing this error with the RMS error of another D415 camera on page 11, the error is five times higher. With a distance of 225 cm to the pig, the RMS error are approximately 7.5 mm. If it is possible to record at a closer distance without a risk of contamination, this should be done. As the distance shortens, the error will decrease. This statement is supported by the characteristics of triangulation and the RMS curve. The referable distance becomes from 150 cm since the pig must be within camera range.

A lot of the pigs were estimated multiple times. The results showed a tendency of which the length decreased gradually as the carcasses were moving to the right. This error can be caused by two factors. Firstly, the trajectory of the feet could have been wrongly inserted. Secondly, if the camera was tilted obliquely, the rail is closer on one side. Thus, the error rate is likely to increase accordingly to the distance. In both cases, if the pigs were photographed at the same place, the inequalities will cease to exist. Besides, there is no need for a trajectory since the feet are fixed in one location. As a consequence, the error would most likely be considerably lower than the extremes of ±4 cm.

A robust measurement system provides a long-lasting and consistent device for industrial use. Furthermore, the device should be installed at a location without the risk of spilled blood or contamination. At Fatland, the camera was installed about the same height as the pigs, over two meters away. There was no interaction between the camera and the pig. It will also remain that way as long they are in the same level. In contrast, sensor technologies with unidirectional signals needs to be placed underneath the pig, exactly where the contamination zone is. Since the camera is small and placed relatively high, it is barely noticeable. This underlines the durability and robustness of the system. Ultimately, the installation and the system as a whole are considered suitable for length measurements in a slaughter line.

6 Conclusion

6.1 Results

A D415 depth camera was found to be the best application for a length measurement system. However, it brought more heavy operations to the process. The first step is to detect and locate the pig on a picture. This can be done by an object detection- or a segmentation process. Based on the content within the different models, the Mask R-CNN gave best results. Consequently, an instant segmentation process was the choice of method. When the pig is located, the corresponding pixels on a depth image are selected. Furthermore, these pixels can be converted to points in a point cloud. The intention is to only include the pig's surface.

Instead of using the location from a model, the whole depth image was turned into a big point cloud. After seeing the trajectory of the pigs, a cuboid was manually constructed to erase the external area. The next step is to filter the remaining point cloud to correct the reassembly of the pig. Then the nose and hind feet are pinpointed, so that the length between them can be calculated. Finally, the length measurement system is ready.

The comparison test was executed with the estimated and real length. An adjustment on the camera standings contributed to a consequential error. Due to this and very few samples, the system's accuracy is not very credible. All of the measured values, however, were adjusted subsequently. With that in mind, this measurement system gave a 4 cm total error. To summarize, the length measurement system is almost complete, but needs improvement.

6.2 Recommendations and Further Work

It is room for improvement in every section of the measurement system. The improvements in this subsection follow a chronological order and focus on decreasing the margin of error. Firstly, the size of the dataset plays a big role in the segmentation process. 200 unique pictures are in this context is few. By increasing the number of pictures and have a variety of pigs and backgrounds improves the quality of segmentation. The lighting, color and position of a pig will also make an impact. Furthermore, the manual drawing of every pig's outline can be more exact. By doing this, the model's algorithm will have a better basis for separating the pig from everything else.

Secondly, it would be ideal to install the camera where the pigs stop. Especially if they stop at a fixed location along the production line. This would improve the accuracy of the length estimation due to better image quality. It would also be easier to determine the distances from the camera to the pig. An even further improvement is to implement a system that accommodate the pigs' movement. Another aspect of reducing error is to find a good filtration process to avoid remote pixels and unnecessary noise.

Finally, the comparison between the real and estimated length should be more comprehensive. A large-scale test could generate a normal distribution curve of the margin of error. This statistical evaluation is necessary to determine the skill ceiling.

7 References

Primary Sources

2. Webster JG, Eren H. Measurement, instrumentation, and sensors handbook: spatial, mechanical, thermal, and radiation measurement: CRC press; 2016.
3. Morris AS, Langari R. Measurement and Instrumentation: Theory and Application: Academic Press; 2016.
6. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. Springer. 2015;9351: 234--241.
8. Uijlings JRR, Sande KEA, Gevers T, Smeulders AWM. Selective Search for Object Recognition. International Journal of Computer Vision. 2013;104.
9. Food and Agriculture Organization of the United Nations. Meat Market Review. Overview of global meat market developments in 2018 2019.
10. Ritchie H, Roser M. Meat and Dairy Production. Our World in Data. 2020.
13. Sánchez R, Verónica S. Models under uncertainty to support sow Herd management in the context of the pork supply chain: Universitat de Lleida; 2010.
14. Alvseike O, Sverdvik H, O'Farrell M, Berg P. Meat Factory Cell – A concept for the future? 2017.
16. Alvseike O, Prieto M, Torkveen K, Ruud C, Nesbakken T. Meat inspection and hygiene in a Meat Factory Cell - An alternative concept. 2018.
17. Brandla N, Jørgensen E. Determination of live weight of pigs from dimensions measured using image analysis. 1996;15(1).
18. Menesattia P, Costaa C, Antonuccia F, Sterib R, Pallottinoa F, Catillob G. A low-cost stereovision system to estimate size and weight of live sheep. 2014;103.
19. Tasdemira S, Urkmezb A, Inalc S. Determination of body measurements on the Holstein cows using digital image analysis and estimation of live weight with regression analysis. 2011;76(2).
20. Matthews SG, Miller AL, Ploetz T, Kyriazakis I. Automated tracking to measure behavioural changes in pigs for health and welfare monitoring. Scientific Reports. 2017;3.
21. Lee J, Jin L, Park D, Chung Y. Automatic Recognition of Aggressive Behavior in Pigs Using a Kinect Depth Sensor. Sensors (Basel). 2016;16.
24. Palmer G. OCM review and Strategy Development. Meat and Livestock Australia; 2015.
25. Psota ET, Mittek M, Pérez LC, Schmidt T, Mote B. Multi-Pig Part Detection and Association with a Fully-Convolutional Network. Sensors. 2019.
26. Purnell G. Robotics and Automation in the meat processing. 2013.
27. Day RA. The Origins of the Scientific Paper: The IMRAD Format. American Medical Writers Association Journal. 1989;4(2):16-8.
29. Howard IP. Perceiving in Depth: Oxford University Press; 2012.
31. Pugh S, Clausing D. Creating Innovative Products Using Total Design: The Living Legacy of Stuart Pugh Hardcover. 75 Arlington Street, Suite 300 Boston, MA, United States: Addison-Wesley Longman Publishing Co., Inc.; 1996.
32. Xie C, Wang J, Zhang Z, Zhou Y, Xie L, Yuill A. Adversarial Examples for Semantic Segmentation and Object Detection. 2017.
33. Zhong Y, Wang J, Peng J, Zhang L. Anchor Box Optimization for Object Detection. 2018.
34. Redmon J, Farhadi A. YOLOv3: An Incremental Improvement. 2018.
35. Dixit KGS, Chadaga MG, Savalgimath SS, Rakshith GR, Kumar MRN. Evaluation and Evolution of Object Detection Techniques YOLO and R-CNN. International Journal of Recent Technology and Engineering (IJRTE). 2019;Volume 8(2S3).
36. Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C-Y, et al. SSD: Single Shot MultiBox Detector. 2016.

37. Ren S, He K, Girshick R, Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2016;39(6):13.
38. Redmon J, Divvala S, Girshick R, Farhadi A, editors. You only look once: Unified, real-time object detection. IEEE conference on computer vision and pattern recognition; 2016.
39. Maor E. The Pythagorean Theorem: A 4,000-year History: Princeton University Press; 2007.
41. Carfagni M, Furferi R, Governi L, Santarelli C, Servi M, Uccheddu F, et al. Metrological and Critical Characterization of the Intel D415 Stereo Depth Camera. Sensors. 2019.

Secondary Sources

1. United Nations. Sustainable Development Goals 2020 [10.05.20]. Available from: https://www.un.org/sustainabledevelopment/.
4. Agro Popular. Francia prohibirá castrar a los cerdos sin anestesia y triturar a los pollitos nada más nacer 2020 [19.05.20]. Available from: https://www.agropopular.com/castrar-cerdos-290120/.
5. Hui J. What do we learn from region based object detectors (Faster R-CNN, R-FCN, FPN)? medium.com2018 [08.05.20]. Available from: https://medium.com/@jonathan_hui/what-do-we-learn-from-region-based-object-detectors-faster-r-cnn-r-fcn-fpn-7e354377a7c9.
7. Grunnet-Jepsen A, Sweetser JN, Woodfill J. Best-Known-Methods for Tuning Intel® RealSense™ D400 Depth Cameras for Best Performance 2018 [Available from: https://www.intel.com/content/dam/support/us/en/documents/emerging-technologies/intel-realsense-technology/BKMs_Tuning_RealSense_D4xx_Cam.pdf.
11. Norges miljø- og biovitenskapelige universitet. Strategi 2019-2023. 2019 [25.03.20]. Available from: https://www.nmbu.no/om/strategi/2019-2023#innledning.
12. CBS News. Why the number of butchers is dwindling 2015 [11.02.20]. Available from: https://www.cbsnews.com/news/why-the-number-of-butchers-is-dwindling/.
15. Animalia. Forsker på framtidens slakterier [11.02.20]. Available from: https://www.animalia.no/no/kjott--egg/automeate--norwegian-food-automation-cluster/aktuelt/forsker-pa-framtidens-slakterier/.
22. SCOTT. X-Ray Primal System [29.04.20]. Available from: https://www.scottautomation.com/products/x-ray-primal-system/.
23. Frontmatec. Beef Classification Center, BCC-3™ [29.04.20]. Available from: https://www.frontmatec.com/en/other/instruments/carcass-grading-traceability/beef-classification-center-bcc-3.
28. Sintef. Piezoelectric materials for sensors, actuators and ultrasound transducers [29.01.20]. Available from: https://www.sintef.no/en/piezoelectric-materials-for-sensors-actuators-and-ultrasound-transducers/.
30. RealSense Technology. Beginner's guide to depth (Updated) 2019 [29.01.20]. Available from: https://www.intelrealsense.com/beginners-guide-to-depth/.
40. Dorodnicov S. Distance to object 2018 [17.04.20]. Available from: https://github.com/IntelRealSense/librealsense/blob/jupyter/notebooks/distance_to_object.ipynb.
42. Intel RealSense. Depth Camera D415 [18.05.20]. Available from: https://www.intelrealsense.com/depth-camera-d415/.

Appendix - A

A comparison of the models that was not included on the result section. It is the same pig at the same location.

A.1

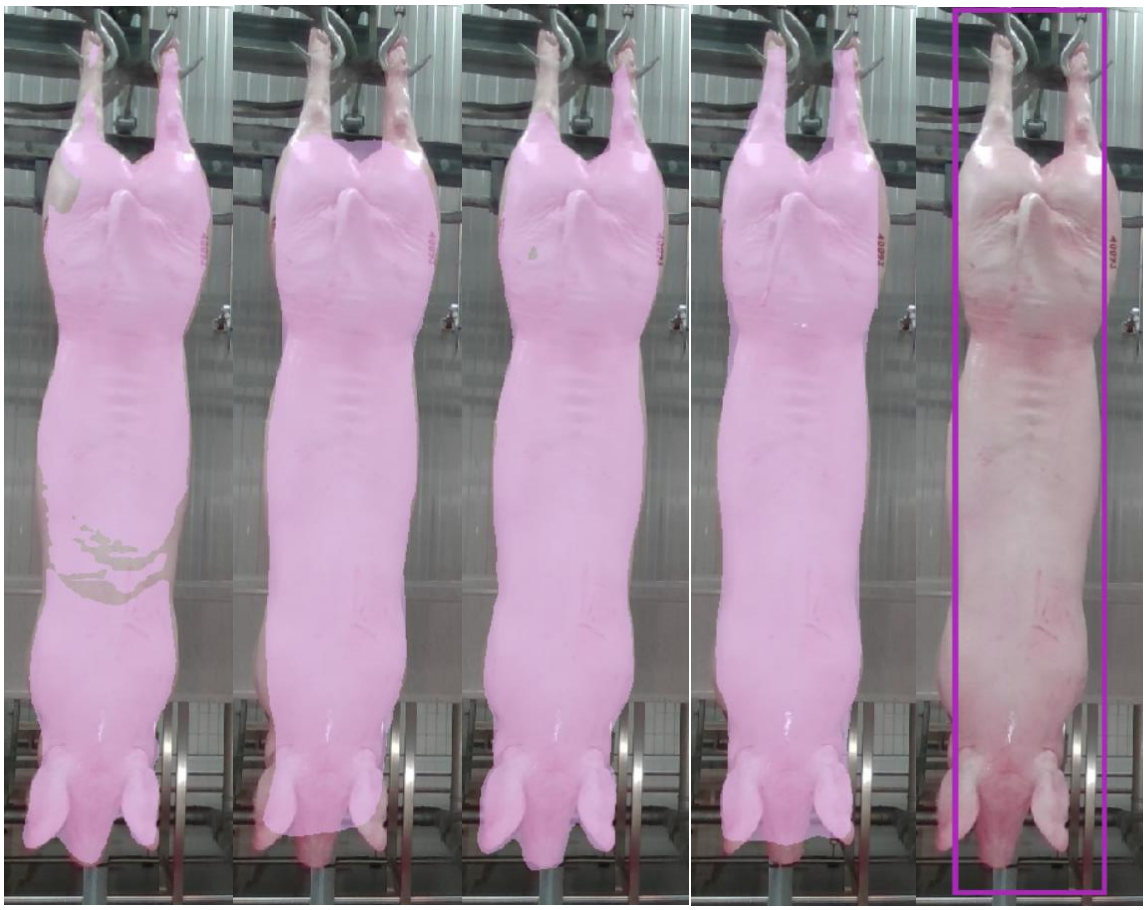


Figure 36. A collection of images from the following models: U-Net V0, V1, V3, Mask R-CNN V1 and YOLO V1.

A.2

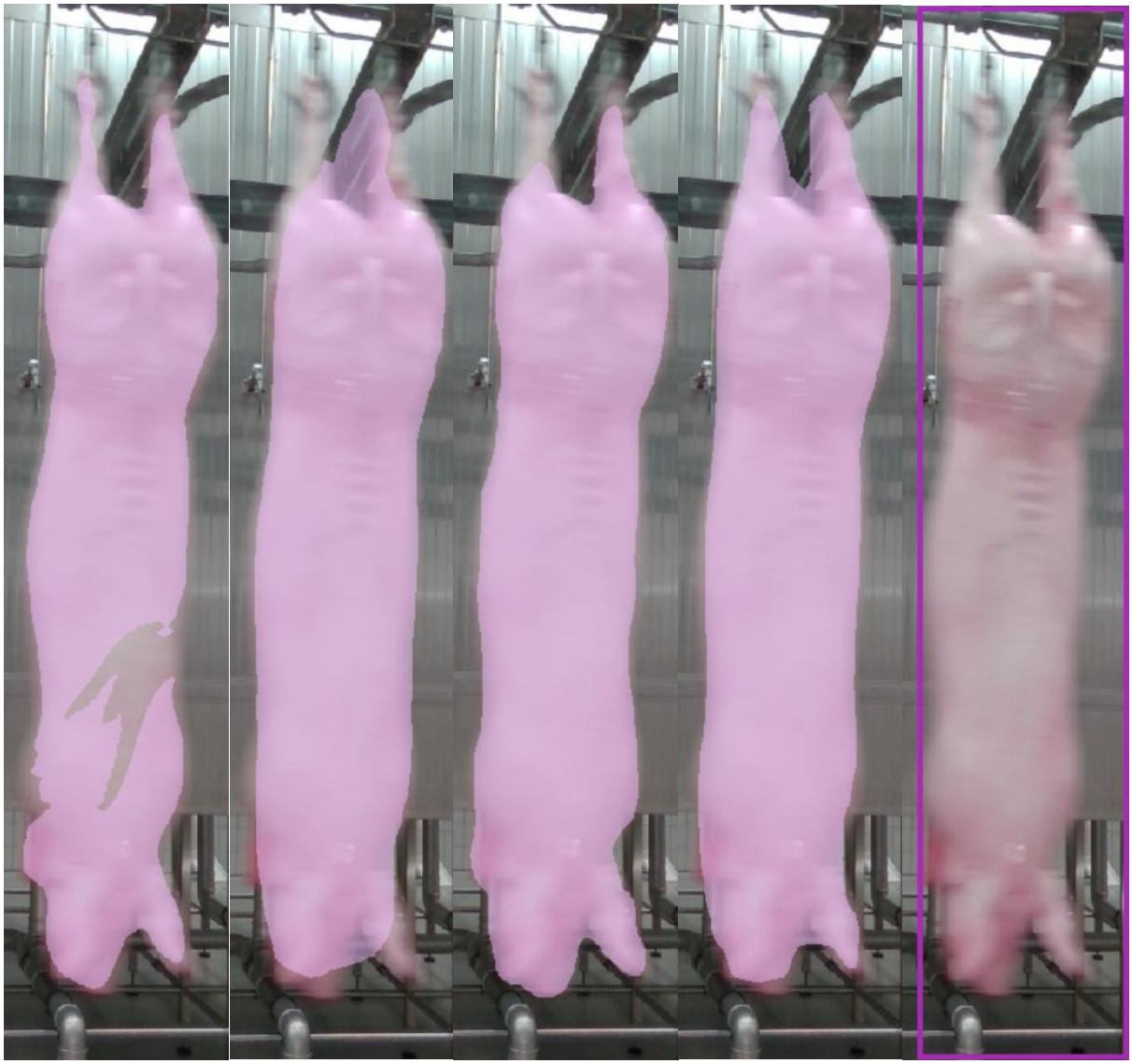
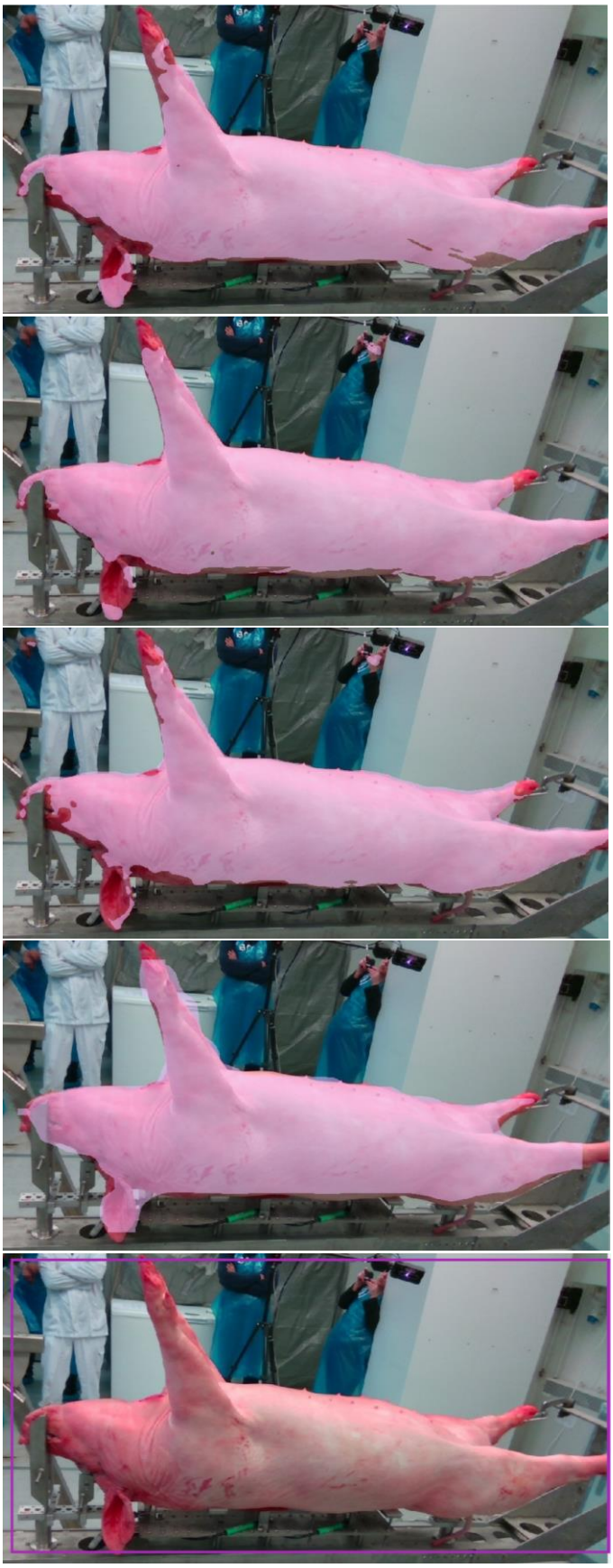


Figure 37. A collection of images from the following models: U-Net V0, V1, V3, Mask R-CNN V1 and YOLO V1.



A.3

Figure 38. A collection of images from the following models: U-Net V0, V1, V3, Mask R-CNN V1 and YOLO V1.

Appendix - B

An indication of the precision, consistency and the margin of error from three different locations. The manually measurements had consequential errors and was adjusted subsequently for every location.

B.1

Table 6. A selection of pigs from the first location. The manual measurement column is already adjusted by a multiplication rate, see the box in the up-right corner. X_L in the same box is a constant of which the length on a screen correspond to the real length. Avg. 30 equals the estimated lengths. Colored cells have the highest (red) and lowest (blue) difference, hence error.

1						x_L	4,9
	Man. Measure	Avg. 30	Difference	Same pig	L (cm)	ratios:	
	1,867	1,886	0,019		16,7	measured	0,53
	1,867	1,888	0,021		16,7	adjusted	0,5477
	1,855	1,876	0,021		16,6		
	1,855	1,841	-0,014		16,6		
	1,855	1,824	-0,031		16,6		
	1,766	1,751	-0,015		15,8		
	1,766	1,780	0,014		15,8		
	1,800	1,800	0,000		16,1		
	1,822	1,808	-0,014		16,3		
	1,822	1,798	-0,024		16,3		
	1,889	1,895	0,006		16,9		
	1,844	1,861	0,017		16,5		
SUM	22,009	22,008	-0,001		196,9		
GJ.SN.	1,834	1,834	0,000		16,4		

B.2

Table 7. A selection of pigs from the second location.

2						Ratios:	0,505
	Man. Measure	Avg. 30	Difference	Same pig	L (cm)		0,5090
	1,807	1,818	0,011		17,4		
	1,839	1,835	-0,004		17,7		
	1,839	1,806	-0,033		17,7		
	1,839	1,843	0,004		17,7		
	1,839	1,845	0,006		17,7		
	1,818	1,824	0,006		17,5		
	1,776	1,781	0,005		17,1		
	1,776	1,779	0,003		17,1		
SUM	14,532	14,531	-0,001		139,9		
GJ.SN.	1,817	1,816	0,000		17,5		

B.3

Table 8. A selection of pigs from the third location.

3						Ratios:	0,51
	Man. Measure	Avg. 30	Difference	Same pig	L (cm)		0,5395
	1,905	1,894	-0,011		17,3		
	1,894	1,926	0,032		17,2		
	1,894	1,889	-0,005		17,2		
	1,850	1,875	0,025		16,8		
	1,894	1,901	0,007		17,2		
	1,894	1,884	-0,010		17,2		
	1,894	1,876	-0,018		17,2		
	1,817	1,807	-0,010		16,5		
	1,861	1,901	0,040		16,9		
	1,861	1,851	-0,010		16,9		
	1,927	1,927	0,000		17,5		
	1,927	1,935	0,008		17,5		
	1,927	1,918	-0,009		17,5		
	1,927	1,913	-0,014		17,5		
	1,927	1,912	-0,015		17,5		
	1,872	1,853	-0,019		17,0		
	1,872	1,880	0,008		17,0		
SUM	32,139	32,142	0,003		291,9		
GJ.SN.	1,891	1,891	0,000		17,2		

Norges miljø- og biovitenskapelige universitet
Noregs miljø- og biovitskapelege universitet
Norwegian University of Life Sciences

Postboks 5003
NO-1432 Ås
Norway