University of BRISTOL

MSc Robotics

Dissertation Project Research Proposal

3D Vision Towards the Robotic Harvest of Shiitake Mushrooms

Thomas Edward Rowland

supervised by
Dr. Mark Hansen

Contents

1 Aims and Objectives

1.1 Aims

The aims of this project are as follows:

- To determine if current computer vision methods can be used to accurately count Shiitake mushroom growing on a sawdust fruiting block
- To investigate whether 3D imaging data will improve the detection of the Shiitake mushrooms
- To research whether 3D imaging can be used to plan the harvesting of the Shiitake mushrooms with a robot arm

1.2 Objectives

To achieve these aims the following objectives have been decided:

- Create a dataset of Shiitake mushrooms using a 3D camera
- Train an object detection model to detect the Shiitake mushrooms on the fruiting block
- Determine the pose of the Shiitake mushrooms in relation to the fruiting block using either 2D or 3D data
- Create a program that will order the detected Shiitake mushrooms and number them in a logical order for harvest
- Create a separate dataset of the mushroom stems and make a cut-point estimation program to determine suitable cut-points.
- **Stretch Objective:** Utilise 3D vision to achieve grasp pose estimation of a robotic end effector at various cut points

2 Motivation

Agriculture 4.0 or 'Smart Agriculture' is the integration of advanced technologies and data-driven methodologies into the agricultural sector. Smart agriculture is an essential development to combat food supply problems in the future. It is estimated that by 2050 the human population will have grown from 7.9 Billion in 2021, to 9.7 Billion in 2050 and that food production must increase by approximately 56% by 2050 to keep up with demand [1, 2]. With much of the worlds habitable land currently used for agriculture further expansion comes at the cost of deforestation and loss of biodiversity. Agriculture has traditionally been a sector that is very wasteful with its use of resources be it land, soil, water, fertilizers or the crops themselves.Two WWF reports [3] and [4] show that 40% of the food grown gets wastes globally 15.3% of all food grown never makes it off of the farm due to damage caused by crops, an inflexible food system and tight specifications set by supermarkets. Difficulties in farmed food production have been exacerbated by unpredictable weather caused by climate change and global labour shortages for farm workers.

Technological improvements in the agricultural sector to reduce this waste could come from producing food in large, heavily automated, greenhouse style farms such as those at AppHarvest [5]. Alternatively the improvements could be integrated by retrofitting existing farms with technology so that they may have more control and increase their yield an efficiency.

In both of these examples, computer vision and robotics will have a part in managing the crops at various stages of their growth cycles. Computer vision and robotics have had an increasing presence in agriculture due to the previously detailed necessity coupled with the increased availability of computing resources, improved techniques and expertise. Research and industrial effort thus far has been focused on specific high-value crops. A review paper of relevant robotic harvesting applications published in 2022 [6] showed that the 61% of robotic picking effort was focused on just Strawberries, Apples, and Tomatoes.

In the mushroom industry, robotic systems have been implemented for harvesting varieties like button mushrooms, which are typically cultivated on elongated flat beds and exhibit vertical growth, perpendicular to the growing medium. However, to date, there have been no industrial applications of robotics for harvesting gourmet mushroom varieties, such as Shiitake mushrooms, which are cultivated on specialized fruiting blocks.

Shiitake mushrooms, *Lentinula edodes*, rank among the world's most popular and extensively cultivated edible mushrooms. While they are considered highly profitable to sell [7], their cultivation and harvesting processes are labor-intensive. The robotic harvesting of Shiitake mushrooms has not yet been realized due to the challenges they present. Both the mushroom fruit and the growing medium are delicate, with any damage rendering the product unsuitable for sale. Moreover, the relatively smaller market size of Shiitake mushrooms, compared to crops like tomatoes, has led to research and industrial efforts being directed elsewhere.

Mushrooms, unlike many fruits and vegetables, are cultivated in highly controlled environments. They are grown indoors with regulated temperature, lighting, and humidity. This reduction in variability simplifies many computer vision challenges, making the exploration of automated harvesting processes more feasible.

By automating the harvesting of Shiitake mushrooms, it is possible to enhance production efficiency, reduce labor costs, and bolster food security. Additionally, if the harvest could be timed optimally, it may result in improved product quality. The insights gained from this project could be applied to numerous other fruits and vegetables that require cutting during the harvesting process, such as tomatoes, grapes, and cucumbers.

This project represents a step towards advancing the integration of computer vision and robotics in the mushroom industry, specifically in the harvesting of Shiitake mushrooms. By addressing this unexplored area, the research has the potential to contribute towards Agriculture 4.0, promoting sustainable and efficient food production practices while reducing waste and resource constraints for the future.

3 Literature Review

3.1 Overview of Shiitake Mushroom Cultivation

Shiitake mushrooms have long been cultivated in Japan and other Far East countries such as China and Korea [8]. Valued for their distinct flavor and numerous medicinal properties, these mushrooms are known to possess anti-cancer, immune-boosting, cholesterol-lowering, and anti-inflammatory benefits [8]. Cultivation of Shiitake mushrooms for hundreds of years has been outside on hardwood logs. In Japan this has commonly been the *shii* tree which is a type of Japanese oak. They have also been cultivated on other hardwoods such as other oak varieties, beech and poplar.

In recent decades, the consumption of Shiitake mushrooms has expanded beyond East Asia and gained popularity worldwide. The global Shiitake mushroom market is projected to grow from USD 2.3 billion in 2021 to USD 4.7 billion by 2030 [9], reflecting the increasing demand for this versatile and health-promoting fungus.

While Shiitake mushrooms are still often grown on logs, large-scale production has transitioned to a more controlled method involving the use of fruiting blocks made from a mixture of hardwood sawdust and nutrient supplements, such as wheat bran [10]. This shift to sawdust-based fruiting blocks has allowed for greater control over the growing process, resulting in increased yield and efficiency for mushroom farmers [11]. Many of the manual tasks traditionally associated with Shiitake cultivation, such as handling logs and drilling and waxing mushroom holes, have been eliminated. However, the adoption of sawdust fruiting blocks has led to a higher concentration of labor in the harvesting stage. Each individual Shiitake mushroom must be carefully hand-cut from the growing substrate to avoid damaging the fruiting block or the mushroom itself [12].

The growing process starts by inoculating the fruiting block with Shiitake mushroom spawn, and once the fruiting block is colonised after 2 - 4 weeks the block is removed from the plastic bag and placed in a fruiting chamber. This chamber normally is regulated to provide the high humidity and heat that is optimal for mushroom cultivation. It then takes a further 9 - 16 days for the mushroom fruiting bodies to appear and to completely grow to the harvest size. Once ready for harvest, the mushrooms will deteriorate if left on the block. Within days their quick growth means that they expose their gills and drop spores rendering them unfit for commercial sale as the quality and shelf life of the mushrooms gets reduced [10]. The fruiting blocks are commonly 2.2kg in weight and the biological efficiency of growing in this method generally varies in the range of 40-80% so that 0.4 to 0.8 kg of Shiitake mushrooms can be grown from each kg of substrate [13].

.

3.2 Robotics and Automation in Agriculture

Agri-tech, a term often used interchangeably with Agriculture 4.0 and Smart Agriculture, represents the intersection of agriculture and technology. This rapidly growing industry has seen significant advancements in recent decades and is projected to continue its growth, with the global market size expected to reach USD 45 billion by 2028, boasting a compound annual growth rate of 14% [14]. Agri-tech encompasses various domains, including precision farming, vertical farming, IoT integration, smart farming, and robotics and automation.

Robotics and automation play a vital role in agri-tech, providing solutions to increase efficiency, reduce labor costs, and improve crop yields. Examples of research and practical applications in this area range from autonomous drones for data collection and livestock monitoring systems to planting robots, sorting systems, and harvesting robots. As technology continues to advance, the integration of robotics and automation into agricultural practices will become increasingly critical for meeting global food demand.

In recent years, researchers and engineers have been continually developing innovative designs for a variety of harvesting robot use cases. Commercial adoption has gained momentum as advancements in computer vision and robotic technology enable more sophisticated and efficient harvesting processes. Developers of picking and harvesting robots have targeted plant types and varieties that are labor-intensive, prone to damage, or have high labor costs associated with manual harvesting. Driven by the growing demand for food and labor shortages, the agricultural sector has accelerated the adoption of these technologies, which were previously stuck in development for extended periods.

The diversity of fruit and vegetable crops presents unique challenges for automated harvesting systems. While some crops grow on the ground, others such as apples, citrus fruits, or olives grow on trees, and still others like tomatoes, strawberries, and grapes are grown hanging from a supporting trellis. Additionally, there is significant variation in the shape and size of each type of fruit and vegetable, as well as within each variety of that fruit or vegetable and differences between individual fruits. Developing an automated system that can robustly handle these variations is a significant challenge. While robots can be optimized for the desired harvest, they must also be able to generalize their abilities to harvest individual fruits across a range of crops.

3.2.1 Horizontal Harvest

Asparagus is a crop that is well-suited for robotic harvesting. Commonly grown in long rows where the only visible plant above the ground is the asparagus spears for harvest. Companies like Muddy Machines [15] and GARotics [16] have created robots that straddle the asparagus rows. Then using object detection and 3D cameras they select the fruits that are ready for harvest and a robot end effector snaps or cuts the asparagus spear similar to how a human would. The asparagus spears are isolated in their environment which makes them more easily identified from the background. This also means that the cut point is generally easily accessible by the end effector for the harvest when compared to other crops. A more challenging horizontal crop is lettuce. Researchers at the University of Cambridge [17] have attempted to create a system that differentiates the lettuce head from its surrounding leaves. The lettuce leaves and the head are similar in colour and texture so the detection system must be able to recognise other features such as the shape or the spatial relationship between the head and the surrounding leaves. For the harvest process the outer leaves must be trimmed off and the cut in the correct place if they are to meet supermarket quality standards. Their robot harvester which was only in he early research phase managed to harvest 88% of the tested lettuces successfully and had a total cycle time of 31.7s demonstrating the potential for future improvements.

3.2.2 Vertical Harvest

More attention has been focused on crops that fruit off of the ground. There are several commercial examples of vertical harvesting robots such as Virgo, originally designed by research company Root AI , which has since been acquired by AppHarvest, a prominent high-tech indoor farming company [18, 5]. The Virgo robot, which is designed to work alongside human pickers, uses advanced computer vision to determine the ripeness of a crop, it then uses a delicate end effector to harvest crops such as tomatoes and strawberries. Virgo autonomously navigates indoor greenhouse environment and uses machine learning to optimize its picking performance over time increasing efficiency.

Other examples of berry harvesting robots are the strawberry harvesting robot, Agrobot [19] and Fieldwork Robotics raspberry picker [20, 21]. Agrobot spans over several strawberry rows and houses up to 12 robot arms with soft-touch picking end effectors to delicately harvest the ripe fruit. The harvesting end effector never contacts the fruit during the picking process, only the stem. This reduces the risk of bruising to the fruit and extends the self life and quality of the strawberry. Similarly, the raspberry harvesting robot developed by Fieldwork Robotics, a product of the University of Plymouth, applies pressure to the raspberry stems to release the fruit without bruising. The robot uses 3D vision and machine learning to accurately detect the ripe fruit and 4 robot arms to encase the raspberry and apply pressure to the stems. This allows each robot to harvest up to 2kg of fruit an hour or up to

25,000 raspberries a day. Comparatively human pickers working an 8 hour shift can pick around 15,000 raspberries.

A collaboration between multiple research institutes lead to the development of the SWEEPER sweet pepper harvesting robot [22] in 2019. This robot end effector shown in Fig. 1 uses an RGB-D camera to locate the peppers and determine the ripeness of the fruit. An end effector which consists of a vibrating knife and a catcher then removes the fruit. The pepper then gets deposited into a bin mounted on the robot. The whole action takes 24s and being late in the research phase the robot managed 61% successful harvest in ideal crop conditions. The most common issue for the robot was the loss of sight because of occlusion caused by the leaves but their approach of utilising an RGB-D camera and onboard illumination has shown their system to be very robust to changes in varying outdoor lighting conditions.

Figure 1: The end effector of the SWEEPER robot housing the vibrating knife, RGB-D camera, LED illuminator and catch mechanism. Image replicated from [22]

Fruit trees in orchards have seen some attention from businesses and research institutes. A commercial example is the FFrobotics apple picking robot. This large robot drives through the rows of apple trees and has multiple simple low DOF (Degree of Freedom) straight line manipulators that pick the apples and as the robot moves along with a 2s cycle time, they then drop the apples onto a conveyor system that stores them into a storage bin. Their platform harvests apples with a 85% success rate of fruits reachable by the straight line effectors after using computer vision to filter out hard to reach fruits [23, 6]. This was a similar method to that used by Abundant Robotics (now shut down)[6, 24]. Their apple harvesting robots had end effectors that sucked the apples off of the trees, whereas FFrobotics would grab the apple and twist.

3.3 Computer Vision Techniques in Agriculture

Discussion so far has centered on the potential for robots to efficiently harvest of fruit and vegetable crops. For the system to be successful, accurate detection and localisation of the crop is essential. This is becomes even more apparent in picking operations such as with Virgo [18] where the harvesting causes the plant to swing. Many of the application previously mentioned such as [22, 18, 15, 20] use vision techniques to determine whether a crops ripeness is suitable for harvest.

3.3.1 Sensors

A system designer can choose from a range of different sensing methods. The simplest being regular 2D RGB (Red-Green-Blue) camera. A RGB camera can be effectively used to as a visual tool capable of differentiating features such as colour, texture, shape and can be easily implemented for use in conventional and machine learning based detection systems. They are susceptible to changes in illumination and without the another dimension of information there can be in difficulty differentiating the harvestable crop from its surrounding [25, 6]. Depth information is an option for combating these limitations. One way of achieving this is by using 2 regular 2D RGB cameras that are separated by a fixed distance in a similar fashion to human eyes. By fusing the two images together depth information can be inferred by triangulating the image disparity [25]. The accuracy of these systems are again plagued by by differing illumination and an object that is nearer than the focal length of the 2 cameras could be construed as 2 separate objects when it is infact just one. The fusion process of the two image streams can also be resource intensive. A superior option to stereo vision is with the use of an RGB-D (RGB-Depth) camera. The fourth input stream is processed simultaneously to the RGB allowing real-time streams. One method by which the RGB-D camera works is ToF (Time-of-Flight). ToF uses the known speed of light to determine the distance to an object. By emitting pulses of IR light and timing the pulses reflection a depth map can be created depicting the distance at each point of the image [26]. With this method, the depth accuracy does not change with distance (within the operating range of the camera) which is useful in an agricultural environment. Disadvantages of RGB-D cameras include problems determining distance on objects that are too close and accuracy can reduce when operating in direct sunlight due to the ambient IR interference. RGD-D cameras have increased localisation accuracy, robustness and computational efficiency when compared with using stereo 3D and this has lead to RGB-D cameras becoming an important and effective component on agricultural robots [25].

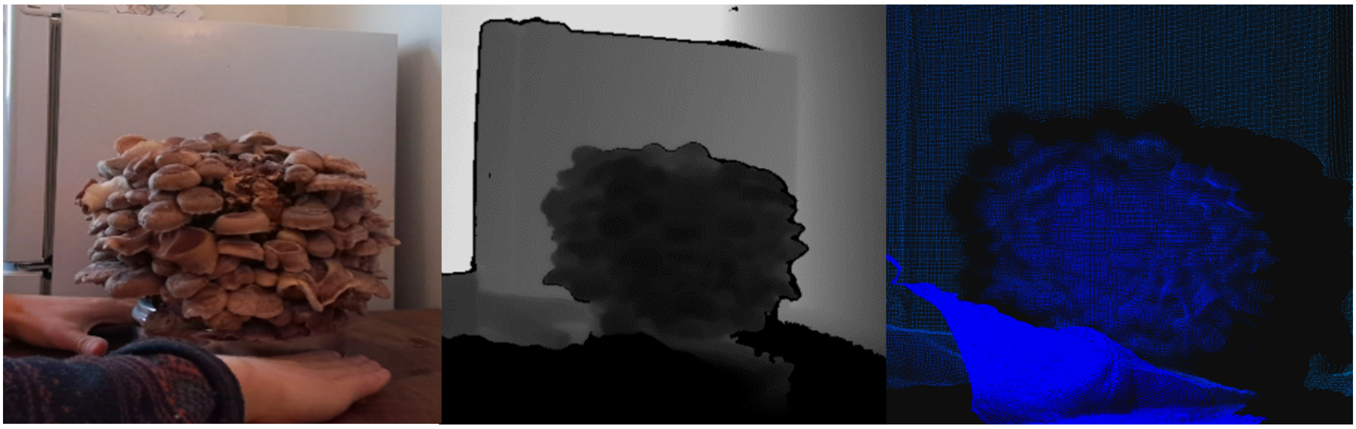

Figure 2: Image representations of a high yielding Shiitake mushroom fruiting block with Shiitake mushrooms ready for harvest. Most fruiting blocks will not fruit this densely and more of the mushroom stems will be visible to see. From left to right there is the 3 Channel RGB image, the 1 Channel Depth image and the Point cloud created from the depth information. They are all at the same frame for comparison.

3.3.2 Detection Algorithms

Detection techniques have improved significantly as deep learning style approaches have become more effective. Traditional vision techniques often relied on using purely the colour information to distinguish the crop from the background. This method can only work with specific crop types and fluctuations in the lighting can cause the performance to diminish. [27] looked to use just colour to determine the ripeness of tomatoes. They found that the RGB colour model could not be relied upon on its own as the greenhouse did not have completely uniform lighting conditions. Where an unripe tomato was in darker lighting conditions, the tomato appeared darker and was more likely to be incorrectly classified as ripe. They found that a two stage process of initially removing the background then classifying ripe from unripe tomatoes led to to improved the accuracy. The 2s processing time and reliability were impressive at the time but by todays standards where generality is valued a deep learning approach is favoured. Classification algorithms such as KNN (K-Nearest Neighbour) and SVM (Support Vector Machine) can be used as a way of classifying an image without deep learning as shown in [28]. SVM and KNN algorithms were compared for detecting mature broccoli heads with an accuracy of 95.2%.

A deep learning approach, using Convolutional Neural Networks (CNNs), extracts features such as color, texture, shape, size, and patterns, as well as contextual information like the typical arrangement of fruits on the plant. CNNs employ multiple filters that convolve over the input image to detect features previously learned from a training dataset. If the detected features match those learned by the model, the corresponding area of the image can be classified as a specific object, such as a tomato. This process enables accurate identification and classification of various crops and objects in an agricultural setting.

There have been many general object detection and deep learning algorithms developed that can be utilised directly or act as a base for use in agriculture. Common among these are YOLO [29], Faster R-CNN and Mask R-CNN [30, 31], and SSD [32]. Many further iterations and improvements of these papers have been developed over the years where developers modify the base architecture for their specific use cases. These can be split into 2-stage detectors and 1-stage detectors [33]. In 2-stage detectors like Mask R-CNN and Faster R-CNN, the first stage uses a RPN (Region Proposal Network) to determine a set of candidate areas in the image where an object might be present. The second stage then uses a CNN to classify each candidate as either one of the recognised classes or the background. Single stage detectors like YOLO and SSD aim to combine these two processes and identify and classify the image in a single pass. This has the advantage of being quicker and more suited to real time applications but it comes at the cost of reduced accuracy when compared to 2-stage detectors, especially when identifying with smaller objects. [29] .

In 2017, [34] introduced the use of Faster R-CNN to detect orchard fruits such as apples, mangoes and almonds. They found that even with a limited number of images in the dataset they could achieve an F1 score (accuracy metric) of > 0.9 or 90% when using pre-trained weights from ImageNet [35]. Shuqin Tu [36] successfully used a modified Faster R-CNN to detect passion fruits in a passion fruit orchard using RGB-D images. They then trained on the RGB and depth data separately using a multi-scale Faster R-CNN after which they fused the RGB and depth detectors together. The combination of RGB and depth data showed that their detiction system was still effective in variable lighting conditions. A modified version of the Mask R-CNN was used in [37] for Asaparagus harvest. They combined RGB and depth data to create a depth-aided Mask R-CNN (DA-MASK R-CNN). The depth information was incorporated into the region proposal network (RPN) so that the system could better differentiate between the asparagus and the rest of the environment and select the correct regions of interest. The depth information was also used to recheck classification from the Softmax function. Thus improving the overall accuracy but at a cost to the processing time.

[38] and [17] focused on modified version of YOLOv3 [39] for their detection networks. [17] concluded that for their aim of lettuce harvest YOLOv3 would be a good option as it would work efficiently on fairly modest hardware. Some draw the drawbacks of YOLO is that is is less effective on objects that are too small or too close to the camera. As their camera system would be a approximately the same distance away from the lettuce this was not a problem. [38] compared the use of 6 detection algorithms for their accuracy detecting mangoes in a mango orchard. They compared, two Faster R-CNN algorithms YOLOv2, YOLOv3, and SSD. They created a new YOLO architecture comprised from YOLOv2 and YOLOv3 features called *MangoYOLO*. This was done to reduce the number of layers so they could optimize the model for speed, memory requirement and accuracy. Their modifications mapped early feature maps from early convolutional layers in the network to later layers with the intention of allowing the network to better use some of the features identified early in the network which would assist the detection of smaller and darker fruit.

3.3.3 Pose Estimation

Pose estimation is the process determining the spatial orientation and position of an object relative to a reference frame [40]. This information can then be used to make a plan about how a robot manipulator would go about the harvesting operation. Pose estimation is important to ensure efficient and gentle harvesting, minimising the chance of damaging the crop.

Pose estimation can be achieved with both deep learning and non-deep learning methods and using 2D or 3D images. Non-deep learning methods of extracting pose information such as SURF (Speeded-Up Robust Features) [41] and SIFT (Scale-Invariant Feature Transform) [42] work by detection and matching features in images. Features are matched across different images to establish correspondence. Geometric algorithms such the Perspective-n-point (PnP) algorithms can then be used to compute the objects pose [40].

Deep learning and the use of CNNs to extract features is another more robust way using features to extract pose. These features can be more adaptable to changes in lighting, appearance and occlusion as the network has learnt patterns in the raw input data. Purpose built models such as OpenPose and PoseNet [43, 44] can be adapted for non-human pose estimation. Alternatively Mask R-CNN and YOLO [31, 29] can be used to detect keypoints for use calculating the orientation.

3D image data can be used to deduce an objects pose. 3D methods will be more robust to variations in lighting, texture and occlusion when compared with using just 2D images. Depth information can be used to construct a point cloud representation of the object from which the pose is established. [45] used this method to establish the pose of clusters of grapes ready for harvest. In their study the grapes where first segmented using a mask R-CNN, the point cloud representation was then extracted from that segmentation, and the RANSAC algorithm [46] used to determine the pose.

Another method of using 3D data for fruit harvesting is shown in [47]. They used a reconstruction of the branch to try and determine the pose by matching the fruit to the nearest point part of the mother branch. Their results suggest another method may have been more effective. Their algorithms relied on the mother branch being detected, it was often too small or the wrong branch was identified resulting in the incorrect pose being calculated.

3.3.4 Cut point estimation

There are many ways by which the fruit can be harvested by a robot end effector. These methods used are dictated by the properties of the crop. They include picking the fruit with a twist and pull motion as shown in RootAIs Virgo robot [18] and FFrobotics apple picking robot [23]. If the fruit or vegetable requires some more delicate extraction or the crop will not easily detach with just pulling, a cutting tool is required for the removal. This sometimes requires precise location of the cut point to be calculated based on the individual geometry of that crop.

The cutting point will be determined by selecting keypoints on the crop as shown in [48]. [49] use RGB-D images to first estimate the pose of a tomato vine then estimate the cut point at the top of the vine. They first use YOLACT++ (You Only Look At CoefficienTs) - a model that can provide real time image segmentation - to segment and create a mask of the whole tomato vine. Due to the geometry of the tomato vine they then used a YOLOACT++ to create a mask for the penduncle (the stalk connecting the tomatoes to the stem) of the tomato. They separated the mask and determined an average pixel point in the 'y' axis to determine the cut point in the middle of the penduncle. Similarly [50] used a trained network to determine canditate areas for cut points on the stems in citrus fruit. They compared their method with two common key point detection methods, the Stacked Hourglass Network [51] and Simple Baselines [52], both originally purposed for human pose estimation. Their proposed model increased the accuracy of the cut point detection.

The SWEEPER robot [22] shown in Fig. 1. first determines the peppers position using deep learning to semantically segment the image and calculating where the centre of the fruit and the stem was located. Using 2 view points the angle of the stem could be calculated and the vibrating knife used to cut the top of the penduncle of the pepper so that the fruit is protected from the knife by the stem. The morphology of the robot end effector allowed a less precise location for the cutpoint when compared to using a snipping end effector.

Researchers from the Bristol Robotics Laboratory's Centre for Machine Vision [53] used the deep learning ResNet object detection model to detect key points on the stem of a lettuce [54]. They adapted a system originally designed for facial keypoint detection, retraining the model to identify the keypoints that would be used as cutpoints. These were used to position the cutting tool along the pinch belt holding the stem and lettuce to efficiently and precisely separate the lettuce head from the stem.

3.4 Computer Vision and Automation in mushroom cultivation

As previously discussed, most research efforts have been directed towards fruit and vegetable harvesting rather than fungi. This could be attributed to the fact that mushrooms are delicate crops that demand labor-intensive cultivation and frequent harvesting. While these factors make a strong case for automation, the complexity involved in growing gourmet mushrooms has necessitated human intervention up until now.

3.4.1 Other Mushrooms Varieties

There have however been efforts some efforts to use computer vision and other automation technologies in mushrooms cultivation. Back in 1994, John Reed [55] developed a robot to robotically harvest button mushrooms. Button mushrooms were grown vertically on a small mushrooms bed and the system first located the mushrooms using traditional image analysis techniques with 88% of the mushrooms being identified. A simple rig with an end effector that applies suction to the mushroom cap and twists would then remove the mushrooms. The picking operation managed to successfully remove 67.5% of the mushrooms found from the image analysis. A more recent effort to locate and analyse button mushrooms has been done by Nathaniel Baisa [56]. Recognising that mushroom caps are circular they first used adaptive contouring to segment the mushrooms from the background and then utilised a Circular Hough Transform [57] to detect circles in the RGB-D images. The depth information was then used to locate the mushroom in 3D space and determine the pose of the mushrooms for harvest.

Since the advent of deep learning, vision systems could be used to locate other mushroom varieties that have less regular shapes. An example is [19], where a SSD [32] was used with RGB-D data to locate Oyster mushrooms in 3D space. On a training dataset of 4000 images the SSD model managed to detect the Oyster mushrooms with an accuracy score of 0.951 and locate the mushroom in 3D space to an average accuracy of 2.43mm. The requirements for their robot stipulated that errors larger than 5mm would lead to the unsuccessful harvest of the mushroom cluster.

3.4.2 Shiitake Mushrooms

Research into automating the growing processes of Shiitake mushrooms is limited at this time. Within the bracket of edible mushrooms Shiitake pose some of their own challenges such as how they grow on all sides of the fruiting block and do not grow in clumps like Oyster mushrooms or Lions Mane. Mushrooms are particular about their growing environment requiring temperature and humidity to be within a specific band. Researchers at the University of Rwanda compared the use IoT (Internet of Things) logic controlled devices to control and react to changes in environmental conditions with a mushroom farmer manually reacting to the conditions. They found that after 8 weeks by automating the environmental changes the yield increased by 70% [58]. A follow up experiment using a fuzzy logic based approach achieved a 50% increase in yield when compared to a farmer controlled system [59]. These impressive results could vary depending on the mushroom farmer.

A vision transformer (ViT) was used in [60] to classify 3 types of Shiitake mushrooms on a conveyer belt. A ViT [61] works by encoding parts of the image onto a vector and processing the image as a sequence similar to how Natural Language Processing tasks are accomplished. Their ViT outperformed 3 CNN architectures to achieve an overall accuracy of 98.5%. Their best performing CNN was Inception V3 which achieved an accuracy of 96.4%.

Both [62] and [63] utilise the YOLO model for Shiitake mushroom detection. [62] propose a lightweight Shiitake mushroom detection model called *MYOLO* which is based of YOLOv3 [39]. With their dataset of 1800 Shiitake mushrooms their MYOLO model achieved a mean average precision of 97.03%, higher than the other compared models which were Faster R-CNN, YOLOv3, YOLOv5 and SSD [30, 39, 64, 32]. Their study, although stated to be a primer for robotic harvest, focuses on the top-down differences of the individual mushroom caps and the dataset only contained images of the top of the mushrooms.

[63] used a modified YOLOv5 [64] algorithm termed *Mushroom-YOLO* to detect Shiitake mushrooms on a fruiting blocks. Their proposed system comprised of an automated 'smart mushroom sheds' trying to increase rate at which the mushroom heads form with cracked ridges called flower mushrooms. These can fetch 5-8 times the standard price and as such are desirable for Shiitake farmers to grow [60]. The smart sheds used machine learning to automate the growing conditions. Their system unlike the previous study [62] has an image angle that is more realistic of a robotic harvester. They modified YOLOv5 throughout to better detect overlapping mushrooms and smaller mushrooms and to improve its detection accuracy at different scales. With a training set of 300 Shiitake mushrooms they achieved a mAP (mean average precision) of 99.24% at identifying flower mushrooms instead of regular Shiitake mushrooms.

4 Impact Assessment

Social: Automated mushroom harvesting could lead to job displacement for farm workers that are currently performing the task manually. The increased agricultural productivity could lead to rural development and food security could increase as harvesting efficiency is increased and food waste caused by damage is reduced.

Technological: The technology utilised in this project will be transferable to other precision agricultural applications. The success of this project could lead to a wider adoption of robotics in the wider agricultural sector and in specific areas of agriculture such a mushroom cultivation.

Economic: Reduced labour cost for mushroom farms and the use of robots could increase the production efficiency and yield leading to higher profits.

Environmental: Precision agriculture could lead to more sustainable farming practices by reducing waste and optimising resource usage. By using efficient harvesting techniques and better handling, food waste from damage could be reduced.

Political: Governments could provide incentives or subsidies to increase the adoptions of precision agricultural techniques to improve productivity and be more environmentally conscious.

Legal: Success in the project could lead to patented technology and the protection of my intellectual property rights.

Ethical: Potential displacement of the manual labour workers raises questions about how new technologies are merges with existing agricultural practices. As the product is a food item food quality and safety must not be compromised if it were to be made into a commercial product.

5 Risk Register

Risk ID	Risk description	Likelihood?	Severity?	Risk level?	Planning and control
1	Not enough training data	5	3	15	The project is proof of concept so I will have to make do with what is available. Data augmentations and creating synthetic data can be used to easily increase the dataset size.
2	Not enough time to annotate data	2	3	6	Ensure data collection is carried out in advance and the annotations are managed appropriately.
3	Limited mushroom availability	2	4	8	Plan accordingly and stagger purchases from different vendors if there looks like there will be a problem. Think about the data I will need in future sections while I have mushrooms available.
4	Lack of exertise	2	4	8	Ensure sufficient research has been conducted before starting the project and that only methods that could be reasonably achieved are started.
5	Technical problems	1	5	5	Make sure that all code and data has been backed up with local hard drives and on cloud based storage platforms.

Figure 3: A table showing project risk scores and the mitigating actions

6 Timeline

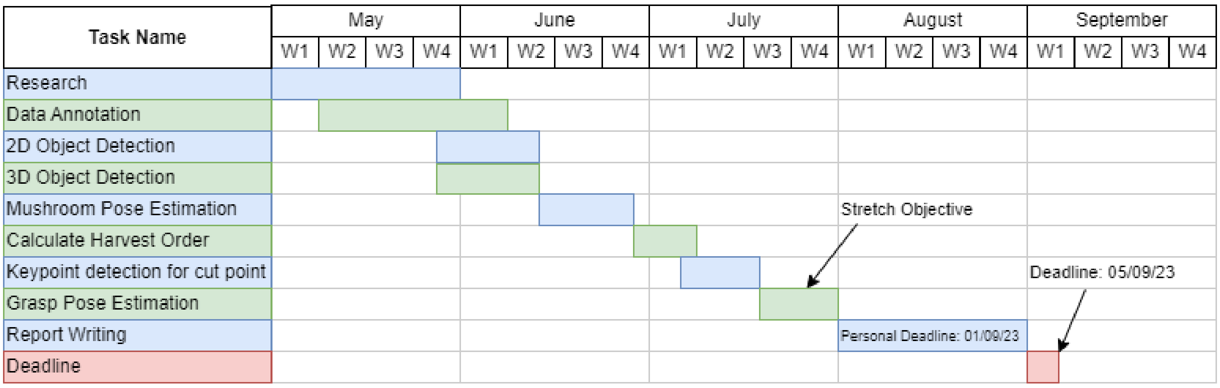

Figure 4: A gantt chart showing the project from research phase through to the deadline

The above diagram shows how the project will progress and the logical steps to achieve the aims and objectives. Research and data collection for this project commenced in the months prior to those shown in Fig.4 but they have mostly been omitted to improve the readability of the gantt chart. Below in the following sections a more substantial explanation of each phase of the project will be given.

1. **Research:** Research will continue to be conducted to refine the scope of the project and to determine the best methods for each part of the project. By this point a full literature review will have been completed and a clear idea of the current state-of-art will have been established. This will be used to inform further decisions as the project progresses.

2. **Dataset Annotation:** Collection of 2D and 3D images of Shiitake mushrooms was started in Feb. 2023. These will be processed and annotated accordingly for the 2D and 3D object detection algorithms that will be used in the next section. The 2D frames from dataset will be annotated and superimposed onto the RGB-D data. The mushrooms will have to be annotated with polygon segmentation as the RGB-D data will not work with bounding box annotations as the system requires each mushroom to be differentiable from each other as well as from the the background. Bounding box annotations will add a lot of noise into the 3D representation and everything within the bounding box gets classified as a mushroom - even if it is not a mushroom.

3. **Object Detection:** Object detection algorithms will be used to determine how effective computer vision can be at detecting Shiitake mushrooms. This will be done with both 2D and 3D algorithms and compared to establish if the 3D system with the addition of depth information is more accurate at detecting individual Shiitake mushrooms.

4. **Mushroom Pose Estimation:** As the mushrooms grow on the fruiting blocks they grow on all of the vertical sides and the top. A program will be developed that can calculate the orientation of the mushroom with respect to the fruiting block.

5. **Calculate Harvest Order:** With the mushroom orientations established the a logical harvest order should be calculated. As the mushrooms generally grow outwards and and then up the logical harvest will be from the bottom. A program will me created that looks at the detected mushrooms and orders them from 1 to n depending on their position relative to each other and the fruiting block.

6. **Keypoint Detection for the cut point:** Depending on the success of the previous systems a decision will be made to determine whether the keypoint detection will be accomplished using machine learning or traditional techniques. A new dataset will have to be curated and the cut points on the mushroom stems annotated. As mentioned previously the cutpoints will have to be at a part on the mushroom stem that a robot arm could safely reach.

7. **Grasp Pose Estimation:** This stretch objective will use grasp pose algorithms to represent the best grasp pose of a robotic cutting end effector.

8. **Thesis Write-up:** 4 weeks will be allocated to focus writing up the thesis.

References

[1] United Nations, Department of Economic and Social Affairs, Population Division, "World population prospects 2019: Highlights," tech. rep., United Nations, 2019.

[2] M. van Dijk, T. Morley, M. L. Rau, and Y. Saghai, "A meta-analysis of projected global food demand and population at risk of hunger for the period 2010–2050," *Nature Food*, vol. 2, no. 7, pp. 494–501, 2021.

[3] World Wildlife Fund and Tesco, "Driven to waste: The global impact of food loss and waste on farms," tech. rep., World Wildlife Fund, 2021.

[4] World Wildlife Fund, "Hidden waste: The food waste in our fields," tech. rep., World Wildlife Fund, 2022.

[5] "Appharvest." `https://www.appharvest.com/`, Feb 2023. Accessed: 2023-03-30.

[6] H. Zhou, X. Wang, W. Au, H. Kang, and C. Chen, "Intelligent robots for fruit harvesting: Recent developments and future challenges," *Precision Agriculture*, vol. 23, no. 5, pp. 1856–1907, 2022.

[7] "How much do gourmet mushroom farmers make? - northstar financial consulting group." `https://nstarfinance.com/how-much-do-gourmet-mushroom-farmers-make/`, Sep 2022. Accessed: 2023-03-30.

[8] S. P. Wasser, "Shiitake (lentinus edodes)," *Encyclopedia of dietary supplements*, pp. 653–664, 2005.

[9] "Shiitake mushroom market, the brainy insights." `https://www.thebrainyinsights.com/report/shiitake-mushroom-market-12716`, Apr 2022. Accessed: 2023-03-20.

[10] T. Shields, "Growing shiitake mushrooms." `https://learn.freshcap.com/growing/growing-shiitake-mushrooms/`, Dec 2020. Accessed: 2023-03-30.

[11] R. Kohsaka, M. Tomiyoshi, O. Saito, S. Hashimoto, and L. Mohammend, "Interactions of knowledge systems in shiitake mushroom production: a case study on the noto peninsula, japan," *Journal of Forest Research*, vol. 20, pp. 453–463, 10 2015.

[12] P. Cong, H. Feng, K. Lv, J. Zhou, and S. Li, "Myolo: A lightweight fresh shiitake mushroom detection model based on yolov3," *Agriculture*, vol. 13, no. 2, 2023.

[13] D. A. Diehle and D. J. Royse, "Shiitake cultivation on sawdust: Evaluation of selected genotypes for biological efficiency and mushroom size," *Mycologia*, vol. 78, no. 6, pp. 929–933, 1986.

[14] A. M. Research, "Agritech market - global industry size, share, growth, trends, competitive analysis, and forecast." https://www.adroitmarketresearch.com/industry-reports/agritech-market, 2023. Accessed: 2023-03-30.

[15] "Meet sprout - muddy machines." https://www.muddymachines.com/technology, 2023. Accessed: 2023-03-30.

[16] A. Leu, M. Razavi, L. Langstädtler, D. Ristić-Durrant, H. Raffel, C. Schenck, A. Gräser, and B. Kuhfuss, "Robotic green asparagus selective harvesting," *IEEE/ASME Transactions on Mechatronics*, vol. 22, no. 6, pp. 2401–2410, 2017.

[17] S. Birrell, J. Hughes, J. Y. Cai, and F. Iida, "A field-tested robotic harvesting system for iceberg lettuce," *Journal of Field Robotics*, vol. 37, pp. 225–245, 3 2020.

[18] "Root AI - Home." https://root-ai.com/#intro, 2023. Accessed: 2023-03-30.

[19] "Agrobot e-series, agrobot." https://www.agrobot.com/e-series, 2023. Accessed: 2023-03-30.

[20] "Fieldwork robotics - home." https://fieldworkrobotics.com/, Sep 2021. Accessed: 2023-03-30.

[21] A. Williams, "Company makes significant progress with raspberry harvesting robots - university of plymouth." https://www.plymouth.ac.uk/news/company-makes-significant-progress-with-raspberry-harvesting-robots, Jun 2022. Accessed: 2023-03-30.

[22] B. Arad, J. Balendonck, R. Barth, O. Ben-Shahar, Y. Edan, T. Hellström, J. Hemming, P. Kurtser, O. Ringdahl, T. Tielen, and B. van Tuijl, "Development of a sweet pepper harvesting robot," *Journal of Field Robotics*, vol. 37, pp. 1027–1039, 9 2020.

[23] R. Courtney and T. Mullinax, "Washington orchards host robotic arms race, good fruit grower." https://www.goodfruit.com/washington-orchards-host-robotic-arms-race/, Dec 2019. Accessed: 2023-03-30.

[24] S. Crowe, "Abundant robotics shuts down fruit harvesting business, the robot report." https://www.therobotreport.com/abundant-robotics-shuts-down-fruit-harvesting-business/, Jul 2021. Accessed: 2023-03-30.

[25] A. Gongal, S. Amatya, M. Karkee, Q. Zhang, and K. Lewis, "Sensors and systems for fruit detection and localization: A review," *Computers and Electronics in Agriculture*, vol. 116, pp. 8–19, 2015.

[26] L. Fu, F. Gao, J. Wu, R. Li, M. Karkee, and Q. Zhang, "Application of consumer rgb-d cameras for fruit detection and localization in field: A critical review," *Computers and Electronics in Agriculture*, vol. 177, 10 2020.

[27] A. Arefi, A. M. Motlagh, K. Mollazade, and R. F. Teimourlou *Australian Journal of Crop Science*, vol. 5, no. 10, p. 1144–1149, 2011.

[28] K. Kusumam, T. Krajník, S. Pearson, G. Cielniak, and T. Duckett, "Can you pick a broccoli? 3d-vision based detection and localisation of broccoli heads in the field," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 646–651, 2016.

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

[30] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Advances in neural information processing systems*, vol. 28, 2015.

[31] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, pp. 2961–2969, 2017.

[32] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*, pp. 21–37, Springer, 2016.

[33] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, pp. 318–327, 2020.

[34] S. Bargoti and J. Underwood, "Deep fruit detection in orchards," in *2017 IEEE international conference on robotics and automation (ICRA)*, pp. 3626–3633, IEEE, 2017.

[35] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision (IJCV)*, vol. 115, no. 3, pp. 211–252, 2015.

[36] S. Tu, J. Pang, H. Liu, N. Zhuang, Y. Chen, C. Zheng, H. Wan, and Y. Xue, "Passion fruit detection and counting based on multiple scale faster r-cnn using rgb-d images," *Precision Agriculture*, vol. 21, pp. 1072 – 1091, 2020.

[37] X. Liu, D. Wang, Y. Li, X. Guan, and C. Qin, "Detection of green asparagus using improved mask r-cnn for automatic harvesting," *Sensors*, vol. 22, no. 23, 2022.

[38] A. Koirala, K. Walsh, Z. Wang, and C. McCarthy, "Deep learning for real-time fruit detection and orchard fruit load estimation: Benchmarking of 'mangoyolo'," *Precision Agriculture*, vol. 20, pp. 1107–1135, 2019.

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

[40] G. Du, K. Wang, S. Lian, and K. Zhao, "Vision-based robotic grasping from object localization, object pose estimation to grasp estimation for parallel grippers: a review," *Artificial Intelligence Review*, vol. 54, no. 3, pp. 1677–1734, 2021.

[41] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in *Computer Vision – ECCV 2006* (A. Leonardis, H. Bischof, and A. Pinz, eds.), (Berlin, Heidelberg), pp. 404–417, Springer Berlin Heidelberg, 2006.

[42] A. Abdel-Hakim and A. Farag, "Csift: A sift descriptor with color invariant characteristics," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, vol. 2, pp. 1978–1983, 2006.

[43] Z. Cao, G. Hidalgo Martinez, T. Simon, S. Wei, and Y. A. Sheikh, "Openpose: Realtime multi-person 2d pose estimation using part affinity fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.

[44] A. Kendall, M. Grimes, and R. Cipolla, "Posenet: A convolutional network for real-time 6-dof camera relocalization," 2016.

[45] W. Yin, H. Wen, Z. Ning, J. Ye, Z. Dong, and L. Luo, "Fruit detection and pose estimation for grape cluster–harvesting robot using binocular imagery based on deep neural networks," *Frontiers in Robotics and AI*, vol. 8, 2021.

[46] K. G. Derpanis, "Overview of the ransac algorithm," *Image Rochester NY*, vol. 4, no. 1, pp. 2–3, 2010.

[47] G. Lin, Y. Tang, X. Zou, J. Xiong, and J. Li, "Guava detection and pose estimation using a low-cost rgb-d sensor in the field," *Sensors*, vol. 19, no. 2, 2019.

[48] B. Zheng, G. Sun, Z. Meng, and R. Nan, "Vegetable size measurement based on stereo camera and keypoints detection," *Sensors*, vol. 22, no. 4, p. 1617, 2022.

[49] J. Rong, G. Dai, and P. Wang, "A peduncle detection method of tomato for autonomous harvesting," *Complex & Intelligent Systems*, pp. 1–15, 2021.

[50] Q. Sun, X. Chai, Z. Zeng, G. Zhou, and T. Sun, "Multi-level feature fusion for fruit bearing branch keypoint detection," *Computers and Electronics in Agriculture*, vol. 191, p. 106479, 2021.

[51] A. Newell, K. Yang, and J. Deng, "Stacked hourglass networks for human pose estimation," in *Computer Vision – ECCV 2016* (B. Leibe, J. Matas, N. Sebe, and M. Welling, eds.), (Cham), pp. 483–499, Springer International Publishing, 2016.

[52] B. Xiao, H. Wu, and Y. Wei, "Simple baselines for human pose estimation and tracking," in *Proceedings of the European conference on computer vision (ECCV)*, pp. 466–481, 2018.

[53] "Bristol robotics laboratory - centre for machine vision." `https://www.bristolroboticslab.com/centre-for-machine-vision`, n.d. Accessed: 29-03-2023.

[54] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, Dec 2015.

[55] J. Reed and R. Tillett, “Initial experiments in robotic mushroom harvesting,” *Mechatronics*, vol. 4, no. 3, pp. 265–279, 1994.

[56] N. L. Baisa and B. Al-Diri, “Mushrooms detection, localization and 3d pose estimation using rgb-d sensor for robotic-picking applications,” *arXiv preprint arXiv:2201.02837*, 2022.

[57] S. J. K. Pedersen, “Circular hough transform,” *Aalborg University, Vision, Graphics, and Interactive Systems*, vol. 123, no. 6, pp. 2–3, 2007.

[58] N. J. de La Croix, M. Didacienne, S. Louis, J. T. Philander, and T. Ahmad, “Internet of things based controlled environment for the production of shiitake mushroom,” in *2022 IEEE International Conference on Blockchain and Distributed Systems Security (ICBDS)*, pp. 1–6, IEEE, 2022.

[59] N. J. De La Croix, M. Didacienne, and S. Louis, “Fuzzy logic-based shiitake mushroom farm control for harvest enhancement,” in *2022 10th International Symposium on Digital Forensics and Security (ISDFS)*, pp. 1–6, IEEE, 2022.

[60] J. Deng, Y. Liu, and X. Xiao, “Deep-learning-based wireless visual sensor system for shiitake mushroom sorting,” *Sensors*, vol. 22, no. 12, p. 4606, 2022.

[61] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, *et al.*, “An image is worth 16x16 words: Transformers for image recognition at scale,” *arXiv preprint arXiv:2010.11929*, 2020.

[62] P. Cong, H. Feng, K. Lv, J. Zhou, and S. Li, “Myolo: A lightweight fresh shiitake mushroom detection model based on yolov3,” *Agriculture*, vol. 13, no. 2, p. 392, 2023.

[63] Y. Wang, L. Yang, H. Chen, A. Hussain, C. Ma, and M. Al-gabri, “Mushroom-yolo: A deep learning algorithm for mushroom growth recognition based on improved yolov5 in agriculture 4.0,” in *2022 IEEE 20th International Conference on Industrial Informatics (INDIN)*, pp. 239–244, IEEE, 2022.

[64] A. S. Glenn Jocher, Ayush Chaurasia, “ultralytics/yolov5: v7.0 - yolov5 sota realtime instance segmentation,” nov 2022.