Multispectral Imaging System Fruit Detection for Green Pepper based on Mask R-CNN and SSIM

Sensing and Mechanism

By

Thatreesophon Ekkorn

Graduate School of Engineering

Kochi University of Technology

A thesis submitted for the degree of

Master of Engineering

Kochi 2023

Abstract

The global agricultural landscape is undergoing a transformative shift driven by technological advancements that aim to address critical issues such as labor shortages and weather variability. Environmental factors, including crop selection, soil quality, and weather conditions, significantly determine agricultural outcomes. Automated harvesting systems incorporating machine vision technology have gained prominence with a declining agricultural workforce and a pressing need for efficiency. This study focuses on developing an autonomous harvesting robot named "pīman" for Japanese sweet peppers, a crucial crop in Kochi prefecture, Japan. The decline in the global agricultural workforce, coupled with the aging demographic of farmers in Japan, underscores the necessity for innovative solutions to sustain productivity. The "pīman" robot integrates RGB and IR camera systems for precise fruit detection, aiming to overcome challenges in synchronizing with optimal environmental conditions, particularly variations in sunlight. Sunlight is crucial in green pepper harvesting, emphasizing the robot's vision system's reliance on accurate identification within the foliage. The study highlights the impact of weather conditions on the optimal harvesting period, affecting the available harvesting time. Considering the declining agricultural workforce and the associated trade deficit, the urgency to support and enhance efficiency in farming practices is evident. The study proposes integrating two different types of cameras to address challenges such as high initial costs and external factors affecting camera efficiency. While this approach may increase accuracy, careful consideration of processing system trade-offs is necessary. Investing in dual-camera systems is expected to enhance efficiency and productivity in modern agriculture.

Artificial Intelligence (AI) systems, particularly those employing deep learning approaches like the Mask Region-based Convolutional Neural Network (Mask R-CNN), showcase proficiency in object recognition and segmentation tasks. The study demonstrates the effectiveness of deep learning techniques in enhancing AI's capacity to discern specific objects, such as sweet peppers. Additionally, the study explores fundamental edge detection techniques in computer vision, which are essential for object recognition and image segmentation tasks, with algorithms like Sobel, Prewitt, and Canny playing pivotal roles. In conclusion, this comprehensive study delves into integrating advanced technologies, including dual-camera systems and AI, to address

challenges the agricultural sector faces. The proposed solutions aim to enhance efficiency, mitigate labor shortages, and ensure sustainable agricultural practices in the face of evolving technological challenges. Shifting gears to image processing and computer vision, the Structural Similarity Index (SSIM) emerges as a fundamental metric. SSIM is critical in quantifying the similarity between two images, surpassing traditional metrics like Mean Squared Error (MSE) by incorporating luminance, contrast, and structure considerations. The holistic approach of SSIM mirrors human visual perception, addressing the limitations of conventional methods that focus solely on pixel-wise differences. SSIM's evaluative scope extends beyond pixel-level disparities, enabling a nuanced assessment of global and local image variations. Using a scale from -1 to 1, with 1 indicating perfect similarity, SSIM captures quantitative distinctions in pixel values and qualitative aspects of structural and textural information within images. The multifaceted utility of SSIM extends to diverse tasks, including image quality assessment, compression optimization, and image restoration. Researchers and practitioners leverage SSIM to refine and optimize image processing algorithms, aligning algorithmic outputs more closely with human perception and enhancing overall visual fidelity. Beyond its foundational role, SSIM finds applications in various sectors, exemplified by its use in biomedicine, radiation therapy, industrial applications, and agriculture. In biomedicine, SSIM aids in iris detection for diagnosing systemic health, while in radiation therapy, it monitors the accurate delivery of doses to the target. In industrial applications, SSIM distinguishes coal from gangue in the coal industry, and in agriculture, it facilitates fruit inspection, overcoming challenges associated with traditional sensor methods. The Structural Similarity Index emerges as a cornerstone in image processing and computer vision, serving as a dynamic and versatile tool with expanding applications across various domains. Its nuanced approach to image similarity assessment and adaptability to diverse challenges position SSIM as a pivotal asset in advancing technological applications and contributing to refining imaging techniques across sectors.

Acknowledgements

I extend my deepest gratitude to my supervisor, Prof. Koichi Oka, whose guidance and unwavering support have been instrumental in shaping the trajectory of my research journey. Prof. Oka's expertise, mentorship, and invaluable insights have been a guiding force, enriching academic experience and contributing significantly to the success of this research endeavor.

I want to express my sincere appreciation to the members of the OKA lab for their collaborative spirit and assistance throughout the research process. Their collective efforts, constructive feedback, and shared dedication to academic excellence created a stimulating and supportive environment that fostered personal and scholarly growth.

Furthermore, I would like to thank all the International Relations Division staff members for their administrative support and facilitation of various aspects of my research. Their efficiency, professionalism, and commitment to ensuring a smooth research experience have been crucial in successfully executing this project.

This acknowledgment extends beyond individual names to encompass a collective acknowledgment of academic research's collaborative and interconnected nature. Each person mentioned has contributed uniquely, and I am genuinely grateful for the collaborative spirit that has defined this research endeavor.

In closing, I am indebted to Prof. Koichi Oka, the OKA lab members, and the International Relations Division staff members for their collective support, encouragement, and commitment to fostering an environment conducive to academic excellence. Their presence has enriched this research journey, and I am thankful for the opportunity to have worked alongside such dedicated individuals.

Table of Contents

List of Figures

List of Tables

1. Introduction

1.1 Background

Agriculture is vital for human survival and is critical in global food production. It faces challenges like labor-intensive tasks and unpredictable weather. Weather variations, such as droughts and floods, can devastate crops, adding risk to farming. Modern agriculture has transformed through technology, including genetics, automation, and precision farming. These innovations enhance efficiency, reduce reliance on manual labor, and promote large-scale cultivation. However, farming remains challenging due to natural variability. Crucial factors like crop selection, soil quality, irrigation, sunlight, and CO2 levels significantly affect agricultural outcomes. These variables are influenced by weather fluctuations, emphasizing the impact of environmental factors. Urbanization draws rural residents to cities, causing a decline in the agricultural workforce. Fallow lands result in reduced output, impacting global food security, as the world relies on agriculture to feed a growing population [1]. In recent decades, extensive global research and resources have focused on automated harvesting systems, often incorporating machine vision technology, to achieve precise fruit and crop harvesting. However, syncing these automated harvesting robots with optimal environmental conditions poses a significant challenge.

The robot's vision system exemplifies the importance of sunlight in green pepper harvesting, which relies on it to accurately identify green peppers within the foliage. The optimal harvesting period typically spans from 8:00 AM to 4:30 PM [2], subject to weather and seasonal variations. Factors like cloud cover, rainfall, and winter's shortened daylight hours can significantly reduce available harvesting time. A research team at Kochi University of Technology (KUT) is dedicated to developing an autonomous harvesting robot specialized for Japanese sweet peppers with both RGB [3] and IR [4] camera systems., known as "pīman" locally. These peppers hold significance in Japanese agriculture, ranking 28th in cultivated land area (3,360 hectares) and 21st in total production (145,300 metric tons) among 41 recognized primary vegetables in fiscal year 2013, according to Japan's Ministry of Agriculture, Forestry, and Fisheries. KUT's location in Kochi prefecture, the third-largest, sweet pepper producer in Japan, highlights the peppers' critical role as a significant income source for local farmers, with 141 hectares of cultivated land and a total harvest of 13,000 metric tons, constituting approximately 8.95% of Japan's sweet pepper production.

1.2 Previous Research

In previous works, there were two types of vision systems for detecting green pepper: An RGB camera to capture green pepper was developed by P. Eizentals [3] and an IR camera to capture images was developed by T. Naoya [4]. The characteristics of the images obtained are different. RGB cameras [Figure 1-1 a] capture images from light, especially sunlight, while IR cameras [Figure 1-1 b] use heat from sunlight and surrounding weather conditions —both experiments are in a greenhouse.

Figure 1-1 Green pepper a) green pepper taken by RGB camera, b) green pepper taken by IR camera.

While conducting greenhouse experiments to regulate environmental conditions, it became evident that specific issues require further attention and resolution. Specifically, challenges arise in the context of RGB cameras due to their sensitivity to light and temporal constraints. RGB cameras rely on adequate lighting for optimal performance, making them susceptible to variations in light availability. Conversely, infrared (IR) cameras encounter challenges, including potential heat-related issues caused by sunlight, particularly during the winter, when these cameras face heightened operational difficulties.

1.3 Research Problem

Continuous research has been primarily directed towards unraveling the complexities surrounding machine vision for green pepper harvesting robots, motivated by various underlying reasons:

- Identifying a target entity presents a substantial challenge, primarily due to the pronounced similarity in color between said entity, in this context, fruits, and the encompassing environmental elements. Numerous approaches have been investigated

to efficiently discriminate green peppers from their immediate surroundings to overcome this challenge. One such method, implemented by Wei Ji, entails the acquisition of RGB images under daylight conditions, yielding an impressive accuracy rate of approximately 89% for the precise identification of green peppers [5]. Additionally, a noteworthy methodology devised by E. Zemmour concentrates on acquiring data specifically concerning yellow peppers during nocturnal operations, encompassing varying lighting conditions. This method showcases exceptional accuracy, with reported rates between 95% and 99% [6]. Furthermore, S. Bachche and K. Oka have undertaken research endeavors endorsing the utilization of the HSV color space for detecting sweet pepper fruits. This approach demonstrates notable effectiveness, particularly with artificial lighting or an IR96 infrared filter during daytime applications. The performance outcomes resulting from these methods exhibit a range between 70% and 80%, with variability contingent upon the presence of occlusions and the prevailing illumination conditions [7]. In addition to these approaches, the deployment of infrared (IR) cameras has been explored as an alternative means of detection. However, it is essential to note that the accuracy of this methodology exhibits significant variability, spanning from 50% to 86% [4]. Moreover, this accuracy is notably susceptible to external factors such as meteorological conditions and temperature fluctuations. In summary, the formidable challenge of target identification in the context of green and sweet pepper detection has prompted the exploration of diverse methodologies, each characterized by its distinct advantages and limitations. The selection of a suitable method is contingent upon factors such as the prevailing lighting conditions, occlusion levels, and environmental variables, all of which influence the accuracy of the detection process.

- The green pepper plant, classified as an annual species, belongs to the same botanical genus as tomatoes. It bears individual fruits that are densely clustered, rendering the harvesting process particularly challenging. The efficacy of green pepper identification is contingent upon the quality of the vision system deployed, as distinguishing green peppers from green chilies becomes problematic without an adept system. This predicament arises due to a partial overlap in visual features between these two entities. Furthermore, green pepper fruits exhibit relatively uncertain characteristics, lacking the pronounced symmetrical attributes observed in other fruits like tomatoes, apples, and oranges, which impede precise detection methodologies.
- A noteworthy challenge in classifying green peppers arises from an abundance of leaves. The high leaf density leads to a compact clustering of green pepper fruits. It introduces

an additional complicating factor by obscuring the precise localization of the green peppers within the plant canopy. Moreover, the uniformity of coloration presents another formidable challenge, as green pepper fruits, leaves, stems, and external structural components all exhibit the same green hue. This shared chromatic characteristic further compounds the difficulty associated with the detection and identification processes. The confluence of factors such as leaf density and uniform coloration significantly amplifies the intricacy of discerning and categorizing green peppers within the agricultural context, necessitating advanced technological solutions for accurate classification and sorting.

- Another noteworthy consideration necessitates attention: the camera's precise positioning within the greenhouse's confines. This strategic placement is of paramount importance, as it is envisaged that the camera will be integrated into an automated robotic system designed to harvest green chilies in future agricultural operations. Consequently, a meticulous calculation of the optimal camera placement distance is imperative. This calculation must align with the harvesting robot's anticipated traversal range and the requisite proximity to the green chili plants to acquire relevant imagery and data. Detailed graphical representations and accompanying explanatory information elucidating the specifics of camera placement are elucidated , providing comprehensive insights into the strategic arrangement of the camera within the agricultural context to support forthcoming robotic harvesting endeavors.

1.4 Research Objectives

The principal research aim is to develop a machine vision system that can effectively detect and accurately identify green peppers under various environmental conditions. The requirements for the machine vision system were as follows:

- The system's competence in precisely discriminating and classifying green peppers, whether partially obscured within a foliage canopy or entirely unobstructed, underscores its adeptness in detecting the target fruit across a spectrum of environmental conditions and amidst varying vegetative contexts.
- The capacity to acquire image data under a wide range of environmental conditions, encompassing situations marked by diminished light levels on cloudy days, lower temperatures during rainy intervals, and varying meteorological circumstances. The ability to collect data images in every condition, including scenarios characterized by

reduced illumination on overcast days, colder temperatures during rainy periods, and diverse meteorological conditions.

- Integrating infrared (IR) and red-green-blue (RGB) data collaboratively improves the precision of green pepper detection and identification processes. This fusion of data sources provides a more comprehensive and reliable basis for distinguishing and categorizing green peppers under varying conditions and scenarios.

2. Machine Vision in Agriculture

2.1 Overview

In recent decades, the agricultural sector has grappled with a persistent issue of labor shortages, with a global decline in the agricultural workforce. As of 2020, the worldwide agricultural labor force numbered 656 million individuals, and this trend is expected to continue, with an anticipated decrease to 624 million by 2030 [1]. Japan, In particular has experienced a substantial reduction in the number of farmers since 1995, declining from 4.14 million to a mere 1.68 million in 2019 *figure 2-1* [9]. This decline shows no signs of abating, as evidenced by the 2022 data, which revealed that most individuals engaged in farming were over 50 years old, numbering 1.086 million, while those between 15-49 years old constituted a mere 0.14 million *figure 2-2*. These demographic shifts have resulted in various consequences, including a trade deficit due to the imbalance between demand and production, leading to rising product prices. In 2023, a survey of various establishments confirmed that they had increased product prices by over 20% to offset the rising costs of raw materials and the persistent shortages. This, in turn, has enabled Japanese exports to be sold at higher prices compared to the domestic market, exacerbating the disparity in domestic product prices [8].

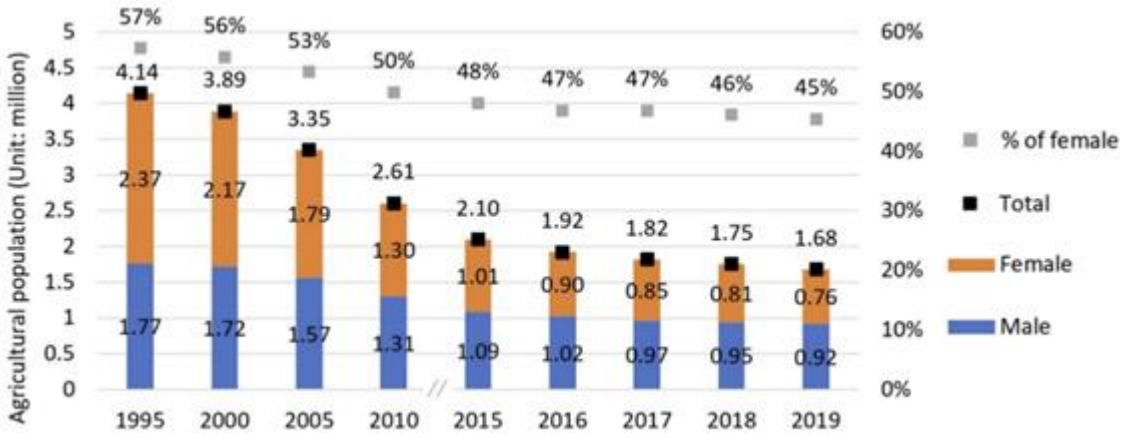

Figure 2-1 Japanese farmer population [9]

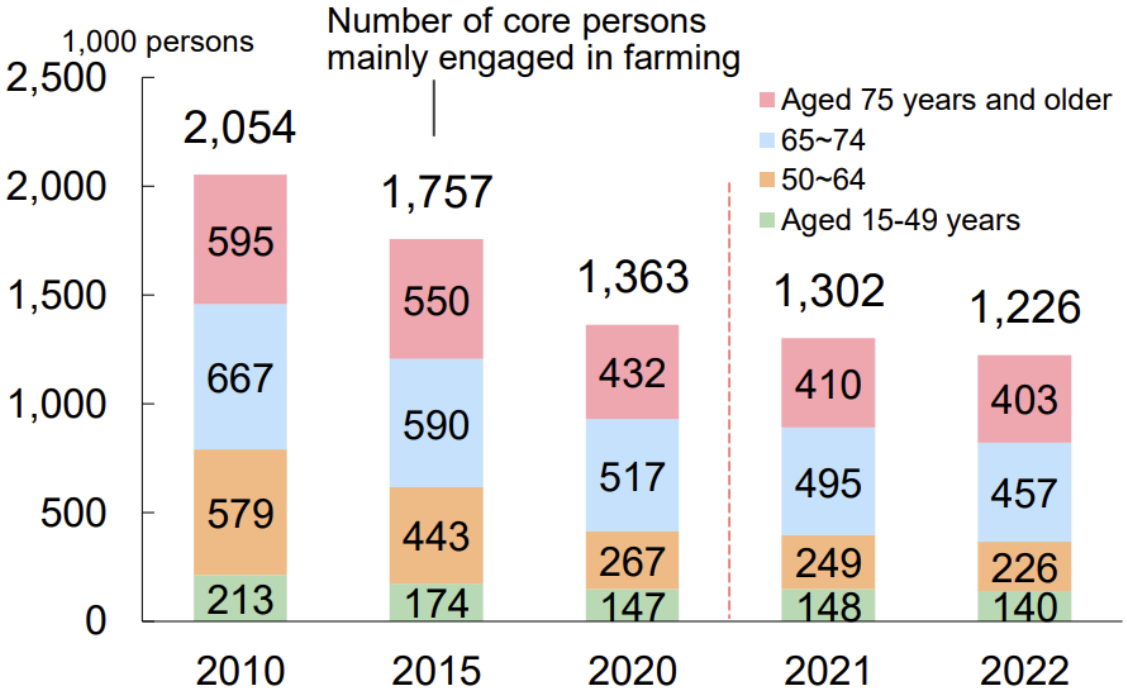

Figure 2-2 Number of core persons mainly engaged in farming in Japan [8]

Attracting a new generation of individuals to pursue careers in farming poses significant challenges. Therefore, it becomes imperative to support the existing farming workforce. The development of agricultural equipment aimed at enhancing efficiency dates to the 1960s when G.E. Coppock introduced a tree shaker to facilitate the harvesting of substantial quantities of produce, thereby eliminating the need for laborious manual picking [10]. However, the effectiveness of such equipment was limited due to fruit damage resulting from the impact during

the harvesting process. Notably, this issue was addressed in 1968 by C. E. Schertz and G. K. Brown [11], who proposed the concept of single fruit harvesting as an alternative to mass harvesting, thereby mitigating tree injuries and fruit loss. Subsequently, researchers worldwide have undertaken endeavors to develop robotic systems for the automated harvesting of diverse fruits and vegetables. Since then, the evolution of harvesting systems has continued to progress up to the present day, enabling the creation of automated harvesting robots. While technology now allows for automated harvesting, specific challenges persist, such as adverse weather conditions and sunlight variations that can affect the performance of cameras utilized for produce detection.

2.2 Target Detection

Camera-based fruit detection plays a pivotal role in developing automated harvesting systems in agriculture. The ability to accurately identify and locate fruits is essential for the efficient operation of harvesting robots and the critical challenges of camera-based fruit detection are:

- *Color Variations* - Fruits exhibit various color variations due to ripeness, species, and environmental conditions. These variations can pose challenges to camera-based detection systems. Researchers have employed color models like RGB *figure 2-3a* [20], *figure 2-3b* HSV [20], and LAB *figure 2-4* [21] to address this issue. Additionally, machine learning techniques, such as convolutional neural networks (CNNs), have been utilized to learn and adapt to varying fruit colors, enhancing the accuracy of detection systems.
- *Reflectance* - Fruits can exhibit different reflectance levels, impacting how they appear in images. This phenomenon is especially pronounced in shiny fruits. Polarized light imaging techniques and multispectral imaging have been explored to mitigate the effects of reflectance. These methods can reduce glare and enhance the contrast between fruits and their surroundings, improving detection accuracy.
- *Occlusions* - In a natural agricultural setting, fruits are often partially hidden by leaves, branches, or other fruits, leading to occlusions in camera images. Addressing occlusions is crucial for complete fruit detection. Research has focused on developing algorithms that can reconstruct partially occluded fruits or predict their positions based on available visual cues, thereby minimizing the impact of occlusions.
- *Illumination* - Changing lighting conditions throughout the day can influence the appearance of fruits in camera images. This inconsistency can lead to variations in color and texture, affecting detection accuracy. Adaptive illumination techniques, such as flash

or LED lighting, have been integrated into camera systems to ensure consistent lighting, enabling more reliable fruit detection.

- *Shadows* - Shadows cast by foliage or other objects can obscure fruits and introduce false positives in detection systems. Researchers have employed shadow detection algorithms to differentiate between actual fruits and shadow artifacts. These algorithms rely on features such as texture, shape, and context to identify and filter out shadowed regions in images.

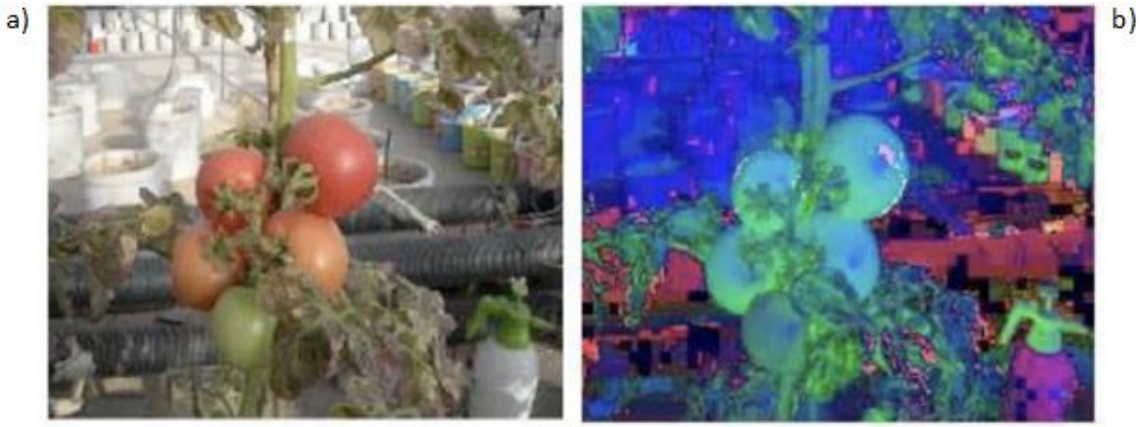

Figure 2-3 a) RGB color image, b) HSV color image [20]

Figure 2-4 Green apples on trees images a) RGB color image, b) LAB color image [21]

2.2.1 Visual Signals

Visual signals play a crucial role in fruit detection, enabling humans and machines to identify and assess the quality of fruits. The human sense of sight is an intricate and versatile mechanism that relies on various visual cues such as colors, greyscale, shapes, depth perception, and spectral imaging. Applying these cues and advancements in hyperspectral imaging technology has significantly enhanced fruit detection processes.

- *Sense of Sight* -The human sense of sight is vital in fruit detection, as it allows us to perceive and differentiate between various fruit characteristics. Colors are the most prominent visual cue, indicating ripeness, quality, and variety. For instance, the vibrant red color of ripe apples contrasts with the green of unripe ones, facilitating visual detection.
- *Colors and Grey Scale* - Colors are essential indicators in fruit detection due to their role in ripeness assessment. Colorimetry, a color-based detection method, has been widely employed to quantify fruit color attributes [12]. Additionally, greyscale images [13] have helped differentiate fruit and non-fruit objects, making them an essential element in image processing techniques for fruit detection [21].
- *Shapes and Depth Perception* - Shapes play a significant role in fruit detection, as each species possesses unique morphological characteristics [22]. Machine vision systems often rely on shape analysis to distinguish between different fruits. Depth perception [14], provided by binocular vision in humans and depth-sensing technologies in machines [25], allows for accurate fruit detection by discerning the spatial arrangement of fruits within a scene.

Visual cues, encompassing sensory perception, chromatic variations, grayscale distinctions, geometric configurations, depth perception, spectral analysis, and hyperspectral examination, play an indispensable role in fruit detection. These visual cues serve as pivotal tools for meticulously evaluating fruit characteristics, encompassing quality, ripeness, and assorted attributes, thereby facilitating the precise categorization and segregation of fruits across various domains, from agrarian harvesting to consumer procurement. The continuous evolution of imaging technologies, particularly in hyperspectral imaging, has notably refined the precision and efficiency of fruit detection processes, propelling the prospect of further innovations within this sphere. Perpetual exploration of these visual signals is anticipated to advance fruit detection methodologies and extend their applicability in agriculture and allied sectors.

2.2.2 Advance Visual Signals

Fruit analysis is a vital component of agriculture and food processing industries, directly influencing quality assessment, sorting, and decision-making processes. Advanced imaging techniques have gained prominence to enhance the accuracy and efficiency of fruit analysis.

- *Hyperspectral Imaging* - Hyperspectral imaging is a sophisticated technique that captures the spectral signature of each pixel in an image, enabling the identification and

characterization of materials based on their unique spectral properties. Hyperspectral imaging has demonstrated its potential for assessing various attributes, including ripeness, defects, and nutritional content in fruit analysis [15]. By examining spectral data across multiple wavelengths, hyperspectral imaging can detect subtle changes in fruit composition, allowing for early detection of defects and optimization of harvesting times [23].

- *Spectral Imaging* - Spectral imaging extends beyond the visible spectrum, offering a broader range of spectral information. This advanced technique aids in the discrimination –of fruit attributes that may not be perceptible to the naked eye. For instance, spectral imaging can detect the presence of diseases or pests by identifying specific spectral signatures associated with these issues [24] *figure 2-5*. Furthermore, it can provide insights into sugar content, moisture levels [16], and other quality parameters critical for fruit processing and marketing.
- *Infrared Imaging* - Infrared imaging is instrumental in assessing fruit's thermal properties, which can indicate their ripeness and quality. By capturing the heat radiation emitted by objects, infrared imaging can identify temperature variations within a fruit *figure 2-6*, helping to distinguish ripe fruit from unripe ones [17]. This technique is beneficial for sorting and grading fruits based on their thermal profiles.
- *3D Imaging Analysis* - 3D imaging techniques, including stereoscopic vision and structured light, facilitate the creation of three-dimensional representations of fruit surfaces. This depth information is valuable for volume estimation, shape analysis, and defect detection [22]. In fruit analysis, 3D imaging aids in accurate size determination, providing crucial data for packaging and storage considerations [18].

Advanced imaging methodologies, encompassing hyperspectral, spectral, infrared, and three-dimensional (3D) imaging, have engendered a transformative paradigm shift in the landscape of fruit analysis. These sophisticated techniques, characterized by their inherent capacity to provide exhaustive, non-intrusive insights into multifaceted fruit attributes, have become integral to optimizing fruit sorting, grading, and quality appraisal procedures. In so doing, they have contributed substantively to refining decision-making processes in the agricultural and food industries. Sustained endeavors in researching and developing these cutting-edge imaging modalities promise further breakthroughs in fruit analysis, with commensurate enhancements in productivity, waste reduction, and consumer satisfaction within this domain.

Figure 2-5 a) Original image, b) Spectral image

Figure 2-6 Japanese green pepper a) Infrared Image, b) Original image

2.3 Image Analysis Methods

Image analysis methodologies elucidate the precise procedures to extract visual cues from images to discern potential fruit locations. This exposition delineates several widely employed techniques, complemented by illustrative instances to enhance comprehension.

- *Thresholding* – Thresholding stands as one of the foundational image analysis techniques in agriculture. It involves segmenting an image into binary regions and distinguishing between foreground and background elements. This method primarily relies on setting a threshold value to separate objects of interest from their surroundings. Thresholding is extensively employed in agricultural applications for crop segmentation, weed detection, and fruit counting [19]. Researchers have explored various thresholding algorithms,

including global and adaptive techniques, to cater to the dynamic nature of agricultural environments. Choosing an appropriate thresholding method depends on lighting conditions, image quality, and the specific crop under consideration.

- *Color-Based Analysis* – Assessing color information in agricultural images is crucial for monitoring plant health, identifying diseases, and assessing fruit ripeness. Color-based analysis relies on color spaces such as RGB [19] (Red et al.), HSV [20] (Hue et al.), and LAB [21] to capture hue, saturation, and brightness variations. Machine vision systems with color cameras are often employed to acquire detailed color information from crops and vegetation. Color-based analysis can assist in diagnosing nutrient deficiencies, spotting pest infestations, and ensuring optimal harvesting times. Additionally, color-based analysis can be coupled with other techniques like texture analysis and machine learning for more comprehensive agricultural image analysis.
- *Shape Analysis* – Shape analysis plays a significant role in characterizing agricultural objects such as crops, fruits, and leaves. By quantifying geometric features, shape analysis facilitates classifying and identifying different agricultural elements. In plant phenotyping, for instance, shape analysis is instrumental in assessing traits like leaf area, fruit size, and canopy structure [18]. Researchers have employed a variety of techniques, including Fourier descriptors, Hu moments [26], and convex hulls, to extract relevant shape information from agricultural images. Additionally, computer vision and machine learning advances have enabled the development of robust shape-based classifiers, allowing for precise discrimination between healthy and diseased plants. Furthermore, the integration of structural similarity analysis [29] (SSIM) and edge detection [30] methods enhance the accuracy of shape analysis. SSIM helps in quantifying the structural similarity between reference shapes and objects in captured images, while edge detection techniques, such as Canny edge detection or Sobel operators, enable the extraction of shape boundaries, aiding in the precise characterization of agricultural shapes. These combined techniques contribute to more comprehensive and accurate assessments of agricultural elements, ultimately benefiting agricultural research and crop management.
- *Segmentation* – Segmentation is a fundamental image analysis method that partitions images into meaningful regions or objects *figure 2-8* [19]. In agriculture, segmentation techniques isolate individual plants, fruits, or other agricultural components from the background, facilitating subsequent analysis and decision-making processes. Segmentation can be performed using various approaches, including threshold-based

segmentation, region-growing, edge detection, and watershed transformation. The choice of segmentation method depends on factors such as image complexity, object shape, and the level of automation required [27]. For instance, in precision agriculture, segmenting crop rows or individual plants allows for precise monitoring and targeted interventions, optimizing resource utilization.

- *Machine Learning* - Machine learning (ML) has revolutionized image analysis in agriculture by enabling the development of sophisticated algorithms capable of learning and extracting complex patterns from images. ML techniques, including deep learning, support vector machines, and random forests, have remarkably succeeded in diverse agricultural applications. These applications encompass crop disease classification, yield prediction, pest detection, and weed management. Convolutional neural networks (CNNs) have mainly gained prominence in plant recognition tasks *figure 2-9* [28], as they can automatically extract hierarchical features from images, enabling high-precision classification and detection. Integrating ML with other image analysis methods, such as feature extraction and data augmentation, further enhances the accuracy and robustness of agricultural image analysis systems. a) b)

In conclusion, image analysis methods are integral to modern agriculture, offering a diverse toolkit for addressing crop management challenges, disease detection, and yield optimization. Thresholding, shape analysis, color-based analysis, segmentation, and machine learning collectively empower researchers and practitioners to harness the potential of agricultural imagery for informed decision-making and sustainable agricultural practices. As technology continues to advance, it is expected that the integration of these methods will play an increasingly pivotal role in shaping the future of agriculture, contributing to improved crop yield, reduced resource wastage, and enhanced food security on a global scale.

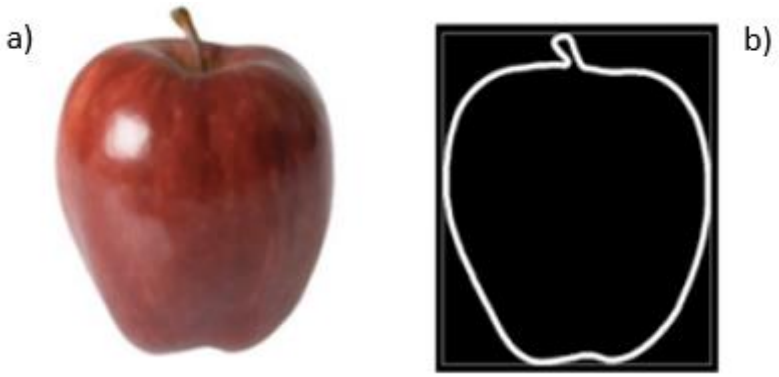

Figure 2-7 a) Original image, b) Apple with Edge detection [30]

Figure 2-8 Apple on trees a) Original image, b) Label image with segmentation [19]

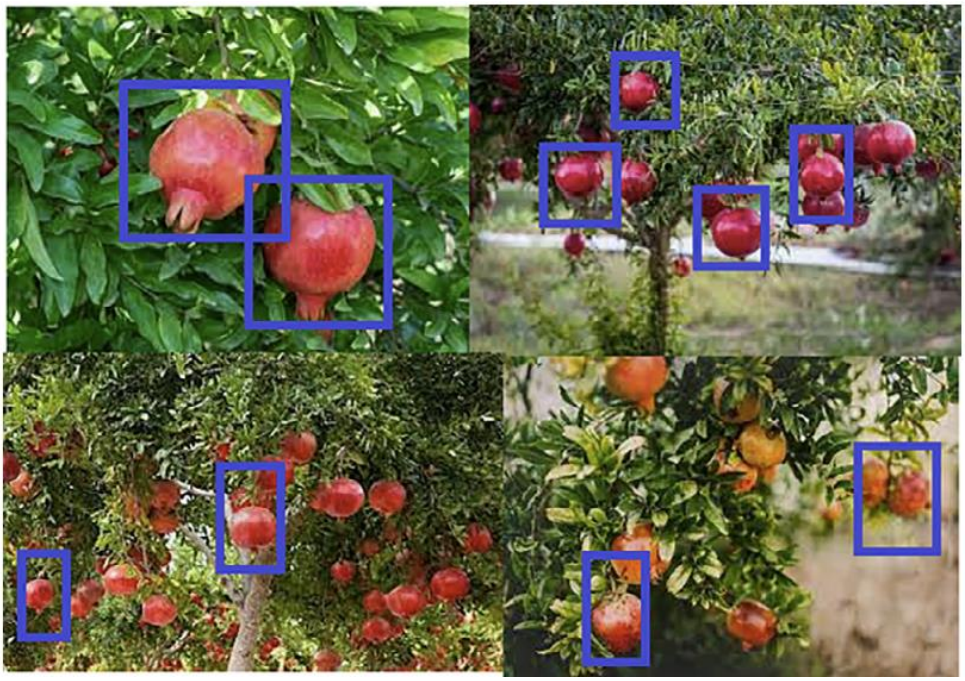

Figure 2-9 Fruit image detection [28]

2.4 Conclusion

In summary, the field of camera-based fruit detection in agriculture is pivotal in developing automated harvesting systems, with accurate identification and location of fruits essential for harvesting robots' efficiency. This conclusion has highlighted several critical challenges and advancements in this domain. Firstly, color variations in fruits due to ripeness, species, and environmental conditions pose significant challenges to camera-based detection systems. Researchers have employed color models such as RGB, HSV, and LAB, along with machine learning techniques like convolutional neural networks (CNNs), to address these variations effectively. Secondly, the issue of reflectance, especially in shiny fruits, affects their appearance in images.

Polarized light and multispectral imaging have been explored to mitigate the effects of reflectance, improving detection accuracy. Thirdly, occlusions caused by leaves, branches, or other fruits can obstruct the view of fruits in images. Researchers are developing algorithms to reconstruct partially occluded fruits or predict their positions, minimizing the impact of occlusions. Fourthly, changing lighting conditions throughout the day can lead to variations in color and texture, affecting detection accuracy. Adaptive illumination techniques, such as flash or LED lighting, have been integrated into camera systems to ensure consistent lighting and more reliable fruit detection. Lastly, shadows can obscure fruits and introduce false positives in detection systems. Shadow detection algorithms have been employed to differentiate between actual fruits and shadow artifacts. Furthermore, the section on visual signals highlighted the importance of human and machine perception in fruit detection, including color, grayscale, shape, depth perception, and spectral imaging. These visual cues are crucial for assessing fruit quality, ripeness, and other attributes. Additionally, advanced visual signal techniques were discussed, such as hyperspectral imaging, spectral imaging, infrared imaging, and 3D imaging analysis. These advanced methods have significantly enhanced the accuracy and efficiency of fruit analysis in agriculture and food processing industries.

3. Detection Assessment

3.1 Overview

The contemporary agricultural landscape has witnessed a notable transformation characterized by the incorporation of various technological tools to enhance agricultural assistance. One prominent technological advancement is the utilization of robots to harvest agricultural products. This shift towards automation has engendered expectations of a reduction in small-scale farmers, a trend that has been steadily declining in recent years. However, despite integrating robotics into farming practices, a substantial hurdle still needs to be solved in the form of high initial costs. The cost factor in robot adoption arises primarily from the need to equip these machines with various specialized components, including advanced cameras. The cost of a camera is intricately linked to its efficiency, with higher-quality cameras typically commanding higher prices. Moreover, the effectiveness of these cameras can be compromised by external factors such as inadequate

lighting conditions or obstructions like leaf shadows, which impede the penetration of light [6][31]. Additionally, certain types of cameras, such as infrared (IR) cameras, are susceptible to temperature-related challenges, particularly in cold weather, where objects' temperatures may closely resemble each other. This can render the detection and identification processes ineffective or unusable [4]. Given these challenges, more than relying on a single type of camera may be required. Consequently, integrating two different types of cameras has emerged as a potential solution to improve accuracy. This approach offers the advantage of a larger dataset for analysis and selection, which can enhance detection capabilities. However, it necessitates a trade-off with the processing system, potentially resulting in longer response times. Additionally, implementing dual-camera systems incurs an additional cost for installing both cameras and the associated detection system. Nevertheless, this investment is anticipated to contribute to greater efficiency and productivity in modern agricultural practices, ultimately aligning with the overarching goal of sustaining and improving agricultural production in the face of evolving technological challenges.

3.2 Equipment Test

This experiment responded to the need for precise data acquisition in a controlled environment. The study site consisted of green chili plants with leaves partially covering the fruits on one side, prompting whether maintaining the same camera distance would yield consistent image results. Consequently, a camera placement test involving diagonal positioning enabled one camera to capture the unobstructed side of the fruit. In contrast, another captured the side obscured by leaves to incorporate diverse data points that could be cross-referenced with information from other perspectives. As a result, prior to conducting concurrent imaging with two distinct camera types, it was imperative to assess the individual capabilities of each camera. This pre-testing phase aimed to preempt any potential issues arising during experimentation. Consequently, the camera experiment was subdivided into three distinct phases:

- Assessment of the RGB camera Intel RealSense D455
- Assessment of the Optris Xi400 IR camera
- Examination of the optimal distance and angle test

3.2.1 Assessment of RGB camera Intel Realsense D455

In order to concurrently employ both camera types, it is imperative to ensure that their positions are aligned in height. Consequently, it is imperative to ascertain the camera's dimensions during the testing phase. The camera's height was determined concerning the Intel Realsense datasheet

[32]. Specifically, the Intel Realsense D455 model possesses a width of 124 mm and a height of 29 mm , as shown in *figure 3-1* [32].

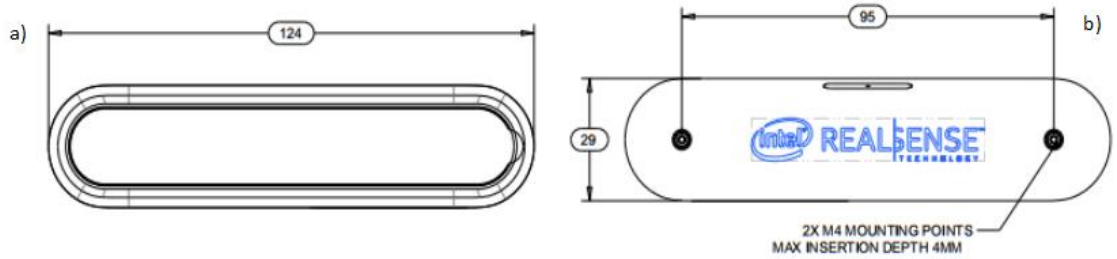

Figure 3-1 Intel Realsense D455 specification [32]

Consequently, the camera's focal point is situated at a height of 124.5 mm with tripod, , as shown in *figure 3-2*.

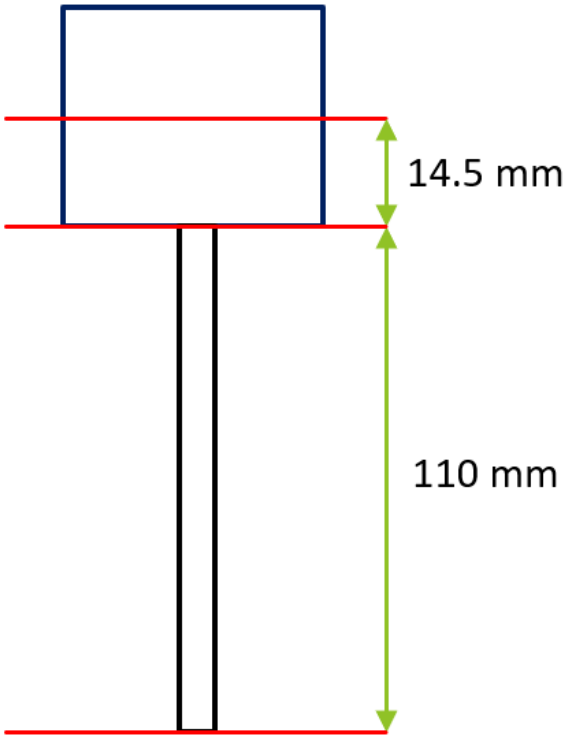

Figure 3-2 Intel Realsense D455 with tripod

Furthermore, the distance between the camera lens and the front mirror amounts to 4.55 mm, , as shown in *figure 3-3* [32]. Consequently, when configuring the camera for simultaneous operation with another camera type, it is necessary to establish an offset in the distance to ensure that both camera types maintain uniform separation.

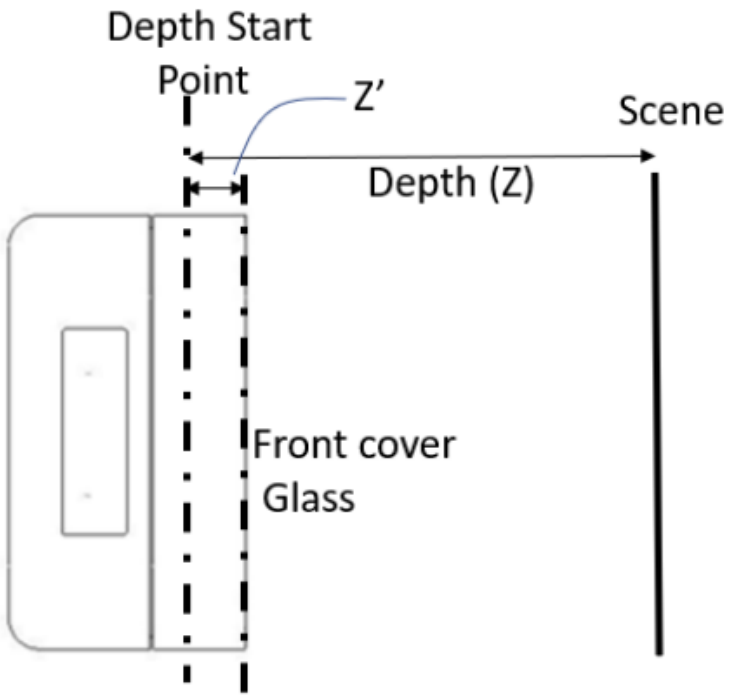

Figure 3-3 Depth of camera start point [32]

As the Intel Realsense camera comprises multiple sensor types within a single housing, it is essential to note that the desired sensor for our purposes is the RGB sensor. However, it is noteworthy that the RGB lens is not located at the optical center of the camera's plane but at 17.5 mm from its center, as shown in *figure 3-4* [32]. Consequently, when conducting simultaneous testing with two cameras, it becomes imperative to compute and establish an equal separation distance for the camera of the other type to maintain uniform alignment.

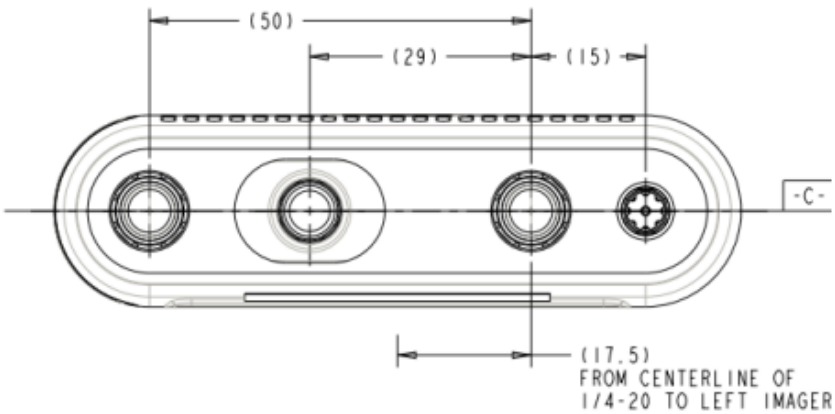

Figure 3-4 Camera position of Intel Realsense [32]

3.2.2 Assessment of IR camera Optris XI400

Compared to RGB cameras, infrared (IR) cameras exhibit distinct object detection characteristics, relying on capturing and processing heat sources to generate images. Optris XI400 camera, in particular, possesses a length of 99.50 mm and a height of 36 mm when excluding the mounting base. However, when the base is installed, the overall height increases to 73 mm, as shown in *figure 3-5* [33].

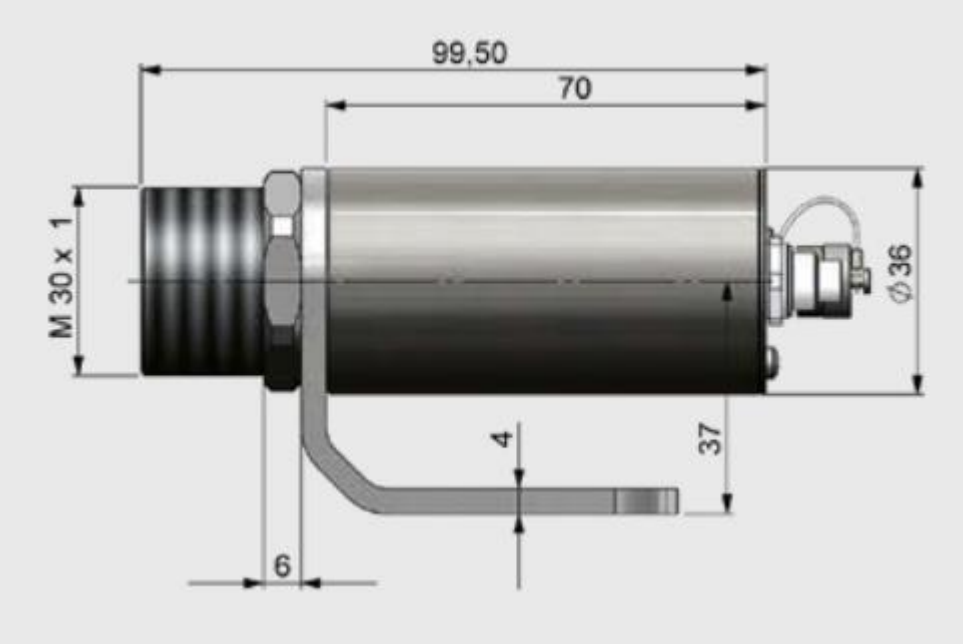

Figure 3-5 Optris XI400 specification [33]

3.2.3 Examination of the optimal distance and angle test

Utilizing multiple cameras to capture identical scenes introduces challenges as the resulting images exhibit distinct characteristics due to variations in camera positions. Despite visual similarities, the inherent spatial differences hinder the images from identical, mainly when utilized for diverse processing needs. This discrepancy poses a significant issue as it can lead to calculation inaccuracies. Camera calibration becomes imperative, enhancing the precision of location detection and minimizing errors. The chosen approach involves applying the principles of triangulation within a stereo-vision system, as shown in *figure 3-6* [34]. This method ensures a more accurate alignment of captured images, facilitating subsequent processing tasks. By adhering to the triangulation principle, the system aims to harmonize the spatial information from multiple cameras, enabling more reliable and consistent results in various applications and addressing the nuanced complexities of utilizing multiple cameras for image capture and subsequent analysis.

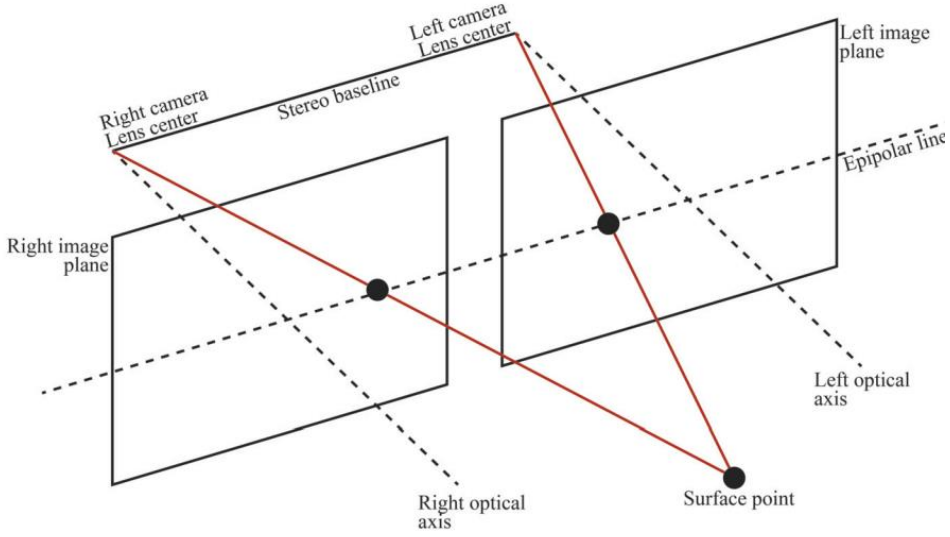

Figure 3-6 Stereo vision system triangulation principle [34]

After calibrating the surface point on the image planes, the subsequent step involves executing the triangulation process, as shown in *figure 3-7* [34]. This procedural step encompasses Equations 3-1, 3-2, and 3-3, designed to calculate the X, Y, and Z coordinates in real-world space. These equations bear a resemblance to the formulations employed in technical vision systems [34], thereby facilitating the computation of precise real-world coordinates based on the calibrated image points.

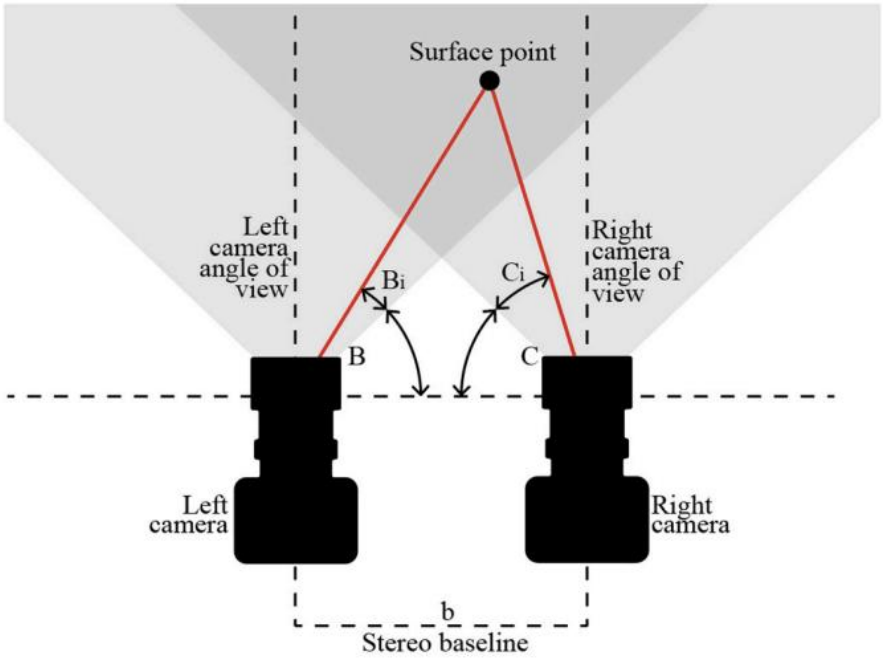

Figure 3-7 Stereo vision system horizontal angles of view (top view) [34]

Here, b represents the stereo baseline, which is the distance between the camera's centers. The variables B, C, and β are derived from the identified match points. In *figure 3-7*, a top view of the stereo vision system illustrates a surface point with angles B and C. These angles are computed concerning the camera's angle of view and the pixel position of the surface point. The calculation of B and C is accomplished through Equations 3-4 and 3-5.

$$X = b\left(\frac{sinB \cdot sinC}{\sin(B+C)}\right) \quad (3\text{-}1)$$

$$Y = b\left(\frac{cosB \cdot sinC}{\sin(B+C)} - \frac{1}{2}\right) \quad (3\text{-}2)$$

$$Z = b\left(\frac{sinB \cdot sinC \cdot tan\beta}{\sin(B+C)}\right) \quad (3\text{-}3)$$

$$B = B_i + B_0 \quad (3\text{-}4)$$

$$C = C_i + C_0 \quad (3\text{-}5)$$

In the context of the stereo vision system, Bi and Ci represent components of the angle within the respective fields of view of the left and right cameras. Their computation is achieved by applying Equations 3-6 and 3-7. Additionally, Bo and Co denote the angles from the baseline (b) to the commencement of their respective fields of view, establishing the starting points for detection angles (B and C). These values are determined using Equations 3-8 and 3-9. The relative orientation of the cameras serves as an extrinsic parameter essential for calibration, ensuring the accuracy of subsequent measurements.

$$B_i = H_{av}\frac{W_{is} - SP_{xl}}{W_{is}} \quad (3\text{-}6)$$

$$C_i = H_{av}\frac{SP_{xr}}{W_{is}} \quad (3\text{-}7)$$

$$B_0 = 90° - \frac{H_{av}}{2} \quad (3\text{-}8)$$

$$C_0 = 90° - \frac{H_{av}}{2} \quad (3\text{-}9)$$

In this context, Hav represents the horizontal angle of view for the respective camera, while Wis denotes the width of the corresponding image measured in pixels. The x value, expressed in pixels, pertains to the coordinates of the surface point on the left image, and $SPxr$ signifies the

x value of the surface point coordinates on the right image. Conversely, the computation of the β angle is facilitated through the utilization of Equation 3-10.

$$\beta = (V_{av}(\frac{\frac{V_{is}}{2} - SP_{ylr}}{V_{is}})) \quad (3\text{-}10)$$

In the given context, Vav represents the vertical angle of view of the camera, Vis denotes the height of the image measured in pixels, and $SPylr$ signifies the y value of the surface point coordinates. Referencing *figure 3-8* [34], which provides a side view of the stereo vision system, one can observe and appreciate both the vertical angle of view (Vav) and the β angle.

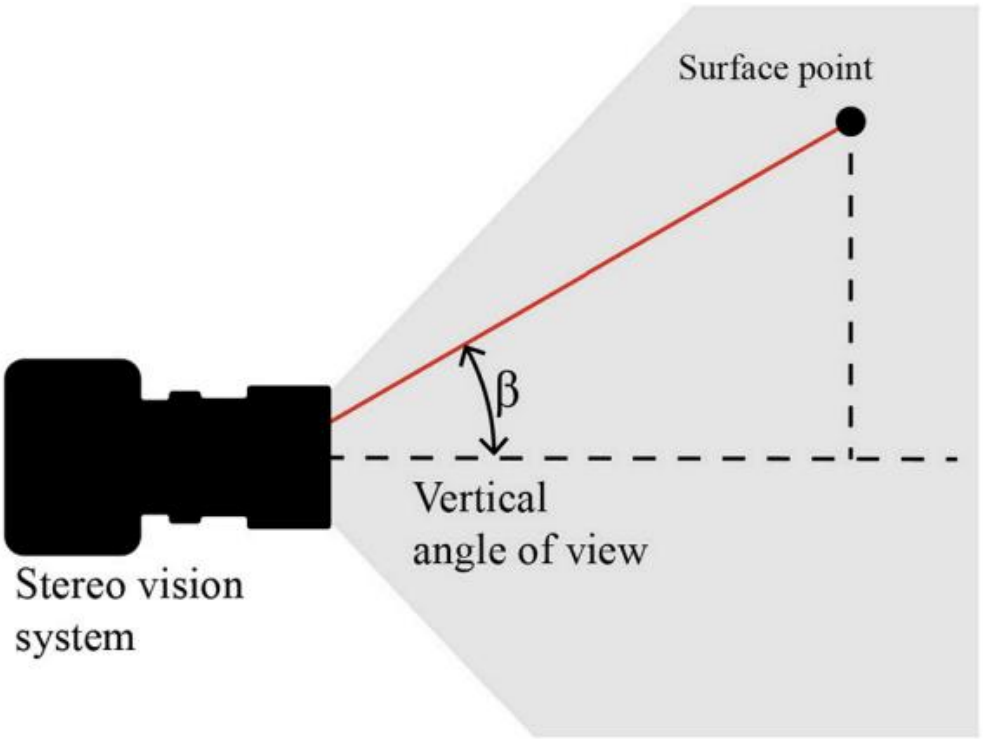

Figure 3-8 Stereo vision system vertical angle of view (side view) [34]

The procedure above delineates a methodology tailored for parallel cameras within the same plane, sharing identical camera types. However, the forthcoming experiment deviates from this configuration, as it involves cameras set parallel to each other but positioned at an inclined angle, as shown in *figure 3-9*.

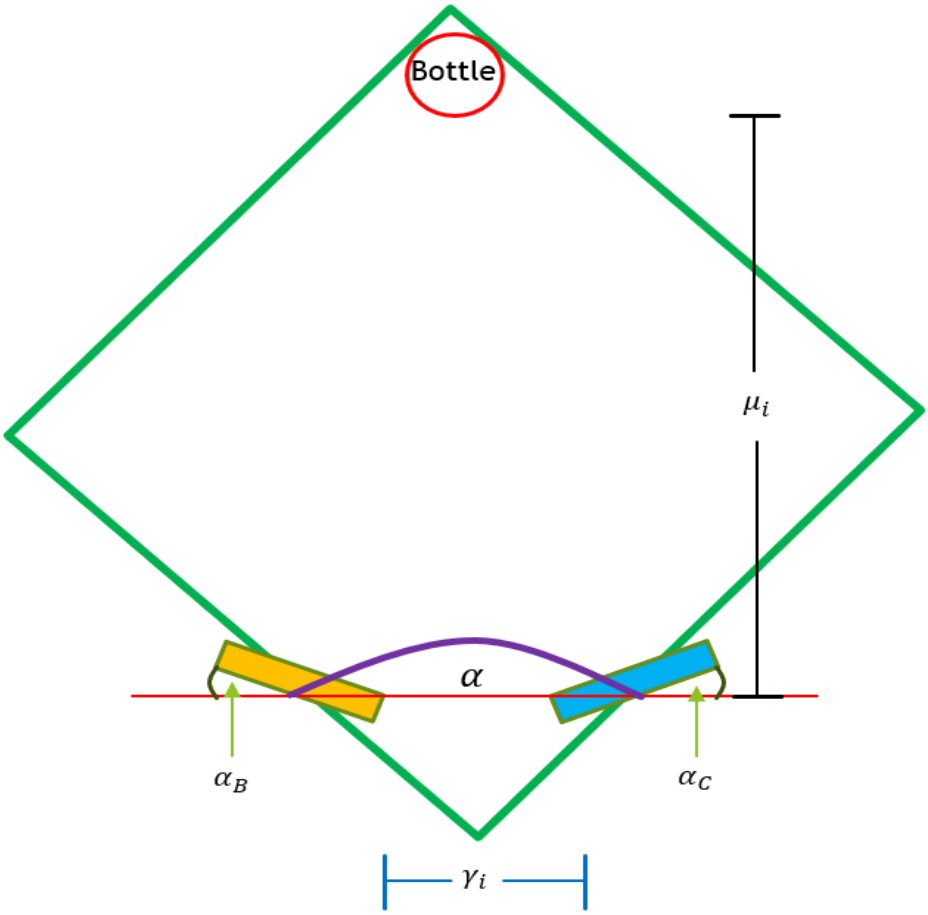

Figure 3-9 RGB Angle and Distance Prototype Test SETUP

Additionally, the cameras employed in this setup are of dissimilar types. Consequently, an additional equation is necessitated. Specifically, this Equation pertains to the angles of the two cameras oriented toward the object, denoted as Equation 3-11

$$\alpha = 180° - (\alpha_B + \alpha_C) \quad (3\text{-}11)$$

Moreover, there will be two additional variables: γ_i which is the distance between the two types of cameras, and μ_i the distance between the camera and the object that collects data.

γ_i = distance between the two types of cameras
μ_i = distance between the camera and the object

3.3 Experiment Method

The experimental procedures are organized into distinct phases:

- Evaluation of the angular orientation of RGB camera during this phase, an isolated evaluation of the RGB camera will be conducted , determining the threshold at which data collection from the object commences. In the experimental evaluation, there is a necessitated adjustment in camera orientation due to the off-centered placement of the RGB camera lens within the camera structure and the angular orientation of each camera is systematically varied within the range of 0 to 90 degrees. Alterations are implemented at increments of 5 degrees, guided by Equation 1, wherein the values α_B and α_C will consistently share the same degree values throughout the testing process.
- Assessment of the maximum distance achievable by each camera for data collection γ_i, ensuring that the camera does not capture images of other cameras within its field of view. The experimental assessment encompasses distances ranging from 15 to 30 centimeters.
- The conclusive phase of the experiment involves utilizing the values of α and γ_i to determine the optimal shooting distance, denoted as μ_i . This optimization considers both the shooting distance and the angles of both camera types. During image capture, it is imperative to ensure no overlap between the field of view of one camera type and another, maintaining distinct visibility for each camera type.

3.4 Results

- The experimental findings indicate that photographing at various angles before and after the side-switching operation proved unproblematic within the range of 0 to 75 degrees. However, deviations arose when the angle exceeded 80 degrees, leading to the object's image extending beyond the frame boundaries.
- Concerning the range of cameras capable of initiating object capture image, this capability extends from γ_i = 0 to 30 cm within the camera's plane at 0 degrees. This ensures unimpeded photographing without including another camera type within the frame of the image.
- The first two experiments were amalgamated and tested collectively in the concluding phase. The outcomes indicate that for the left side, with the RGB camera on the left and

the IR camera on the right, images can be captured without interference from other camera types within the range of γ_i = 15.5 - 22 cm. The parameters α = 120°, α_B and α_C are set within 0-30 degrees. Similarly, for the right side, with the RGB camera on the right and the IR camera on the left, images can be taken without incorporating other camera types within the range of γ_i = 15.5 - 22 cm, with the same angular constraints.

- Regarding μ_i, the RGB camera exhibits no issues and can capture images within the 15-30 cm range. However, the IR camera encounters challenges when the camera-object distance exceeds 28 cm, manifesting barrel distortion symptoms. The subsequent summary section will provide further elucidation on barrel distortion symptoms.

3.5 Summary

The experimental findings suggest that the optimal shooting angle for capturing images without other cameras in the frame is 30 degrees. Maintaining a distance between cameras ranging from 15.5 to 22 cm is recommended, with a preference for the lower limit of 15.5 cm for space efficiency in potential installations on an automated harvesting robot. The permissible range for camera-object distance extends from 15 to 30 cm without encountering issues in the RGB camera. However, in the case of IR cameras, complications arise beyond 28 cm, leading to a phenomenon known as barrel distortion *figure 3-10* , akin to a fisheye lens effect, as shown in *figure 3-111* [37]. Commonly associated with wide-angle lenses, this distortion can be rectified through algorithms, such as the Correction of Barrel Distortion in Fisheye Lens Images Using Image-Based Estimation of Distortion Parameters by M. Lee [35] or T. Hwan Kim's An Efficient Barrel Distortion Correction Processor for Bayer Pattern Images [36]. However, due to the significantly lower resolution of the IR camera 382x288 pixels [33] compared to the RGB camera 1280x800 [32], image editing complexities arise in the IR domain. Consequently, employing the IR camera within a shooting distance of less than 28 cm is recommended for expeditious and straightforward resolution, thereby mitigating challenges and noise during subsequent image processing.

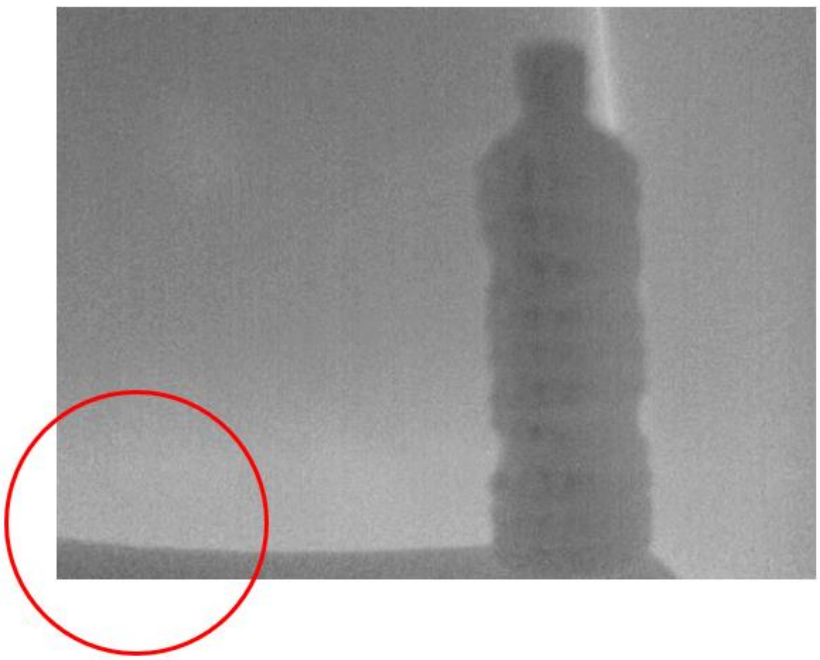

Figure 3-10 Barrel distortion of IR image when range distance exceeds 28 cm

Figure 3-11 Image of Brick wall captured with wide angle lens [37]

4. First Detection Method

4.1 Mask R-CNN

Artificial Intelligence (AI) systems represent advanced computational frameworks that emulate and supplement human capabilities through continuous learning. These systems excel in executing intricate tasks that can substitute human involvement, particularly in domains requiring discernment. The significance of AI in contemporary society is paramount, especially in tasks reliant on visual discrimination. While the human eye exhibits remarkable accuracy in distinguishing objects, its limitations become apparent during prolonged periods of continuous activity, leading to an escalating error rate. In stark contrast, AI systems, including robots, can operate ceaselessly, 24 hours a day, without succumbing to fatigue, and maintain consistent accuracy over time. Achieving such proficiency in AI involves a preliminary phase known as training, wherein the system learns from labeled datasets, a process referred to as machine vision. Machine vision encompasses diverse methodologies like neural networks [38], deep learning [39], and machine learning [40]. This study adopts the deep learning approach, explicitly employing the Mask Region-based Convolutional Neural Network (Mask R-CNN) [41]. This network architecture, a subset of Convolutional Neural Networks (CNNs) [42], is particularly adept at segmentation, a technique that isolates desired objects by applying masks. The targeted objects are defined within a specified scope in the segmentation process, as shown in *figure 4-1* [43]. The subsequent phase involves training the model with data to distinguish the desired objects. Upon completing the training data, the AI system becomes proficient in discerning specific objects, as exemplified by its capability to differentiate between sweet peppers, as shown in *figure 4-2* [44].This underscores the efficacy of deep learning techniques, specifically Mask R-CNN, in enhancing AI's capacity for object recognition and segmentation tasks.

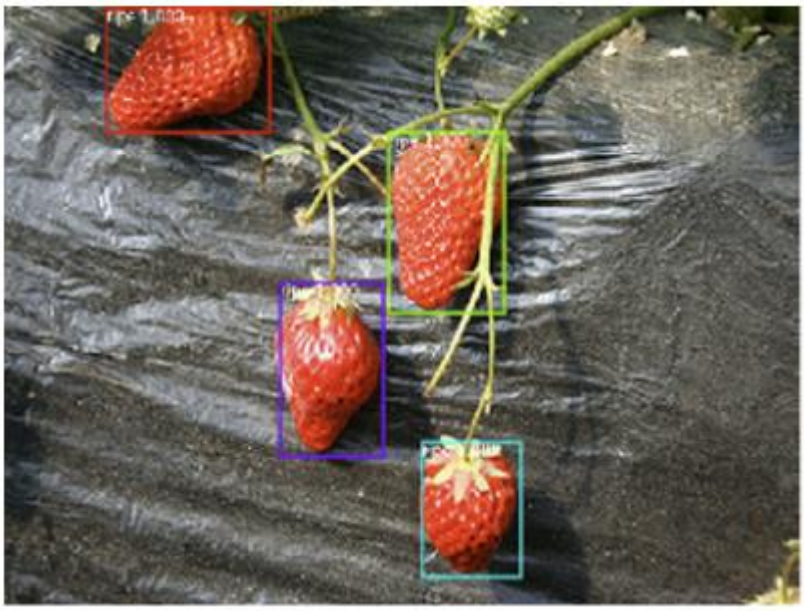

Figure 4-1 Strawberry detection with mask label [43]

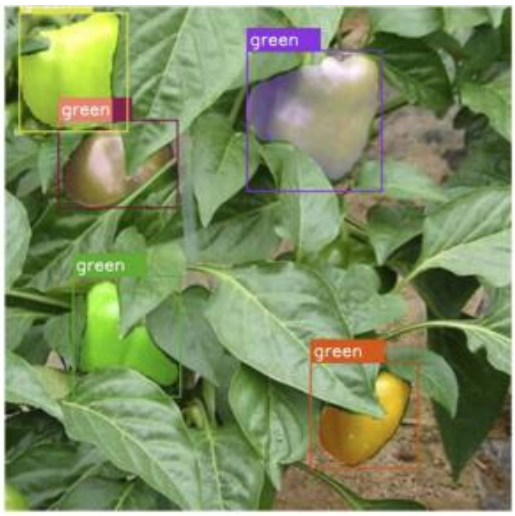

Figure 4-2 Green pepper detection with mask label and segmentation [44]

4.2 Material and Methods

4.2.1 Image Acquisition

This investigation gathered a dataset comprising greenhouse green pepper images generously provided by KUT. The dataset incorporated diverse environmental conditions, encompassing sunny and cloudy days, and captured the subjects from various perspectives. The dataset employed in this study comprised a total of 4320 images, encompassing four distinct green pepper types and two different image modalities (RGB and IR) , as shown in *figure 4-3*. The authors judiciously distributed each category into training and validation sets to ensure a comprehensive evaluation, employing a randomized allocation method, as shown in Table 1.

Specifically, the training sets were designated for utilization during the model training phase, serving as the original input images for the training process. In contrast, the validation sets were reserved for assessing the model's performance after training. This meticulous dataset division into training and validation subsets facilitates a robust evaluation of the developed model under varying conditions, ensuring its efficacy and generalizability beyond the training data. Table 1 provides a succinct representation of the randomized allocation of images across the training and validation sets for each category.

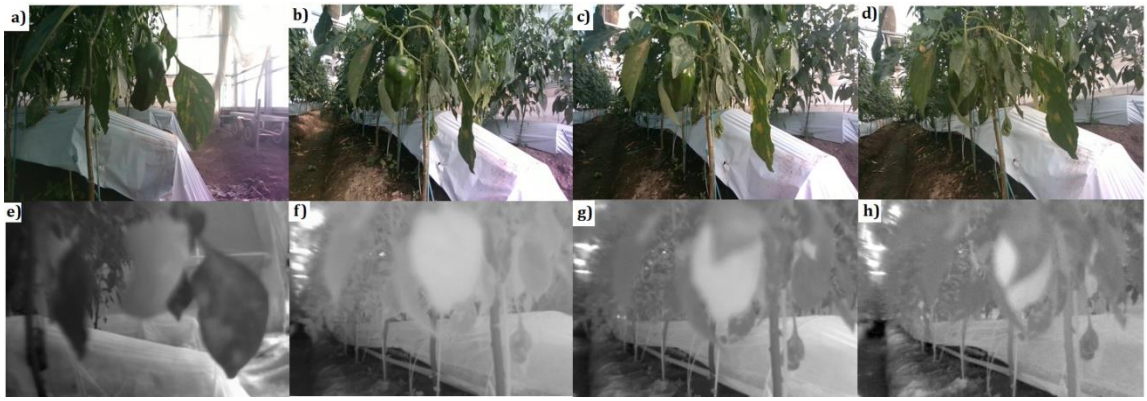

Figure 4-3 Dataset of green pepper a), b), c), d), e), f), g), h)

Table 1. Green pepper types and dataset

Category	Training set	Validating set	Testing set
RGB : left side covered by foliage 0% (a)	378	108	54
RGB : right side covered by foliage 0% (b)	378	108	54
RGB : right side covered by foliage 10-30% (c)	378	108	54
RGB : right side covered by foliage >30% (d)	378	108	54
IR : left side covered by foliage 0% (e)	378	108	54
IR : right side covered by foliage 0% (f)	378	108	54
IR : right side covered by foliage 10-30% (g)	378	108	54
IR : right side covered by foliage >30% (h)	378	108	54

4.2.2 Image Preprocessing

Data augmentation was embraced to expand the sample size further to enhance the dataset's comprehensiveness, augment the feature information across various levels within the images, and improve the algorithm's adaptability to real-world scenarios. Specifically, the data augmentation technique employed in this investigation incorporated Laplacian sharpening. The utilization of Laplacian sharpening serves to enhance image sharpness, rendering object edge

details within the image more distinct. Additionally, it addresses issues arising from unclear images due to low resolution. The application of the Laplace operator, integral to Laplacian sharpening, is instrumental in achieving these improvements. Laplacian sharpening encompasses the application of the Laplacian operator to an image [44]. The Laplacian operator, symbolized as ∇^2, constitutes a second-order derivative and is frequently expressed in mathematical terms as follows equation 4-1.

$$\nabla^2 f(x,y) = \frac{\partial^2 f(x,y)}{\partial x^2} + \frac{\partial^2 f(x,y)}{\partial y^2} \quad (4\text{-}1)$$

Within the domain of image processing, the technique of Laplacian sharpening entails the deduction of the outcome derived from applying the Laplacian operator to the original image from the original image. This process generates a sharpened image, denoted as g and can be formally articulated as follows equation 4-2.

$$g(x,y) = f(x,y) - \nabla^2 f(x,y) \quad (4\text{-}2)$$

$g(x,y)$ is the sharpened image.

$f(x,y)$ is the original image.

In this context, where x and y represent pixel coordinate values, $g(x,y)$ signifies the resulting sharpened image, $f(x,y)$ denotes the original image, $\nabla^2 f(x,y)$ represents the Laplace transform of the original image, and the Laplace mask is visually presented in equation 4-3 and results after image sharpening, as shown in *figure 4-4*

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix} \quad (4\text{-}3)$$

Figure 4-4 Image sharpening a) Original RGB image, b) RGB After sharpening
c) Original IR image, d) IR After sharpening

4.2.3 Dataset Annotation

The image annotation process holds pivotal significance in training models, as it involves delineating object boundaries to ensure the specificity of model training towards desired objectives. In this context, the annotation tool employed is Labelme. The experimental dataset was annotated using Labelme to produce mask images corresponding to the delineation of green peppers within the images. Furthermore, evaluating the trained model's performance in instance segmentation involved a comparative analysis between the annotated mask images and the model's predicted mask outputs. Specifically, regions of the images corresponding to green peppers were meticulously labeled, while the remaining areas were designated as background. The resultant annotated images depicting the labeled regions of green peppers are depicted in *figure 4-5*.

Figure 4-5 RGB and IR image with label and mask box

4.2.4 Target Detection of Mask R-CNN

Mask R-CNN, an abbreviation for Mask Region-based Convolutional Neural Network, stands at the forefront of contemporary computer vision research, exhibiting remarkable prowess in instance segmentation. Introduced as an extension of the Faster R-CNN architecture, Mask R-CNN seamlessly integrates object detection and segmentation, enabling precise delineation of object boundaries and identifying distinct instances within an image [44]. The fundamental innovation within Mask R-CNN lies in its ability to concurrently generate pixel-level masks for each object instance while performing object detection. This task amalgamation is accomplished by incorporating a dedicated mask branch parallel to the existing branches for object classification and bounding box regression. Leveraging a two-stage approach, Mask R-CNN initially proposes region proposals through the Region Proposal Network (RPN) and subsequently refines these proposals with refined bounding box coordinates and corresponding instance masks. The architecture's robustness is underscored by its capacity to handle various object scales and shapes, rendering it highly adaptable to complex scenes. Mask R-CNN has proven instrumental in various applications, ranging from medical image analysis to autonomous vehicles, owing to its proficiency in extracting fine-grained spatial information. As a testament to its efficacy, Mask R-CNN has emerged as a cornerstone in instance segmentation, embodying the paradigm shift towards comprehensive visual scene understanding and semantic segmentation in academic ,industrial spheres and an overview of Mask R-CNN, as shown in *figure 4-6*.

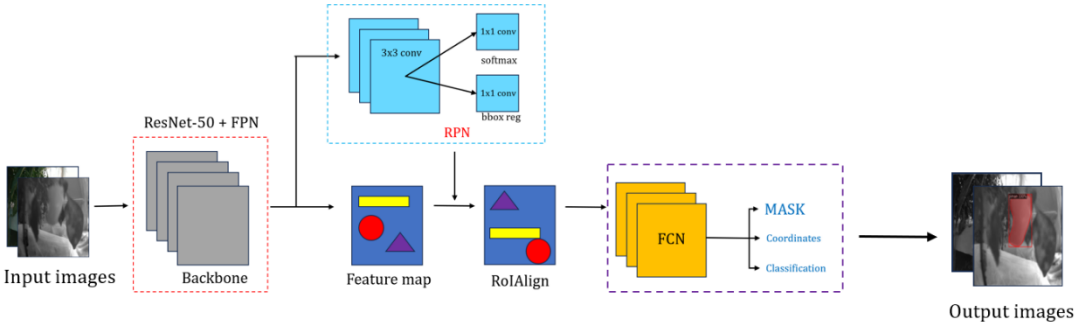

Figure 4-6 Overview of Mask R-CNN

4.2.5 Feature Extraction and ROI

The establishment of deep neural network models with varied depths is accomplished through the design of different weight layers. AlexNet, ZF, VGG, GoogleNet, and ResNet currently stand as prominent models in the domain of deep neural networks [43]. Although deeper networks have the potential to yield higher accuracy, a trade-off exists with a reduction in model training and detection speeds. ResNet, notable for its residual structure mitigating challenges such as gradient disappearance and training degradation without increasing model parameters, has been chosen as the foundational network for feature extraction in this research. Image feature extraction is based on shared convolution layers within this framework. The underlying network captures low-level features, such as edges and angles, while higher-level features describing target categories are extracted at elevated levels. To enhance the representation of fruit targets across multiple scales, the Feature Pyramid Network (FPN) is introduced, extending the backbone network. This is particularly effective for detecting small targets. The FPN architecture merges top-level features with underlying features through up-sampling, independently predicting feature maps for each layer [45]. This research employs two types of cameras, resulting in two distinct pixel configurations for RGB and IR images of a single green pepper. FPN outputs for different levels are designed to accommodate these varying image scales. The study focuses on analyzing single green peppers using two distinct imaging modalities: RGB and Infrared (IR). The base image size is standardized for the RGB images at 720x1280 pixels. Feature Pyramid Network (FPN) outputs at different levels are meticulously tailored to accommodate these specifications:

FPN output for level 2: 1/4 of the input size, resulting in dimensions of 180x320 pixels.

FPN output for level 3 is 1/8 of the input size, yielding dimensions of 90x160 pixels.

FPN output for level 4: 1/16 of the input size, presenting dimensions of 45x80 pixels.

FPN output for level 5 is 1/32 of the input size, culminating in dimensions of 22.5x40 pixels.

Concurrently, Infrared (IR) images of single green peppers are acquired with a base image size of 288x382 pixels. The FPN outputs at different levels for IR images are adjusted accordingly:

FPN output for level 2 is 1/4 of the input size, resulting in dimensions of 72x96 pixels.

FPN output for level 3 is 1/8 of the input size, yielding dimensions of 36x48 pixels.

FPN output for level 4 is 1/16 of the input size, presenting dimensions of 18x24 pixels.

FPN output for level 5 is 1/32 of the input size, culminating in dimensions of 9x12 pixels.

These meticulously designed imaging scales and FPN outputs are integral to generating Region of Interest (RoI) for subsequent analysis and facilitate the effective representation and detection of features in the study's context of single green peppers. In RoI generation, the aspect ratio of labeled rectangular boxes for single and occluded green peppers is approximately 1:1, determined by bounding box definition using minimum and maximum coordinates in both x and y directions. The FPN outputs play a crucial role in this process, offering tailored information for generating RoIs.

4.3 Image Segmentation and Loss Function

The RoIAlign-generated feature maps were subsequently processed through fully convolutional network. The utilization of fully convolutional network was threefold, encompassing classification, bounding box regression and coordination. Conversely, applying fully convolutional network was dedicated to segmenting individual instances of green peppers. The classification task involved feeding the outputs from the fully connected network into a Softmax layer, thereby obtaining the classification probabilities. Simultaneously, the convolutional layers were employed for the intricate instance segmentation process. The training of the network entailed the establishment of a loss function, which quantified the disparities in the network prediction. Assuming P_{class} , P_{bbox}, and P_{mask} represent the predicted class probabilities, bounding box coordinates, and mask predictions, respectively, and T_{class}, T_{bbox} , and T_{mask} represent the corresponding values, the loss function can be written as equations 4-4 , 4-5, and 4-6.

Classification Loss (Cross-Entropy):

$$L_{class} = -\frac{1}{N_{roi}} \sum_{i=1}^{N_{roi}} \sum_{c=1}^{C} T_{class,i,c} \log\left(P_{class,i,c}\right) \quad (4\text{-}4)$$

Bounding Box Regression Loss (Smooth L1):

$$L_{bbox} = \frac{1}{N_{roi}} \sum_{i=1}^{N_{roi}} \sum_{j\in\{x,y,w,h\}} smooth_{L1}(P_{bbox,i,j} - T_{bbox,i,j}) \quad (4\text{-}5)$$

Mask Segmentation Loss (Binary Cross-Entropy):

$$L_{mask} = -\frac{1}{N_{roi}} \sum_{i=1}^{N_{roi}} \sum_{p=1}^{(H)(W)} [T_{mask,i,p} \log\left(\sigma\left(P_{mask,i,p}\right)\right) + (1 - T_{mask,i,p}) \log\left(1 - \sigma(P_{mask,i,p})\right)] \quad (4\text{-}6)$$

N_{roi} denotes the count of regions of interest (RoI), C signifies the number of distinct classes, and $H$and W correspond to the height and width of the precited mask, respectively. Additionally, σ represents the sigmoid function. Comprehensive loss is formulated as the cumulative sum of individual losses, incorporating potential weighting coefficients is presented in equation 4-7.

Total Loss: $L_{total} = \lambda_{class} L_{class} + \lambda_{bbox} L_{bbox} + \lambda_{mask} L_{mask}$ (4-7)

It is noteworthy that the user has the flexibility to fine-tune the weighting coefficients (λ_{class}, λ_{bbox}, λ_{mask}) based on the relative significance of each constituent in the specific context of their application. This flexibility allows for the customization of the loss function to best align with the prioritized aspects of the given task.

4.4 Structural Similarity (SSIM)

The Structural Similarity Index (SSIM) stands as a fundamental metric in the realm of image processing and computer vision, serving the critical purpose of quantifying the degree of similarity between two images. Traditional metrics like Mean Squared Error (MSE) typically focus on pixel-wise differences, but SSIM distinguishes itself by incorporating luminance, contrast, and structure considerations. This holistic approach mirrors critical aspects of human visual perception and addresses the limitations of conventional methods. SSIM's evaluative scope extends beyond mere pixel-level disparities, enabling a nuanced assessment of global and local image variations. The index generates values on a scale from -1 to 1, with 1 indicating perfect similarity. This scalar output encapsulates SSIM's ability to capture the quantitative distinctions in pixel values and the qualitative aspects of structural and textural information within images.

Consequently, SSIM finds application in diverse tasks, including image quality assessment, compression optimization, and image restoration. Its multifaceted utility has led researchers and practitioners to leverage SSIM in refining and optimizing image processing algorithms. This utilization ensures that algorithmic outputs align more closely with human perception, ultimately enhancing the overall visual fidelity of digital imagery. The broad adoption of SSIM underscores its significance as an indispensable tool for objectively gauging image similarity and quality. Its impact reverberates across various applications within computer science and multimedia domains, emphasizing its role in advancing the state of the art. Beyond its foundational role, SSIM's relevance is expanding across diverse sectors. In biomedicine, D. J. Vresdian's work exemplifies SSIM's application in iris detection. This biomedical technique aids in diagnosing systemic health based on the patterns and characteristics of the iris. Notably, normalization plays a pivotal role in providing image data that facilitates more straightforward observation of iris patterns, contributing to enhanced diagnostic precision [46]. In the realm of radiation therapy, J. Peng utilizes SSIM to monitor the accurate delivery of doses to the target. This application is crucial, as ensuring the consistency between planned and delivered doses is paramount in patient-specific quality assurance. SSIM provides an intuitive means of evaluating the overlap between planned and delivered dose profiles, contributing to the meticulous management of radiation therapy processes [47]. In industrial applications, F. Hu leverages SSIM to detect and distinguish coal from gangue in the coal industry, employing multispectral imaging for heightened accuracy [48]. Meanwhile, in the agricultural sector, Y. Hong incorporates SSIM for fruit inspection, overcoming challenges associated with traditional sensor methods. The implementation of SSIM proves particularly valuable in scenarios where conventional approaches face limitations, exemplifying its adaptability and effectiveness in addressing diverse challenges across sectors [49]. The Structural Similarity Index emerges as a cornerstone in image processing and computer vision and as a dynamic and versatile tool with expanding applications across various domains. Its nuanced approach to image similarity assessment and its adaptability to diverse challenges position SSIM as a pivotal asset in advancing technological applications and contributing to the refinement of imaging techniques across sectors.

4.4.1 SSIM Algorithm

In the comparative analysis of two images facilitated by a software system, the Structural Similarity Index (SSIM) principles are applied to ensure a comprehensive evaluation that transcends mere correlation coefficients. The first step involves mitigating the impact of brightness on structural information. Luminance information is subtracted during the calculation

of structural information, and subsequently, the mean value of the image is subtracted. This initial adjustment aims to preserve the inherent structural characteristics of the fruits depicted in the images. Subsequently, the structural information is further refined to eliminate the influence of image contrast. Normalization of the variance of the images is undertaken during the computation of structural details. This step ensures that the structural features are assessed independently of variations in image contrast, contributing to a more precise analysis. The final phase involves the comprehensive calculation of structural information, incorporating the outcomes of brightness and contrast comparisons [49]. The conventional approach of calculating correlation coefficients is augmented to account for the nuanced impact of brightness and contrast on image dissimilarity. The overarching workflow of this SSIM process is delineated in *figure 4-7*, emphasizing the sequential application of these principles in achieving a holistic evaluation of image similarity. By systematically addressing the influence of luminance and contrast, the SSIM methodology ensures a refined and nuanced assessment, offering a more accurate depiction of the similarity between two images within the software system.

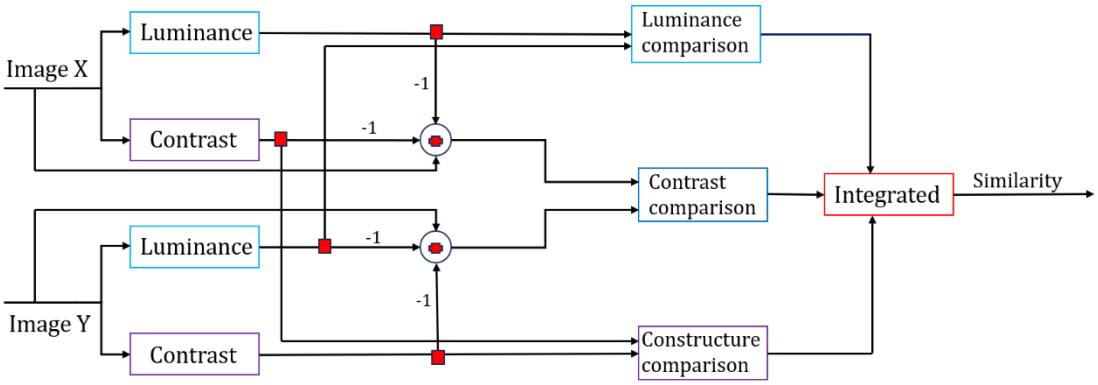

Figure 4-7 Workflow structure of SSIM

4.4.2 Calculation Process

The Structural Similarity Index (SSIM) comprises three constituent sub-indices: the luminance index, contrast index, and structure index. Luminance, within the context of the SSIM index, pertains to the intensity of the object portrayed in the image, delineated by pixel values. The luminance index, therefore, serves as a metric for capturing the inherent brightness characteristics of the recorded object within the image. The contrast index encapsulates the discernible difference in luminance or the extent of luminance variation across the image. This index provides a quantitative measure of the variability in luminance values, offering insights into the image's overall contrast properties. As a component of the SSIM, the structure index

reflects the Pearson correlation of luminance between two images, namely, image X and image Y. This index evaluates the similarity in the structural patterns of luminance across corresponding points in the images. The comparison functions for luminance, contrast, and structure at each point in the images are expressed through equations 4-8, 4-9, and 4-10, respectively. These equations encapsulate the mathematical formulations employed to quantify the luminance intensity, contrast variation, and structural correlation, forming the basis for a comprehensive evaluation of image similarity within the SSIM framework.

Luminance: $l(x,y) = \frac{2\mu_x\mu_y+C_1}{\mu_x^2+\mu_y^2+C_1}$ (4-8)

Contrast: $c(x,y) = \frac{2\sigma_x\sigma_y+C_2}{\sigma_x^2+\sigma_y^2+C_2}$ (4-9)

Structure: $s(x,y) = \frac{\sigma_{xy}+C_3}{\sigma_x\sigma_y+C_3}$ (4-10)

In the Structural Similarity Index (SSIM) formulation, μ_x and μ_y represent the local means, σ_x and σ_y denote the standard deviations, and σ_{xy} signifies the cross-covariance between image X and image Y. The equations 4-11, 4-12, and 4-13 express the mathematical definitions of μ_x, σ_x, and σ_{xy}. To ensure computational stability and prevent division by minute denominators, C_1, C_2, and C_3 function as regularization constants with diminutive values. The introduction of these constants is imperative for mitigating potential numerical instabilities in SSIM calculations, thereby reinforcing precision and robustness in diverse image-processing contexts.

$\mu_x = \frac{1}{N}\sum_{i=1}^{N} x_i$ (4-11)

$\sigma_x = ((\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu_x)^2)^{\frac{1}{2}}$ (4-12)

$\sigma_{xy} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu_x)(y_i-\mu_y)$ (4-13)

The index i represents all points within a localized region. At the same time, N signifies the total number of points encompassed by this area, including the evaluating point and its N neighboring points. The configuration of the local area is adaptable, allowing for adjustments in shape and size through the selection of filter types, such as the Gaussian filter and filter size. Ultimately, the Structural Similarity Index (SSIM) integrates three sub-functions, culminating in its final

formulation as shown in equations 4-14 and 4-15. These equations encapsulate the SSIM index's mathematical representation, a composite measure derived from the interplay of these sub-functions.

$$SSIM(x,y) = [l(x,y)]^{\alpha}[c(x,y)]^{\beta}[s(x,y)]^{\gamma} \quad (4\text{-}14)$$

$$SSIM(x,y) = \frac{(2\mu_x\mu_y+C_1)(2\sigma_{xy}+C_2)}{(\mu_x^2+\mu_y^2+C_1)(\sigma_x^2+\sigma_y^2+C_2)} \quad (4\text{-}15)$$

An SSIM index attaining 1 signifies perfect concordance between two images, whereas an SSIM index below 1 indicates a disparity between the compared images. The overall SSIM and its sub-indices for the compared images are computed as the mean values of their respective index maps. This analytical approach allows for a quantitative assessment of image similarity, with 1 indicating a complete match and values lower than one indicating deviations or dissimilarities between the compared images. Using mean values in the computation contributes to a comprehensive and representative evaluation of the images' structural, luminance, and contrast attributes.

4.5 Experimental Method

Following the exploration detailed in the third chapter involving camera experimentation, the subsequent phase of the study involves practical assessments utilizing an authentic camera within a greenhouse environment to acquire empirical data for further investigation. Data collection transpired on the dates 11/29, 11/30, and 12/1. The initial step encompassed the systematic recording of data as per Table 1. This involved capturing images in four RGB types (a-d) and four IR types (e-h), each associated with specific locations as shown in *figure 4-8 and 4-9*, and green pepper 540 images were recorded for each type, resulting in 4320 images. Furthermore, each image type was subdivided into three sets: 1) training set, 2) validating set, and 3) testing set. The experimental procedures are methodically organized into distinct phases.

- The initial phase of the study involved data annotation through the utilization of labelme. Subsequently, the annotated data was employed to assess the accuracy of the Mask R-CNN system algorithm. This iterative process facilitated evaluating the algorithm's performance in processing annotated data, providing insights into its efficacy and precision in handling the specific task. The systematic data annotation through labelme was a foundational step in preparing the dataset for algorithmic testing and performance evaluation.

- In the second procedural step, images obtained from the segmentation process in step 1 were utilized for testing purposes. Specifically, the acquired data were employed in two distinct scenarios: first, employing infrared (IR) images as the testing dataset for models trained with red-green-blue (RGB) images, and second, using RGB images as the testing dataset for models trained with infrared (IR) images. This methodology aimed to assess the efficacy of the segmentation algorithm in identifying green peppers under varying conditions, specifically evaluating its ability to generalize across different imaging modalities and ascertain the robustness of the model in pepper identification.
- In the third procedural step, after obtaining images from the initial segmentation step, an analytical procedure was employed utilizing the Structural Similarity Index (SSIM) method. This method entailed a comparative analysis akin to the approach in step 2, involving the juxtaposition of infrared (IR) images with red-green-blue (RGB) images and vice versa. The objective was to ascertain the discernibility of green peppers through the SSIM method, thereby evaluating the effectiveness of the segmentation process in distinguishing these peppers based on variations in imaging modalities. This analytical step contributed to the comprehensive assessment of the algorithm's performance in identifying green peppers across different image representations.
- In the fourth procedural step, this process resembles step 3, albeit explicitly focusing on a designated point of interest. Unlike the comprehensive image analysis in the prior step, the comparison is selectively confined to the region delineated by the bounding box. This targeted approach aims to streamline the computational workload by restricting the assessment to solely those images within the specified area. By doing so, the intent is to mitigate data volume, enhance computational efficiency, and diminish extraneous noise that might arise from considering the entire image, thereby refining the precision of the evaluation.

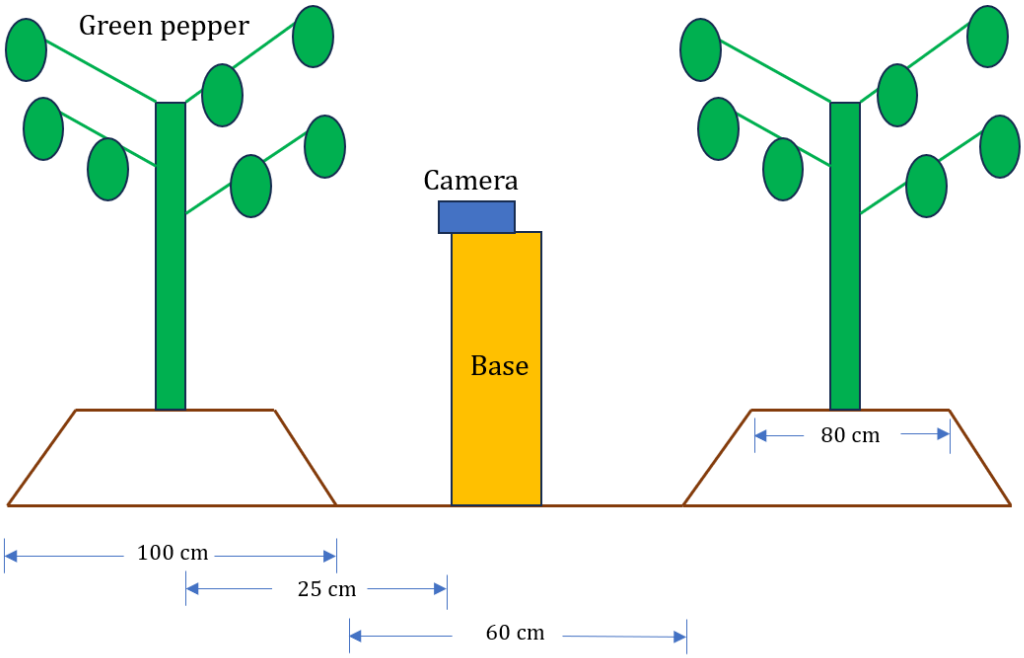

Figure 4-8 Data collection information setup in greenhouse

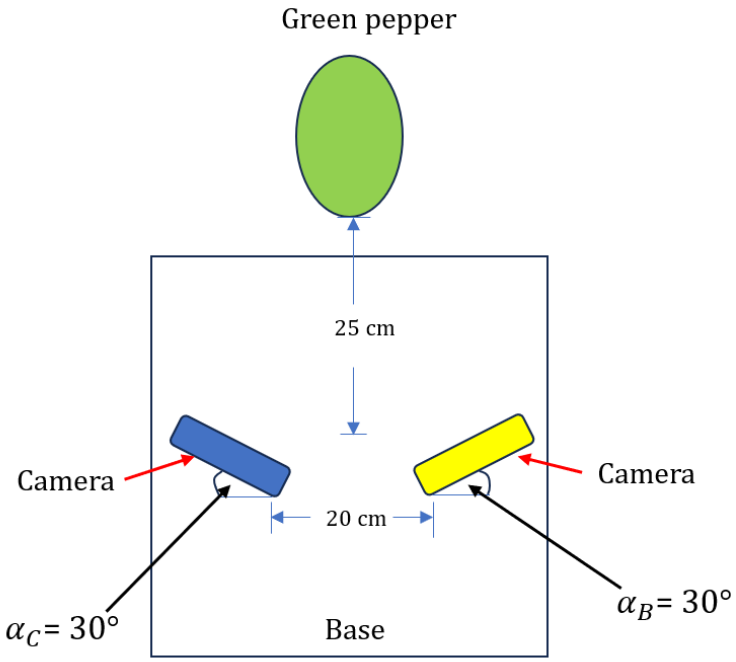

Figure 4-9 Top view of data collection

4.6 Validation and Analysis

The F1 score, a pivotal metric in image processing, is prominently employed in binary classification scenarios such as object detection, image segmentation, and classification tasks. It amalgamates the principles of precision and recall, offering a comprehensive assessment of a model's efficacy in delineating and identifying objects within images. In image processing, particularly in tasks involving segmentation or object detection, precision denotes the accuracy with which the model correctly identifies relevant regions. At the same time, recall measures the model's ability to encompass all pertinent instances. Precision is the ratio of accurate optimistic predictions to the sum of true and false positives. At the same time, recall is expressed as the ratio of true positives to the sum of true positives and false negatives. The F1 score, the harmonic mean of precision and recall, is a harmonized metric that encapsulates the impact of false positives and false negatives. Its formula delineates a balanced evaluation considering both precision and recall. Symbolically, the F1 score equations are represented as equation 4-16 to 4-18

Precision (P): Precision quantifies the ratio of true positive predictions to the aggregate number of positive predictions generated by the model. In the context of image processing, precision is indicative of the model's precision in correctly identifying pertinent regions (objects) within the image.

$$P = \frac{True\ Positives}{True\ Positives + False\ Positives} \quad (4\text{-}16)$$

Recall (R): Recall, alternatively known as sensitivity or the true positive rate, characterizes the ratio of true positive predictions to the comprehensive count of actual positive instances within the dataset. In image processing, recall gauges the model's proficiency in capturing all relevant instances.

$$R = \frac{True\ Positives}{True\ Positives + False\ Negatives} \quad (4\text{-}17)$$

F1 score: The F1 score represents the harmonic mean of precision and recall, offering a harmonized metric that encompasses the influence of false positives and false negatives. The mathematical formulation for the F1 score is expressed as:

$$F1 = 2\frac{Ture\ Positives}{True\ Positives + False\ Negatives} \quad (4\text{-}18)$$

The F1 score, bounded between 0 and 1, provides a consolidated evaluation of a model's performance, with higher values denoting superior efficacy. This metric proves particularly advantageous in scenarios characterized by an uneven distribution of positive and negative instances. In the domain of image processing, an elevated F1 score signifies the model's adeptness in accurately delineating and excluding regions of interest within images.

4.7 Results

- In the initial phase, employing the Labelme annotation method and testing with the Mask R-CNN algorithm on red-green-blue (RGB) images yielded an annotation accuracy of 0.976. In contrast, a corresponding accuracy of 0.989 was achieved when testing on infrared (IR) images. This quantitative assessment reflects the algorithm's proficiency in accurately identifying and annotating regions of interest within RGB and IR images, with higher values indicating greater precision in the annotation process. The numeric outcomes provide quantitative insights into the algorithm's performance during the image annotation stage, contributing to the overall evaluation of its efficacy under different imaging conditions.
- During the second step, employing infrared (IR) images as testing data for red-green-blue (RGB) models, and vice versa, yielded outcomes with a discernible absence of accuracy, registering a score of 0. This denotes a need for more precision in the models' ability to correctly classify and identify objects when confronted with testing data from an alternate imaging modality. The null accuracy values underscore the challenges and limitations encountered when attempting cross-modal testing, signifying the necessity for further refinement and adaptation of the models to enhance their capacity for generalized object recognition across different spectral domains.
- In the third step, the introduction of the Structural Similarity Index (SSIM) into the image comparison methodology resulted in SSIM scores ranging approximately between 0.20 and 0.25 when comparing infrared (IR) and red-green-blue (RGB) images. This quantitative assessment reflects the degree of structural similarity between the two modalities, with the SSIM scores measuring the likeness in structural patterns. The observed scores in this range suggest a moderate level of similarity, indicating that the structural characteristics of the IR and RGB images exhibit discernible differences while still possessing certain standard features, as quantified by the SSIM.

- In the fourth step, following a methodology analogous to the third step, the emphasis was explicitly directed toward a designated object within a bounding box. The application of the Structural Similarity Index (SSIM) to compare infrared (IR) and red-green-blue (RGB) images, limited to this defined region, yielded SSIM scores ranging approximately between 0.4 and 0.45. This targeted comparison within the bounding box indicates a moderate increase in the SSIM scores compared to the comprehensive image assessment in the third step, suggesting a higher level of structural similarity when focusing solely on the specified object within the bounding box.

4.8 Summary

Upon conducting testing, several observations emerged regarding the efficacy of the methodology. In the second method, the interchangeability of infrared (IR) and red-green-blue (RGB) images for testing proved unfeasible due to numerous constraints and image-specific conditions. However, in the third step, where the Structural Similarity Index (SSIM) was incorporated for image comparison, some degree of success was achieved despite the relatively low scores. Notably, focusing exclusively on the region of interest within a bounding box in the fourth step yielded favorable outcomes. The SSIM score exhibited a significant improvement, escalating from 0.25 to 0.45. This progression underscores the pivotal role of noise reduction in enhancing SSIM scores, emphasizing the importance of concentrating solely on the target object within the bounding box for optimal results, particularly when assessing structural similarities in images.

5. Second Detection Method

5.1 Overview

From previous first detection methods, building upon the insights gleaned from the preceding methodological steps, the paramount importance of noise reduction in amplifying Structural Similarity Index (SSIM) scores becomes evident. Focusing attention on the target object within the bounding box is a pivotal strategy for achieving optimal results, particularly in evaluating structural similarities within images. Considering these observations, a deliberate decision is

made to employ the edge detection method, a technique renowned for its efficacy in mitigating noise and isolating pertinent features within images. The rationale behind edge detection is its intrinsic capability to eliminate extraneous information, thereby streamlining the visual data to its essential components. This method aligns with the research findings of S.M. Mangaonkar, as shown in *figure 5-1* [50], where edge detection was successfully employed to identify fruits amidst a myriad of superfluous elements, such as hands holding fruit and various utensils within the image. Despite these complexities, edge detection offers a promising avenue to accurately determine the spatial coordinates of the fruit, contributing to a more precise and refined analysis. Integrating edge detection into the methodology is grounded in recognizing that this approach facilitates noise reduction, rendering the images more amenable to accurate structural similarity assessments. By leveraging edge detection capabilities, the study endeavors to enhance the discernment of pertinent features while mitigating the influence of irrelevant and distracting elements. In essence, this methodological refinement represents a strategic response to the challenges posed by noise in the images. The adoption of edge detection aligns with the overarching objective of optimizing the accuracy and reliability of the analysis, particularly when confronted with diverse and intricate visual environments. As the study advances, the integration of edge detection is anticipated to yield a more robust and nuanced understanding of the structural similarities within images, affording a comprehensive and accurate depiction of the objects of interest.

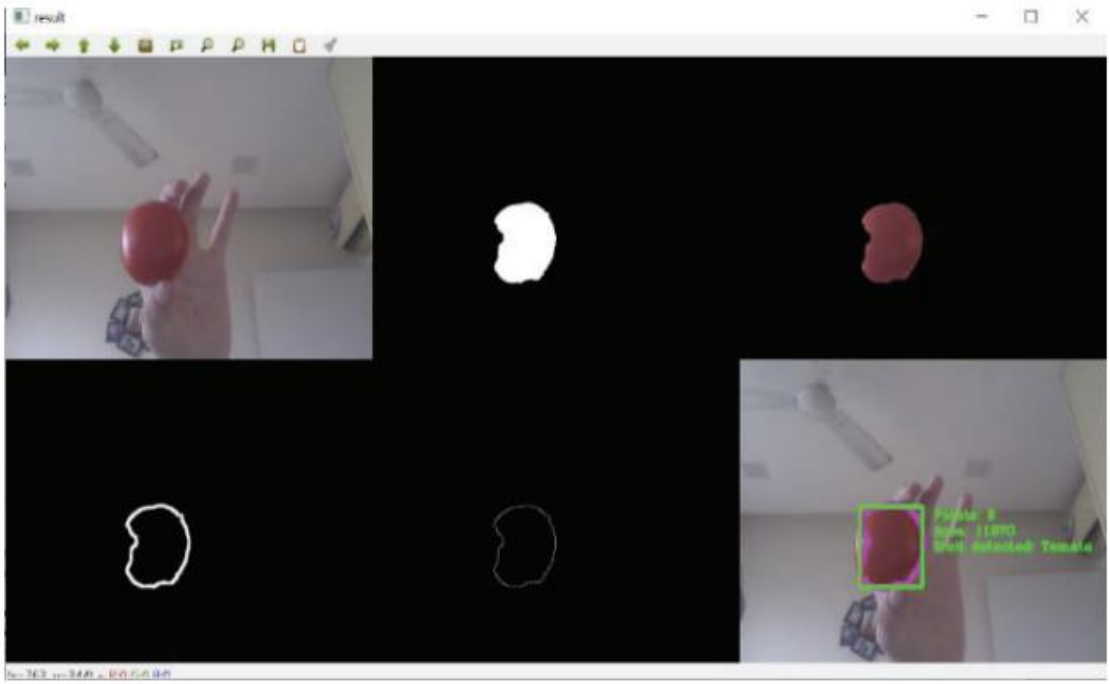

Figure 5-1 Image processing with Edge detection methods [50]

5.2 Edge Detection

Edge detection is a fundamental technique in computer vision and image processing that aims to identify boundaries and transitions within images, highlighting regions were intensity or color changes sharply. These boundaries represent the contours or edges between distinct objects or structures in the visual content. The primary objective of edge detection is to enhance the visibility of these essential features, enabling subsequent analysis, segmentation, and recognition tasks in computer vision applications. Various algorithms are employed for edge detection, each with their approach and characteristics. Popular methods include the Sobel and Prewitt operators, which emphasize gradient changes in horizontal and vertical directions, and the Canny edge detector, known for its multi-stage process that minimizes false positives. Other techniques, such as the Laplacian of Gaussian (LoG) and Kirsch operator, leverage convolution and mathematical operations to identify edges based on image intensity variations. Edge detection is a critical step in image processing pipelines, serving as a foundation for tasks like object recognition, image segmentation, and feature extraction. Its application is widespread in fields such as medical imaging, autonomous vehicles, surveillance, and industrial quality control, where accurate delineation of objects and structures within images is essential for robust and precise computer vision analyses.

5.2.1 Edge Detection Algorithm

Edge detection algorithms are essential components in computer vision, focusing on identifying abrupt variations in image intensity. The Canny edge detector, a prominent method, employs gradient calculation through convolution filters, emphasizing both horizontal and vertical changes. Subsequent non-maximum suppression isolates local maxima, and hysteresis-based edge tracking discerns solid and weak edges, producing a binary edge map delineating structural boundaries. Other methodologies, including Sobel and Prewitt operators, utilize gradient information to accentuate directional changes. These algorithms play a crucial role in image processing tasks, such as object recognition and segmentation, where accurate demarcation of object boundaries is imperative. By enhancing the visibility of significant transitions in visual data, edge detection algorithms contribute significantly to feature extraction and pattern recognition within diverse computer vision applications. The following enumeration delineates noteworthy edge detection algorithms

- *Roberts* – Roberts Cross edge detection is a simple and computationally efficient method for detecting edges in images. It involves convolving the image with a pair of 2x2 convolution kernels [51]. The Roberts Cross operator equations can be written as equations 5.1 to 5.3

$$G_x = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} * I \quad (5\text{-}1)$$

$$G_y = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} * I \quad (5\text{-}2)$$

Where: I is the input image , $*$ denotes the convolutional operation

The resulting gradient images, Gx and Gy, capture intensity changes in the horizontal and vertical directions, respectively. The gradient magnitude G at each pixel is calculated using the formula:

$$G = \sqrt{Gx^2} + Gy^2 \quad (5\text{-}3)$$

The Roberts Cross operator is straightforward and particularly useful for quick edge detection tasks. However, it can be sensitive to noise due to its simplicity, and more advanced methods like the Sobel or Canny edge detectors are often preferred for applications where noise robustness is crucial.

- *Sobel* – Sobel edge detection is a widely used method in computer vision for highlighting edges in an image by emphasizing changes in intensity in both the horizontal and vertical directions [51]. The Sobel operator involves convolving the image with 3x3 kernels, one for detecting changes in intensity in the horizontal direction Gx and the other for the vertical direction G_y. The resulting gradient images, Gx and G_y, are combined to obtain each pixel's gradient magnitude G and direction θ, The Sobel operator can be written as equations 5.4 to 5-7

$$Gx = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} * I \quad (5\text{-}4)$$

$$Gy = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} * I \quad (5\text{-}5)$$

Where: I is the input image , $*$ denotes the convolutional operation

The gradient magnitude G at each pixel is calculated using the formula:

$$G = \sqrt{Gx^2} + Gy^2 \quad (5\text{-}6)$$

and the gradient direction θ is given by:

$$\theta = \arctan\left(\frac{G_y}{G_x}\right) \quad (5\text{-}7)$$

The Sobel edge detection algorithm effectively highlights edges by accentuating intensity changes in the image along both the horizontal and vertical axes.

- *Laplacian* – The Laplace operator ∇^2 is a mathematical operator commonly utilized for edge detection in image processing. It is applied through convolution with a Laplacian kernel, represented by specific convolution matrices [52]. The 3x3 Laplacian kernel can be written as equations 5.5 and 5.6

$$\nabla^2 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix} * I \quad (5\text{-}5)$$

Here, the central element (-4) represents the weight assigned to the pixel being processed, while the neighboring elements (1) indicate the weights of surrounding pixels. For the 5x5 Laplacian kernel, the formulation is:

$$\nabla^2 = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 2 & 1 & 0 \\ 1 & 2 & -16 & 2 & 1 \\ 0 & 1 & 2 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} * I \quad (5\text{-}6)$$

In this case, the central element (-16) represents the weight assigned to the pixel being processed, while the surrounding elements (1 or 2) denote the weights of neighboring pixels. The convolution operation * is applied to the image I, producing the Laplacian response. This response emphasizes regions where intensity changes abruptly, facilitating effective edge detection in image analysis and processing tasks.

- *Canny Edge detection* – Canny edge detection is a sophisticated image processing technique designed to identify and highlight edges in an image, minimizing the influence of noise. Proposed by J. Canny in 1986 [53], this method involves multiple stages to achieve robust edge detection. The key steps include:

  Gradient Calculation: Compute the image gradient using convolution with Sobel filters to emphasize changes in intensity in the horizontal and vertical directions.

  Gradient Magnitude and Orientation: Determine the gradient magnitude and orientation at each pixel.

  Non-Maximum Suppression: Suppress non-maximum gradient values to retain only local maxima along the edges.

  Edge Tracking by Hysteresis: Establish high and low thresholds for gradient magnitudes, identify pixels with gradient magnitudes above the high threshold as strong edge points,

connect weak edge points to strong edge points if they are part of the same edge structure.

Mathematically, the gradient magnitude G is calculated as $G = \sqrt{Gx^2} + Gy^2$, where Gx and Gy are the horizontal and vertical gradients. The gradient orientation θ is determined as $\theta = \arctan\left(\frac{G_y}{G_x}\right)$. The Canny edge detection equation incorporates these principles, providing an effective approach for accurate edge localization and noise reduction in various computer vision applications.

- *Prewitt* – Prewitt edge detection is a method commonly employed in image processing for highlighting edges by emphasizing changes in intensity along both horizontal and vertical directions. Proposed by Judith M. S. Prewitt [54], this technique employs convolution with Prewitt kernels to calculate image gradients, offering a simplified alternative to more complex operators. The Prewitt operator equations for horizontal Gx and Gy gradients are represented by equations 5-7 to 5-8

$$Gx = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix} * I \quad (5\text{-}7)$$

$$Gy = \begin{bmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix} * I \quad (5\text{-}8)$$

here, I represents the input image, and $*$ denotes the convolution operation. The gradient magnitude G at each pixel is computed as $G = \sqrt{Gx^2} + Gy^2$, providing a measure of intensity changes in the image. Prewitt edge detection proves valuable for its simplicity and efficiency in capturing edge information along multiple directions, contributing to applications such as image segmentation and feature extraction in computer vision tasks.

5.3 Experimental Method

Given the multitude of edge detection operators, the selection process for our study involved careful consideration of several methods, including Roberts, Sobel, Laplacian3x3, Laplacian5x5, and Canny edge detection. Prewitt was omitted due to its structural similarity to Sobel, with the latter demonstrating distinct advantages. The comparative analysis revealed that the Sobel edge enhancement filter holds a notable advantage, concurrently providing differentiation for edge response and smoothing for noise reduction. This strategic choice was made to ensure optimal

performance and relevance to the specific objectives of our research, aligning with the requirements and nuances of the targeted image-processing tasks. The experimental procedure encompasses two sequential steps and the identical methodology, as outlined in Chapter 4, Section 4.6, will be employed for validation and analysis.

5.3.1 Initial Step Method

In the initial phase, the procedure involves the demarcation of object boundaries and the application of a masking technique to isolate the object box. This methodology remains consistent with the framework expounded in Chapter 4, supplemented by an additional procedural step: the specification of the region of interest within the bounding box, accompanied by the normalization of pixel values between the images. Following this, all five distinct edge detection methods are applied, and the resulting images undergo a comparative evaluation utilizing the Structural Similarity Index (SSIM). The intricate details of this inaugural step are visually elucidated in *figure 5-2*, offering a graphical representation of the implemented methodology for the precise definition of object boundaries and subsequent analytical processes involving diverse edge detection techniques.

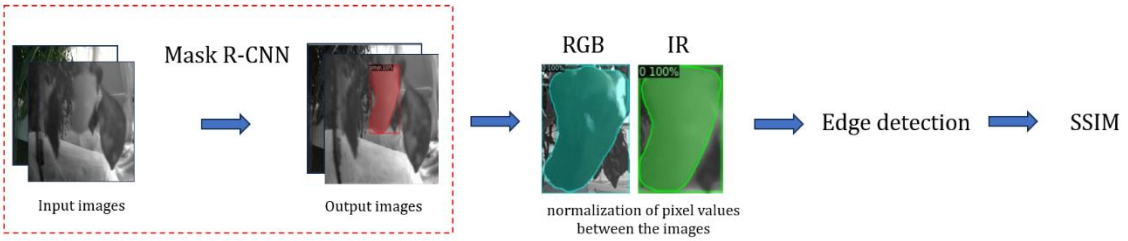

Figure 5-2 Initial step method overview

Upon completion of the initial testing phase, the obtained results are as follows: Canny Edge Detection, Roberts, Laplacian 3x3, Sobel, and Laplacian 5x5. The corresponding Structural Similarity Index (SSIM) scores, arranged in descending order, are 0.577, 0.552, 0.551, 0.44, and 0.269. These results are visually presented in figure 5-3, showcasing the highest scores. Notably, the application of Edge Detection in this method yielded an SSIM score of 0.577, surpassing the score of 0.45 obtained in the previous method detailed in Chapter 4. This discrepancy underscores the efficacy of utilizing Edge Detection to manage noise in the image effectively. Based on the experimental outcomes in the initial step, there is a conviction that mitigating unnecessary noise is imperative to obtain precise object delineation during Edge Detection. The emphasis on noise reduction ensures that the subsequent application of Edge Detection yields

images primarily featuring the targeted objects, and the SSIM comparisons further affirm the extraction of object shapes with enhanced accuracy.

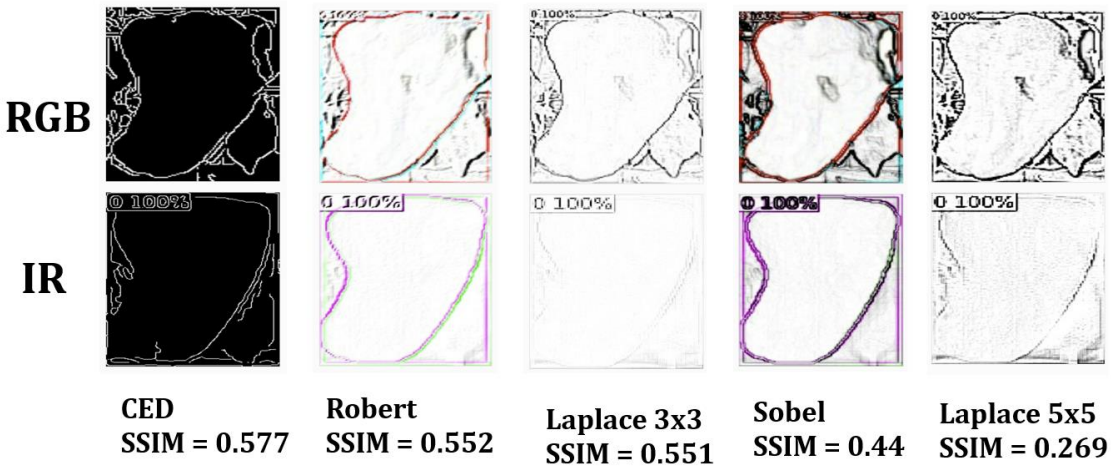

Figure 5-3 SSIM score comparison with Edge detection methods

5.3.2 Secondary Step Method

Building upon the insights gained in the initial step, where the emphasis was placed on noise reduction for precise object delineation, the subsequent phase follows a strategic guideline of further noise elimination to optimize results. Based on the outcomes of the initial experiment, three methods emerged with the highest Structural Similarity Index (SSIM) scores: Canny Edge Detection, Roberts, and Laplacian 3x3, scoring 0.577, 0.552, and 0.551, respectively. In the second step, these top-performing methods from the first step will be reapplied in the Edge Detection process after initial noise reduction. This phase involves a nuanced comparison with the Mask R-CNN step. Unlike the conventional approach in Mask R-CNN, which consists in masking objects and applying label painting, the modified procedure in this step employs the Mask and label without painting over the object. The image size is adjusted uniformly, and a crop operation is executed to retain only the desired objects, as shown in figure 5-4. Subsequently, the Edge Detection process is initiated, followed by the SSIM evaluation, as shown in figure 5-5. This systematic approach aims to capitalize on the superior performance of selected Edge Detection techniques while incorporating insights from the initial noise reduction steps, thereby refining the overall image processing methodology.

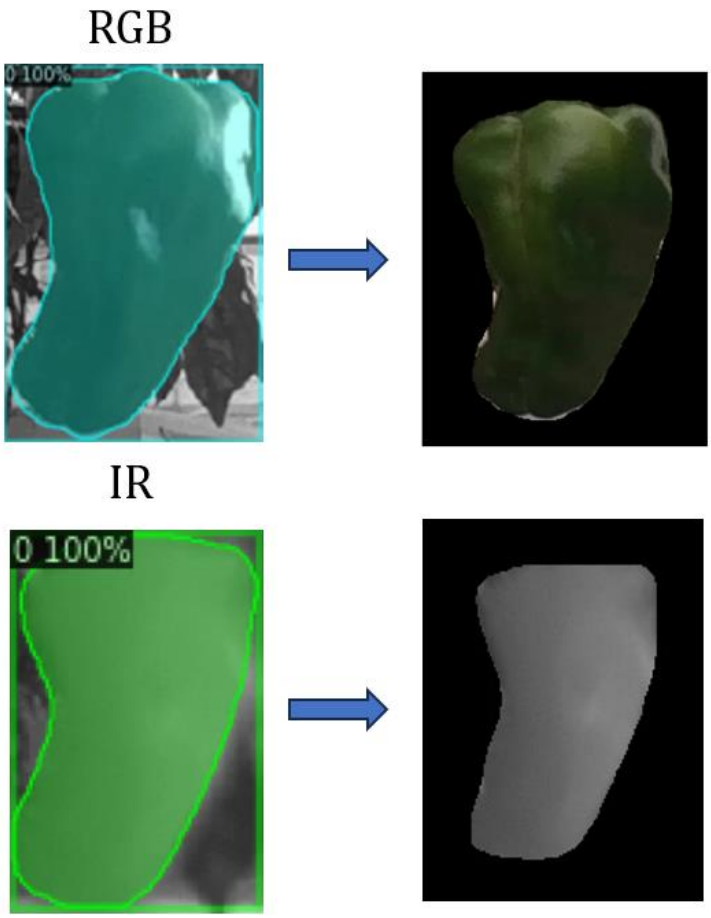

Figure 5-4 Crop Object in bounding box

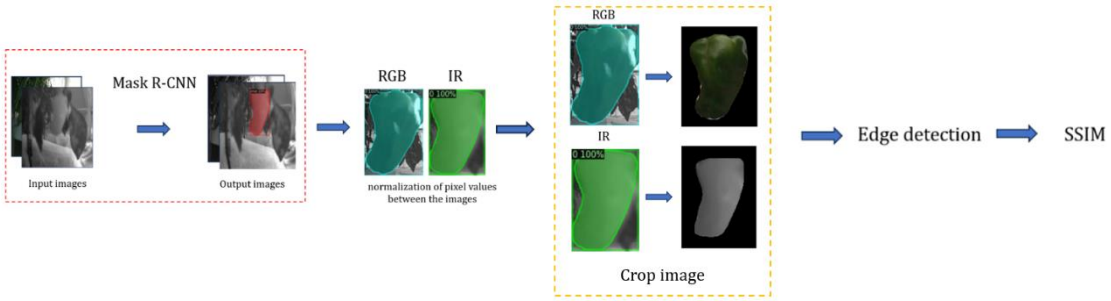

Figure 5-5 Secondary Step method overview

The outcomes derived from the second experiment assess the performance metrics associated with the recently introduced Mask R-CNN. Within the RGB and IR cropping segments, the acquired F1 scores stand at 0.7894 and 0.9815, respectively. Furthermore, applying the new method for edge detection results in images, as exemplified in *figure 5-6*.

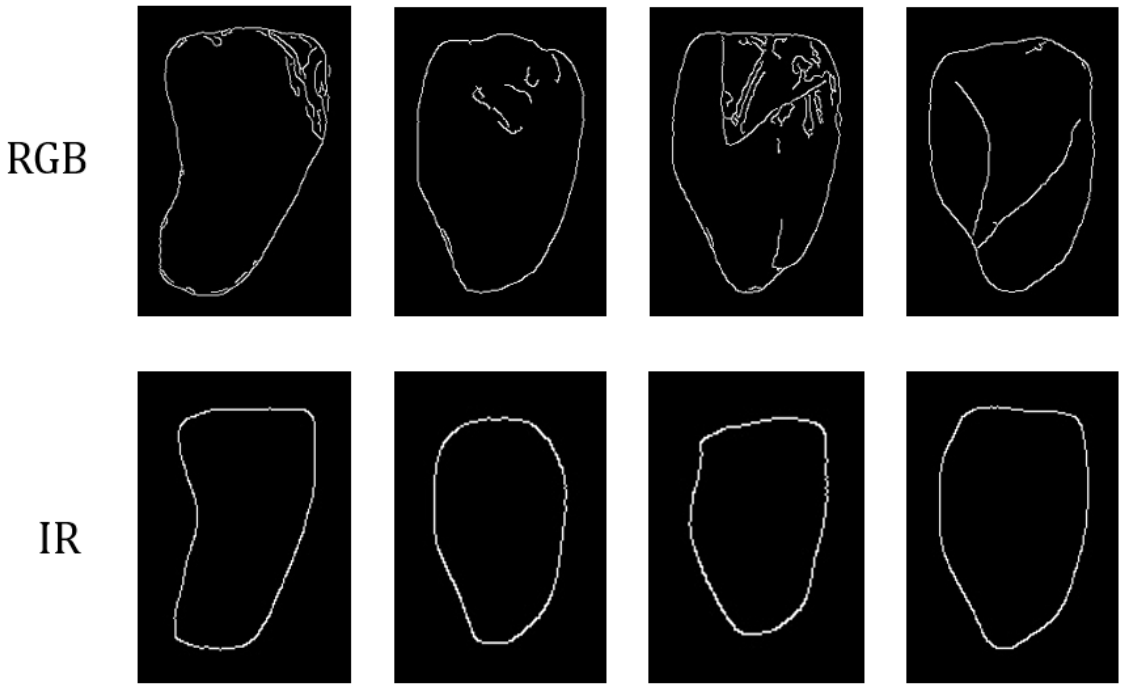

Figure 5-6 RGB and IR with Canny Edge Detection , From left to right are: capture from the left side, capture from the right side without a foliage, capture from the right with a foliage of 10-30%, capture from the right with a foliage more than 30%.

In contrast, the SSIM segment yields analogous scores of 0.7894 and 0.9815. Notably, the recorded scores predominantly concentrate within the range of 0.790 to 0.810., thereby encapsulating a substantial portion of the experimental outcomes, as illustrated in *figure 5-7*

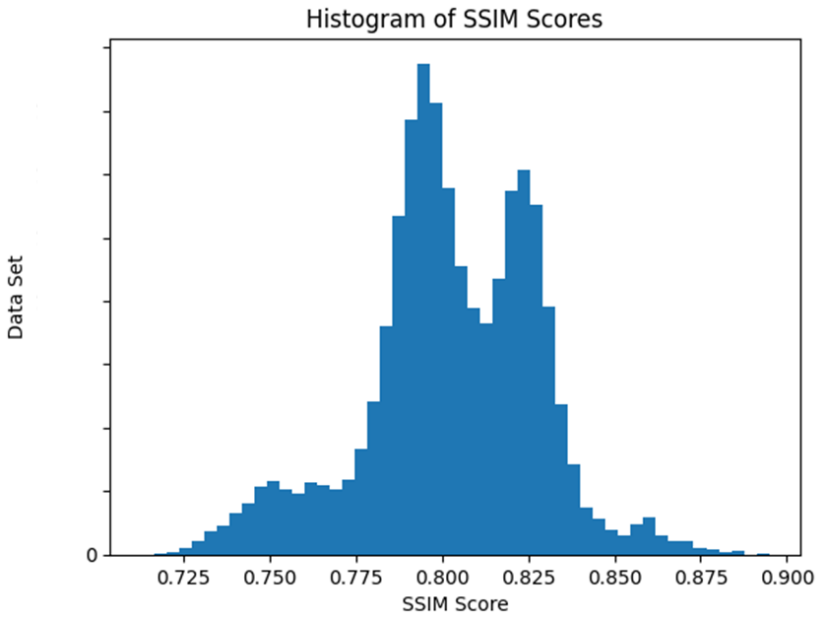

Figure 5-7 Second step method SSIM score

5.4 Summary

Both experiments revealed that the optimal approach involves mitigating image noise before SSIM-based comparisons. Highly detailed RGB images exhibit a modest correctness probability of merely 0.7984. In contrast, IR images, characterized by reduced image intricacies, achieve a notably higher score of 0.9815, surpassing the RGB score of 0.1831. Despite the superior IR score, a notable issue arises during the edge detection phase: the images acquired from the IR camera struggle to detect leaves due to insufficient image details. This contrasts RGB images that boast discernible leaf lines, contributing to object coverage. Such intricacies may pose future challenges, particularly if objects become conglomerated and need clear demarcation.

6. Conclusion

This research is focused on advancing methodologies for effectively detecting green peppers, particularly within the controlled environment of a greenhouse. A comprehensive analysis of green peppers grown in the greenhouse environment has revealed various challenges that necessitate nuanced solutions. One of the primary challenges identified is the varied positioning of green pepper fruits throughout the plant. Given that green pepper plants can produce fruit across the entirety of the plant, the resulting fruits exhibit a range of heights, some elevated and others at lower levels. This inherent variability in fruit positioning introduces complexities in subsequent detection processes. Moreover, the abundance of leaves on green pepper plants and their dense arrangement further complicates distinguishing individual green peppers from the foliage. The inherent tendency of green pepper fruits to grow nearby, forming clusters, poses an additional layer of difficulty in achieving accurate and precise detection. The challenge is exacerbated by the fact that individual fruits must be isolated during the data collection, preventing them from being clustered with other fruits. The research underscores the importance of meticulous data collection to address these challenges. The optimal approach involves maintaining 25cm from the object of interest and tilting the camera at a 30-degree angle. These parameters ensure the collected data is well-positioned, avoiding clustering issues and enabling accurate separation of individual green peppers from the surrounding foliage. However, even with careful data collection, challenges persist in accurately discerning the color of green peppers,

leaves, and fruits, mainly when relying solely on a conventional RGB camera. The similarity in color poses difficulties in differentiation. As a potential solution, the research explores using an infrared (IR) camera, which exhibits promise in classification but encounters challenges related to accuracy, particularly in cases where the temperature of the fruits and leaves is similar.

The study employs the Mask R-CNN process to analyze and detect green peppers, achieving commendable accuracy with scores of 0.976 and 0.989 for different process aspects. However, the subsequent structural similarity index (SSIM) process presents distinct challenges. Despite the RGB image scoring marginally lower 0.1831 than the IR image, it encounters more intricate challenges. The high resolution of the RGB image facilitates the differentiation of fine details in the leaves, which may obscure the green peppers. Conversely, the images obtained from the IR camera struggle to distinguish leaves that may obscure the green pepper fruits. Practical challenges also extend to the physical setup within the greenhouse. Walkways composed of dirt necessitate frequent adjustments to the camera position to ensure a level and consistent perspective. The computational aspect of the process introduces another layer of complexity, with the calculation process involving numerous steps and substantial processing time. For instance, calculating a single result requires up to 16 hours, underscoring the need for more efficient computational methodologies for practical applications. In conclusion, the research highlights the multifaceted challenges of detecting green peppers in a greenhouse environment. Each aspect requires careful consideration and innovative solutions, from nuanced data collection to color differentiation and computational efficiency. Addressing these challenges is pivotal for practically implementing the methodology in real-world scenarios, were efficiency and accuracy.

References

[1] Erenstein, O., Chamberlin, J., & Sonder, K. (2021). “Farms Worldwide: 2020 and 2030 Outlook”, in Proceeding of the Outlook on Agriculture, Vol. 50(3), pp. 221-229.

[2] Nan, Y., Zhang, H., Zeng, Y., Zheng, J., & Ge, Y. (2022). Faster and Accurate Green Pepper Detection Using NSGA-II-based Pruned YOLOv5l in the Field Environment. Computers and Electronics in Agriculture, Dec 2022.

[3] Eizentals, P. (2016). Picking System for Automatic Harvesting of Sweet Pepper.

[4] Tada, N. (2022, December). Recognition of Sweet Pepper Fruit in Greenhouse Using Far-Infrared Camera.

[5] Ji, W., Chen, G., Xu, B., Meng, X., & Zhao, D. (2019). Recognition Method of Green Pepper in Greenhouse Based on Least-Squares Support Vector Machine Optimized by the Improved Particle Swarm Optimization. IEEE Access, Vol. (7),pp. 119742-119753,Aug 2019.

[6] Zemmour, E., Kurtser, P., & Edan, Y. (2019). Automatic Parameter Tuning for Adaptive Thresholding in Fruit Detection. Sensor, Access MDPI, Vol.19(2130), May 2019.

[7] Bachche, S., & Oka, K. (2013). Distinction of Green Sweet Peppers by Using Various Color Space Models and Computation of 3-Dimensional Location Coordinates of Recognized Green Sweet Peppers Based on Parallel Stereovision System. Journal of System Design and Dynamics, Vol.7, No2, pp. 178-196.

[8] MAFF (Ministry of Agriculture, Forestry, and Fisheries). (2023). FY2022 Summary of the Annual Report on Food, Agriculture, and Rural Areas in Japan. May 2023.

[9] Satake, A. (2020). Number of Women Farmers in Japan Continues to Decline. USDA Foreign Agricultural Service, May 2020.

[10] Coppock, G. E. (1961). Picking Citrus Fruit by Mechanical Means. Associate Agricultural Engineer Florida Citrus Commission Citrus Experiment Station, pp. 247-251.

[11] Schertz, C. E., & Brown, G. K. (1968). Basic Considerations in Mechanizing Citrus Harvest. Transactions of the ASAE, 11(3), pp. 343-346.

[12] Kang, H., Wang, X., & Chen, C. (2022). Accurate Fruit Localization for Robotic Harvesting using High-Resolution LiDAR-Camera Fusion. Access ResearchGate, Dec 2022.

[13] Siripatrawan, U., & Makino, Y. (2024). Hyperspectral Imaging Coupled with Machine Learning for Classification of Anthracnose Infection on Mango Fruit. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, Vol.309, Dec 2023.

[14] Chen, M., Tang, Y., Zou, X., Huang, K., Huang, Z., Zhou, H., Wang, C., & Li, G. (2022). Three-dimensional Perception of Orchard Banana Central Stock Enhanced by Adaptive Multi-vision Technology. Computers and Electronics in Agriculture, Vol.174, May 2020.

[15] Liu, X., Yu, J., Kurihara, T., Xu, L., Niu, Z., Zhan, S. (2022). Hyperspectral Imaging for Green Pepper Segmentation Using a Complex-valued Neural Network. Optik - International Journal for Light and Electron Optics, Vol.265, June 2022.

[16] Xuan, G., Gao, C., Shao, Y. (2022). Spectral and Image Analysis of Hyperspectral Data for Internal and External Quality Assessment of Peach Fruit. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, Vol.272, Feb 2022.

[17] Chen, H., Qiao, H., Feng, Q., Xu, L., Lin, Q., & Cai, K. (2021). Rapid Detection of Pomelo Fruit Quality Using Near-Infrared Hyperspectral Imaging Combined with Chemometric Methods. Frontiers in Bioengineering and Biotechnology, Vol.8 (616943), Jan 2021.

[18] Gené-Mola, J., Sanz-Cortiella, R., Rosell-Poloa, J. R., Morrosb, J.-R., Ruiz-Hidalgob, J., Vilaplanab, V., & Gregorio, E. (2020). Fruit Detection and 3D Location Using Instance Segmentation Neural Networks and Structure-from-Motion Photogrammetry. Computers and Electronics in Agriculture, Vol.169,

[19] Fukuda, M., Okuno, T., & Yuki, S. (2021). Central Object Segmentation by Deep Learning to Continuously Monitor Fruit Growth through RGB Images. Sensor, Access MDPI, Vol.21(6999), Oct 2021.

[20] Malik, M. H., Zhang, T., Li, H., Zhang, M., Shabbir, S., Saeed, A. (2018). Mature Tomato Fruit Detection Algorithm Based on Improved HSV and Watershed Algorithm. IFAC-Papers Online, Vol. 51(17), pp. 431-436.

[21] Wachs, J. P., Stern, H. I., Burks, T., & Alchanatis, V. (2010). Low and High-Level Visual Feature-Based Apple Detection from Multi-modal Images. Precision Agriculture, Access ResearchGate, Dec 2010.

[22] Kang, H., & Chen, C. (2020). Fruit Detection, Segmentation, and 3D Visualization of Environments in Apple Orchards. Computers and Electronics in Agriculture, Vol.171, Feb 2020.

[23] Xiao, F., Wang, H., Xu, Y., & Zhang, R. (2023). Fruit Detection and Recognition Based on Deep Learning for Automatic Harvesting: An Overview and Review. Agronomy, Access MDPI, Vol.13, pp. 1625.

[24] Best, S., Ringdahl, O., Oberti, R., & Evain, S. (2015). CROPS: Clever Robots for Crops. Engineering & Technology Reference, Access ResearchGate, Vol.10, pp. 1049.

[25] Kang, H., Wang, X., & Chen, C. (2022). Accurate Fruit Localization Using High-Resolution LiDAR-Camera Fusion and Instance Segmentation. Computers and Electronics in Agriculture, Vol.10, pp. 1016.

[26] AbuRass, S., Huneiti, A., & Al-Zoubi, M. B. (2020). Enhancing Convolutional Neural Network using Hu's Moment. International Journal of Advanced Computer Science and Applications (IJACSA), Vol. 11, No. 12, pp. 130-137.

[27] Septiarini, A., Hamdani, H., Sari, S. U., Hatta, H. R., Puspitasari, N., & Hadikurniawati, W. (2021). Image Processing Techniques for Tomato Segmentation Applying K-Means Clustering and Edge Detection Approach. In Proceedings of the 2021 International Seminar on Machine Learning, Optimization, and Data Science (ISMODE). Vol10, pp. 92-96.

[28] Indira, D.N.V.S.L.S., Goddu, J., Indraja, B., Challa, V. M. L., Manasa, B. (2021). A Review on Fruit Recognition and Feature Evaluation Using CNN. Materials Today: Proceedings, Vol.80 , pp. 3483-3443.

[29] Hong, Y. (2016). Intelligent Detection Method of Fruit Based on Improved SSIM Algorithm. Advance Journal of Food Science and Technology, Vol.10 (4) , pp. 309-312.

[30] Zarate, V., Gonzalez, E., Caceres-Hernandez, D. (2023). Fruit Detection and Classification Using Computer Vision and Machine Learning Techniques. In Proceedings of the 2023 IEEE 32nd International Symposium on Industrial Electronics (ISIE), Vol. 10.

[31] Sheng, X., Kang, C., Zheng, J., Lyu, C. (2023). An Edge-Guided Method to Fruit Segmentation in Complex Environments. Computers and Electronics in Agriculture, Vol.208, Mar 2023.

[32] Intel Corporation. (2020). Intel RealSense Product Family D400 Series Datasheet.

[33] Optris infrared measurements. Optris Xi 400 TECHNICAL DATA.

[34] Real-Moreno, O., Rodríguez-Quiñonez, J. C., Flores-Fuentes, W., Sergiyenko, O., Miranda-Vega, J. E., Trujillo-Hernández, G., & Hernández-Balbuena, D. (2024). Camera Calibration Method Through Multivariate Quadratic Regression for Depth Estimation on a Stereo Vision System. Optics and Lasers in Engineering, Vol.174, Nov 2023.

[35] Lee, M., Kim, H., & Paik, J. (2019). Correction of Barrel Distortion in Fisheye Lens Images Using Image-Based Estimation of Distortion Parameters. IEEE Access, Vol.7, pp. 45723-45733, Apr 2019.

[36] Kim, T.-H. (2018). An Efficient Barrel Distortion Correction Processor for Bayer Pattern Images. IEEE Access, Vol.6, pp. 28239-28248.

[37] Darvatkar, S., & Bhandari, S. U. (2017). Implementation of Barrel Distortion Correction on FPGA. IEEE, 2017.

[38] Kan, N. H. L., Cao, Q., & Quek, C. (2024). Learning and Processing Framework using Fuzzy Deep Neural Network for Trading and Portfolio Rebalancing. Applied Soft Computing, Vol.152, Jan 2024.

[39] Zhou, W., Cui, Y., Huang, H., Huang, H., Wang, C. (2024). A Fast and Data-Efficient Deep Learning Framework for Multi-class Fruit Blossom Detection. Computers and Electronics in Agriculture, Vol.217, Jan 2024.

[40] Sun, Z., An, G., Yang, Y., Liu, Y. (2024). Optimized Machine Learning Enabled Intrusion Detection System for Internet of Medical Things. Franklin Open, Vol.6, Nov 2023.

[41] Ganesh, P., Volle, L., Burks, T. F., Mehta, S. S. (2019). Deep Orange: Mask R-CNN based Orange Detection and Segmentation. IFPA Conference Paper Archive, Vol.52-30, pp.70-75.

[42] Passos, D., & Mishra, P. (2023). Deep Tutti Frutti: Exploring CNN Architectures for Dry Matter Prediction in Fruit from Multi-fruit Near-Infrared Spectra. Chemometrics and Intelligent Laboratory Systems, Vol.243, Nov 2023.

[43] Yu, Y., Zhang, K., Yang, L., Zhang, D. (2019). Fruit Detection for Strawberry Harvesting Robot in Non-Structural Environment Based on Mask-RCNN. Computers and Electronics in Agriculture, Vol. 163, June 2019.

[44] Cong, P., Li, S., Zhou, J., Lv, K., & Feng, H. (2023). Research on Instance Segmentation Algorithm of Greenhouse Sweet Pepper Detection Based on Improved Mask RCNN. Agronomy, MDPI, Vol.13, Jan 2023.

[45] Wang, D., & He, D. (2022). Fusion of Mask RCNN and Attention Mechanism for Instance Segmentation of Apples under Complex Background. Computers and Electronics in Agriculture, Vol.196, Mar 2022.

[46] Vresdian, D. J., Al-Yousif, S., Pratama, L. P., Hapsari, A. A., Islami, A. Y., Dionova, B. W. (2022). SSIM as Validation Technique on Normalization Segmented Iris. FORTEI-International Conference on Electrical Engineering (FORTEI-ICEE), pp. 87-90.

[47] Peng, J., Shi, C., Laugeman, E., Hu, W., Zhang, Z., Mutic, S., & Cai, B. (2020). Implementation of the Structural SIMilarity (SSIM) Index as a Quantitative Evaluation Tool for Dose Distribution Error Detection. American Association of Physicists in Medicine, Vol.47(4), pp. 1907-1919.

[48] Hu, F., Hu, Y., Cui, E., Guan, Y., Gao, B., Wang, X., Wang, K., Liu, Y., Yao, X. (2023). Recognition Method of Coal and Gangue Combined with Structural Similarity Index Measure and Principal Component Analysis Network under Multispectral Imaging. Microchemical Journal, Vol.186, Dec 2022.

[49] Hong, Y. (2016). Intelligent Detection Method of Fruit Based on Improved SSIM Algorithm. Advance Journal of Food Science and Technology, Vol.10(4), pp. 309-312.

[50] Mangaonkar, S. M., Khandelwal, R., Shaikh, S., Chandaliya, S., Ganguli, S. (2022). Fruit Harvesting Robot Using Computer Vision. Proceedings of the 2022 International Conference for Advancement in Technology (ICONAT), Jan 2022

[51] Burnham, J., Hardy, J., Meadors, K. (1997). Image Processing Group: Comparison of Edge Detection Algorithms - Comparison of the Roberts, Sobel, Robinson, Canny, and Hough Image Detection Algorithms. MS State DSP Conference.

[52] Wang, X. (2007). Laplacian Operator-Based Edge Detectors. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.29(5), May 2007.

[53] Canny, J. (1986). A Computational Approach to Edge Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-8(6), pp. 679-698 Nov 1986.

[54] Prewitt, J. M. S. (1970). Object Enhancement and Extraction. Picture Processing and Psychopictorics, pp.75-149.