SCARLETT: Towards Scalable, Structured and Resource-Efficient Harvesting of Lettuce with Collaborative Biomimetic Robots

Roberto Mendivil Castro

Thesis submitted for the degree of
Masters by Dissertation in Computer Science

School of Computer Science and Electronic Engineering
University of Essex
Supervisor: Dr Vishwanathan Mohan

October 2024

Acknowledgement

This thesis was developed under SCARLETT, a DEFRA and UKRI-funded project that aimed to incorporate collaborative robots into hydroponic farms. The emphasis on collaboration is a key aspect of this project. I am grateful to my supervisors, Vishwanathan Mohan and Jonathan Dove, for their support in the development, allowing me to successfully implement all aspects of the project, from farm experiments to on-site results. I want to thank my family, parents, and sisters for their love and support. Special acknowledgement to you, Andrea Carlos, for being there, always inspiring me and supporting me to grow with every single challenge, no matter the long distance or any circumstance in the path.

Abstract

In recent years, there has been a significant shortage of labour in agriculture, particularly in manual and repetitive tasks that require a high level of dexterity, task-specific adaptivity and cognition. Robotic automation in such environments can have a multidimensional impact- from mitigating labour shortage, increasing productivity and efficiency in the harvesting workflows. This thesis focuses on scalable robotic harvesting of Lettuce in state-of-the-art deepwater pool hydroponic farm. The primary contributions of this thesis are 1) development of deep learning based 2D/3D perception system for precise identification/localization of plants in varying growing stages, picking/placing points in floats of different dimensions in noisy/unstructured conditions; 2) development of an ANN based adaptive motion controller to coordinate the CR5 and CR10 Cobot arms to automate three critical harvesting tasks- seedling transplantation, float handing and harvesting; 3) Embedding of the robotic framework into the farm operation taking into account food safety, scalability and human-robot collaboration. The SCARLETT framework has been evaluated in the state-of-the-art 1.1Ha hydroponic farm of JEPCO, producing 1.95 million plants annually. While this thesis focuses specifically on Lettuce, the framework itself is adaptable to other ‘crop types, tasks and growing environments’, hence opening further opportunities for robotics in Smart Farming.

Contents

Chapter 1 Introduction 1

1.1 Motivation for robotics in agriculture 1

1.2 Main Contributions 1

1.2.1 End-user scenario addressed in this Thesis 3

1.2.2 Technical Challenges for Robotics/AI addressed in this Thesis 8

1.3 Organization of this Thesis 9

1.4 Summary of Achievements 11

1.4.1 Summary of Technical Contributions 11

1.4.2 Awards and Research Archievements 11

1.4.3 Publications 12

Chapter 2 SCARLETT- Survey of state of the art 14

2.1 Perception Systems for Crop 14

2.1.1 From classic computer vision to deep learning detectors 14

2.2.1 Optimal and Impedance control motion planning 20

Chapter 3 Configurable 2D-3D Perception for Robotic Lettuce harvesting 24

3.1 Vision methods for agriculture vision 24

3.1.1 Object detection and feature extraction 24

3.2 Object detection and Key point localisation 27

3.3 Evaluation from field trials 29

Chapter 4 Robot Motion Control for Lettuce Harvesting 34

4.1 Passive Motion Paradigm 34

4.1.1 Internal body of the body - Collaborative DOBOT arms 34

4.1.2 Constraints and Reaching on real environment 39

Chapter 5 SCARLETT - Farm Deployment and Impact Analysis 42

5.1 Goal-Directed grasping and transplanting under constrain on-site 44

5.2 Analysis of Economic Benefits 51

Chapter 6 General Conclusions and Future Work 53

6.1 Summary 53

6.2 Research directions 54

List of Figures

Figure 1 – Hydroponic farm of JEPCO, 1.1Ha space producing 1.95 million plants annually 3
Figure 2 – Production workflow of JEPCO. 1) Transplanting of seedlings,2) Transportation of floats between water flume to pond for growth. 3) Placing of full-grown lettuce on flume to the conveyor belt, 4) Root cutting and 5) Packaging closing the workflow from seedling to consumption 3
Figure 3 – Transplanting of seedlings involves dexterous manipulation for handling soft, delicate seedlings with accuracy on placement. 4
Figure 4 – Hydroponic floats. A)Seed floats, where the root starts growing. B) Propagation floats for full-grown spacing. 5
Figure 5 – Seedling placing with root alignment. If not aligned, the task is considered a failure 6
Figure 6 – The transportation task of the fully grown lettuce float in the pool (with 8KG payload and water resistance) is to a water flume that transports the float to the root cutting/packaging station. 7
Figure 7 – Dobot CR5 for transplanting in JEPCO. Mounted on a stainless steel stand with OnRobot RG2 Gripper 8
Figure 8 – Classic Computer Vision Approach. Template matching and colour filtering and 3D localisation from pixel point 24
Figure 9 – Feature Extraction ad image level and template matching for detection 25
Figure 10 – Point Cloud thickness approximation of crops 26
Figure 11 – dataset preparation process, including the collection, annotation, export, enhancement, and final segmentation into training and testing sets, with datasets created to cater to different 27
Figure 12 – Lef panel Object Detection. Right OBB Oriented bounding Box 28
Figure 13 – Keypoint localization from object detection. Ideal method 28
Figure 14 – Keypoint Localization mean average precision (mAP) 90% 29
Figure 15 – Perception Laboratory Setup. Vicon Tracking System as reference for 3D localisation 31
Figure 16 – Farm Setup. ARUCO markers system as a reference for 3D localisation 31
Figure 17 – Metrics of vision models 32

Figure 18 – JADE - Joint admittance from human pose estimation for specific joint constraints from human manipulation 34
Figure 19 – CR10 2mm accuracy ANN-based motion planning 35
Figure 20 – DH Parameters of Dobot CR10 37
Figure 21 – Human pose extraction at joint level and reference for robot 6dof 41
Figure 22 – Human Pose Estimation trajectory to find individual contribution of human joints 41
Figure 23 – Side by side human and robot pick and place performance for transplanting operations 42
Figure 24 – Software Architecture of SCARLETT. Closed-loop solution 44
Figure 25 – The transition between Lab and Farm 45
Figure 26 – Harvesting room electrical and mechanical adaptation 46
Figure 27 – Close-loop key point detection, robot coordination and manipulation on real environment 47
Figure 28 – Harvest Room pick and place for packaging. A rotational gripper was used to handle the full-grown lettuces and manipulate them from conveyor belt to packaging station 48
Figure 29 – Flot handling adter transplanting 49
Figure 30 – Robot transplanting in farm 49
Figure 31 – Future work, extension as agricultural robotics framework 55

Chapter 1
Introduction

1.1 Motivation for robotics in agriculture

In recent years, there has been a significant shortage of labour in the agriculture field, particularly in manual and repetitive tasks, due to several demographic and economic factors; among them, urbanisation and socioeconomic dysfunction have led to decreased demand and financial resources, complicating availability in the agricultural sector and producing a low quality of life. [1] There is also a lack of work security, considering not only the current ageing workforce but also new generations avoiding the sector and looking for options to expand the workload internationally may not be feasible due to implications in migration policies that disable the growth of that labour. Automation is needed to address these labour implications, which affect society, the economy, and the environment. [2].

Robotics has emerged as a revolutionary trend, transforming tasks and reshaping traditional methods while impacting several fields. It has led to significant advances in precision, efficiency, and scalability, incorporating intelligence where systems can adapt and learn, improving the results over time.[3], [4]. Thus, robotics represents an attractive solution for Agri-tech applications. In the United Kingdom, for instance, departments like DEFRA and UKRI are leveraging technology and innovation to tackle the labour shortage in agriculture. They aim to create a more sustainable and efficient agricultural sector and encourage collaboration between researchers, industry stakeholders, and farmers to develop innovative solutions to address workload challenges[5]. These initiatives address immediate production challenges and contribute to the long-term transformation of the industry.

1.2 Main contributions

The current work focuses on the research and development of Robotics automation for deep-water hydroponic systems as part of the DEFRA-funded SCARLETT project: Scalable, structured, and resourced efficient indoor robotics harvesting of lettuce.
Building on the workloads of industrial cells as inspiration and backed by the UK government's Farming and Innovation program initiative, the SCARLETT project has been designed to address the main problem of labour shortage in the indoor hydroponic agriculture sector. [5] This project is a significant step towards a long-term solution to this pressing issue. This research-funded initiative by the University of Essex, with support from Innovate UK (UKRI), with the primary purpose of the automation for JEPCO Geble as the leading end-user, aims to revolutionise indoor robotics harvesting of lettuce through a collaborative, adaptive, and intelligent solution with robot arms and 3D vision sensors. The long-term goal is the creation of a scalable solution which can be easily reimplemented in similar scenarios, contributing to the central labour shortage problem and revolutionising food production in the United Kingdom [6].

The challenges, scalability, and well-defined workflows in the harvesting loop while embedding robotics, sensing, and machine learning in such an environment to automate a range of well-defined but repetitive, laborious tasks in the harvesting loop with high speed and efficiency where human-robot interaction is crucial due to hygiene standards in a food-humid environment, where constant intervention is intrinsic. Thus, the robotics system must be able to work under those challenging constraints [8], [9].

The JEPCO indoor lettuce harvesting workflow is structured around 11 production ponds and three seeding pools, each accommodating a specific number of plants. The space between the pools is designed for walking, and a water flume, which serves as a conduit, is also part of the setup (Figure 1). Lettuces are grown in two types of hydroponic polystyrene floats, each with a defined structure, where the seed floats can accommodate 24 plants. In contrast, the propagation floats can accommodate eight spaces for eventual growth. The water flume, which has an adjustable slope, uses gravity to control the flow and serves as a pathway for the floats during the harvesting workflow.
In this setting, the technological advancement in SCARLETT has successfully tested and deployed two collaborative robots as a feasibility study to perform three critical tasks presently requiring significant manual labour.

Figure 1 – Hydroponic farm of JEPCO, 1.1Ha space producing 1.95 million plants annually.

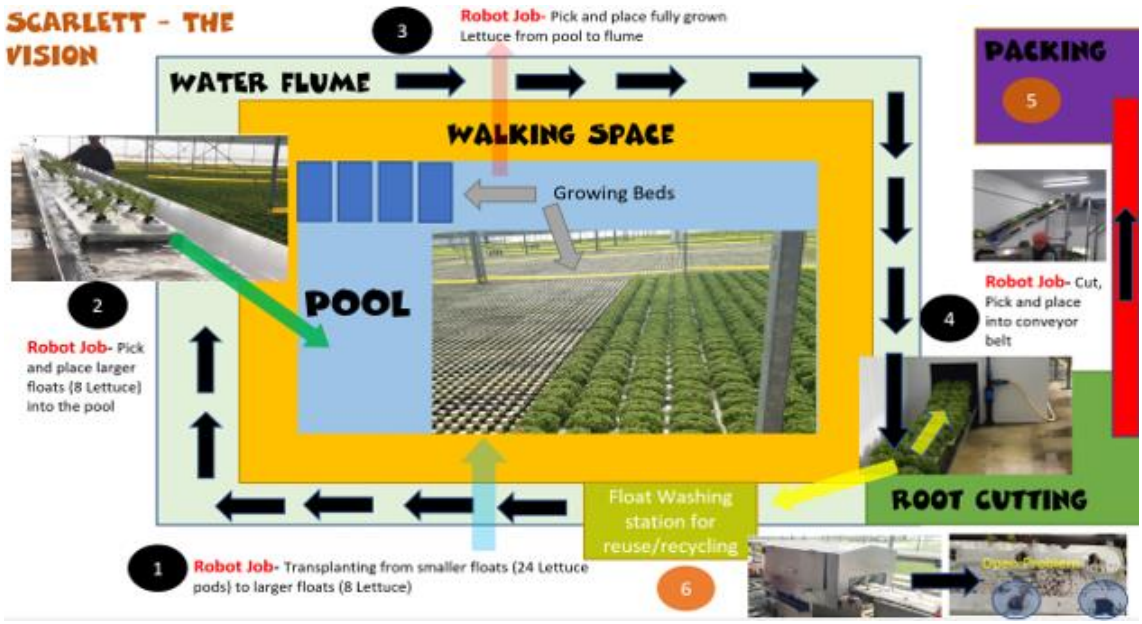


Figure-2 –Pictorially depicts the farm (September 2024) and planned tasks in SCARLETT. The harvesting workflow revolves around deep-water pool with growing beds, walking space and water flume. The water flume serves a conduit where floats (with known geometry) move around during the harvesting workflow. SCARLETT aims to deploy robotic automation in 5 critical tasks presently requiring manual labour (labelled 1-5), deploying two agricultural robots operating in parallel. Task 6 (orange) is also amenable to robotic automation but out of the scope of SCARLETT, due to specialized equipment needed for scrubbing and removal of peat stuck in the float after root cutting, which is manually done before machine float wash.

1.2.1 End-user scenario addressed in this Thesis

This study approaches how SCARLETT, a robotic perception-action system, aims to solve the main tasks of the hydroponic facility. The tasks are described in Figure-2, where all the JEPCO workflow is divided into five main tasks: initial growth, handling of hydroponics floats, spacing the lettuces for propagation, and then moving to the growth cycle to be ready to be cut and packaged, finally. In this thesis, the focus will be mainly on the transplanting tasks but its implementation was tested on three main tasks: A) Transplanting of lettuces, B) Float Handling, and C) Lettuce packaging, where the system can be applied to the tree tasks.

A) Transplanting of Lettuce seedlings from Seed Floats to Propagation Floats.

The first step in the automation process is seeding transplanting, where lettuce seedlings are transplanted from seedling floats to propagation floats. The main challenge is the precision of placing those lettuces, dipping them in water to avoid static, and inserting the root in the end hole, which determines whether the tasks were successful or not.

The Vision system for this task focuses on structure (known geometry, markers) and counteract variance (crop, object motion) under a robust adapting perception system under environment changes such as light, solid, and occlusion. The main challenge is soft grasping and gentle insertion without damaging the roots, considering that the destination float is in water. The strategy of human labourers deployment inspires the ANN-based motion planning system.

Figure -3 - Transplanting of seedlings involves dexterous manipulation for handling soft, delicate seedlings with accuracy on placement.

The innovation achieved in this research project SCARLETT has been in multiple directions.

I. CNN-based Pose 2D/3D vision system for detection/localisation of picking and placing points in small and large floats in noisy conditions.

II. ANN-based motion planning for the CR10 and CR5 [7] collaborative robot arms with precise control of motion trajectory, wrist pose learning from motion analysis of humans performing the transplanting task

III. As external development from the current work, the design of a novel 4X low-cost transplanting gripper- A human hand or a conventional parallel gripper can grasp/transplant only one seedling at a time. Still, this novel 3D printable gripper can simultaneously grasp and transform four seedlings. With a modular architecture, in-house electronics, and 3D printable design, the gripper costs only £200 (compared to a parallel gripper that can cost £3K or more) while doing 4x the job simultaneously. This also saves the cost of 3 additional robot arms, the complex motion control, vision system, and collision avoidance needed (if four robot arms were to transplant seedlings in a shared workspace). This is approximately a 75x reduction in cost to automate a highly complex/dexterous task.

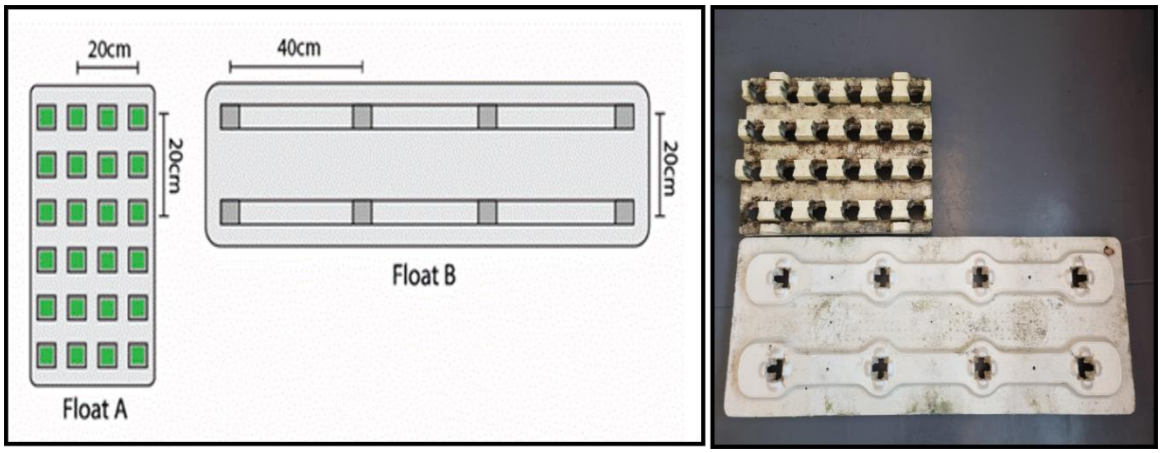


Figure 4 - Hydroponic floats. A)Seed floats, where the root starts growing. B) Propagation floats for full-grown spacing.

B) Float Transportation- From flume to pounds (two-picking tasks)

Once the seedlings have been transplanted from the seed tray to the propagation float, they must be picked and transported to a water flume, where the task can be seen as trivial through the transplanting methodology. Two kinds of pick and place tasks are

implemented: first, by moving the propagation float travelling in the water flume after transplantation into the pool; then, after approximately 21 days [10], the seedling converts into a fully grown Lettuce, which has to be transported from the Pool to the water. The design of a novel float-handling gripper that can grasp both small and large floats with varying payloads is a practical innovation. The arm controls the gripper directly, eliminating any additional control or wiring. This novel design can be directly deployed in the existing facility and is compatible with the CR10 and CR5 collaborative robot arms, making it a feasible and cost-effective solution.

Figure 5 - Seedling placing with root alignment. If not aligned, the task is considered a failure.

C) Root cutting, inspection and packing the lettuce head into the conveyor belt.

In the last stage, the robot picked up fully grown lettuces from a float tray and placed them onto a conveyor belt. At this stage, the lettuces are much larger and much heavier. Thus, a specific type of commercial gripper had been contemplated to manage the right grip and strength to pick up the lettuce. However, to do this the gripper would need to have some additional elements designed for it to better adapt to the task. The root cutting, a crucial part of the process, is currently a work in progress. We are optimistic that this will be achieved in the coming months, further enhancing the efficiency of our system. These innovations are not just innovative; they are pushing the boundaries of what's possible.

Hydroponic farming. As outlined in the previous section, this feasibility project has led to significant innovations in multiple directions ranging from 2D/3D vision system, Collaborative robot motion planning, Design of novel actuators (for transplanting, float handling, harvest room), Modification of the farm with robot workstations, novel

harvesting workflows, onsite system integration. The innovations are beyond the state of the art and offer significant opportunities for further scaling up The deep learning-based 2D/3D vision system is a standout feature of this project. It is designed to detect floats, pick, and place points in a range of noisy conditions such as lighting, soil, motion in water, and occlusion. This system can be applied to all kinds of floats and other plants grown in hydroponic farms. The low-cost camera, priced at £300, makes it a cost-effective computer vision system for monitoring and detection, providing accurate sensory information to drive robotic action. The novel 3D printable 4X transplanting gripper and float handling gripper is a game-changer in the field. It surpasses both the state of the art in robotic grippers and human capabilities in performing the same task. Notably, it presents a remarkable 75x reduction in cost, making it a highly efficient and cost-effective solution. The design of the Cobot workstation in the farm is not just about functionality, but also about safety. It takes into account a range of health and safety considerations, setting a high standard for further robot installations both in JEPCO and other facilities. This design can be a model for replication, ensuring safety and efficiency in all future installations.

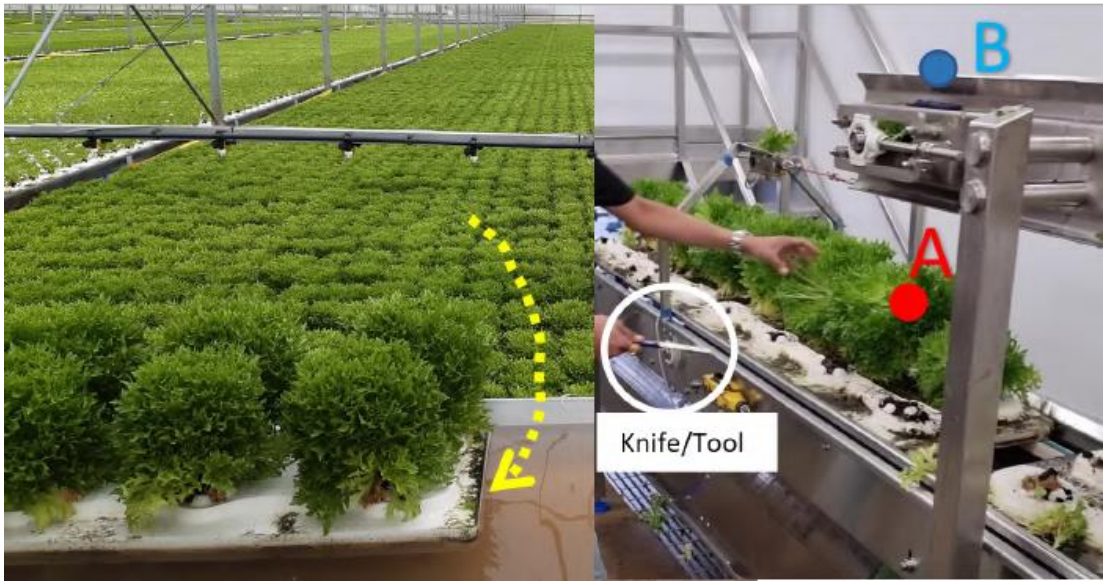


Figure 6 - The transportation task of the fully grown lettuce float in the pool (with 8KG payload and water resistance) is to a water flume that transports the float to the root cutting/packaging station.

All the individual subsystems, i.e. vision, motion planning, gripper control, robot behaviour planning/decision making, have been integrated into SCARLETT software architecture, which can be deployed easily in other farms (both individual subsystems and full architecture- hence opening up multiple directions for exploitation). The industry is facing a critical labour shortage that is preventing it from catching up with the scale and production requirements. JEPCO is developing a 10Ha facility (7.15 times larger) where handling the number of plants at quicker process rates is critical to avoid wastage.

Figure 7 - Dobot CR5 for transplanting in JEPCO. Mounted on a stainless steel stand with OnRobot RG2 Gripper.

1.2.2 Technical Challenges for Robotics/AI addressed in this Thesis

This application has been designed to be deployed on a real industrial environment inside a hydroponic lettuce farm which comes not only with complexity in the robotics side of things but the logistic around it. The key challenges can be listed as:

- **Perception complexity**, since the environment is uncontrolled in the sense of soil presence, occlusion, light environment and changes in position of different components around the robot. Due to this challenges, a deep learning approach with adaptability, accuracy and task constraints in mind needed to be designed and implemented with 3D perception through industrial-grade stereo cameras.
- **Cleaning and food standards requirements**. The workstations as transplanting, handling and harvesting on JEPCO, where the robot is placed, must be sanitized after every work shift, which means both the stereo camera and the arm must be taken out of place to clean completely. This robots are IP54 which means they can resist water jets but are unable to resist a fully humid environment for a long time. This comes with other challenges as the camera and arm calibration, thus more methods were implemented as ARUCO codes to re calibrate the reference frames and get the target coordinate from the camera to the robot correctly.
- **Safety and risk assessment**, the Robot needs to be collaborative and not industrial since while dealing with food processes needs to involve human inspection to reach the food standards. To address this issue, the CR series was

selected due the safe skin feature, that is an electromagnetic layer that is able to detect human contact within the range of 10 cm with 10 mS response.

- **End-tool custom task and development**, to deal with the specific lettuce seedling, float handling and full-grown lettuce head grasping, conventional grippers do not offer a complete solution for both tasks efficiency and cost effectiveness. This motivated the inside development of grippers from the mechanism, electronic development and embedded systems/integration with the robots through RS485 interface with Modbus Protocol. Once the individual lettuce seedling grasping was achieved, a 4X gripper solution was developed and at the time of this thesis it has been filled as patent.
  This specific point is treated outside the content of this thesis due to intellectual property of University of Essex.

1.3 Organization of this Thesis

This thesis begins by exploring the growing significance of robotics across various domains, particularly in agriculture. It highlights the pressing need for innovation due to labour shortages, economic challenges, and environmental concerns in the agricultural sector. The introduction chapter presents the project SCARLETT, an initiative that applies collaborative robotics and 3D vision systems to address these challenges. The critical contributions of the thesis are outlined, emphasising the novel methodologies and technologies developed. Additionally, the chapter provides an overview of the thesis structure and summarises the key achievements realised during the research.

The second˙ chapter offers a comprehensive review of the current state of the art in robotic perception and motion planning, focusing on their application within Project Scarlett. It examines the latest advancements in perception systems, mainly 3D perception techniques for plant detection in agricultural environments, and delves into pose estimation methods that enhance the precision of robotic movements. The discussion then shifts to motion planning, introducing the Passive Motion Paradigm (PMP) as an alternative to traditional Optimal Control methods. The chapter concludes by presenting the Vision-Action System, which integrates perception and motion planning to improve accuracy and safety in robotic operations.

The third chapter focuses on the development of single perception systems tailored for hydroponic lettuce farming. It discusses the application of deep learning techniques for object detection and feature extraction, which is critical for accurately identifying and handling. The chapter also explores hand-eye coordination and calibration processes

necessary for precise robotic operations in controlled agricultural environments. These systems' effectiveness is evaluated through field trials, demonstrating their potential for enhancing agricultural automation.

The fourth chapter presents the task-adaptive motion control systems developed for hydroponic farming, emphasising the Passive Motion Paradigm (PMP). It describes the implementation of collaborative DOBOT arms and the challenges of operating in real-world agricultural environments. The discussion includes managing physical constraints and optimising robotic reach and movement, ensuring safe and effective operations. The chapter underscores the scalability and adaptability of these systems to various agricultural tasks.

The fifth chapter integrates the perception and action systems developed in previous chapters into a cohesive closed-loop system for hydroponic farming. It details the object detection and pose estimation techniques that enable precise plant manipulation using 3D vision. The chapter also addresses the challenges of goal-directed reaching under real-world constraints, particularly in on-site agricultural environments. Additionally, it includes an analysis of the economic benefits of adopting these robotic systems, demonstrating their potential to enhance productivity and sustainability in agriculture.

The final chapter provides a comprehensive summary of the research, highlighting the essential findings and contributions of the thesis. It reflects on the advancements made in robotic perception and motion planning, particularly within the context of Project Scarlett. The chapter also discusses the implications of these findings for the broader field of agricultural robotics. It suggests potential directions for future research, including the development of collaborative robotic systems and the exploration of new applications in agriculture and beyond.

1.4 Summary of Achievements

1.4.1 Summary of Technical Contributions

This thesis highlights the application of deployed vision-action robotics systems in real scenarios combining motion control, 3D computer vision and system integration for real implications for a hydroponic facility. This is summarized as follows:

1) This study provides data generation, training and deployment of the Passive Motion Paradigm (PMP) model for DOBOT CR-Series 6DoF robot arms. This action system was used for goal-directed picking and placing tasks under different constraints and different tasks across the farm. Developing closed-loop systems for this low-cost class of robots represents a leap forward in applied research for Agri-robotics, which presents a shortage of labour.

2) Development towards the automation of the first hydroponic facility in the UK for lettuce transplanting, showing as a case study of what vision-action can accomplish.

3) Introduction of JADE building block as a solution to calculate constraints for the PMP model based on human trajectories. This can be used as a comparison for tasks during a given time where the robot's performance remains constant, in contrast to humans.

4) Development of 3D Key point detection for hydroponic floats, that is robust against occlusion, light-changing conditions, soil and colour, which are crucial conditions for the system to work on a real-world scenario.

5) System Integration and field trials combining constrain motion planning and 3D computer vision, analysing time, and accuracy across three different tasks: seedling transplanting, float handling and packaging. The overall accuracy of the vision action system is under 10 mm error.

1.4.2 Awards and Research Achievements

Awards through this thesis:

- UKRI - AI and Robotics Research Award Winner in the Best Demonstration category (March 2025)
- Essex Innovation Award Rising Star 2024 Best Research Officer of University of Essex working on business-research projects. (September 2024).

Patent Filled March 2025 – Application Number 2503218.6

- 4x Multi-seedling gripper for agriculture.
- Robotic Float handling for hydroponic systems.

Both patient applications have been developed and reviewed and are currently being filed.

1.4.3 Publications

- **Roberto Mendivil-Castro**, Rodolfo Cuan-Urquizo, Tianyue Qin, Fuli Wang, Mohan Vishwanathan (2024). Manipulating Lettuce Seedlings in Hydroponic Farms. "From Expert Humans to Adaptive Robots". The 5th UK Robot Manipulation Workshop, University of Oxford.
- **Roberto Mendivil-Castro**, Jiayu Luo and Mohan Vishwanathan (2024) "Lettuce Seedling Transplanting with Robot Pose Estimation for Cultivation Guidance on Hydroponic Floats". ICRA 40, Rotterdam.

Submitted publications

- Roberto Mendivil Castro, Fuli Wang, Leo Geer, Mohan Vishwanathan (2024) "Task Configurable Adaptive Framework for Robotics in Agriculture, From Soft Fruit Harvesting to Lettuce Seedling Transplanting". Under review in 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2024).
- Roberto Mendivil Castro, Jonathan Dove, Vishwanathan Mohan (2024) "Biomimetic and Adaptive Motion Control Framework for delicate robotic transplanting of Lettuce seedlings in Hydroponic Systems". The draft has been prepared for submission to the Journal IEEE RA-L.

Publications in preparation

- Towards transforming an unstructured farm into a structured assembly line with robotic automation, Target Journal Precision Agriculture, Springer Nature.

- Human-Robot Symbiosis, collaboration, and labour upskilling in Smart Farms-Case study of Robotic Lettuce harvesting, Target Journal Precision Agriculture, Springer, Nature.

Chapter 2
SCARLETT- Survey of the state of the art

2.1 Perception Systems for Crop

2.1.1 From classic computer vision to deep learning detectors

Traditional computer vision, especially in 2D and 3D imaging, has played a transformative role in precision agriculture, enhancing production processes from crop health monitoring to harvesting automation. Low-level computer vision is based on analysing an image as a function of pixels, reading, correlating, and manipulating those pixel values to get specific features from that image. In object detection, classic vision generates a mask, extracting necessary features while operating on the pixel level and identifying what and where the given object is[8]. Then, that detection is referenced in 3D space through 3D vision methods and sensors such as RGB-D and stereoscopic cameras. [9] This is crucial for robotics applications since the main purpose is to interact with the real elements around the robot, which have been detected through vision algorithms.

However, as effective as these technologies have been in monitoring crop health, detecting diseases, and estimating yields by analysing standard RGB images, they have inherent limitations that necessitate integrating more advanced techniques like deep learning. Traditional computer vision refers to those techniques to process an image as a function of pixels and operate over that, while high level refers to deep learning for computer vision where the use of artificial intelligence and neural networks to estimate the content of an image given a supervised learning approach[10], [11]. This leads to a more robust system that can generalize under variable constraints as they are in robotics applications.

Occlusion, light environment and soil are common challenges in the context of robotics in agriculture where the initial conditions of the image may change in the real world, so the system must be able to deal with those changes. Additionally, 2D vision systems are crucial for assessing the quality of agricultural products such as fruits and vegetables. They can detect visible defects and ensure consistency in product grading, thus reducing manual labour. However, these systems are often limited by their reliance on superficial features, such as colour and texture, which can vary significantly under different lighting conditions and other environmental factors.

Moving beyond 2D, 3D computer vision offers greater precision, particularly in field mapping and analysis. By generating detailed 3D models, farmers can gain insights into plant growth, soil conditions, and the topography of their fields, which aids in more accurate planning and resource management[8], [9]. Moreover, 3D vision is vital for robotic systems that handle fruit picking and crop management tasks. These systems rely on depth perception to interact accurately with crops, distinguishing ripe from unripe produce and navigating complex field environments. Despite these advantages, 3D vision systems also face challenges. They often struggle with issues like occlusion (when one object interferes with the vision field), variable lighting conditions, and the high computational demands required to process complex data.

These limitations in classic computer vision highlight the need for more sophisticated methods, such as deep learning, in agriculture. Deep learning models, particularly image-based classifiers and those using convolutional neural networks (CNNs) excel at identifying complex patterns in data that classic vision systems might miss. For instance, while traditional 2D systems might struggle to differentiate between similar shades of green in healthy versus diseased plants, a deep learning model can analyse subtle differences and make more accurate predictions. Deep learning enables more robust image recognition under varying conditions, such as lighting or weather, which is crucial for real-world agricultural applications.[12].

Moreover, deep learning can process vast amounts of data faster and more accurately than classic computer vision methods. This capability is critical as agricultural systems increasingly rely on large datasets, such as those generated by drones or satellite imagery, to make informed decisions. Deep learning models can be trained on these datasets to recognise patterns that are not immediately obvious to identify specific features.

These models require large amounts of labelled data for training, which can be time-consuming and expensive to collect. Additionally, deep learning systems are computationally intensive, often necessitating specialised hardware like GPUs, which can be costly for smaller farms. However, integrating deep learning with classic computer vision represents a powerful hybrid approach that can address many of the limitations of each technology individually, leading to more efficient and accurate agricultural practices.

In conclusion, while classic computer vision has significantly advanced precision agriculture, its limitations—such as sensitivity to environmental conditions and superficial feature reliance—highlight the need for deep learning. Deep learning offers

the ability to analyse complex patterns, improve accuracy under variable conditions, and handle large datasets, making it an essential tool for the future of agriculture. By combining the strengths of both classic computer vision and deep learning, we can create more resilient, efficient, and intelligent agricultural systems that better meet the demands of a growing global population.

2.1.2 Image-based object detection and feature extraction

Object detection has long been a fundamental task in computer vision, requiring the identification and localisation of objects within digital images. Early approaches to this problem primarily relied on manually designed features and region proposal methods. Techniques such as sliding windows and Histogram of Oriented Gradients (HOG) were pivotal in the initial stages of object detection development. However, these methods proved inefficient and inaccurate, significantly when scaled to larger datasets or applied to a broader variety of object categories.

Introducing Convolutional Neural Networks (CNNs) marked a revolutionary advancement in object detection. The advent of the R-CNN (Regions with CNN features) and its improved versions, like Fast R-CNN and Faster R-CNN, significantly improved detection accuracy and speed by leveraging deep learning to extract image features and enhance candidate region selection. These developments represented a logical evolution in object detection technology, enabling more precise and faster performance. [13], [14], [15].

A significant turning point in the field came with the emergence of the YOLO (You Only Look Once) algorithm. The core idea behind this algorithm is to predict the categories and locations of multiple objects simultaneously in a single forward pass, thereby drastically enhancing processing speed. This approach optimised real-time object detection and opened new possibilities for applications such as real-time video analysis. The ongoing improvements in the algorithm series have pushed the boundaries of object detection technology, making it more efficient and accurate for various practical applications.

This work redefined object detection by presenting it as a single-shot regression problem that starts with image pixels and progresses to bounding boxes and class probabilities. The "unified" concept of YOLO allows for the simultaneous prediction of multiple bounding boxes and class probabilities, greatly enhancing both speed and accuracy. Although Redmon ceased his work on the model name after version 3, other researchers

have further developed the effectiveness and potential of the "unified" approach, creating several subsequent versions, including version 8 which this research covers. [16]

The 7th version of this algorithm, launched shortly after version 6, brought significant architectural reforms to improve accuracy while maintaining high detection speeds. These included the Enhanced Efficient Layer Aggregation Network (E-ELAN) [17] Which was inspired by research on network efficiency. Version 8th, released by Ultralytics in January 2023 [18], represents the latest and most advanced iteration in the YOLO series.
Preliminary comparisons with previous versions showed that YOLO-v8 offers better throughput and hardware efficiency, solidifying its status as the new state-of-the-art in object detection technology.[19]

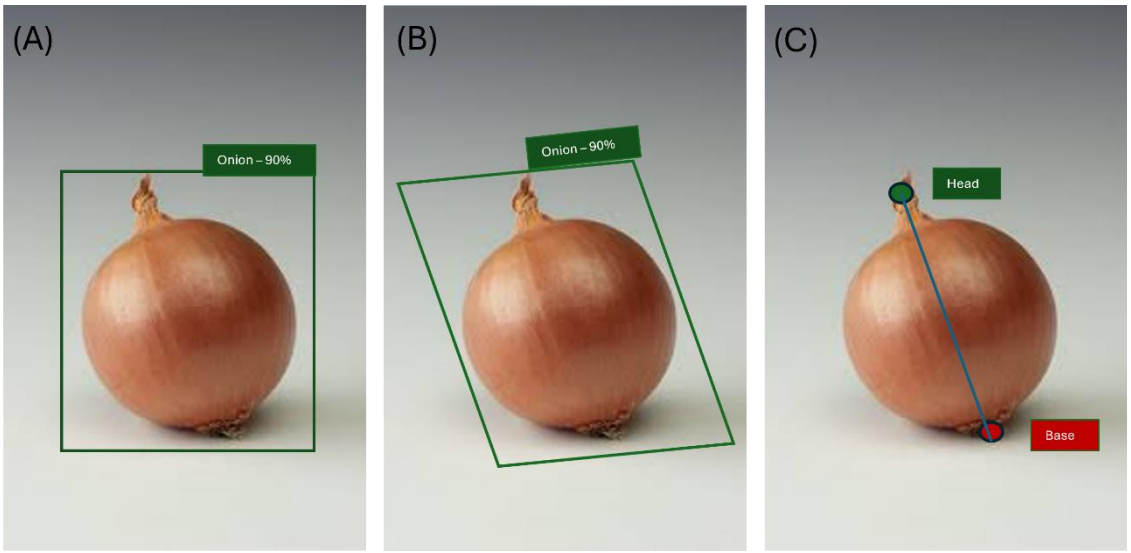


Figure 8 – (A) Indicates traditional object detection, where the ROI is determined by a bounding box from the entire image. (B) Illustrate the Oriented Bounding Box (OBB) where ROI includes a rotation which can be used to extract the pose. (C) The last panel shows key point localisation, where the Head and Base of the onion are localised and the correlation of both can be used to get the entire pose estimation or as directly treat them as grasping points.

A solution to find specific grasping points from a detection is to post-process the object detection given by the CNN model and feed that information into a second network where we specifically categorize the dimensions and grasping, combining YOLO and VGG as [20]. While this may be seen as an elegant solution, reducing the vision pipeline to a standalone neural network and getting all the data from it, represent a better alternative for a low-cost and real-time solution as agriculture applications require.

Beyond object detection but not considering object segmentation, we can find Oriented Bounding BOX (OBB)[21] and key point localisation as two interesting and innovative solutions to get the whole data from the data. Key point detection also known as pose estimation [22], [23] involves identifying specific points that typically represent joints of the human body or interesting features from an object. Each point represents $P\{x,y,c\}$ and

those can be interconnected to find the pose of a given object as can be seen in Figure 8: object detection by itself may not be sufficient to get enough pose, and specific features for robotics grasping hence it will be studied over the rest of this thesis and how can be used in lettuce harvesting, extracting the detected 2D pixels to 3D coordinates with a stereo camera. Key point localisation its been tested with the latest robotics manipulation approaches where the points are embedded as reasoning sentences powering a large language model LLM [24] since the specific grasping point can be passed through. This is a huge example of exploring key point detection for real-world applications focused on robotics grasping can be a valuable research topic to explore.

2.1.3 3D perception for grasping in robotics

Once the ROI (Region of interest) has been found and placed inside the image on a bounding box (BB) described as a vector *[x,y,w,h]* where the pixel position *P(x,y)* is expanded plus with and height relatives to the original image sized, displaying a box. In robotics, detecting the bounding box may not be sufficient to interact with the given object, since we need to know A)3D position relative to the robot reference frame and B) the Grasping point which the robot will reach, interact with and manipulate.

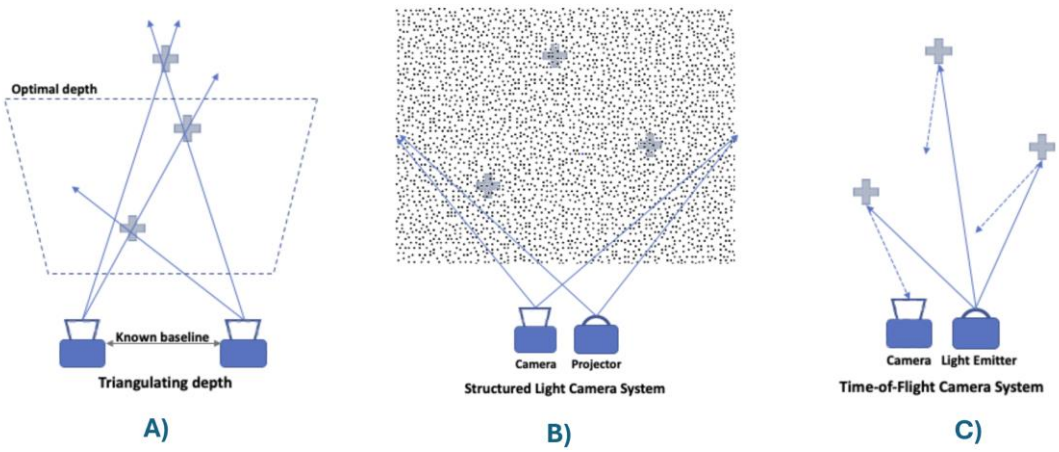


Figure-9 - 3D Cameras technologies. A) Stereoscopic image, which uses the dual camera configuration and the known baseline to triangulate the left camera pixel position to the 3D space. B) Structured Light camera which is a dual configuration of a projector and RGB camera. C) ToF uses the time of light reflection to estimate the distance[25]

The 3D position of a pixel *P(x,y,z)* is obtained by processing the original 2D point with a specific sensor, which can be a stereoscopic or RGB-D camera, where the first one is a dual camera configuration to add a depth map from the 3D Space. An RGBD camera technology can be either structured light where the camera produces a narrow band of light which allows one to see the object from another perspective and estimate the global

position on the 3D pace with light reflection, or time of flight (ToF) using the time that takes light to reflect.

RGBD cameras are accurate, cost-effective and not power processing hungry since the same outputs the process depth map. The trade-off of these technologies is that any abrupt change in the light environment produces a chance for the 3D reconstruction, which makes them unreliable for outdoor applications, and in this case, agriculture tasks. For that reason, stereoscopic cameras are ideal for outdoor, high-precision 3D vision applications, since the 3D position is obtained by a triangulation between the known baseline or distance between the two cameras. The only trade-off is the computational cost due to the need for an online calculation made by the device where the camera is connected. That is the reason stereo cameras are considered a 3D passive sensor while RGB-D is active since each pixel already has the depth estimation included.[26], [27]

Object detection in computer vision refers to solving the problem of what and where the object is in the image relative to the maximum dimensions of the given image. This may not be sufficient to accomplish a robot-vision system, since for a grasping application more information as the grasping points given the specific class. Once we identify the ROI of the grasping, then we convert those pixels in 3D space with the used sensors to finally achieve real-time grasping.

2.2 Motion Planning in Agricultural Robotics

2.2.1 Optimal and Impedance control motion planning

Motion Planning is a critical component for robotics systems particularly in agriculture applications where dynamic and challenging environments are a fundamental requirement. It is defined as the calculation of all the trajectories and the optimal to get the robot planning from 2 or more set points, reaching those at a minimum error value considering external factors such as task constraints and collision avoidance.

One of the main traditional approaches for robot motion control is Optimal Control (OCT), which involves determining an optimal control scheme from a class of allowable control variables by cost function definition. This can be further described mathematically as obtained cost function minimum value under motion equation constraints and allowable control variables, such as the minimization of energy, time or errors while determining optimal control trajectories.

Mathematically, optimal control aims to minimize the cost function over a time horizon with discrete conditions which are not expected to change. Where J represents Cost function, x the stated (initial and desired), u the control input and T the time horizon. As shown in (1.1):

(1.1)

$$J = \int_0^T \left[\left(x(t) - x_d(t)\right)^T Q \left(x(t) - x_d(t)\right) + u(t)^T R u(t) \right] dt$$

The concept of robotic motion control predicated on the Optimal Control Theory (OCT) encompasses identifying a superior control strategy derived from a set of permissible control variables. This objective can be realised by formulating an objective or const function that mathematically can be articulated as the limitations imposed by the equations of motion and the permissible control variables determine the extreme value of the objective function (specifically, the minimum value of the cost function).

A substantial corpus of literature emerged, comprising analogous investigations that employed diverse objective functions, including integrated torque change and minimal object crackle. [28], and the minimum acceleration criterion [29]. Contemporary advancements suggest that OCT has progressively established itself as a robust theoretical framework for elucidating a spectrum of motor behaviours, facilitating online movement corrections, and analysing the structure of motor variability. Furthermore, various inverse kinematics algorithms have been integrated into specialised motion planning software to

ascertain the optimal solution for the manipulator. [30] [31] Nevertheless, a principal challenge inherent in this methodology lies in deriving the optimal control signal for a nonlinear time-varying system predicated on a specific objective function and assumptions about the noise structure[32]. The mathematical intricacies involved in the computation of an optimal feedback controller are exceedingly complex [33]. Moreover, the propensity to become trapped in local optima constitutes a prevalent issue within optimisation algorithms. The nonlinear optimization framework pertinent to inverse kinematics demands further investigation.

An impedance control is a class of controller that focus on the relation between the system, its environment the effect that produces to it and how the external condition affects itself. It can be defined as the dynamic of the interaction between physical systems. For the case of a robot manipulator, it is important to consider and distinguish both the impedance and also the admittance as the inertial objects involved in such manipulation.

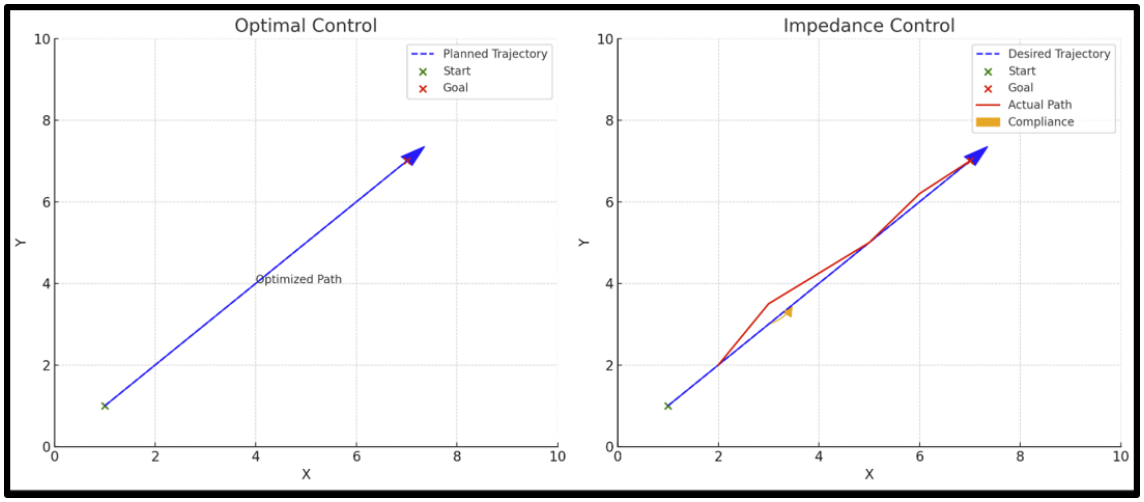


Figure 10 In Optimal Control (left), the robot follows an optimized trajectory from start to goal (pre-computed), minimizing a cost function without deviation. In Impedance Control (right) the path is updated by external forces (contact with the environment, obstacles), shown by the deviating red path from the desired behaviour, representing compliant behaviour needed in adaptive robotics applications.

Then, impedance control establishes a dynamic relation of interaction between physical elements, mostly of position, velocity and acceleration $(x, \dot{x}, \ddot{x})$ and desired counterparts, with the force F and control variables to adjust the compliance of the system established by stiffness K, Damping D and Inertia M. ((1.2)

(1.2)

$$F = M(\ddot{x} - \ddot{x_d}) + D(\dot{x} - \dot{x_d}) + K(x - x_d)$$

To shift the cost function in OCT to the force field in impedance control, a neural network implementation of the PMP [34] has been developed for robot manipulation based on the EPH [35], [36]. Qualitatively, the process by which the brain determines the distribution of work across a redundant set of joints when the end-effector is assigned the task of reaching a target point in space can be represented as an "internal simulation process" that calculates how much each joint would move if an externally induced force (i.e., the goal) pulls the end-effector by a small amount toward the target.

The mechanism labelled "passive" aligns with the EPH because the brain does not explicitly specify the equilibrium point; instead, it contributes to the activation of "task-related" force fields detailed in this novel perspective of viewing motor control and summarises the principle of a neural network implementation of the PMP.

Recently, PMP has been applied in different contexts, such as combining postural and focal synergies during whole-body reaching tasks [37] and coordination of the movements of the upper body of the iCub along with the paintbrush to derive motor commands for drawing the shapes [38]. Regarding the application of the agricultural robot, this thesis applied the PMP for goal-directed reaching considering various task constraints (e.g., gripper pose, joint limits, timing, bimanual coordination, and alignment of the gripper/cutter to the stem). The action system is a forward/inverse model that can simulate the consequences of actions for predictive planning and extend to a range of tools coupled to the arm.

A huge advantage of PMP is motion under the passive dynamics since it's a model-based controller that does not depend on external hardware, such as force and torque sensors placed on the End-tool or individual input per joint, enhancing the use of cost-effective robots. This type of system contributes on creating robotics frameworks independent of specific hardware.

2.3 Related 3D Robot-Vision systems for agriculture

Over the past years, with the rise of artificial intelligence, computer vision and the price reduction on robotics systems, the impact of robots in agriculture has been emerging as current autonomous robots for crop harvesting that combine 3D Vision systems and motion planning consolidate the research field. To achieve a harvest application in robotics, the above concepts as stereo vision, object detection feature extraction as well and impedance control play a fundamental role in leveraging robots in real-world agriculture workstations.

A 3D vision with an impedance control system applied to agriculture is the dual-arm robot developed for harvesting tomatoes in a greenhouse [39].However, the DoF of this type of double manipulator represents restrictions under uncertain conditions. To improve the success rate, a computer vision approach was implemented for optimal sorting and fruit nearest neighbour positioning algorithms were developed for determining the position of the tomato fruit and estimating the grasping pose [40], [41].

Specifically for lettuce harvesting applications, the literature is constrained since this approach of adapting a farm and converting it to a manufacturing line embedding robotics can adapt under the uncertainties of dealing with food, moisture and living tissues (plants). Similar approaches can be found for iceberg lettuces in [42] The challenge was to detect the crop in 2D space, align the robot base with the end tool, and repeat the process over the cultivation line until all the crops had been grasped. Another example is in [43] where a computer vision-based robot harvests lettuces on hydroponic setups, which are placed on plastic tubes, that are related to the hydroponic floats, this thesis will be exploring. The setup is a simplified version of what this research tries to accomplish since the lettuces are static and always in the same place, where the challenge is identifying the crops with traditional computer vision, and then correlating the lettuces between them by incorporating linear padding. The grasping is done with just 2D vision and traditional position control, calibrating robot and camera frames. While this is an interesting reference point, the setup is a prototype workstation and not an industrial facility with ongoing production as this thesis is focused on.

Chapter 3
Configurable 2D/3D Perception for Robotic Lettuce harvesting

3.1 Vision methods for agriculture vision

3.1.1 Object detection and feature extraction

The central objective of this project is to develop a sophisticated deep learning-based 2D/3D vision system designed to automate the identification and localisation of lettuce plants grown on floats in hydroponic farms. The system's architecture is meticulously crafted to manage the entire workflow, encompassing image acquisition, data preprocessing, feature extraction, object detection, and, ultimately, the precise localisation and classification of objects. This architecture is structured into three distinct stages: the 3D camera integration and data acquisition, the deep learning models at the intermediate level, and dexterous manipulation by goal-directed reaching with grasping.

The system then relies on robust hardware facilities. Central to this layer are ZED stereo cameras, which are employed to capture high-resolution images of the farm environment. These cameras are chosen not only for their ability to produce high-quality images but also for their stability across varying lighting conditions, ensuring the reliability and consistency of the data captured. The quality of the hardware at this foundational layer is crucial, as it forms the basis for all subsequent processing and analysis.

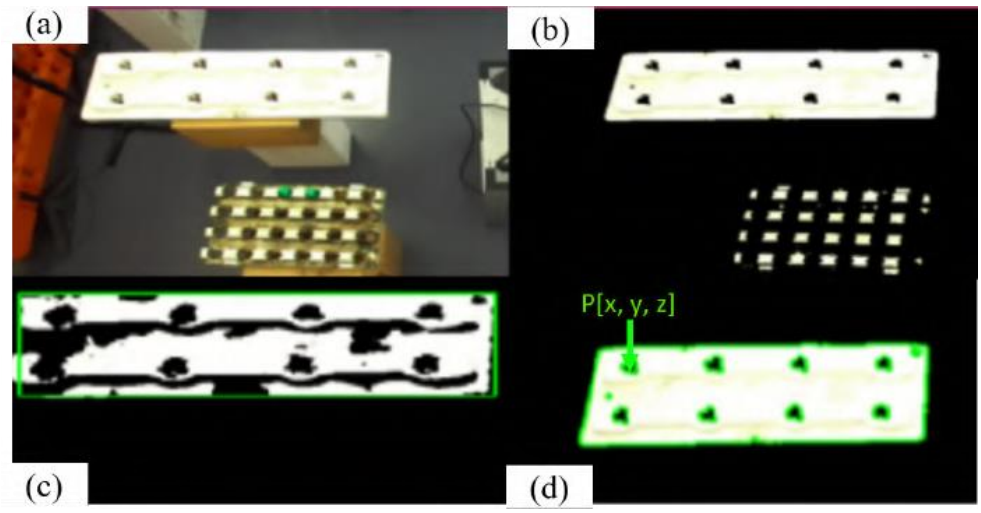


Figure 11 Classic Computer Vision Approach. Template matching and colour filtering and 3D localisation from pixel point.'

The image data gathered from the bottom layer is processed and analysed here. The middle layer executes critical functions such as data preprocessing, feature extraction, object detection, and object localisation. By leveraging models from the YOLOv8 series, the system achieves a high degree of precision in identifying the exact locations of the lettuce floats and their specific planting points. Integrating these deep learning models is essential for transforming raw image data into actionable insights, which are then used to guide the system's operations.

This layer is responsible for interfacing with automated robotic arms and other automation equipment to perform specific tasks, such as harvesting and transplanting lettuce. The control and application layer ensures that the insights generated by the deep learning models are effectively translated into real-world actions, completing the automation cycle. This layer is where the theoretical and computational elements of the system converge with practical agricultural applications, enabling a seamless transition from data analysis to physical operations in the farm environment.

By structuring the system in this layered manner, each component—hardware, deep learning models, and control mechanisms—creates a comprehensive, efficient, and scalable solution for automating critical processes in hydroponic farming. This architecture not only enhances the precision and efficiency of farm operations but also represents a significant step forward in integrating advanced robotics and AI into agriculture.

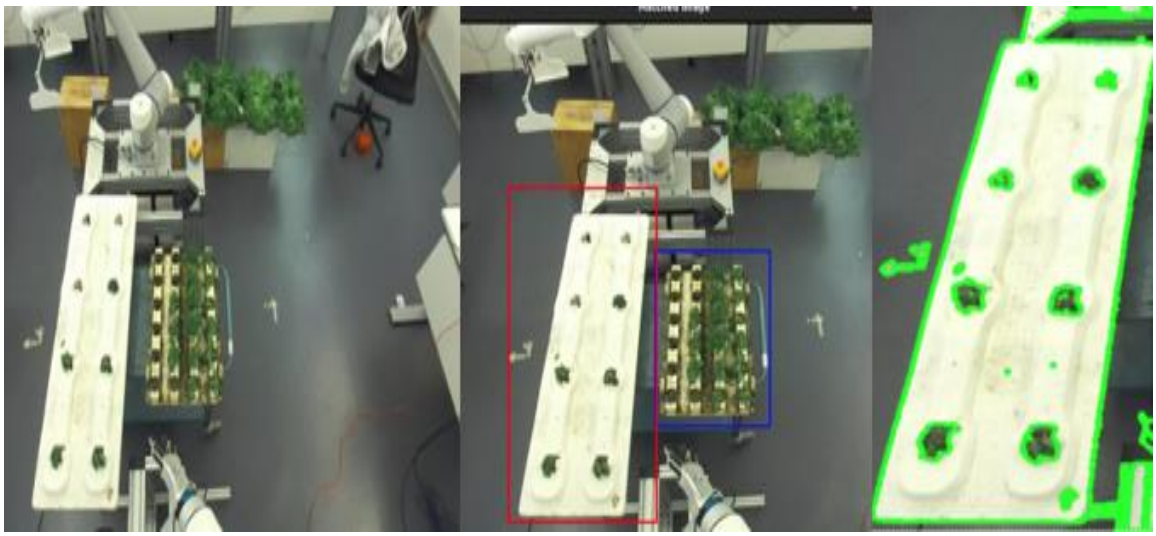


Figure 12 Feature Extraction ad image level and template matching for object detection. After the Bounding Box is detected, a contour detection is performed to extract the placing holes on the float.

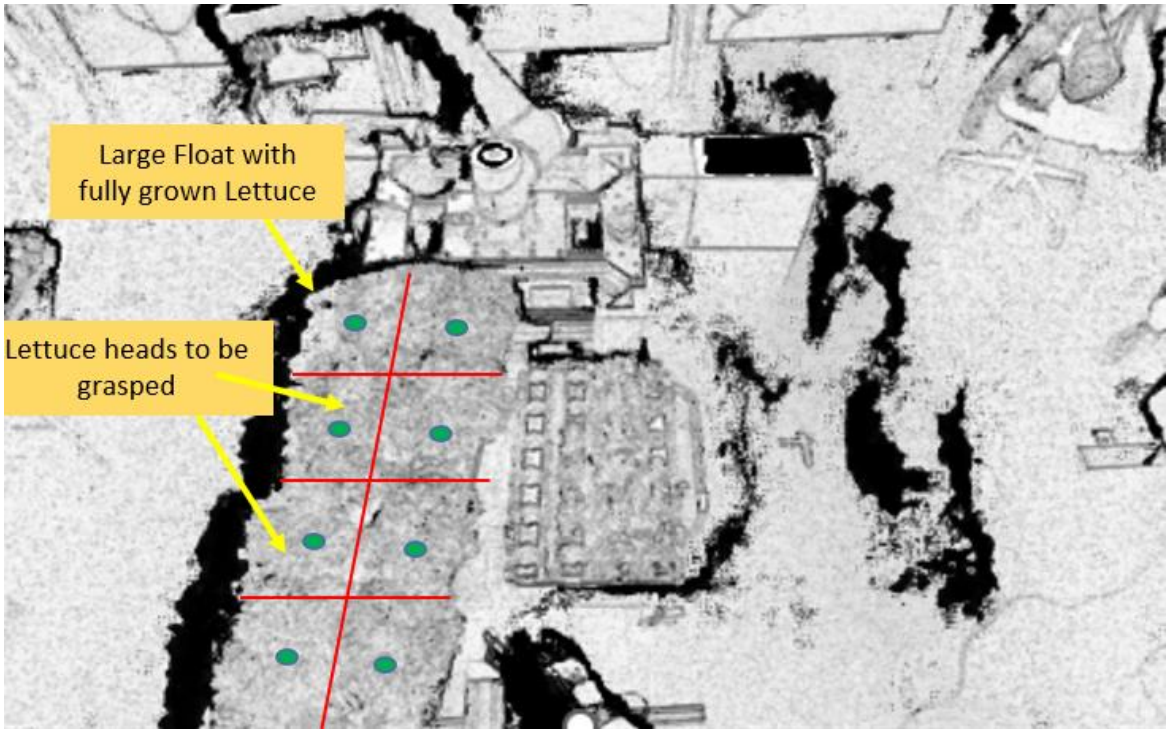


Figure 13 Point Cloud thickness approximation of crops.

System Flexibility and Scalability

The primary goal of this architectural design is to enhance the system's flexibility and scalability, enabling it to adapt to various crops and complex agricultural environments. The layered architecture allows each component to be independently optimized and upgraded without disrupting the overall system's operation. For instance, when new image recognition technologies become available, the deep learning models in the middle layer can be upgraded independently without replacing hardware facilities or redesign the top-level applications. This modular approach ensures that the system remains sustainable and efficient, providing a robust foundation for future developments in agricultural technology.

This project has adopted the open-source object detector series as the core deep learning model to precisely identify and localise the floats and their planting points within a hydroponic farm. This includes the foundational object detection model, the OBB (Oriented Bounding Box) model, and the pose model. The YOLO (You Only Look Once) series is renowned for its exceptional real-time object detection efficiency and accuracy, and it has been widely applied across various real-world scenarios.

Regarding specific model applications, YOLOv8-obb can output rotated bounding boxes, which help recognise and locate floats that appear at various rotational angles under the camera. Meanwhile, YOLOv8-pose, which integrates key point detection, can

precisely identify the specific locations of planting points on the floats. This capability is crucial for precisely operating automated devices such as robotic arms, ensuring accuracy in tasks like planting and harvesting.

3.2 Object detection and Key point localisation

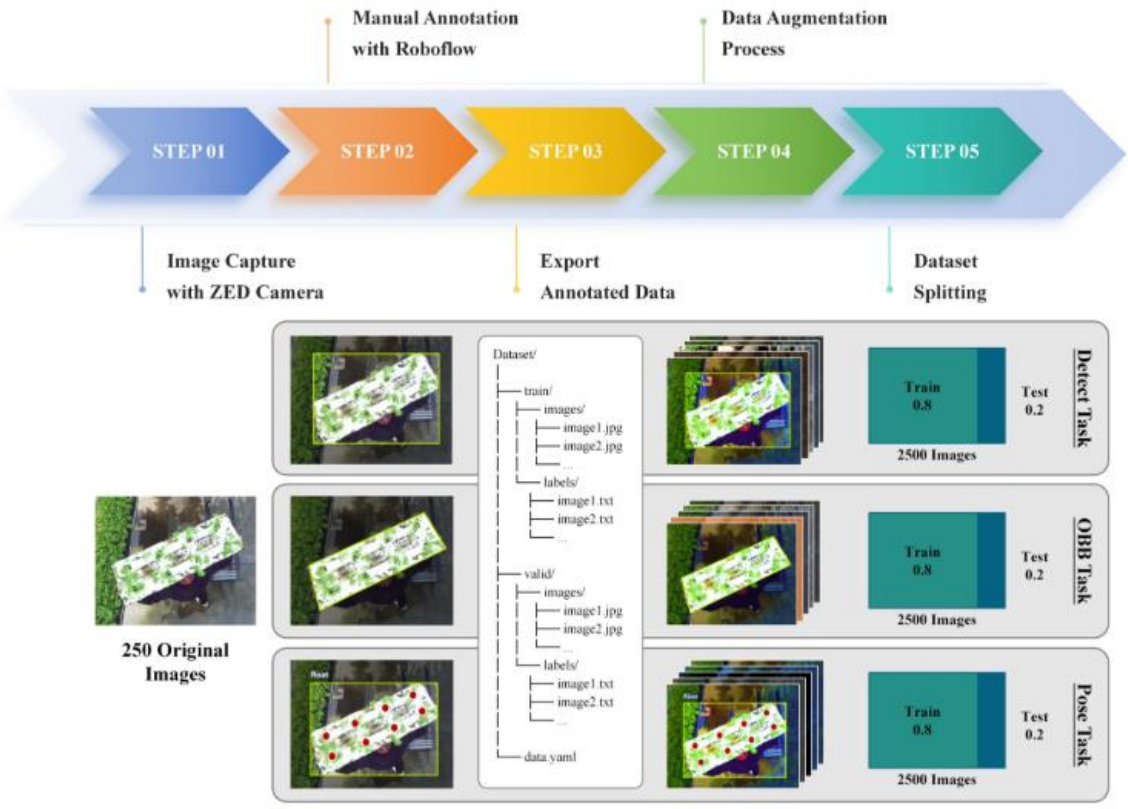


Figure 14 dataset preparation process, including the collection, annotation, export, enhancement, and final segmentation into training and testing sets, with datasets created from 250 original images from the farm and augmented to 2500 in total.

Initially, 250 scene images captured by a ZED stereo camera were imported for manual annotation, as shown in Figure 14. By applying five augmentations per image, five new images from each original, calculated and saved the new bounding box information, thereby expanding the dataset size. In total, including the

original images, a dataset of 2,500 images was constructed. The dataset was then randomly split into a training set (2,000 images) and a validation set (500 images) in an 8:2 ratio, thus completing the dataset construction. Subsequently, the dataset was converted to the YOLO format and exported to the local system. Using the imaging library; the images were resized to maintain their original aspect ratios, resulting in a resolution of 640x360 pixels. A series of data augmentation techniques were also applied, including but not limited to random rotations, scaling, flipping, brightness and contrast

adjustments, and random occlusions, with the detailed parameters of these image enhancement techniques

Figure 15 Lef panel Object Detection. Right OBB Oriented bounding Box.

Figure 16 Keypoint localization from object detection. Ideal method.

Although the initial target detection model was capable of identifying the positions of floats in a hydroponic farm with considerable accuracy, it only provided the bounding boxes without revealing specific directional information within these boxes, such as the rotation angles of the floats, as shown in Figure 16. To obtain more precise positional information on the floats to aid subsequent robotic operations, we shifted to using the OBB model. This model can recognise the orientation of objects and provides detailed parameters of the rotated bounding boxes.

Given the model transition, it was necessary to re-establish a dataset compatible with the model. The overall setup process was similar to the previous one, with the main differences being in the manual annotation and data augmentation stages. During the annotation process on Roboflow, we drew polygonal frames that fitted the edges of the floats instead of the regular rectangular frames used previously. We also ensured that the exported format was compatible with the YOLO format adapted for the OBB model. The data augmentation stage was also adjusted due to changes in the structure of the label files. We modified and recalculated the positions of the labels after applying various enhancement techniques. Specifically, the model processes bounding boxes in the format, where *[x,y,w,h,r]* represents the coordinates of the centre point of the bounding box, width, height, and rotation angle,

3.3 Evaluation from field trials

The key-point detection method was the most precise and overall suitable for this application since it is an open-ended approach. Since we don't need to classify the float but get the picking and placing points.

Metrics including Precision, Recall, and average Precision (MAP). Precision refers to the proportion of true positive samples among all samples identified as positive by the model. It is a metric that measures the accuracy of the model in predicting positive classes.

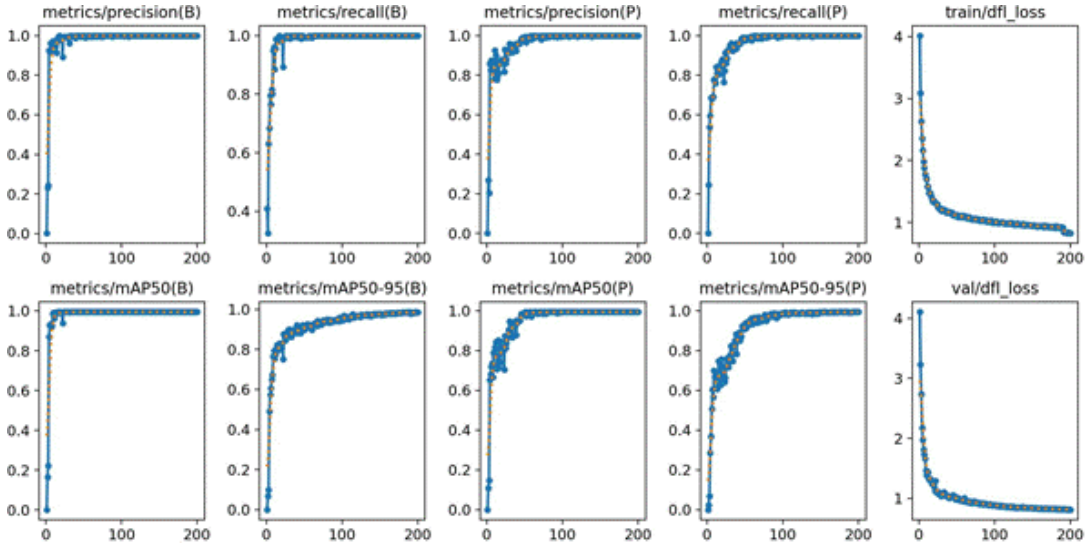


Figure 17 Keypoint Localization mean average precision (mAP) 90%.

(3.1)

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

where, TP is the number of true positives, i.e., the number of samples correctly identified as positive; FP is the number of false positives, i.e., the number of samples incorrectly identified as positive. Recall refers to the proportion of samples correctly identified as positive by the model out of all actual positive samples. It measures the model's ability to capture positive samples.

(3.3)

$$Recall = \frac{TP}{TP + FN}$$

In object detection tasks, model performance is often measured by calculating the Average Precision(AP) and the Mean Average Precision (mAP). AP is obtained by integrating the area under the Precision-Recall curve generated at different confidence thresholds, reflecting the model's accuracy at various recall levels. mAP is the average of the AP values across all categories, used to comprehensively evaluate the model's accuracy at various recall levels.

(3.3)

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i$$

The loss and performance metrics for the object detection model were used during training on both the training and validation datasets. In this figure 17, the blue points represent the loss value for each epoch, while the orange dashed line shows the trend of the smoothed loss values. The losses include box_loss for bounding box discrepancies, cls_loss for classification errors, and dfl_loss for directional discrepancies. Specifically, λ box_loss represents the discrepancy between the predicted bounding boxes and the actual bounding boxes, measuring both location accuracy and size accuracy. A lower box loss indicates more precise object localization and size prediction by the model. λ Classification loss, cls_loss, focuses on the accuracy of the model's predictions of object categories. It is calculated based on the difference between the model's output probability distribution for classifications and the actual labels. A reduction in classification loss signifies improved performance in distinguishing between different object categories. λ Directional loss, dfl_loss, is specific to models designed for tasks involving estimating direction or angle, measuring the discrepancy between the model's predicted object orientations and their actual orientations. The loss curves for the training and validation sets show the trends in bounding boxes, classification, and directional losses throughout the training process. Initially, the losses are relatively high, indicating lower accuracy in object boundary recognition by the model.

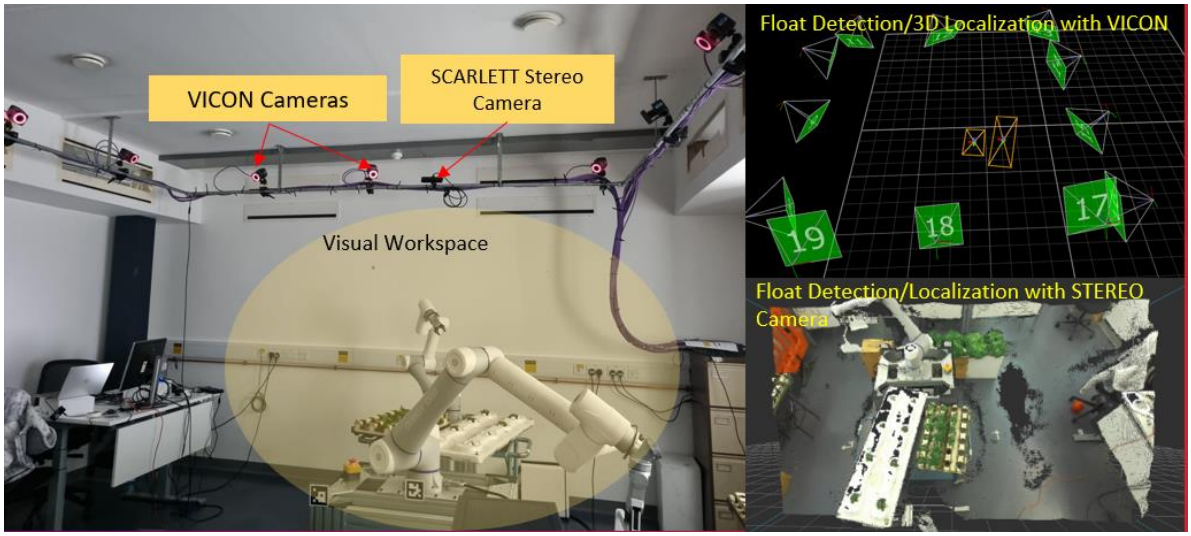


Figure 18 Perception Laboratory Setup. Vicon Tracking System as ground truth to estimate the accuracy of 3D localisation.

Figure 19 Farm Setup. ARUCO markers system as a reference for 3D localisation.

The test results reveal that all three models accurately identified the float regardless of whether in the laboratory or on the farm under varying conditions. The object detection model was able to frame the float fully, but it could not pinpoint the specific location of the float within the bounding box. The OBB(Oriented Bounding Box) model provided rotated bounding boxes based on the float's orientation, 29 enhancing the accuracy of position recognition; however, since the float does not always appear as a standard rectangle under the camera, even with rotated bounding boxes, it was still impossible to determine the precise locations of the planting spots from the outputted bounding boxes.

In contrast, the key-point detection model demonstrated greater utility in identifying the float and picking and placing grasping points' positional information. This is immensely beneficial for guiding robotic arms in operations such as transplanting and harvesting.

Metric	Training	Validation	Metric	Bounding Box	Pose
Box Loss	0.21792	0.2067	Precision	0.99971	0.99971
Pose Loss	0.08183	0.05978	Recall	1	1
Keypoint Object Loss	0.01076	0.00243	mAP@50	0.995	0.995
Classification Loss	0.17677	0.1646	mAP@50-95	0.9839	0.99452
Directional Loss	0.82428	0.813			

Figure 20 Metrics of vision models, detection only on floats,

This study successfully developed a deep learning-based two-dimensional/three-dimensional visual system designed for automated robots harvesting lettuce in hydroponic farms. The system effectively addressed the challenges of detecting and locating lettuce plants at various growth stages on floating rafts in the water. It estimated the picking and placing points for seedling transplantation under noisy conditions. The visual system was trained using the YOLOv8 series models, incorporating actual image data from the farm and synthetically generated images to adapt to varying lighting conditions, obstructions, and real-world noise factors such as impurities and soil. Collecting real farm images and applying image enhancement techniques, three distinct datasets were constructed tailored for object detection, OBB, and critical point detection models. These models demonstrated high accuracy in the laboratory and real farm environments, with precision on the validation set exceeding 0.99. While the object detection model provided highly accurate predictions, it failed to display specific positions of the rafts within the bounding boxes, such as rotation angles. The OBB model provides rotated bounding boxes to better fit the edges of the rafts and improves alignment despite some shape distortion issues.

The key point detection model, by integrating bounding boxes and key point 31 localisation, perfectly resolved the challenges of accurately detecting picking/placing points. Specifically, precision and recall reached optimal values of 0.99971 and 1, respectively. The mean Average Precision(mAP) was 0.995 at an IOU threshold of 50%, and ranged from 0.9839 to 0.99452 between 50% to 95%. The visual system was successfully field-tested at JEPCO Glebe’s 1.1-hectare lettuce using a ZED stereo camera as the visual input. The system's outputs are currently used to control a collaborative robot arm (DOBOT CR10), automating labour-intensive, repetitive tasks such as seedling

transplantation, raft handling, and packaging of mature lettuce, significantly enhancing the farm's production efficiency and operational precision.

On Figure 20, metrics and evaluations of the computer vision model are expressed were the Pose Lose is rounded to 1 (0.9998), which was cause the evenly distribution of the points from the bounding box was constant, since YOLO model performs object detection first over the whole image and then get the distribution of the points. Unlike human pose estimation where body composition can vary, placing points on floats is evenly distributed and considering 3D estimation, the loss may be seen unaltered.
These mathematical losses are described as follows: in (3.4) Object loss, which correlates the discrepancy between the detected bounding box and the ground truth, using the distribution focal loss [44].

(3.4)

$$\mathcal{L}_{\text{bo}} = \text{DFL}\left(B_{\text{pred}}, B_{\text{gt}}\right)$$

Key point loss is defined by the computation of the weighted Euclidean distance between detections and ground truth considering key point scalability defined on annotations.

(3.5)

$$\mathcal{L}_{\text{k}} = \sum_{i=1}^{K} v_i \cdot |\boldsymbol{k}_i^{\text{pe}} - \boldsymbol{k}_i^{\text{g}}|_2$$

The directional loss has the purpose of angle correction based on object orientation. In this case, Pose loss is 1 due to the same angle between the evenly distributed key points and this ensures the correct spatial relation instead. Is defined on (3.6) where the cosine ensures smaller angular deviations contribute less to the loss.

(3.6)

$$\mathcal{L}\text{dr} = \sum_{i=1}^{K} v_i \cdot \left(1 - \cos\left(\theta_i^{\text{pred}} - \theta_i^{\text{gt}}\right)\right)$$

Chapter 4

Configurable Action: Task Adaptive Robot Motion Control for Lettuce Harvesting

4.1 Passive Motion Paradigm

4.1.1 Internal body of the body - Collaborative DOBOT arms

The biomimetic aspect of this thesis involves the mapping of human features performing the given tasks to the robot's body composition. This is achieved by mapping the robot's internal joint space and the external 3D space and then training an artificial neural network (ANN) that encodes this relationship is essential. An ANN is a mathematical model that mimics the structure and function of biological neural networks to achieve artificial intelligence. Commonly referred to as a neural network, it can efficiently estimate or approximate functions [32].

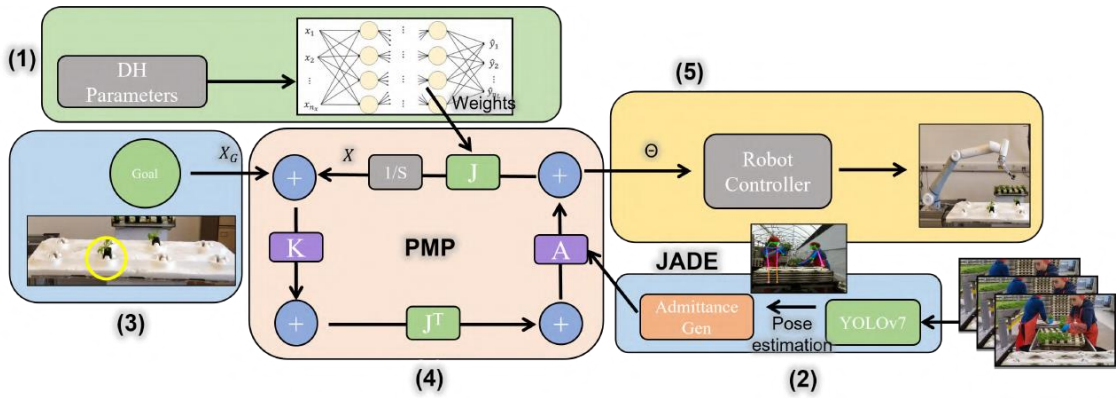


Figure 21 JADE - Joint admittance from human pose estimation for specific joint constraints from human manipulation.

A single-layer neural network, also known as a perceptron, consists of only one layer of neurons between the input and output. This type of network can only express linear relationships, which limits its ability to represent more complex logic. In contrast, a multi-layer neural network is composed of stacked single-layer networks, creating a hierarchical structure. Typically, a multi-layer neural network includes three main components: the input layer, which receives non-linear input information from numerous neurons to form an input vector; the output layer, which generates an output vector by transmitting, analysing, and weighing information through links between neurons; and the hidden layer, which consists of multiple layers of neurons and links located between the input and output layers The hidden layer may contain one or several layers with an arbitrary number

of nodes. As the number of nodes increases, the non-linear characteristics of the neural network become more pronounced, enhancing its robustness and making it more capable of handling complex tasks.

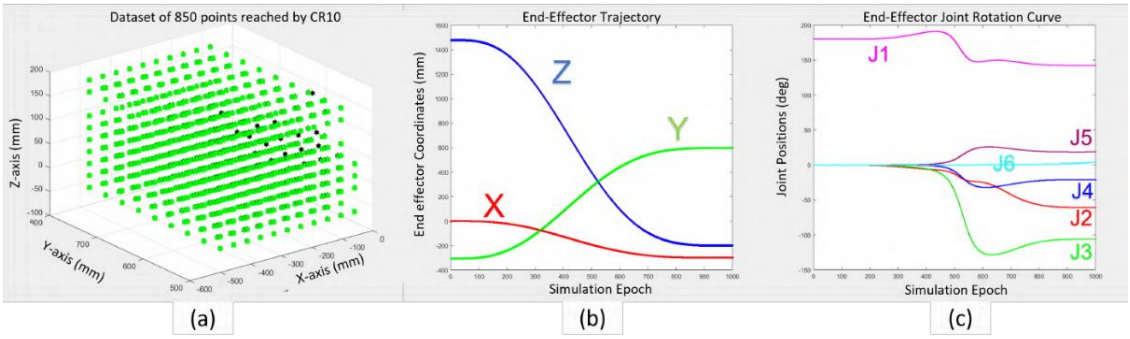


Figure 22 CR10 2mm accuracy ANN-based motion planning.

In this project, a biomimetic neural network based on the Passive Motion Paradigm (PMP) is employed to implement impedance control for robotic manipulation. The PMP algorithm is a control method grounded in dynamic modelling, which treats the robot as a passive object and utilises a passive motion model for motion planning and control. The dynamic model can be approximated by leveraging PMP to train the neural network, transforming the robot control problem into a weight calculation problem. This approach enhances the accuracy and stability of robot control and offers strong robustness and adaptability to varying conditions.

The following steps are used to establish a neural network for PMP within the CR10 architecture:

Generate the Dataset: Using the Denavit-Hartenberg (D-H) parameters of the CR10 robot, the kinematic and dynamic models of the robot are established. This process generates a large dataset of motion trajectory data, including the position and orientation information of the robot's end effector and the corresponding joint angle data. This dataset serves as the training data for the neural network model.

Establish the Neural Network Model: MATLAB is used to create the neural network model, which includes the input layer, hidden layers, and output layer. The input layer receives the end effector's target position information, while the output layer generates the corresponding joint angle data needed to move the end effector.

Train the Neural Network Model: The generated dataset is input into the neural network model as the training set. The backpropagation algorithm is employed to adjust the network weights, ensuring that the joint angle data output by the network accurately positions the robot's end effector at the target coordinates.

Verify and Optimize the Neural Network Model: A test set is used to verify and further optimise the well-trained neural network model. Parameters and the network structure are adjusted to improve the accuracy and robustness of the robot's motion planning and control. This neural network implementation plays a crucial role in the precise transplantation operation of the CR10 manipulator, enabling the robot to execute complex tasks with high accuracy and stability.

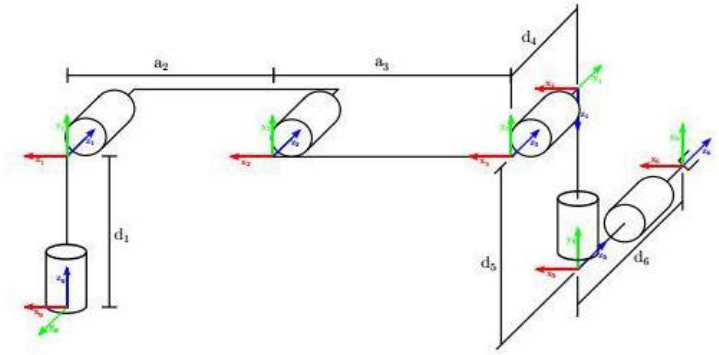
Kinematic Analysis of the CR10 Robot Arm Before performing a kinematic analysis of the CR10 robot arm, it is important to understand the method used to describe the robot arm's pose. The Denavit-Hartenberg (D-H) parameters provide a mathematical model of the arm, utilising a coordinate system to express the positional and angular relationships between pairs of joint links through four parameters. In the D-H parameter system, the axis of a joint is defined as the z-axis, and the common normal between joints is defined as the x-axis, with the x-axis direction pointing towards the next joint.

Using the coordinate system established on the previous joint as a reference, the motion of one joint drives the motion of the subsequent joint, leading to changes in the position and orientation of the coordinate system fixed on the joint. Consequently, as each of the six joints rotates by specific angles around their respective joint coordinate systems, a cumulative effect occurs, ultimately controlling the motion of the end effector.
Understanding the D-H parameters is crucial for effectively controlling the CR10 robot arm and ensuring that the end effector achieves the desired position and orientation through precise manipulation of the joint angles.

Kinematic Analysis and Dataset Generation for the CR10 Robot Arm
Utilizing the mathematical capabilities of MATLAB, the kinematic analysis of the CR10 robotic arm begins with defining the Denavit-Hartenberg (D-H) parameters. Once these parameters are established, the transformation matrix for each joint can be derived. By sequentially multiplying the transformation matrices for all the joints, the homogeneous transformation matrix for the end effector is obtained. This matrix is essential for constructing the forward kinematic equation of the CR10 robotic arm, which establishes the relationship between the joint angles and the position and orientation of the end effector.
In addition to this kinematic analysis, MATLAB's built-in random number generator was employed to create a comprehensive dataset consisting of 50,000 random sets of CR10 joint angles along with their corresponding end effector positions. It is important to note that the angle of joint 6, which does not significantly influence the end effector

coordinates, was fixed at either 90 degrees or 0 radians to improve the efficiency of training the neural network.

Generating this extensive dataset of random joint angles and their associated end effector positions is a critical step in training a neural network for the CR10 robotic arm. The dataset serves as a foundation for the network, providing a vast array of input-output pairs that enable it to learn the complex relationship between the joint angles and the resulting end effector positions. By including a wide variety of joint angle configurations in the dataset, the network is better equipped to generalise across different scenarios and accurately predict the corresponding positions of the end effector. This process enhances the network's predictive capabilities and ensures that it can handle the diverse range of movements required for effective robotic operation.



	Theta[rad]	a[m]	d[m]	Alpha[rad]
Joint1	0	0	d1 = 0.176.5	0
Joint2	-π/2	0	0	π/2
Joint3	0	a2 = -0.607	0	0
Joint4	-π/2	a3 = -0.568	d4 = 0.193	0
Joint5	0	0	d5 = 0.125	π/2
Joint6	0	0	d6 = 0.108.4	-π/2

Figure 23 DH Parameters of DOBOT CR10.

The necessary computational and plotting functions were executed using MATLAB in the implementation of the simulation based on the passive motion paradigm (PMP) control model. The results obtained through these simulations were used to conclude.

For the simulation setup, the CR10 robotic arm was securely mounted on a solid frame base, and the position of the float and workspace was calibrated to match the

specifications that would be used in the hydroponic farm deployment. The robotic arm was aligned to a reference frame coinciding with the working area, with its origin positioned on one side of the workspace. The X, Y, and Z coordinates were defined relative to the base of the CR10 platform. During the movement of the CR10 through the workspace, comparative data was generated, including the target coordinates of the end effector and the actual coordinates reached.

To simulate the spatial relationships within the lettuce hydroponic working area, the XYZ axes were constrained to the following ranges: X(-500mmto-105mm), Y(500mmto 850mm), and Z (-50mm to 200mm). This constraint allowed for an evaluation of whether the trained neural network met the required accuracy and facilitated the computation of the gap between the expected and actual end effector positions.

The simulation involved directing the CR10 robotic arm to reach 850 evenly distributed localization points within the workspace. These points were plotted, with the green markers representing the target coordinates and the black markers indicating the actual coordinates reached by the end effector. The plotted data showed that the majority of the green markers overlapped with the black markers, indicating minimal discrepancies between the target and actual positions.

Although a small number of black markers did not align perfectly with their green counterparts, the corresponding error did not exceed 3mm. Statistical analysis of the 850 coordinate pairs revealed a mean square error of 1.276mm and a standard deviation error of 2.2396mm. These results confirm that the difference between the target positions and the actual positions was minimal, demonstrating the CR10 robotic arm's ability to accurately reach its intended targets.

Taking the target coordinates (-300, 600, -200) as an example, the trajectory of the end effector, along with the corresponding joint angle data, was plotted using MATLAB's visualization tools. The results demonstrated a continuous trajectory executed by the CR10 robotic arm, smoothly transitioned from the starting coordinates to the intended target. The smooth and continuous angular changes observed in the six joints, along with the uniform velocity, indicated that there were no discontinuities or abrupt changes in the system's motion.

The simulation results confirmed that the trajectory planning not only ensured the continuity of the path but also maintained the smooth rotation of the joints. During the lettuce hydroponic process, this method of trajectory planning for the CR10 robotic arm

effectively prevented vibration-induced damage to the lettuce, ensuring the system's reliability and robustness.

4.1.2 Constraints and Reaching on real environment

This study introduces a biomimetic motion control framework designed specifically for transplanting lettuce seedlings in hydroponic systems. The primary objective of this framework is to replicate the nuanced motion constraints observed in human operators, ensuring that robotic actions maintain the precision and adaptability required in unstructured agricultural environments.

The development of this motion control framework centred around an impedance control-based neural motion planning architecture. This architecture allows a robotic arm to dynamically incorporate task-specific constraints such as wrist pose, motion trajectories, and the delicate handling necessary for transplanting living tissues like lettuce seedlings. The framework was implemented on the CR10 collaborative robot, a 6-degree-of-freedom (DOF) robotic arm, selected for its versatility and capability to achieve high precision.

To replicate human-like precision in robotic motion, the methodology began with the extraction of human motion data using the YOLOv7 model for pose estimation. Video footage of human operators performing transplanting tasks in a hydroponic farm was analysed, focusing on key joints including the shoulders, elbows, and hands. The pixel coordinates of these joints were mapped and used to calculate the admittance values for the robot's joints, which are critical for mimicking the natural motion constraints observed in human operators.

The CR10 robotic arm was trained using a bio-inspired neural network implementation of the Passive Motion Paradigm (PMP). This model calculates the admittance for each joint based on the analysed human motion data, enabling the robot to replicate the smooth, precise movements necessary for transplanting tasks. The training dataset was generated using forward kinematic equations derived from the Denavit-Hartenberg(D-H) parameters of the CR10, encompassing 50,000 random joint configurations and their corresponding end-effector positions.

The layered control structure, particularly the integration of the JADE (Joint Admittance Data Extractor) building block, provides significant advantages. JADE allows for the systematic calculation of the PMP model's inner parameters by analysing human operators' movements. This approach offers a solution to the challenge of determining the

admittance values, which were previously calculated empirically. By observing the human performance of the task, the JADE module adjusts the robot's admittance settings to mirror the dexterity and subtlety of human movements.

Implementing biomimetic motion constraints through the PMP control model on the CR10 robotic arm demonstrated significant effectiveness. In both simulated and real-world environments, the robot achieved millimetric precision, with a mean square error of 1.276 mm and a standard deviation of 2.2396 mm. This level of precision was validated through tests in which the robot successfully reached 850 evenly distributed points within its workspace. The vast majority of target coordinates overlapped with the actual coordinates achieved by the robot, indicating negligible discrepancies.

The CR10 robotic arm's smooth and accurate movements were further validated during the transplanting task. The robot successfully picked and placed lettuce seedlings, demonstrating a seamless transition between picking and placing points without abrupt changes or discontinuities in joint movement. This smooth operation is crucial, as it prevents damage to the delicate seedlings during handling, ensuring the reliability and effectiveness of the system.

The systematic use of the JADE building block significantly enhances the robot's performance by optimizing the admittance matrix of the joints. By fine-tuning these parameters based on observed human behaviour, the CR10 robotic arm can perform transplanting tasks with a level of precision that closely mirrors that of skilled human workers. This human-inspired approach to robotic motion control not only improves the task's accuracy but also increases the system's overall adaptability to varying environmental conditions and task requirements.

The biomimetic motion constraints framework developed for the CR10 robotic arm offers a highly effective solution for automating the transplanting of lettuce seedlings in hydroponic systems. Integrating human-inspired motion planning with the PMP control model allows the robotic system to achieve the high precision necessary for such delicate agricultural tasks. The systematic approach provided by the JADE building block ensures that the robotic movements closely replicate human dexterity, making this framework a valuable tool for advancing automation in agriculture. Future developments will focus on further optimizing the joint admittance data extraction algorithm and adapting the current framework for additional agricultural tasks, such as the transportation and processing of

fully grown lettuce, thereby enhancing the automation capabilities of hydroponic farming systems.

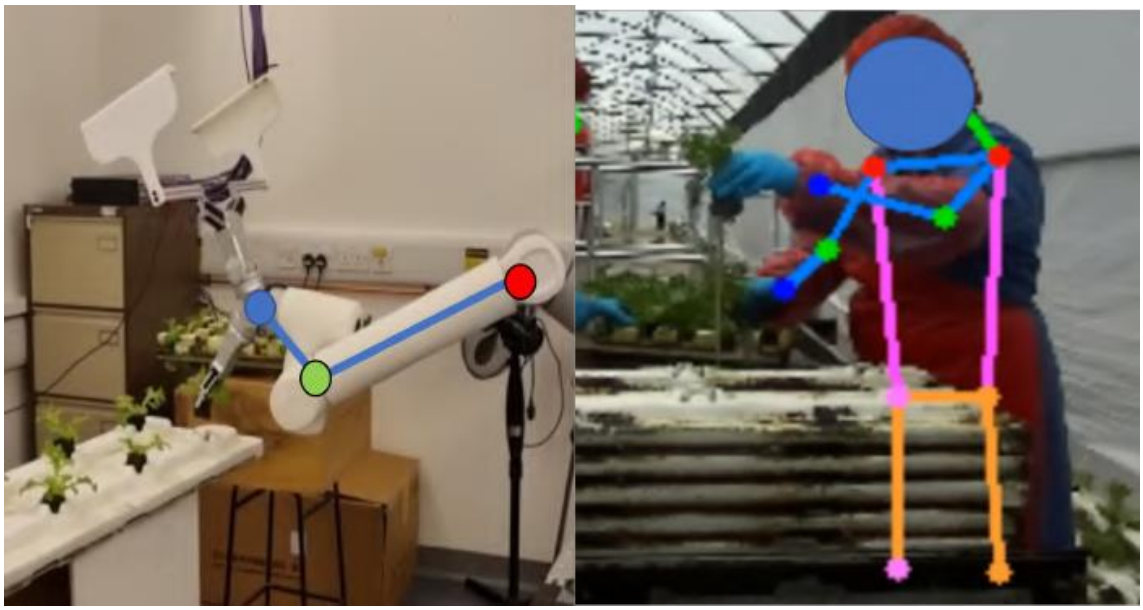


Figure 24 Human pose extraction at joint level and reference for robot 6dof.

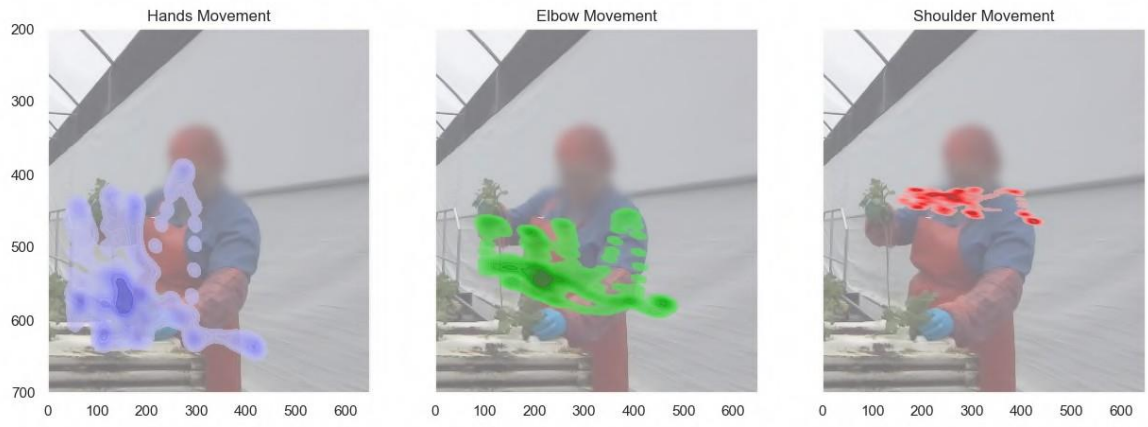


Figure 25 Human Pose Estimation trajectory to find individual contribution of human joints.

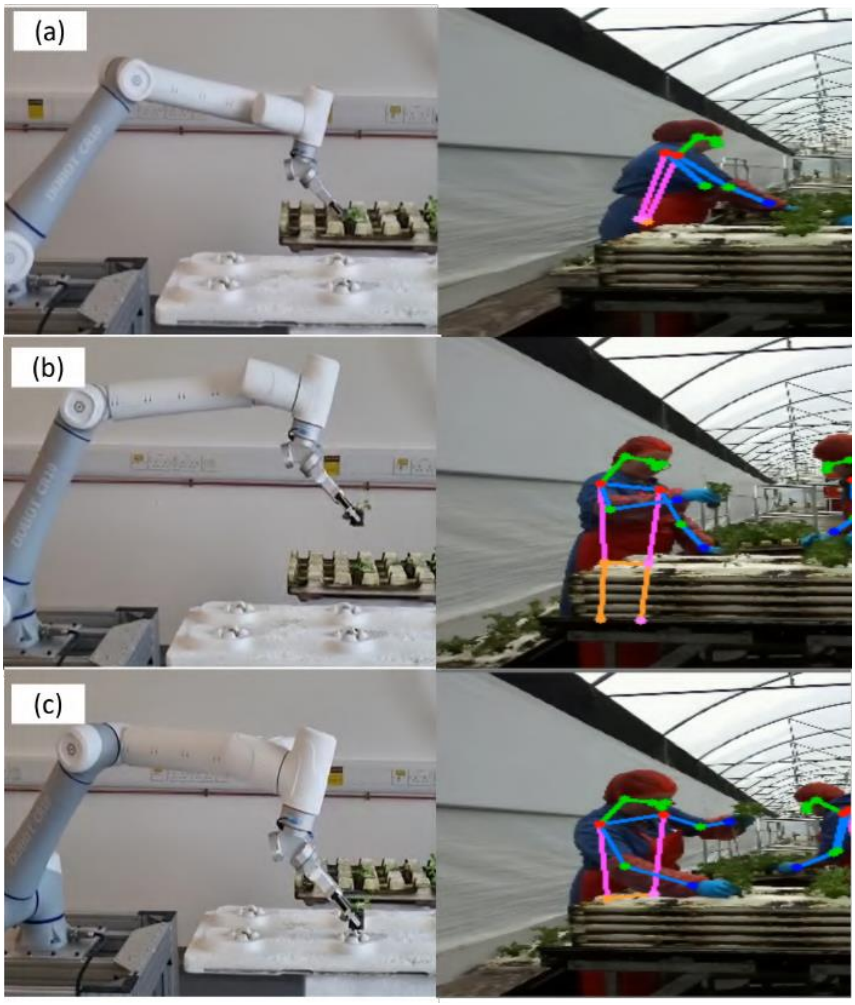


Figure 26 Side by side human and robot pick and place performance for transplanting operations.

The plotted results in Fig. 23 illustrate one example of a continuous trajectory executed by the CR10 reaching one point to another. The graphs show how the angular changes of the six joints are smooth and continuous, and the velocity is also uniform, indicating the absence of discontinuities or sudden changes in the system's motion. The simulation results show that this planning not only achieves the continuity of the path trajectory but also ensures the continuity of the joint rotation. During the lettuce hydroponic process, CR10 based on this trajectory planning method ensures that the end-effector does not vibrate during work, which effectively prevents vibration-induced damage to the quality of lettuce and ensures the reliability and robustness of the system. To test the accuracy of the PMP in the CR10 Robot, we generated a set of 8 different picking and placing point pairs by first determining the origin of each of the two floats by localizing the cavity in the left corner (shown in yellow in 5). Then, the rest of the points are calculated by moving an offset in the X and Y axes, using the evenly spaced geometry of the cavity matrix.

The 3D localization of the origin can be obtained in two ways: manually if the floats are placed in a fixed position relative to the robot's origin or by using point cloud stereo vision readings and a detection computer vision algorithm (as shown in Fig 6) if the floats present slight movement. In our case, the origin is manually determined for testing. The robot was controlled using the PMP controller to move first to the picking point, and then use the end position as the initial for the calculation of the final pose and trajectory towards the placing point. The results show accuracy within a 2 mm mean-square error in both, the simulation and robot.

The implementation of the PMP in the CR10 Robot arm has been demonstrated to reach different points within the workspace with millimetric accuracy, and precision demanded for the transplanting of lettuce seedlings between floats. The methodology was tested in simulation, as well as with the robot arm in a realistic setup with real crops. The introduction of the JADE building block, as a way to analyse the task's movement, provides a good starting point from where to tune the admittance matrix of the joints. Future work will include the optimization of joint admittance data extraction algorithm performance and continuing with the hydroponic farm automation by adapting the current architecture for specific tasks such as transporting the full-grown lettuces on the same base and cutting them (Fig 8), finishing the production line. This will require different end tool adaptations which is one of the main advantages of this framework.

Chapter 5
SCARLETT - Farm Deployment and Impact Analysis

5.1 Goal-Directed grasping and transplanting under constrain on-site

This thesis section integrates the Passive Motion Paradigm (PMP) with key point detection for goal-directed reaching, tailored explicitly for lettuce seedling transplanting on hydroponic floats. Combining these advanced methodologies ensures the robotic system can operate with the precision and adaptability required in real-world agricultural environments. Also, on-site modification were needed in the farm to archive a long-term full installation to optimise the workflow

Figure 27 Software Architecture of SCARLETT. Closed-loop solution

The methodology begins with robot pose estimation, utilising key point detection to accurately identify and localise specific planting points on the hydroponic floats. Unlike conventional object detection methods that rely on bounding boxes, key point detection provides a more refined approach by identifying particular coordinates on the floats,

which is crucial for tasks that require high precision, such as transplanting lettuce seedlings. This method is particularly effective in challenging conditions, such as varying lighting, soil coverage, and object occlusion, which are common in agricultural settings.

The captured key points, representing the positions of the planting holes on the float, are then converted into 3D space using stereo cameras. This 3D positional data is essential for guiding the robotic arm during the transplanting process, ensuring that the robot can accurately reach and manipulate the seedlings.

The motion planning component of the system is driven by the Passive Motion Paradigm (PMP), which is integrated into the control framework of the robotic arm. PMP serves as a neural control framework that facilitates goal-directed reaching while learning the internal model of the arm. This approach allows the robot to adapt its movements based on the specific constraints of the task, such as avoiding collisions in the workspace and maintaining the integrity of the seedlings during transplanting.

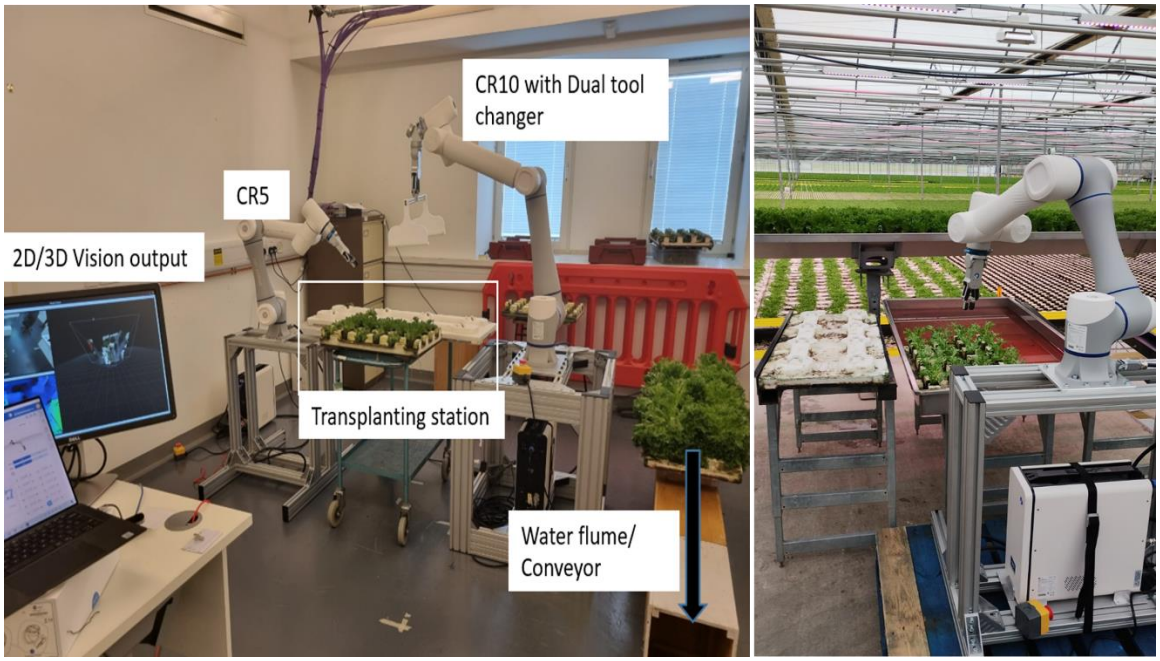


Figure 28 Panel A shows the set up in the lab

The system calculates the admittance per joint to implement PMP, which introduces specific constraints on the robot's movements. These constraints are derived from the training data generated using forward kinematic equations based on the Denavit Hartenberg parameters of the CR10 robotic arm. The dataset includes 50,000 random joint configurations and their corresponding end-effector positions, providing a comprehensive training basis for the neural network.

Figure 29 Harvesting room electrical and mechanical adaptation

The combination of key point detection and PMP has proven highly effective in real-world applications. The system was tested in both controlled laboratory environments and actual farm settings, with results showing that the robot could achieve millimetric precision in transplanting tasks. The accuracy of the key point detection was validated using a VICON tracking system in the lab, while on-site accuracy was measured by placing markers on the floats. Despite the complexity of the task, including challenges like root alignment and the need for precise placement, the system achieved a 90% accuracy in the closed-loop pick-and-place. The system was tested under 8 full transplanted floats, which is equivalent to 240 lettuces

One of the key advantages of this approach is its ability to handle the specific requirements of the transplanting process, which cannot be addressed as a simple pick-and-place task. The integration of PMP ensures that the robotic arm's movements are smooth and continuous, preventing damage to the delicate seedlings during handling. This is critical for maintaining the lettuce's quality and ensuring the transplanting operation's success.

In conclusion, integrating the Passive Motion Paradigm with key point detection offers a robust solution for goal-directed reaching in the context of lettuce transplanting on hydroponic floats. The system's ability to accurately identify and reach specific planting

points, combined with its adaptive motion planning, makes it a valuable tool for advancing automation in smart agriculture. Future work will enhance the system's speed and precision in large-scale production environments, further bridging the gap between robotic automation and traditional farming practices. Overall taking in consideration controller accuracy, stereo camera 2D-3D pixel reprojection and real reaching produces under 10mm error, taking 3-7mm accuracy.

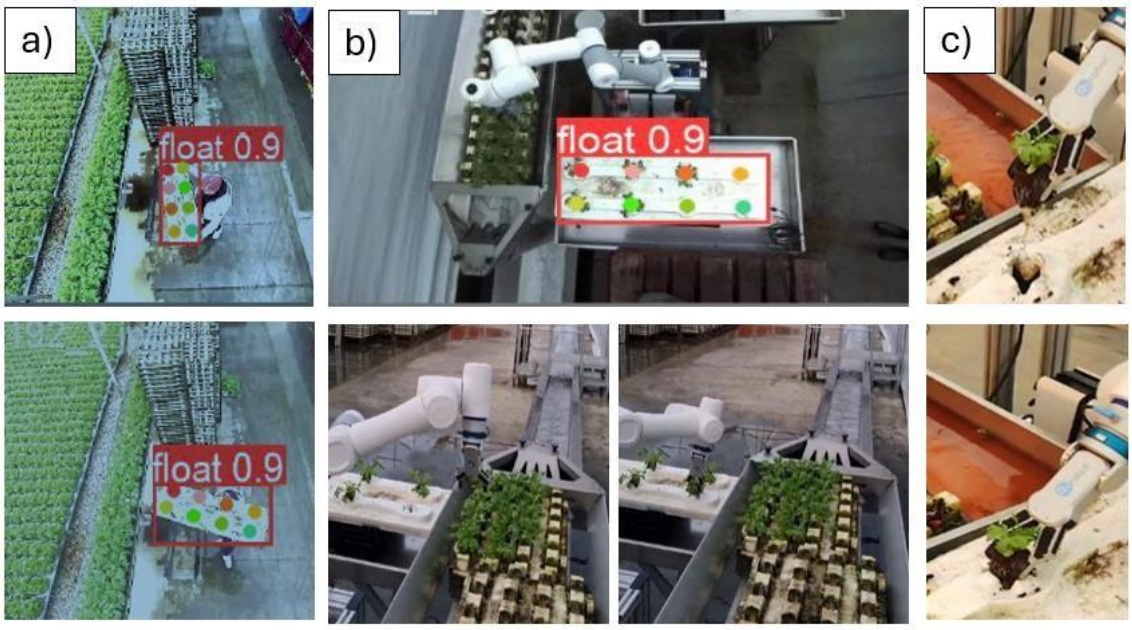


Figure 30 Close-loop key point detection, robot coordination and manipulation in a real environment.

Figure 31 Harvest Room pick and place for packaging. A rotational gripper was used to handle the full-grown lettuces and manipulate them from conveyor belt to packaging station

Figure 32 Flot handling after transplanting

Figure 33 Robot transplanting in farm

The methodology was demonstrated inside the real scenario of JEPCO facility, which is a hydroponic station composed by a water flume and two different float stacks. The small float has 24 lettuce seedlings, which must be transplanted on batches of 8 lettuces. This allows the necessary spacing to grow in a water pond. The small float is coming from a guided stack inside the water flume, which locations are being simplified using the stereo camera depth estimation. The proposed method focused on the large float which must be transported to another flume, hence the need for an accurate system. Key Point Detection Accuracy: The network was tested in two setups, a control environment inside a laboratory, and the farm facility for real- world production. For the ground truth inside the laboratory, a VICON tracking system was used to measure the actual position of each hole which represented a 7mm accuracy, while the precision on-site was measured by markers placed on top of each placing point.

Due to root alignment and dipping constraints, the transplanting problem cannot be addressed as a simple pick-and-place task. This process is essential for this application and requires millimetric precision for placing. If the root is placed outside the float target hole, the crop transplanting is considered a failure operation since the lettuce will not grow on the hydroponic system. This is where a reliable vision system and accurate body model are crucial. For the seedling transplanting speed testing, the human operator takes around 2.5-3.5 seconds per lettuce, while the current CR5 on-site performs around 7.5-8.5 seconds per lettuce with a speed factor of 60, around 100 degrees/second per joint. Since the focus has been on accuracy, the area of improvement for fast application is promising. Given a set of 8 lettuces for transplanting, repeating the process for the other 30 batches, the precision of the transplanting was successful in terms of reaching and grasping. The external factors for the placing is the over absorption of water in the seedling soil which leads to the lettuce to be stuck and due the second grasping the soil is crushed. Analysing in terms of accuracy and success, the reaching represents a 90% since this specific work focuses on goal-directed reaching accuracy rather than deep grasping analysis, where slippery, break of seedling blocks due excess of moisture and drops during the robot motion were real-world factors that affected the overall success rate.

The proposed robot pose estimation for goal-directed reaching for lettuce floats. The first part is getting the pose and picking points and transform then in 3D space and the second stage is using the internal body model adding constraints in different spatial tasks. This research proposes the use of the robot pose estimation and reaching in a structural approach. Such robots are being currently used by the farm in other task besides transplanting, as floats transport and fully grown lettuces packaging farms. Future work

will be focused on field precision on real production, with over 100 batches and speed comparison with humans, to enhance the production and reach the future of automation.

5.2 Analysis of Economic Benefits

Expected returns within 12 months of the end of the project:
The most significant commercial returns related to automating well-defined tasks will improve crop handling accuracy and efficiency. The 24/7 operation and control afforded by robotics will reduce waste and increase the rate of production, which is currently limited by the rate/efficiency at which manual operators can process produce.
The most critical benefit will be alleviating the critical shortage of unskilled labour to perform repetitive tasks. Utilising two robots (SCARLETT project), we estimate saving 2500 hours in harvest flume Pick Place, 2000 hours transplant to flume PnP, 7000 hours transplanting, and 3000 hours packing, which is a total of 14,500 hours per annum of human labour time and cost that can be saved.

Based on an average of £15/hr (basic pay + NI + pension), that is an annual saving of £217,500 in 2024 money £. This will only increase year on year assuming overseas labour can be sourced, which is another key factor for future-proofing UK food production and security. Additionally, there will be savings on recruitment, HR, and pastoral care costs estimated at around £20,000 per annum.

Not in the scope of this thesis, but anticipated future commercial returns:
Expected returns within 36 months of the end of the project:
The robotics system will be rolled out to our larger site, which is 10x larger than the Thorrington site, with savings above replicated.

Annual savings, wages: £217,500 x 7 = £1,522,500 (this will only increase year on year)
Annual savings, recruitment, and HR costs: 7 x £20,000 = £140,000
Licensing income from patenting the system cannot be quantified at this stage, but the returns from replication at license sites could be transformational. The robotics-as-a-service model could be charged at a per-hour cost, including maintenance, servicing, and know-how/training. In this context, the UoE steering group has already approved the creation of an AgriTech spin-off called Versatile Robotics to develop products and tap into the growing agricultural robotics market.

Summary of wider benefits from the Project for those outside the consortium.
There are several benefits beyond the commercial gains for the company and academic benefits for the university which this innovation will deliver:
Reduced reliance on increasingly scarce foreign labour for highly repetitive manual tasks (potentially across a range of sectors, but with initial focus on agriculture) through advances in robotic perception and manipulation. Reduction in resource use through increasing efficiency of the harvesting process Increasing UK's food production capability and food security without increasing land needed for agriculture Provision of skilled jobs all year round and reducing need for unskilled seasonal labour Reduced carbon footprint from food miles as more food is produced locally Reduced wastage and increased supply chain freshness from more UK-produced salad crops being harvested 52 weeks of the year.

The JEPCO team will engage with both the existing customer base and emerging markets through demonstrations at the Thorrington facility, industry events and participation in Agri Robotics events, trade shows and social media PR.
The immediate exploitation plan is to deploy both the subsystems and the full integrated robotic solution in JEPCO's 10 times larger facility in Lincoln. Multiple stages of transplanting are required in this new facility where the 4X transplanting gripper can be deployed/further scaled up. JEPCO's use of the Ellepot system, reduces carbon footprint and enables them to grow at scale, 900 seedlings/m2 to 225 plants/ m2 and finally 25 fully grown lettuces/m2. The dart-shaped structure makes the plant more amenable to rapid/precise robotic transplanting at multiple stages, handling multiple seedlings in parallel through the 'modular and novel' 3D printable salad grippers (>20x cost-reduction), associated 2D/3D vision, robot arm motion planning system developed at Essex. The innovation is cost-effective and configurable to other crop types.
To guarantee food safety standards, these applications aim to be collaborative were JEPCO or the given grower must receive a 3-hour induction, as well as following risk assessment for health and safety for each tasks were the robot is being used.

Chapter 6
General Conclusions and Future Work

6.1 Summary

There are several benefits beyond the commercial gains for the company and academic benefits for the SCARLETT project delivered by this innovation:
Reduced reliance on increasingly scarce foreign labour for highly repetitive manual tasks (potentially across a range of sectors, but with an initial focus on agriculture) through robotic perception and manipulation advances.
Reduced resource use through increasing efficiency of the harvesting process increases the UK's food production capability and security without increasing the land needed for agriculture.

The provision of skilled jobs all year round reduces the need for unskilled seasonal labour and the carbon footprint produced by food miles, as more food is produced locally and wasted. There is also increased supply chain freshness from more UK-produced salad crops being harvested 52 weeks of the year.
This project, funded by UKRI and DEFRA and in collaboration with JEPCO and Essex University, achieved several milestones and proposed and solved new ways to revolutionise agriculture automation.

The main challenge of this implementation was designing an adaptive and scalable framework that could adapt to different constraints in the real world. The transition between lab and farm brings external challenges for actual implementation and crucial engineering, industrial, and safety aspects. One of the first tasks, although not focused on the current research, was the choice, optimisation and development of different end tools for the system to adapt to the various crops and beyond human capabilities. For instance, in the transplanting stations, methods were approached to expand the number of lettuce seedlings the robot was able to grasp per movement, which enhanced the overall profits and minimised the labour.

The research's main contribution is the adaptive vision system that gets the 3D grasping points for the defined set of objects. In this particular case, specifically on the propagation hydroponic float, this is a set of placing points. Then, this system needed a robust implementation in a natural environment where electrical and mechanical installation was

required for the stereo camera and communication with the main computing platform. Also, the specific lens of the stereo camera was polarised and designed for outdoor applications since the harvest facility has a changing light environment. For this reason the need of not only a deep learning model but with 3D perception and robust against the background was needed which was solved by the key-point localization technique. The detections had millimetric accuracy under 2 different reference systems and was tested in the lab and in the farm. The last challenge was the calibration of the cameras and a method for easy calibration, since the camera is removed after every shift due to the facility water cleaning. This was the reason of using markers in non-adapt areas and docking stands where the position of the robot to the camera is always fixed. With this, the computer vision methodology was successfully tested on the farm.

The next step was the manipulation system of the series of collaborative robots. This was achieved by solving the inverse kinematic problem with an ANN-based impedance control under physical constraints known as the passive motion paradigm. The development of the action system had scalability and versatility in mind, where the model of the robot can be easily changed, and the solution as a whole remains really accurate for broader applications outside of lettuces and outside of this specific series of robots.
The ANN-based was trained for three series or 6DOF robots, CR10, CR5 and CR3, where the number of the model specifies the KG payload and the deployment of each robot depends on the specific application.

6.2 Research directions

Future work can exploit the perception-action system's learning component by implementing state-of-the-art reinforcement learning and imitation learning techniques. Getting accurate data from sensors used by expert humans can produce accurate datasets that approximate the actual behaviour of the systems. This technique can be leveraged using physics simulations considering external factors such as human collaboration and safety constraints. Future work involves the creation of a digital twin of the sensorimotor data and the 3D reconstruction of the given environment, and specifically for the JEPCO farm, physic simulators are being implemented to enhance the system identification, as lettuce weight, water resistance and movement.

Once the accurate simulation of the system is concluded, state-of-the-art methods in the domain of vision language models and time-series large language models can be applied, where the combination of the vision as detecting grasping points, collision objects and humans, as well as the scene with the ideal trajectories under specific constraints as

compliance, making it a complete solution which could be a mayor agricultural robotics framework.

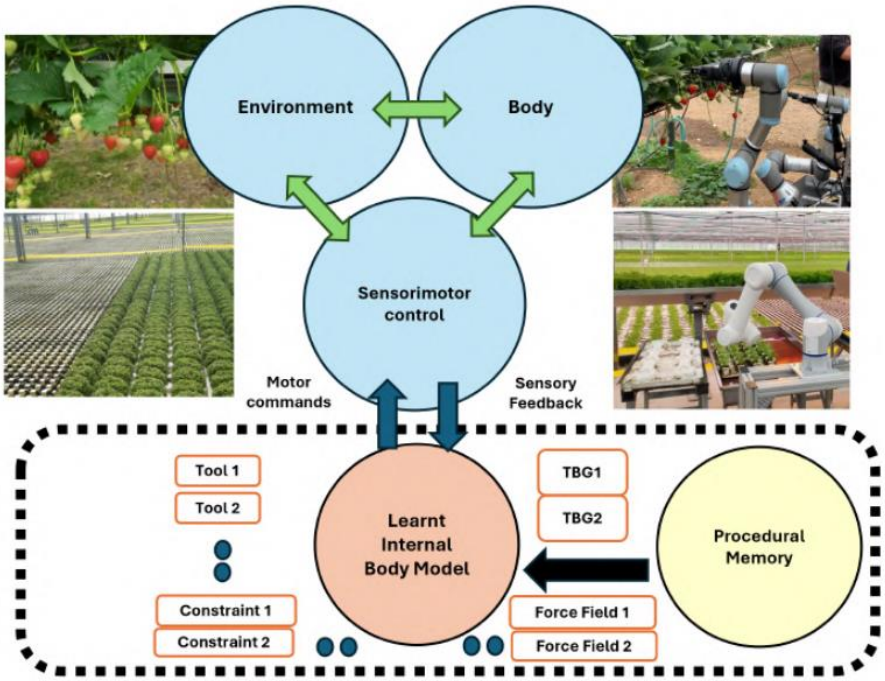


Figure 34 Future work, extension as an agricultural robotics framework

Another way to introduce the current solution as a general framework outside of physics simulators and learning models is the multi-robot operation with scalability in mind and shared per personal space under adaptive environments, focusing on the passive motion paradigm with adaptive models for perception and action. One current work proposes Swift Configurability to multiple crop types, robots and harvesting tasks. Procedural Memory Configuring the internal body model based on the goal of sensorimotor control and action simulation. TGB refers to Time Base generators that allow the synchronisation of multiple arms. The learnt model includes constrained and force fields tailored for each application. From the environment, we can adapt and transfer to the Body respectively.

Finally, adaptive perception robotic systems will bring the next era of automation, which is needed due to the labour shortage in agriculture. This will inevitably transform the way food is produced.

Bibliography

[1] V. Marinoudi, C. Sørensen, S. Pearson, and D. Bochtis, 'Robotics and labour in agriculture. A context consideration', *Biosystems Engineering*, 2019, doi: 10.1016/J.BIOSYSTEMSENG.2019.06.013.

[2] C. Mitaritonna and L. Ragot, 'After Covid-19, will seasonal migrant agricultural workers in Europe be replaced by robots', *CEPII Policy Brief*, vol. 33, pp. 1–10, 2020.

[3] M. Kondoyanni, D. Loukatos, C. Maraveas, C. Drosos, and K. G. Arvanitis, 'Bio-inspired robots and structures toward fostering the modernization of agriculture', *Biomimetics*, vol. 7, no. 2, p. 69, 2022.

[4] T. Duckett *et al.*, 'Agricultural Robotics: The Future of Robotic Agriculture', Aug. 02, 2018, *arXiv*: arXiv:1806.06762. Accessed: Aug. 20, 2024. [Online]. Available: http://arxiv.org/abs/1806.06762

[5] 'UK: Experts help assess potential for robotics and automation to alleviate labor shortages in UK food chain'. Accessed: Aug. 20, 2024. [Online]. Available: https://www.hortidaily.com/article/9625575/uk-experts-help-assess-potential-for-robotics-and-automation-to-alleviate-labor-shortages-in-uk-food-chain/

[6] 'UKRI awards more than £8m to innovative new farming concepts'. Accessed: Aug. 20, 2024. [Online]. Available: https://www.ukri.org/news/ukri-awards-more-than-8m-to-innovative-new-farming-concepts/

[7] 'DOBOT CR Series Robotic Arms | Safest Cobots for Industrial Ssage'. Accessed: Sep. 03, 2024. [Online]. Available: https://www.dobot-robots.com/products/cr-series/dobot-cr-series.html

[8] 'Computer vision in agriculture: The ultimate guide', Software Development Company - N-iX. Accessed: Feb. 20, 2025. [Online]. Available: https://www.n-ix.com/computer-vision-in-agriculture/

[9] Y. Tang *et al.*, 'Recognition and Localization Methods for Vision-Based Fruit Picking Robots: A Review', *Frontiers in Plant Science*, vol. 11, 2020, doi: 10.3389/fpls.2020.00510.

[10] S. Varastehpour, H. Sharifzadeh, and I. Ardekani, 'A Comprehensive Review of Deep Learning Algorithms', Unitec ePress, 2021. doi: 10.34074/ocds.092.

[11] G. Boesch, 'Computer Vision in Agriculture - Valuable 2025 Use Cases', viso.ai. Accessed: Feb. 20, 2025. [Online]. Available: https://viso.ai/applications/computer-vision-in-agriculture/

[12] N. O. Mahony *et al.*, *Deep Learning vs. Traditional Computer Vision*, vol. 943. 2020. doi: 10.1007/978-3-030-17795-9.

[13] R. Girshick, 'Fast R-CNN', Sep. 27, 2015, *arXiv*: arXiv:1504.08083. doi: 10.48550/arXiv.1504.08083.

[14] K. C. Z. Zou, 'Object Detection in 20 Years: A Survey', *http://arxiv.org/abs/1905.05055*.

[15] F. H. Ankile, M. and K. Krange, 'Deep Convolutional Neural Networks A survey of the foundations, selected improvements, and some current applications.', Nov. 2020, doi: 10.48550/arXiv.2011.12960.

[16] J. Terven and D. Cordova-Esparza, 'A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS', *MAKE*, vol. 5, no. 4, pp. 1680–1716, Nov. 2023, doi: 10.3390/make5040083.

[17] ‘Comparative Study of YOLOv5, YOLOv7 and YOLOv8 for Robust Outdoor Detection | Journal of Applied Electrical Engineering’. Accessed: Oct. 03, 2024. [Online]. Available: https://jurnal.polibatam.ac.id/index.php/JAEE/article/view/7207
[18] Ultralytics, ‘YOLOv8’. Accessed: Feb. 20, 2025. [Online]. Available: https://docs.ultralytics.com/models/yolov8
[19] A. Tripathi, M. K. Gupta, C. Srivastava, P. Dixit, and S. K. Pandey, ‘Object Detection using YOLO: A Survey’, in *2022 5th International Conference on Contemporary Computing and Informatics (IC3I)*, Dec. 2022, pp. 747–752. doi: 10.1109/IC3I56241.2022.10073281.
[20] H. Zhao *et al.*, ‘Real-time object detection and robotic manipulation for agriculture using a YOLO-based learning approach’, Jan. 28, 2024, *arXiv*: arXiv:2401.15785. doi: 10.48550/arXiv.2401.15785.
[21] Ultralytics, ‘OBB’. Accessed: Feb. 20, 2025. [Online]. Available: https://docs.ultralytics.com/tasks/obb
[22] ‘YOLO-NAS Pose Keypoint Detection Model: What is, How to Use’. Accessed: Feb. 20, 2025. [Online]. Available: https://roboflow.com/model/yolo-nas-pose
[23] Ultralytics, ‘Pose’. Accessed: Feb. 20, 2025. [Online]. Available: https://docs.ultralytics.com/tasks/pose
[24] W. Huang, C. Wang, Y. Li, R. Zhang, and L. Fei-Fei, ‘ReKep: Spatio-Temporal Reasoning of Relational Keypoint Constraints for Robotic Manipulation’, Sep. 03, 2024, *arXiv*: arXiv:2409.01652. doi: 10.48550/arXiv.2409.01652.
[25] ‘Depth cameras and RGB-D SLAM’, Kudan global. Accessed: Feb. 16, 2025. [Online]. Available: https://www.kudan.io/blog/depth-cameras-and-rgb-d-slam/
[26] ‘(PDF) Robot Arms with 3D Vision Capabilities’, in *ResearchGate*. doi: 10.5772/9668.
[27] ‘How does the ZED work?’, Help Center | Stereolabs. Accessed: Feb. 16, 2025. [Online]. Available: https://support.stereolabs.com/hc/en-us/articles/206953039-How-does-the-ZED-work
[28] ‘Experimentally Confirmed Mathematical Model for Human Control of a Non-Rigid Object | Journal of Neurophysiology’. Accessed: Oct. 03, 2024. [Online]. Available: https://journals.physiology.org/doi/full/10.1152/jn.00704.2003
[29] ‘Minimum Acceleration Criterion with Constraints Implies Bang-Bang Control as an Underlying Principle for Optimal Trajectories of Arm Reaching Movements | Neural Computation | MIT Press’. Accessed: Oct. 03, 2024. [Online]. Available: https://direct.mit.edu/neco/article-abstract/20/3/779/7291/Minimum-Acceleration-Criterion-with-Constraints
[30] ‘Error Correction, Sensory Prediction, and Adaptation in Motor Control | Annual Reviews’. Accessed: Oct. 03, 2024. [Online]. Available: https://www.annualreviews.org/content/journals/10.1146/annurev-neuro-060909-153135
[31] ‘Evidence for the Flexible Sensorimotor Strategies Predicted by Optimal Feedback Control | Journal of Neuroscience’. Accessed: Oct. 03, 2024. [Online]. Available: https://www.jneurosci.org/content/27/35/9354.short
[32] ‘Frontiers | Passive Motion Paradigm: An Alternative to Optimal Control’. Accessed: Oct. 03, 2024. [Online]. Available: https://www.frontiersin.org/journals/neurorobotics/articles/10.3389/fnbot.2011.00004/full

[33] ‘Optimal feedback control and the neural basis of volitional motor control | Nature Reviews Neuroscience’. Accessed: Oct. 03, 2024. [Online]. Available: https://www.nature.com/articles/nrn1427

[34] V. Mohan, P. Morasso, G. Metta, and G. Sandini, ‘A biomimetic, force-field based computational model for motion planning and bimanual coordination in humanoid robots’, *Auton Robot*, vol. 27, no. 3, pp. 291–307, Oct. 2009, doi: 10.1007/s10514-009-9127-x.

[35] V. Mohan and P. Morasso, ‘Passive Motion Paradigm: An Alternative to Optimal Control’, *Front Neurorobot*, vol. 5, p. 4, Dec. 2011, doi: 10.3389/fnbot.2011.00004.

[36] V. Mohan, A. Bhat, and P. Morasso, ‘Muscleless motor synergies and actions without movements: From motor neuroscience to cognitive robotics’, *Physics of Life Reviews*, vol. 30, pp. 89–111, Oct. 2019, doi: 10.1016/j.plrev.2018.04.005.

[37] P. Morasso, M. Casadio, V. Mohan, F. Rea, and J. Zenzeri, ‘Revisiting the Body-Schema Concept in the Context of Whole-Body Postural-Focal Dynamics’, *Front Hum Neurosci*, vol. 9, p. 83, Feb. 2015, doi: 10.3389/fnhum.2015.00083.

[38] ‘Frontiers | Social Cognition for Human-Robot Symbiosis—Challenges and Building Blocks’. Accessed: Feb. 20, 2025. [Online]. Available: https://www.frontiersin.org/journals/neurorobotics/articles/10.3389/fnbot.2018.00034/full

[39] X. Ling, Y. Zhao, L. Gong, C. Liu, and T. Wang, ‘Dual-arm cooperation and implementing for robotic harvesting tomato using binocular vision’, *Robotics and Autonomous Systems*, vol. 114, pp. 134–143, Apr. 2019, doi: 10.1016/j.robot.2019.01.019.

[40] S. Lem and J. Mackey, ‘Field Performance of a Dual Arm Robotic System for Efficient Tomato Harvesting’, *JRS*, pp. 66–75, May 2024, doi: 10.53759/9852/JRS202402007.

[41] J. Rong, P. Wang, T. Wang, L. Hu, and T. Yuan, ‘Fruit pose recognition and directional orderly grasping strategies for tomato harvesting robots’, *Computers and Electronics in Agriculture*, vol. 202, p. 107430, Nov. 2022, doi: 10.1016/j.compag.2022.107430.

[42] 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, no. 2, pp. 225–245, 2020, doi: https://doi.org/10.1002/rob.21888.

[43] V.-C. Pham, H.-G. Nguyen, T.-K. Doan, and G.-B. Huynh, ‘A Computer Vision Based Robotic Harvesting System for Lettuce’, *IJMERR*, pp. 1526–1531, 2020, doi: 10.18178/ijmerr.9.11.1526-1531.

[44] X. Wang, L. Kong, Z. Zhang, H. Wang, and X. Lu, ‘Keypoint regression strategy and angle loss based YOLO for object detection’, *Sci Rep*, vol. 13, p. 20117, Nov. 2023, doi: 10.1038/s41598-023-47398-w.