

Received 2 December 2023, accepted 27 December 2023, date of publication 3 January 2024,  
date of current version 11 January 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3349418

## TOPICAL REVIEW

# A Comprehensive Review on Deep Learning Assisted Computer Vision Techniques for Smart Greenhouse Agriculture

JALAL UDDIN MD AKBAR<sup>1</sup>, (Member, IEEE),  
SYAFIQ FAUZI KAMARULZAMAN<sup>1</sup>, (Senior Member, IEEE), ABU JAFAR MD MUZAHID<sup>2</sup>,  
MD. ARAFATUR RAHMAN<sup>3</sup>, (Senior Member, IEEE),  
AND MUEEN UDDIN<sup>4</sup>, (Senior Member, IEEE)

<sup>1</sup>Faculty of Computing, Universiti Malaysia Pahang Al-Sultan Abdullah, Pekan, Pahang 26600, Malaysia

<sup>2</sup>Department of Mechanical, Aerospace, and Biomedical Engineering, The University of Tennessee, Knoxville, TN 37996, USA

<sup>3</sup>School of Mathematics and Computer Science, University of Wolverhampton, WV1 1LY Wolverhampton, U.K.

<sup>4</sup>College of Computing and Information Technology, University of Doha for Science and Technology, Doha, Qatar

Corresponding authors: Syafiq Fauzi Kamarulzaman (syafiq29@ump.edu.my) and Mueen Uddin (mueen.uddin@udst.edu.qa)

This open access research is supported by the Qatar National Library (QNL).

**ABSTRACT** With the escalating global challenges of food security and resource sustainability, innovative solutions like deep learning and computer vision are transforming agricultural practices by enabling data-driven decision-making. This paper provides a focused review of recent advancements in deep learning-enabled computer vision techniques tailored specifically for greenhouse environments. First, deep learning and computer vision fundamentals are briefly introduced. Over 100 studies from 2020 to date are then comprehensively reviewed in which these technologies were applied within greenhouses for growth monitoring, disease detection, yield estimation, and other tasks. The techniques, datasets, models, and overall performance results reported in the literature are analyzed. Tables and figures showcase real-world implementations and results synthesized from current research. Key challenges are also outlined related to aspects like model adaptability, lack of sufficient labeled greenhouse data, computational constraints, the need for multi-modal sensor fusion, and other areas needing further investigation. Future trends and prospects are discussed to provide guidance for researchers exploring computer vision in the niche greenhouse domain. By condensing prior work and elucidating the state-of-the-art, this timely review aims to promote continued progress in smart greenhouse agriculture. The focused analysis, specifically on greenhouse environments, fills a gap compared to previous agricultural surveys. Overall, this paper highlights the immense potential of computer vision and deep learning in driving the emergence of data-driven, smart greenhouse farming worldwide.

**INDEX TERMS** Agricultural automation, computer vision, deep learning, convolutional neural networks(CNN), controlled-environment agriculture (CEA), greenhouse farming, smart farming, smart agriculture, precision agriculture, image classification, image segmentation, object detection.

## I. INTRODUCTION

In the face of escalating global challenges such as population growth, climate change, and urbanization, the

The associate editor coordinating the review of this manuscript and approving it for publication was Yongjie Li.

call to reimagine our agricultural and horticultural systems becomes more critical than ever. The shadow of these issues hovers ominously over future food security, intertwined with the sustainable usage of our dwindling natural resources [1]. Yet, in this realm of concerns, a beacon of promise emerges from technological advancements, holding the potential to

revolutionize the agricultural and horticultural landscapes [2]. Our pursuit, therefore, centers on the strategic leverage of these cutting-edge technologies to create more sustainable, productive, and resilient farming ecosystems worldwide.

Sustainable access to high-quality food is not just a predicament for developing countries; even developed nations are grappling with this issue. Current agricultural practices, conducted predominantly in open fields, are not sufficiently productive to meet the escalating demand. As the World Health Organization projects, food production will need to increase by 70% by 2050 to meet the needs of an estimated global population of 10 billion [3], about 7 billion of whom will be living in urban areas [4]. Thus, exploring alternative production systems becomes not just an option but a necessity to ensure a sustainable food supply chain.

One promising alternative is Controlled Environment Agriculture (CEA), which includes a variety of methods such as greenhouses, high-tunnels, vertical farms, and plant factories [5]. In particular, greenhouse farming has emerged as a method offering greater control over the growing conditions of crops, thereby enhancing yield and quality. This technique leverages controlled environments to optimize plant growth, promising higher production rates compared to traditional farming [6]. However, despite the potential of greenhouse farming within the CEA framework, it faces certain challenges. Economic sustainability remains a significant concern due to high operational costs, complex microclimate controls, and the need for continuous labor [7], [8]. These factors can inhibit the scalability and efficiency of greenhouse farming. To overcome these obstacles, it is imperative to incorporate advanced technologies such as artificial intelligence (AI) and deep learning (DL) into these farming systems. These cutting-edge tools can potentially transform the operational aspects of greenhouse farming. For instance, they can be leveraged for improved micro-environment monitoring and root-zone control, creating the optimal conditions for plant growth while minimizing resource wastage [5]. Additionally, AI and DL can facilitate the automation of labor-intensive tasks, thereby enhancing overall operational efficiency. In this context, the adoption and integration of AI and DL in greenhouse farming becomes not only a strategic advantage but a necessity to realize the full potential of Controlled Environment Agriculture (CEA) in addressing our global food security challenges.

Simultaneously, the sphere of horticulture presents a distinct set of challenges and opportunities. Here, we face the necessity of balancing the cultural and economic value of crops with the labor-intensive nature of their cultivation. Thankfully, modern advancements offer promising solutions. Deep learning, in particular, holds the potential to streamline and revolutionize horticultural practices. The sheer volume of data that can now be collected from digital horticulture necessitates efficient processing and analysis. Deep learning algorithms can effectively handle such 'big data', enabling precise and timely decision-making in crop management [9].

In light of these advancements, it is crucial to remember that the applicability of such technologies is not limited to large-scale farming. Small indoor farms, which require significant labor year-round, can also benefit from the integration of intelligent automation. This inclusive approach to technological application in agriculture is critical to ensure food security in the long term.

Through this review, we strive to explore and underline the recent advances in deep learning-assisted computer vision technologies in the domains of agriculture and horticulture in greenhouse setup. We aim to shed light on the challenges, opportunities, and future prospects of these technologies, underlining their potential role in securing global food supply chains, improving horticultural productivity, and propelling us towards a sustainable future.

## A. REVIEW SCOPE

In our extensive review of over 100 research papers sourced from esteemed scientific databases such as ScienceDirect, Web of Science, IEEE Xplore, and Scopus, we noticed a trend: while many surveys tackled the broader agricultural domain, there was a distinct gap in the literature specifically focusing on greenhouse farming. Recognizing this limited attention to greenhouse environments, we sought to make a significant contribution by narrowing our scope and offering an in-depth analysis of computer vision applications within greenhouse farming. This work emphasizes the state-of-the-art in computer vision techniques for this specialized agricultural setting. Table 1 provides a comparison between existing survey papers on deep learning-based computer vision applications in the broader agricultural domain and our niche exploration into greenhouses. With our paper, our aim was to shed light on every potential application of computer vision in greenhouse agriculture. This review is intended not only for agricultural researchers keen on understanding nuances in greenhouse setups but also for general computer vision enthusiasts curious about its specific applications in such controlled environments. Furthermore, we have pointed out the real-world impacts and challenges of scaling up these innovative solutions within greenhouses.

## B. CONTRIBUTION

Numerous research has explored the application of deep learning and computer vision techniques in the field of agriculture. However, there is a noticeable gap in the existing literature about the comprehensive implementation of these techniques, particularly within the context of greenhouse farming. This work aims to address the existing knowledge gap by providing a comprehensive analysis of the implementation of deep learning techniques in the realm of greenhouse farming, specifically emphasizing the application of computer vision. The paper makes several notable contributions. Key contributions are summarized as follows:

- Provides an in-depth literature review focused specifically on deep learning-enabled computer vision

**TABLE 1.** Comparison of Deep learning-based Computer Vision applications in agriculture-related survey papers. GM Growth monitoring, RCC Recognition and classification of crops, DM Disease Monitoring, AH Automatic Harvesting, QI, Quality Inspection, YE Yield Estimation, PHA Plant Health Analysis, PIM Pest and Insect Management, SQA Seed Quality Analysis, WM Weed Management.

Ref.	Year	DL based CV application in agriculture									
		GM	RCC	DM	AH	QI	YE	PHA	PIM	SQA	WM
[9]	2021	✓	✓	×	×	✓	✓	✓	×	×	×
[5]	2022	✓	×	✓	✓	×	✓	✓	✓	×	×
[10]	2022	×	×	×	×	×	✓	✓	×	✓	✓
[11]	2022	×	✓	✓	×	×	×	✓	✓	✓	✓
[12]	2020	✓	×	✓	✓	✓	×	×	✓	×	×
Ours	-	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

applications in greenhouse environments, addressing a gap in existing research.

- Investigates and synthesizes techniques, capabilities, limitations, and future work needed for computer vision across major greenhouse application areas including growth monitoring, disease detection, yield estimation, etc.
- Analyzes performance of computer vision techniques based on results in current literature and summarizes findings in coherent tables.
- Explains common deep learning architectures leveraged in agricultural research to enhance reader comprehension.
- Highlights challenges that need to be addressed for effective real-world implementation in greenhouses.
- Provides visualizations of real-world systems and schematics to showcase practical applications.
- Outlines future directions for advancements in computer vision technologies tailored to controlled agriculture settings.
- Focuses the scope on an understudied niche area of computer vision in greenhouses to fill a literature gap and make novel contributions.

Overall, this review paper makes significant contributions by providing a focused technical synthesis, performance analysis, and future outlook specifically targeted to greenhouse applications of computer vision. The findings aim to catalyze advancements in this promising domain.

The paper is organized as follows. Section I provides an introduction covering the background, objectives, and scope of the literature review. Section II presents an overview of smart greenhouses and controlled environment agriculture. Section III delves into the fundamentals of computer vision and deep learning, the key technologies explored. Section IV, the core of the paper, investigates various applications of computer vision in greenhouse farming based on current literature, spanning areas like growth monitoring, disease detection, yield estimation, and more. For each application, the techniques, results, limitations, and future work are discussed. Section V summarizes the key challenges faced in implementing computer vision in greenhouses and provides perspectives on the future direction of these technologies in advancing greenhouse agriculture. Finally, Section VI gives concluding remarks and highlights the contributions made by

this focused review paper. Overall, the logical organization facilitates comprehension and showcases the in-depth analysis of computer vision capabilities, specifically in controlled greenhouse environments.

## II. SMART GREENHOUSE: AN OVERVIEW

Controlled Environment Agriculture (CEA), a modern farming method that uses greenhouses, is a powerful tool in today’s agriculture. This method is not new; in fact, greenhouses first started to be used in farming in The Netherlands and France back in the 19th century [13]. Since then, the technology has improved, and its use has spread worldwide. So, what exactly is a greenhouse? Essentially, a greenhouse is a structure - often built from glass or plastic that allows for year-round crop production, irrespective of the season. The glass or plastic walls and roof let sunlight in while keeping pests, diseases, and bad weather out. Depending on the outside weather, different variations of greenhouses can be used. For example, in very cold places, smaller greenhouses called “cold frames” can be used. These structures trap heat from the sun to keep the plants warm. In hot and dry areas, “shade house” greenhouses can be used. These provide shade and help to keep the plants moist [7]. As illustrated in figure 1, a modern greenhouse is a structure that provides a controlled environment ideal for crop cultivation. The image depicts a typical setup, including the protective coverings and the internal layout. This practical example gives a clear understanding of how greenhouse farming can defy external weather conditions and provide suitable growth conditions for diverse crops.

The main advantage of greenhouses is the ability to manipulate environmental parameters such as temperature, light intensity, moisture, and nutrient levels, adapting them to specific crop needs [15]. This facilitates an extended growing season, improved crop quality, and efficient use of resources. However, the management and maintenance of greenhouses can be resource-intensive, and fine-tuning conditions to the optimal range for various crops can be a complex task. This is where smart greenhouses, equipped with Internet of Things (IoT) technology, come into the picture. Smart greenhouses integrate sensors and embedded controllers that collect real-time data and relay it to a cloud server. The system can then make automatic adjustments to the internal greenhouse conditions based on this data, minimizing human intervention [16], [17], [18]. Smart greenhouses offer



**FIGURE 1.** A real-life example of a modern greenhouse. This image offers a practical perspective of the controlled environment within which diverse crops can be cultivated irrespective of external weather conditions. Source: [14].

automatic regulation of critical factors like temperature, light, and irrigation, as well as control over other mechanical operations [19]. This brings a new level of efficiency to farming, optimizing resource use and potentially improving crop yields.

Moreover, smart greenhouses provide farmers with valuable insights into the most suitable harvesting times, soil quality, nutrient requirements for plants, and water quality [20]. This data-driven approach allows for more informed decision-making, making farming more reliable and cost-effective. The future of farming appears even more promising with the integration of Artificial Intelligence (AI) and Computer Vision. These technologies can further automate the greenhouse processes and increase their precision [21], [22], [23]. For instance, AI can analyze the vast amounts of data collected, predicting future needs and helping make even more accurate adjustments [24], [25]. Computer Vision can monitor plant growth, detect diseases early, and even identify when crops are ready to be harvested [12], [26], [27], [28], [29]. In essence, the marriage of greenhouse farming with IoT, AI, and Computer Vision is ushering in a new era of efficient, data-driven agriculture, transforming the traditional greenhouse into a 'smart' one (see figure 2). This integration, as illustrated in figure 2, provides a complete picture of how smart greenhouses are designed with various components working

harmoniously. The ultimate vision is a fully automated and remotely controlled farm, optimizing resources, enhancing crop yield, and paving the way for a sustainable agricultural future.

### III. VISION THROUGH LEARNING: A DEEP DIVE INTO COMPUTER VISION AND DEEP LEARNING

Computer vision (CV) represents a sophisticated interdisciplinary field deeply rooted in both biological science and engineering, converging human-like perception with machine efficiency. Its core idea revolves around replicating how humans see and understand their surroundings and translating this understanding into computational models that machines can utilize. Historically, CV began as an effort to mimic human visual faculties. Early research aimed at understanding how humans perceive the world and then endowing machines with similar visual abilities. This required extensive collaboration between neuroscience to decode the human visual system and computer science to develop corresponding algorithms for machines [30]. Groundbreaking applications such as optical character recognition (OCR) [31] and vehicle plate detection marked the initial steps in this journey, impacting diverse areas like traffic control, law enforcement [32], and even retail [33]. These achievements were furthered by the use of deep learning (DL) and neural networks, allowing for the automatic

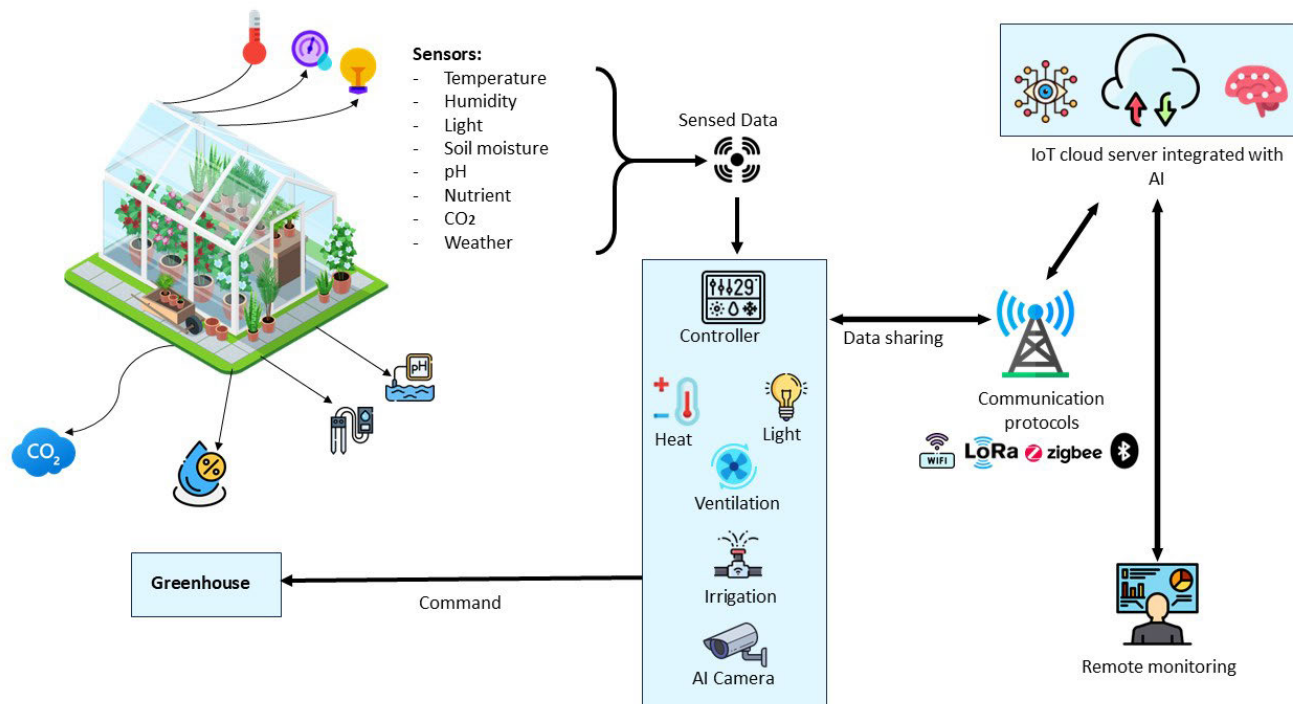


FIGURE 2. Schematic Representation of a Smart Greenhouse components Integrated with IoT, AI, and Computer Vision Technologies.

extraction of high-level features from data, reducing the need for human intervention.

The evolutionary development of CV expanded into more advanced areas like medical imaging, for detecting and diagnosing diseases [34], and autonomous driving, where it guides self-driving cars [35]. In manufacturing, CV aids in quality control with machine consistency, surpassing human ability [36]. It even breathes life into virtual reality in entertainment and social interaction [37] and enhances personalized safety measures through facial recognition systems in security [38], [39]. A vivid representation of these diverse applications of computer vision across various sectors can be seen in figure 3. An intriguing expansion of CV can be seen in the agricultural domain [12], [40], [41], [42]. Here, CV assists not only in crop-related tasks but also plays a significant role in livestock management. It contributes to monitoring animal health, tracking behavior, and managing resources, thereby increasing efficiency and sustainability in animal farming. For crop health, CV enables the detection of diseases, identification of pests, and optimization of resource allocation. With the analysis of visual data, CV tools can automate tasks like fruit picking, guide precision farming techniques, and even assist in tasks related to animal husbandry, such as recognizing individual animals, monitoring their movements, and observing their health status.

Despite these advances, achieving a computer’s understanding of an image at a human child’s level remains unattainable. This underscores the ongoing challenges in a field with nearly limitless possibilities. The continuous progress in machine

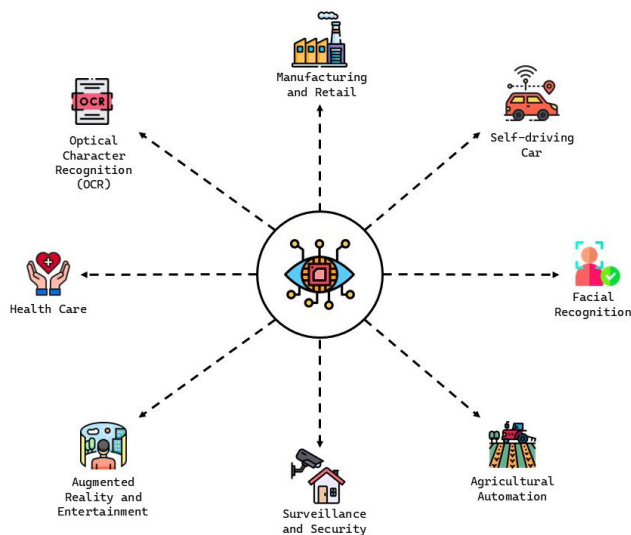
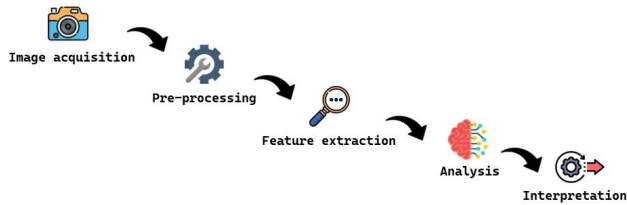


FIGURE 3. Diverse Applications of Computer Vision Across Various Sectors.

learning and DL further amplifies what CV can accomplish. Deep neural networks enable machines to recognize patterns and make decisions once believed to be exclusive to humans. CV is more than a field of study; it is a continually evolving area that harmoniously blends human perception with machine precision. Its applications are enriching various societal sectors, sometimes even surpassing human capabilities.

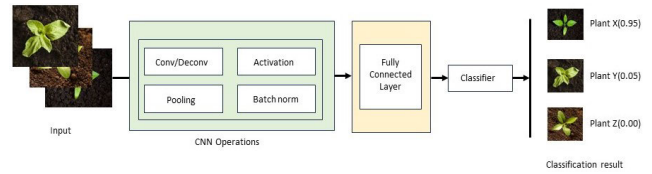


**FIGURE 4. Flowchart of the Computer Vision Process.** This diagram illustrates the sequential stages involved in computer vision applications, from the initial acquisition of image data to the final interpretation and utilization of the processed information.

### A. FROM BASIC SHAPES TO COMPLEX PATTERNS

In the initial stages of computer vision, the primary focus was on identifying simple and basic geometric forms such as edges, curves, and corners. These early techniques were often grounded in methods like gray-level segmentation [43], which involves dividing an image into different regions based on variations in brightness or color. However, these primitive methods had significant limitations and were not robust enough to handle more complicated visual tasks where understanding and interpreting complex patterns were required. To address these limitations and enhance the interpretation of visual data, researchers sought new approaches and began to integrate artificial neural networks into computer vision systems [44]. Artificial neural networks, inspired by the human brain's interconnected neuron structure, allowed for a more sophisticated analysis of visual data [45]. Unlike the pixel-by-pixel analysis used in more rudimentary methods, neural networks provided a way for computer vision systems to analyze entire sections of an image in context, providing a more holistic understanding of the image content. This shift in approach led to substantial improvements in both performance and accuracy, enabling computer vision systems to understand and recognize more intricate patterns and shapes. Furthermore, these new methods allowed computer vision systems to analyze dynamic visual data, such as videos, enabling a higher level of interpretation that considered not only the shapes and patterns within a single frame but also how these elements changed over time [46]. The integration of neural networks marked a turning point in the field, transforming it from a discipline that could only handle rudimentary visual tasks to one that could take on complex challenges. This evolution also laid the groundwork for the modern, advanced applications of computer vision that we see today.

Today's computer vision applications follow a process that includes the acquisition of image data, the processing and analysis of that data using sophisticated algorithms and neural networks, and ultimately, the interpretation of the image as shown in figure 4. These applications range from medical imaging to autonomous vehicles, demonstrating the field's progress from understanding basic shapes to interpreting complex, dynamic patterns.



**FIGURE 5. Diagram of CNN's image classification architecture mechanism.**

### B. THE TOOLBOX OF COMPUTER VISION: TECHNIQUES AND TASKS

Computer vision is a diverse and exciting field that tries to mimic how humans perceive and understand the world around them. Its primary goal is to make sense of images, just like our own eyes and brains do. By understanding what is depicted in an image, computer vision systems can provide valuable information to guide the actions of robots, AI, and other automated systems. In the following sections, we will delve deeper into some of these critical techniques: image classification, object detection, segmentation, and 3-D reconstruction, that make it all possible.

#### 1) IMAGE CLASSIFICATION

Image classification, a fundamental technique in computer vision, assigns predefined labels to images, categorizing complex visuals. Its essence is the ability to capture the essential features of a scene—such as differentiating crops or identifying plant diseases—without focusing on each tiny detail. The advent of Convolutional Neural Networks (CNNs) has led to a significant transformation in this process. These robust networks use a variety of mathematical processes to efficiently learn from labeled images of crops (In figure 5). The revolution began in 2012 with AlexNet [47], a novel CNN architecture, setting a benchmark in the ImageNet Large Scale Visual Recognition Challenge, ushering a new era for image classification via CNNs. Between 2014 and 2017, there was a surge in refining CNN architectures, as seen with advancements in models like ResNet [48], VGG [49], and DenseNet [50]. These models, by employing innovative algorithms and connectivity patterns, heightened the ability to discern intricate details in agricultural imagery. Post-2017 witnessed the incorporation of reinforcement learning into CNNs, enabling these models to auto-determine optimal architectures, enhancing adaptability in dynamic agricultural environments. Contemporary studies further improve upon this with variants like ZFNet [51], which leverages visualization tools to understand neural activity, translating this into pixel-space insights. Such advancements not only enhance classification abilities but also unveil the working mechanisms of CNNs, making image-based agricultural decision-making more precise and insightful. Currently, these technologies are of extreme significance in improving the accuracy of image-based agricultural decision-making.

#### 2) OBJECT DETECTION

Object detection in computer vision aims to locate and classify all possible objects within a given image. Central to this

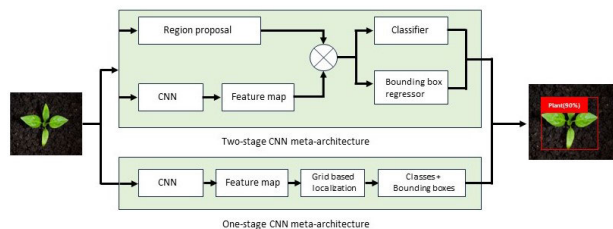


FIGURE 6. CNN architecture mechanism for Object detection.

task are Convolutional Neural Networks (CNNs), which have predominantly been employed in two main architectures: one-stage and two-stage detection systems (in figure 6). Two-stage models, exemplified by the RCNN family, first pinpoint potential object regions (known as region proposals) and then classify these regions into distinct object categories. The evolution of this approach has led to various iterations, such as the original RCNN [52], Fast RCNN [53], and the more popular Faster RCNN [54]. Three pivotal advancements in the Faster RCNN are the Region Proposal Network (RPN), which efficiently creates object regions; ROI pooling, which extracts consistent features from regions of varying dimensions; and a multitask loss function, which consolidates the training process. While Faster RCNN is acclaimed for its accuracy, its processing speed is a limiting factor, especially for real-time applications. In contrast, one-stage models like YOLO [55]

and SSD [56] are designed for faster processing. These architectures generate candidate object regions from each pixel in feature maps, which are then classified and adjusted for accurate object boundaries. However, a key challenge with one-stage models is the significant imbalance between object and background regions in images. To address this, the RetinaNet [57] framework introduced a focal loss function, which emphasizes detection accuracy, balancing out the disproportionate number of irrelevant regions. While RetinaNet offers a blend of accuracy and efficiency, the choice between one and two-stage models depends largely on the specific application needs, with the former being ideal for real-time tasks and the latter prioritized when accuracy is paramount. Surprisingly, these state-of-the-art object identification methods have significant uses outside of traditional fields. The agricultural industry has embraced these technological marvels, especially in the context of greenhouse farming. Precision agriculture has entered a new era by adopting such advanced object detection technologies in greenhouse farming. These improvements have allowed for exact monitoring of plant growth, effective detection of pests, and efficient allocation of resources.

### 3) SEGMENTATION

Image segmentation is a fundamental computer vision task which categorizes each pixel of an image based on the object it belongs to. Early techniques used region division and merging, and later algorithms leveraged metrics like intra-regional consistency and inter-regional dissimilarity.

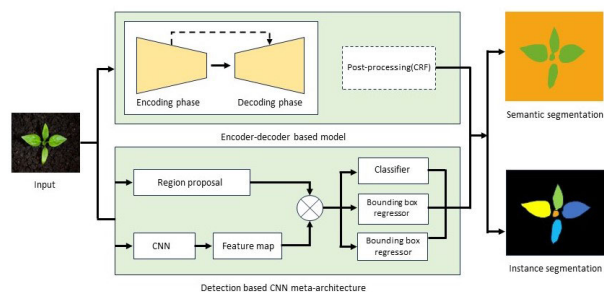


FIGURE 7. CNN architecture mechanism for Segmentation.

Recently, machine learning has greatly enhanced segmentation, with key advancements being Mask RCNN [58], dual attention network [59], and particularly, U-Net [60] and its variants such as Attention U-Net [61], U-Net++ [62], ResUNet++ [63], and TransUNet [64]. These models have not only advanced the state of image segmentation but have also demonstrated exceptional efficacy in specialized tasks like medical image segmentation. Building on these semantic and instance segmentation techniques, such as the encoder-decoder framework and detection-based methods, further expands the versatility and precision of computer vision applications. Semantics and instance segmentation are vital techniques in computer vision. At a high level, semantic segmentation assigns masks to groups of objects with the same meaning in an image, such as all plants, while instance segmentation focuses on individual objects. Two primary frameworks are employed in these types of segmentations: encoder-decoder-based frameworks and detection-based frameworks. In the encoder-decoder approach, models typically consist of two main phases, as shown in figure 7. The encoder extracts meaningful feature maps from images using convolutional neural networks (CNNs). The decoder, on the other hand, upsamples these feature maps into per-pixel labels using transposed convolution. To enhance the precision of segmentation, these models often employ a lateral connection scheme. This connects feature maps between the encoder and decoder phases, ensuring the preservation of the image’s semantic meaning throughout the process. Additionally, post-processing methods like conditional random fields (CRFs) are utilized to refine object boundaries. Notable models in this category include U-Net, fully convolutional network (FCN) [65], FastFCN [66] and DeepLab [67].

Detection-based frameworks, in contrast, pivot on CNN architectures tailored for object detection. Some early efforts attempted to leverage object detection models, for instance, segmentation, such as simultaneous detection and segmentation (SDS) [68] and DeepMask [69]. However, these approaches struggled to achieve desired performance levels. The game-changer in this arena was the Mask RCNN, which integrated an FCN with a Faster RCNN to create masks for individual objects. This model has consistently

demonstrated top-tier performance in both semantic and instance segmentation, solidifying its reputation in the field.

#### 4) 3-D MODELING

3-D modeling in computer vision is fundamentally about stereo correspondence and 3-D reconstruction [70], [71]. Stereo correspondence generates a 3-D model from multiple images of the same object or scene by finding matching pixels across these images and mapping their 2-D positions to 3-D. Techniques such as epipolar geometry, sparse correspondence, and dense correspondence are commonly used [72]. On the other hand, 3-D reconstruction creates a 3-D model from a single image [73], [74]. The earliest approach involved predicting object shape from visual shading, pioneered by Horn in 1970 [75]. This was followed by other “shape from X” methods like shape from texture and shape from focus. Active range finding and model-based reconstruction, which are often used in architectural 3-D modeling, are among other methods. Designing an effective loss function for evaluating predicted 3-D point clouds against ground truth remains challenging, with options including evaluating the coverage of the ground truth object’s silhouette by the projected 3-D point clouds. Deep-learning-based algorithms have recently led to significant enhancements in the performance of 3-D reconstruction systems. In summary, the world of computer vision is as complex as it is fascinating, comprising an array of techniques and tasks that together help the system understand and interact with its environment. With the rapid advancements in machine learning and artificial intelligence, we can expect to witness further evolution in these techniques, expanding the possibilities of what computer vision can achieve.

### C. THE ADVENT OF DEEP LEARNING

Machine learning offers a significant advancement in data processing. Traditional methods typically necessitate manual feature extraction. With the surge in large data sets and the introduction of graphics processing units (GPUs), algorithmic methodologies have seen considerable refinement. Deep learning, an evolution from traditional machine learning, incorporates some “deeper” (more complex) structures, enabling automatic feature extraction from unprocessed data. It often surpasses the efficacy of its predecessor in various classification and prediction tasks [76]. By integrating multiple layers of abstraction, it allows hierarchical data representation [77], [78]. This multi-layered approach enhances the analytical performance for numerous large-scale data processing tasks [79], [80]. Essentially, deep learning is an advanced non-linear data processing technique grounded in representation learning and pattern analysis. Typically, deep learning models refine data representations through multi-layered neural networks. These networks comprise several neurons structured in layers. Neurons in adjacent layers connect based on weights, which are adjusted during learning. These neurons represent diverse non-linear functions, facilitating the creation of complex models. By connecting

multiple layers, deep learning provides solutions for intricate real-world challenges efficiently [78].

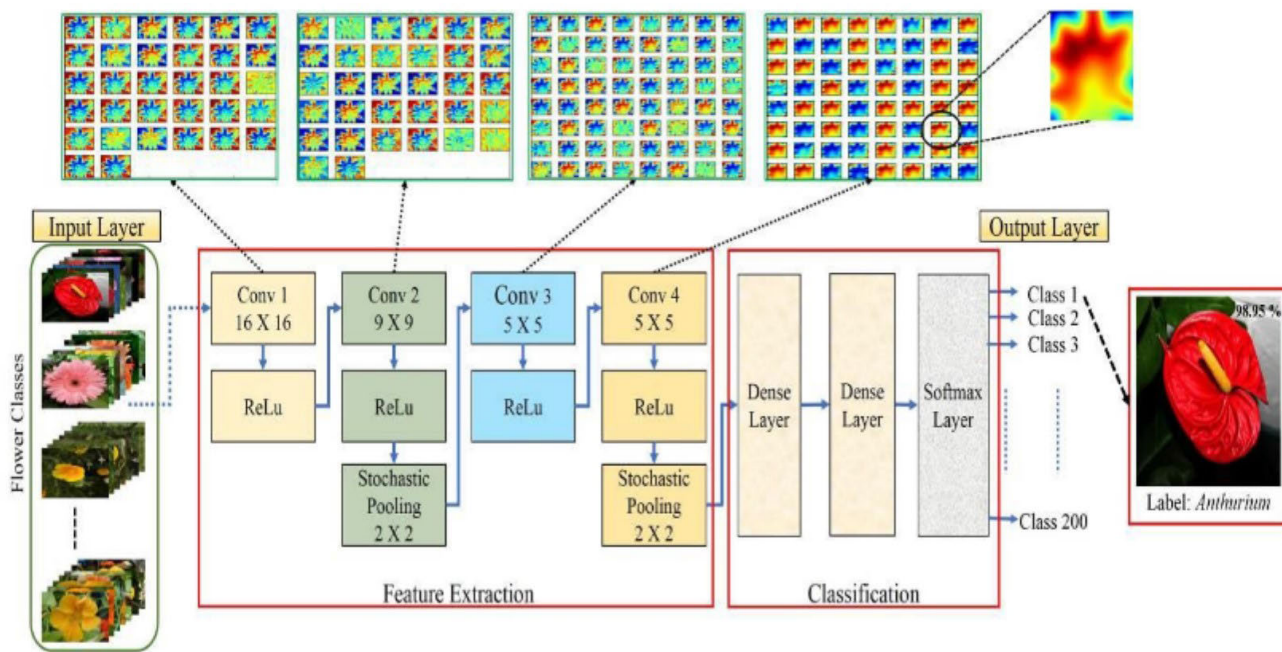
The complex architecture and massive learning capability of deep learning models equip them with exceptional prediction and classification capabilities. This enables them to adapt effectively to complex data analyses. Leveraging its innate ability for automatic feature extraction, deep learning addresses several challenges in agriculture, such as various recognition, growth monitoring, yield estimation, quality assessment, stress detection, and more. These applications are discussed in detail in the next section.

Convolutional neural networks (CNNs) and their derived models play a pivotal role in artificial intelligence and have led to breakthroughs in image processing and analysis. CNNs are a class of deep, feed-forward artificial neural networks (ANNs) that are a family of multi-layered neural networks that have been effectively used in computer vision applications. They constitute a prominent method for analyzing vast amounts of data. Our analysis indicates that a significant majority of horticulture-related papers utilize CNNs. Typical CNNs incorporate convolution, pooling, and fully connected layers in various configurations to perform complex learning tasks. A model CNN architecture displays the process of classifying different flower species using multiple layers and components in figure 8.

In the convolutional layer, local patterns within an image are identified using the process of convolution. Here is how it works: A kernel is initially positioned on the top-left part of the image as shown in figure 9. Each pixel under this kernel gets multiplied by the respective kernel value. The resulting products are then summed, with a bias added afterwards. The kernel shifts by a pixel, and the process continues until the entire image undergoes this filtering. Following the convolutional process, the pooling layer steps in, aiming to down-sample and extract prominent features from the obtained feature map. It also brings invariance to minor translations, rotations, and scaling in the image. Two prevalent pooling methods are max and average pooling, as shown in figure 10. While max pooling considers the highest value from a designated portion of the image, average pooling computes the mean of that portion. In most CNN architectures, convolutional and pooling layers alternate. The last key layer is the fully connected layer, where every neuron is interlinked to the preceding neuron. In this stage, the various features that were obtained from earlier layers are combined and condensed into a one-dimensional representation, preparing them for detection or classification objectives.

The efficacy of a deep learning model depends on the appropriate choice of hyperparameters. These include aspects like the network architecture, number of layers, number of neurons in the hidden layer, convolution and pooling layer structure, learning rate, weight initiation, and activation function. While custom architectures can be revolutionary and innovative, they typically demand an advanced level of computational expertise that may be beyond the capabilities of ordinary agricultural researchers. As a result, researchers





**FIGURE 8.** A Deep Convolutional Neural Network (DCNN) layout: This structure comprises an initial input layer, followed by four convolutional layers and their corresponding ReLU activations. It also features two stochastic pooling layers, a pair of fully connected layers, and concludes with a softmax regression output layer. Source: Prasad et al. [81].

frequently start with a pre-trained architecture that has proven to perform well over a wide range of data structures and challenges and then modify it to fit the issue at hand. This approach is reliable and effective.

CNN architectures like LeNet, AlexNet, VGGNet, MobileNet, Inception V3, EfficientNet, GoogLeNet, and ResNet have been deployed for many computer vision tasks discussed in the next section. These networks process input through multiple convolution and pooling layers before utilizing fully connected layers for classification or detection tasks. The choice of parameters within these layers and model selection should be tailored to the specific problem as illustrated in Fig 11, graphical representations of the SSD deep learning architecture have been used to detect cherry tomatoes. The SSD utilized VGG16 as its base network, with multiple layers appended to its tail end. During the training process, both the image and its associated Ground-Truth Box were simultaneously fed into the system. Following this, each feature map across 6 layers generated a default box set, as well as the confidence of object category within the box.

Another prevalent deep learning model used for sequential data is the recurrent neural network (RNN). This model excels in predicting prices [83], processing natural languages [84], recognizing speech [85], among other applications [86], [87]. An RNN’s unique attribute is its ability to retain previous data, which influences current outputs. Figure 12 demonstrates the basic architecture of the RNN applied to fruit quality assessment, illustrating how prior outputs serve as inputs due to the presence of the hidden layer. This layer effectively acts as a memory cell, ensuring every prior result informs the subsequent iteration. Consequently, each unit within the hidden layer is termed a recurrent cell.

Image segmentation tasks also benefited from deep learning techniques. The R-CNN method, for instance, combines CNNs with region proposals for object detection. Currently, Mask R-CNN leads the segmentation task, adding a branch to the Faster R-CNN for precise, high-quality segmentation mask generation for each region of interest. We already covered this in the previous section. There are many other neural network architectures that have been applied in the realm of deep learning. These include the Single Shot Multibox Detection (SSD), Long Short-Term Memory networks (LSTM), the “You Only Look Once” (YOLO), Region-based CNN (R-CNN), Fast R-CNN, and its successor, the Faster R-CNN etc. Beyond RGB imagery, these structures also support varied data formats, including videos, hyperspectral visuals, and spectral datasets. Moreover, for ease of reference, we have compiled a summary of the models reviewed. This includes details such as the year of development, key concepts, and links to source code or third-party implementation if available (See Table 2). Furthermore, to assess the effectiveness of these neural networks, various evaluation metrics have been used. Here are some commonly utilized metrics in the studies we have reviewed:

- **Classification Accuracy (CA):** This metric calculates the ratio of accurately identified images or classes to the overall number of images or classes. For problems involving multiple classes, the CA is computed as the average across all categories.

$$CA = \frac{\text{Number of Correctly Classified Samples}}{\text{Total Number of Samples}} \quad (1)$$

- **Precision:** Defined as the ratio of True Positives (TP) to the combined count of TP and

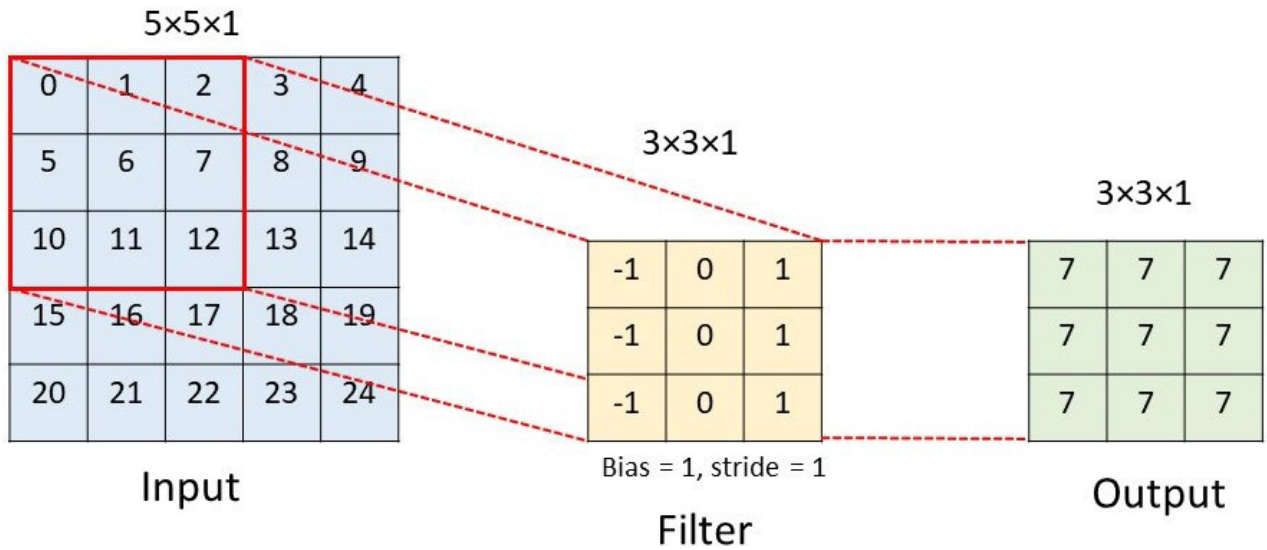


FIGURE 9. The process of convolutional operation.

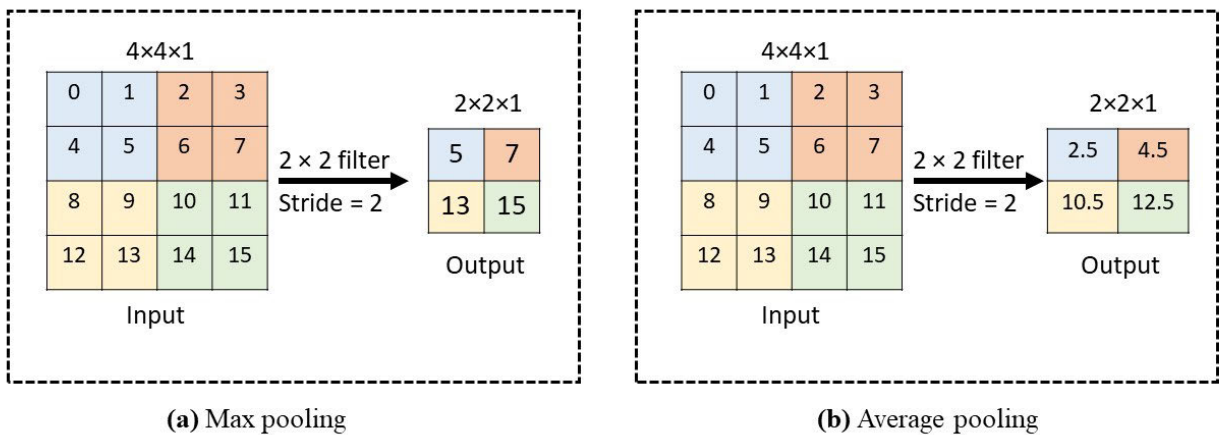


FIGURE 10. The process of pooling operation.

False Positives (FP).

$$\text{Precision (P)} = \frac{TP}{TP + FP} \tag{2}$$

- **Recall:** Denoted as the ratio of TPs to the collective sum of TPs and False Negatives (FN).

$$\text{Recall (R)} = \frac{TP}{TP + FN} \tag{3}$$

- **F1-score:** This is the harmonic mean of precision and recall, providing a balance between the two.

$$F1 = \frac{2 \times (\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}} \tag{4}$$

- **Root-Mean Square Error (RMSE):** It measures the root-mean square of the differences between predicted and actual values, offering insight into the model’s prediction accuracy.

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \tag{5}$$

#### IV. DEEP LEARNING ASSISTED COMPUTER VISION APPLICATIONS IN GREENHOUSE FARMING

The integration of deep learning-assisted computer vision techniques into greenhouse environments has produced promising results, indicating significant effectiveness. Deep learning-enhanced computer vision has a wide range of applications within the greenhouse industry, including recognition and classification of crops, crop yield estimation, crop quality inspection, crop growth monitoring, automatic harvesting, disease and pest management and so on. A small taxonomy of these applications is illustrated in figure 13. In this section of this paper, we explore the specific problems addressed in academic literature across these application areas. We investigate various computer vision techniques, models and architectures that have been utilized, the performance of the model adopted, discuss the underlying constraints and challenges, and offer insights to guide future research endeavors.

##### A. CROP GROWTH MONITORING

Ensuring the healthy growth of crops is essential in greenhouse farming in order to achieve optimal yield, quality, and efficient

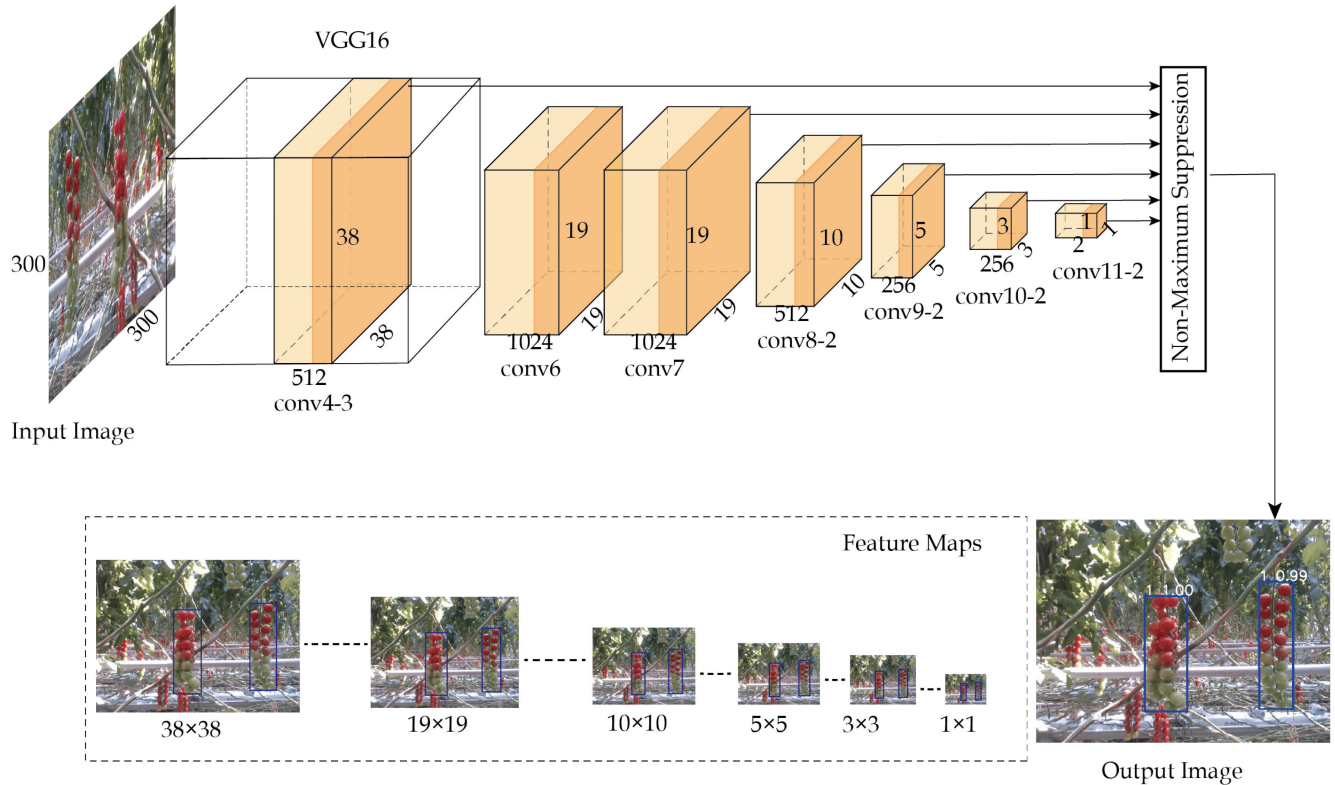


FIGURE 11. Classical SSD deep learning network architecture based on VGG-16. Source: [82].

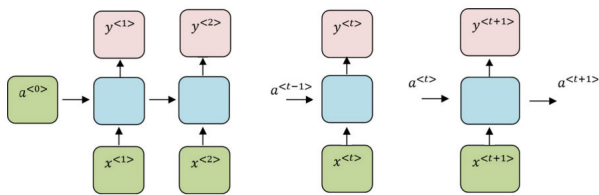


FIGURE 12. Architecture of RNN used for fruit quality assessment. Source: [88].

use of resources, all of which contribute to profitable outcomes in agricultural production. Crops require a variety of nutrient elements, including macronutrients, secondary nutrients, and micronutrients, for optimal growth. Traditionally, monitoring crops has always relied heavily on human observations, which frequently leads to inaccurate and delayed results. Precision agriculture, especially in controlled environments like greenhouses, emphasizes the importance of consistent and accurate crop monitoring at varied growth stages. By integrating computer vision into greenhouse management, we can achieve real-time, precise monitoring. It provides timely interventions for optimal growth by detecting minute changes in crop health due to nutritional deficiencies far earlier than manual inspections.

Utilizing innovative technologies to enhance agricultural processes, especially in the realm of greenhouse farming, is increasingly prevalent. Lee et al. [92] embarked on an exploration using clip-type Internet of Things (IoT) cameras to monitor tomato growth in a greenhouse setting, using deep

learning-based object detection to track flower blooming and immature fruit development. By integrating the flower's bloom date with temperature forecasting, their system could predict harvest dates with an average error of two days. While the clip-type design effectively addressed challenges like plant overlap in densely populated greenhouses, reliance on wired power sources and occasional occlusions posed limitations. The introduction of battery-powered IoT cameras and more adaptive positioning might enhance its applicability. In a different study, Gang et al. [93] implemented a convolutional neural network (CNN)-based model to estimate growth indices for greenhouse lettuce, utilizing RGB-D data from stereo cameras. Their dual-stage CNN architecture, building upon the ResNet50V2 layers, demonstrated remarkable accuracy. Similarly, Zhang et al. [94] showcased the capabilities of convolutional neural networks in relation to digital images. The study demonstrated how CNN can be useful in monitoring growth indicators like leaf fresh weight, leaf area, and leaf dry weight, delivering results that outperform traditional methodologies, especially for specific lettuce cultivars. However, while these models can extract depth information to enhance accuracy, their real-time application demands considerable computational power. Nevertheless, the speed of image processing, particularly with edge devices like Jetson SUB mini-PC, implies a potential future in real-time monitoring.

The growing field of computer vision and deep learning has opened up new possibilities in crop management. Pre-trained models like YOLO, ResNet, VGG16, MobileNet,

**TABLE 2. Summary of major CNN architectures.**

Model	Task	Key Concept	Source Code
AlexNet [47]	Image Classification	Five-layer CNN	TensorFlow: <a href="https://github.com/TensorFlow/models/blob/master/research/slim/nets/alexnet.py">https://github.com/TensorFlow/models/blob/master/research/slim/nets/alexnet.py</a> PyTorch: <a href="https://github.com/pytorch/vision/blob/master/torchvision/models/alexnet.py">https://github.com/pytorch/vision/blob/master/torchvision/models/alexnet.py</a>
ZFNet [51]	Image Classification	Feature visualization	TensorFlow: <a href="https://github.com/amir-saniyan/ZFNet">https://github.com/amir-saniyan/ZFNet</a> (Third-party)
VGGNet [49]	Image Classification	Small 3x3 filters, up to 19 layers to increase the depth of CNNs	PyTorch: <a href="https://github.com/pytorch/vision/blob/main/torchvision/models/vgg.py">https://github.com/pytorch/vision/blob/main/torchvision/models/vgg.py</a> Caffe: <a href="http://www.robots.ox.ac.uk/~vgg/research/very_deep/">www.robots.ox.ac.uk/~vgg/research/very_deep/</a>
Inception [89]	Image Classification	Inception modules for increasing the width of CNNs	TensorFlow: <a href="https://github.com/tensorflow/models/blob/master/research/slim/nets/inception_v3.py">https://github.com/tensorflow/models/blob/master/research/slim/nets/inception_v3.py</a> PyTorch: <a href="https://github.com/pytorch/vision/blob/main/torchvision/models/inception.py">https://github.com/pytorch/vision/blob/main/torchvision/models/inception.py</a>
ResNet [48]	Image Classification	Residual representation and a skip connection strategy enable the training of very deep CNNs (up to 1000 layers).	TensorFlow: <a href="https://github.com/tensorflow/models/blob/master/research/slim/nets/resnet_v1.py">https://github.com/tensorflow/models/blob/master/research/slim/nets/resnet_v1.py</a> PyTorch: <a href="https://github.com/pytorch/vision/blob/main/torchvision/models/resnet.py">https://github.com/pytorch/vision/blob/main/torchvision/models/resnet.py</a>
DenseNet [50]	Image Classification	Dense blocks, fewer params	PyTorch: <a href="https://github.com/pytorch/vision/blob/main/torchvision/models/densenet.py">https://github.com/pytorch/vision/blob/main/torchvision/models/densenet.py</a>
NASNet [90]	Image Classification	Reinforcement learning using a small dataset to determine ideal convolutional cells used to design a CNN architecture for a large dataset	TensorFlow: <a href="https://github.com/tensorflow/models/tree/master/research/slim/nets/nasnet">https://github.com/tensorflow/models/tree/master/research/slim/nets/nasnet</a>
RCNN [52]	Object Detection	Two-stage, ROIs, class label, bounding box for each ROI	PyTorch (Faster RCNN): <a href="https://github.com/pytorch/vision/blob/main/torchvision/models/detection/faster_rcnn.py">https://github.com/pytorch/vision/blob/main/torchvision/models/detection/faster_rcnn.py</a>
YOLO [55]	Object Detection	One-stage, class labels, bounding box	<a href="https://github.com/ultralytics/yolov5#yolov8--new">https://github.com/ultralytics/yolov5#yolov8--new</a>
SSD [56]	Object Detection	One-stage, class labels, bounding box	PyTorch: <a href="https://github.com/pytorch/vision/blob/main/torchvision/models/detection/ssd.py">https://github.com/pytorch/vision/blob/main/torchvision/models/detection/ssd.py</a>
RetinaNet [57]	Object Detection	A one-stage approach using focal loss, a novel loss function, to address the issue of foreground-background class imbalance	PyTorch: <a href="https://github.com/pytorch/vision/blob/main/torchvision/models/detection/retinanet.py">https://github.com/pytorch/vision/blob/main/torchvision/models/detection/retinanet.py</a>
FCN [91]	Semantic Segmentation	Fully convolutional architecture for pixel-by-pixel training and class prediction	PyTorch: <a href="https://github.com/pytorch/vision/blob/main/torchvision/models/segmentation/fcn.py">https://github.com/pytorch/vision/blob/main/torchvision/models/segmentation/fcn.py</a>
U-Net [60]	Semantic Segmentation	Encoder-decoder architecture	TensorFlow: <a href="https://github.com/jakeret/tf_unet">https://github.com/jakeret/tf_unet</a> PyTorch: <a href="https://github.com/milesial/Pytorch-UNet">github.com/milesial/Pytorch-UNet</a>
DeepLab [67]	Semantic Segmentation	Atrous convolution to increase the receptive field and reduce computational complexity for better segmentation; a post-processing method with fully connected CRF enhances accuracy	PyTorch: <a href="https://github.com/pytorch/vision/blob/main/torchvision/models/segmentation/deeplabv3.py">https://github.com/pytorch/vision/blob/main/torchvision/models/segmentation/deeplabv3.py</a> TensorFlow: <a href="https://github.com/tensorflow/models/tree/master/research/deeplab">https://github.com/tensorflow/models/tree/master/research/deeplab</a>
Mask RCNN [58]	Instance Segmentation	Integrating a masking head using the ROI align procedure on top of the Faster RCNN model to enhance segmentation precision	TensorFlow: <a href="https://github.com/tensorflow/models/blob/master/official/vision/tasks/maskrcnn.py">https://github.com/tensorflow/models/blob/master/official/vision/tasks/maskrcnn.py</a> PyTorch: <a href="https://github.com/pytorch/vision/blob/main/torchvision/models/detection/mask_rcnn.py">https://github.com/pytorch/vision/blob/main/torchvision/models/detection/mask_rcnn.py</a>

Detectron2, etc., have showcased remarkable capabilities in object detection. Integrating these advanced techniques into horticultural research has yielded impressive performance and opened avenues for innovative applications. For instance, Moysiadis et al. [95] highlighted the potential of using the YOLOv5 pre-trained model for mushroom growth monitoring. The complexities of mushroom growth patterns raise the significance of this research. However, the accuracy of detection underscores the challenges of using computer vision in dense environments with overlapping objects. Similarly, in a recent study, Shinoda et al. [96] introduced the “RoseTracker”, a system that combines YOLOv5, SORT, and a regression model aimed at monitoring the growth of roses in cultivation environments. The dataset they provided stands out due to its specific focus on the unique stages of rose growth. Though the results demonstrate remarkable accuracy, the system’s adaptability across diverse global cultivation conditions and

its versatility with other floral species will be crucial factors in determining its broader acceptance and utility.

These examples of applying deep learning-based computer vision to greenhouse crops demonstrate its utilization for various aspects of research and production, including species classification, organ detection, growth stage monitoring, and localization. Studies have implemented techniques like convolutional neural networks, YOLO, and other object detection models to track developmental indicators and phases precisely. While occlusion and adaptability across diverse environments remain challenges, deep learning enables real-time, accurate crop monitoring to support data-driven decision-making for optimal yields, quality, and resource efficiency in greenhouse farming. The technology displays remarkable potential in replacing time-consuming and imprecise manual inspection. Overall, deep learning-based computer vision facilitates timely, precise interventions for optimal greenhouse

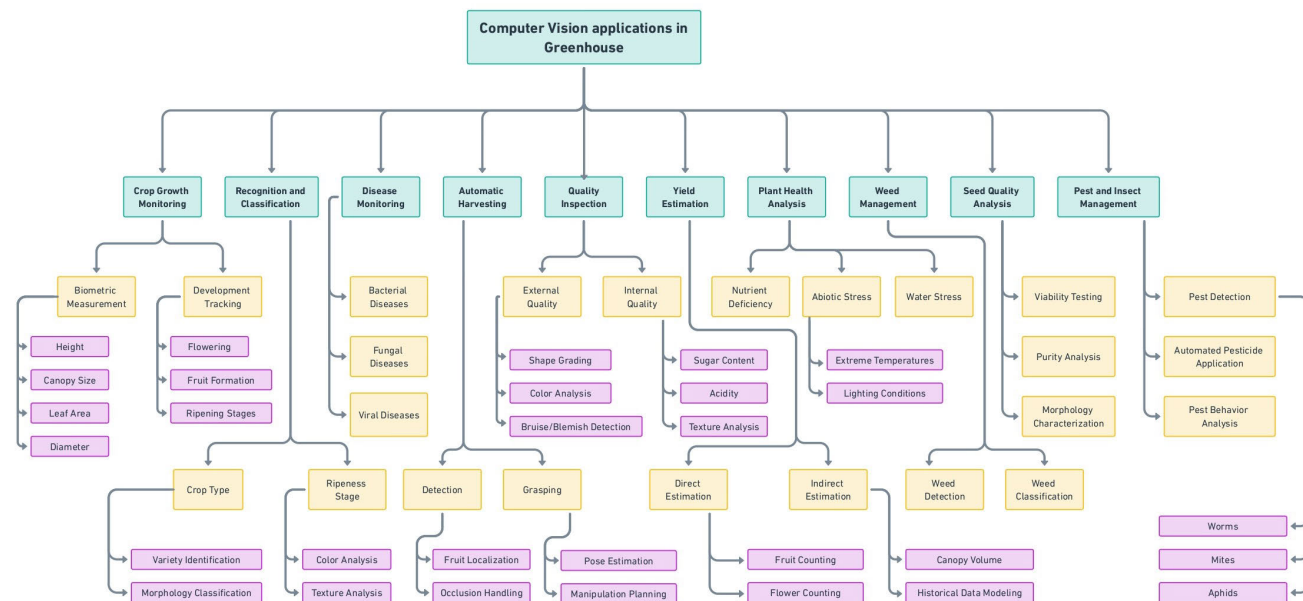


FIGURE 13. Taxonomy of computer vision applications in greenhouse farming.

crop growth by detecting growth changes earlier than humanly possible. Further technological advances could transform greenhouse management and contribute to more sustainable, productive agricultural practices. Some of the studies in this area are precisely summarized in table 3

**B. RECOGNITION AND CLASSIFICATION OF CROPS**

Automated recognition and classification of greenhouse crops is a challenging task due to the vast diversity and continuous evolution of plant species. Moreover, crops can undergo various mutations, leading to significant variations within a single class. As a result, flowers from different species may share similar features such as shape, colour, and general appearance. Therefore, the recognition and classification of greenhouse crops presents a complex challenge with multi-class problems. Although manual classification is possible, it is typically labour-intensive and prone to errors, particularly when handling large numbers of samples. Hence, the application of computer vision-based deep learning techniques for the purpose of identifying and classifying species or cultivars holds immense potential as a groundbreaking advancement in the field of smart agriculture and greenhouse farming, thanks to their outstanding speed and accurate recognition capabilities.

In recent advancements, the adoption of deep learning-based computer vision techniques in the domain of greenhouse agriculture has witnessed notable developments, particularly when it comes to enhancing precision and efficiency. Chen et al. [97] attempted to resolve the challenge posed by the dense intertwining of cucumber canopy vines. By proposing an image recognition model based on an enhanced YOLOv5, they aimed to increase the accuracy of detecting cucumber

canopy vine tops. Notably, they introduced the CA (Coordinate Attention) mechanism module and transitioned from GIOU to EIOU for loss regression, resulting in a commendable accuracy of 97.1% in recognizing cucumber canopy tops in varied conditions. While the methodology is outstanding, there is potential to further adapt it for broader agricultural applications, considering the dynamism of outdoor conditions that might affect accuracy.

Islam et al. [98] addressed the complications in separating leaf pixels from backgrounds in thermal images due to factors like thermal radiation and greenhouse humidity. They proposed TheLNet270v1, achieving a remarkable 91% accuracy in distinguishing canopy pixels. This innovative approach underscores the capability of deep learning in analyzing thermal images within greenhouses. While the results are promising, it is essential to make sure the model works well with different pixel size and adapt to the diverse environmental shifts commonly observed in greenhouses.

Zhou et al. [99] proposed an “improved Faster-RCNN” architecture to detect strawberries from ground-level RGB images. This mechanism not only aids in efficient harvesting but also plays a pivotal role in selecting high-yield strawberry varieties. Their method achieved a notable fruit extraction accuracy of 86%, which surpassed the three other methods tested. However, the complexity and environmental adaptability of these models still necessitate further research. Future work in this domain needs to address training complexity and refine models to ensure robustness across various environmental conditions.

Cong et al. [100] innovatively incorporated the Swin Transformer attention mechanism into Mask RCNN, enhancing the model’s feature extraction capabilities. Their model efficiently segmented sweet peppers even in complex

TABLE 3. Summary of vision-based crop growth monitoring studies.

Reference	Plant Type	Growing Medium	Task	Method Applied	Dataset	Imaging Sensor	Feature Type	Performance
[96]	Rose Plant	Soil-based	Detecting the number of roses & their blooming stages	YOLOv5, SORT and regression	Rose-Blooming dataset comprises a training set of 312 images, a validation set of 106 images, and a test set of 101 images.	Gopro MAX	Color and Shape	Precision = 97% Recall = 93% F1 Score = 95%
[95]	Mushrooms (genus Pleurotus)	Soil-based (Substrate bags in greenhouse)	Mushroom growth monitoring & to detect mushrooms at their harvesting stage.	YOLOv5 and Deelectron2	1128 images for different stages and, 231 images for growth rate	Not Specified	Shape	Best model (YOLOv5l with hyperparameters derived from evolution and batch size of eight) achieved mAP@0.5 of 0.79353 and mAP@0.5:0.95 of 0.54582. Best F1-Score of 76.65% with YOLOv5l, hyperparameters derived from evolution, and batch size of eight. Average growth days for mushrooms was 5.22 days.
[93]	Lettuce (Four varieties)	Greenhouse hydroponics system	Estimating growth indices (fresh weight, dry weight, height, leaf area, diameter)	Two-stage CNN based on ResNet50V2	Online autonomous greenhouse challenge data Contains 388 pairs of RGB and depth images.	Intel RealSense D415 stereo camera	RGB color space and Depth data	Coefficients of determination ranging from 0.88 to 0.95. RMSE: 6.09% (fresh weight), 6.30% (dry weight), 7.65% (height), 7.92% (diameter), 5.62% (leaf area)
[92]	Tomato	Soil-based	Growth monitoring & harvest date prediction	YOLO v3-Tiny, DarkNet19, DarkNet53 And Tem	1350 images	M5Camera (Clip-type IoT camera)	Not Specified	Accurate detection of anthesis, number of immature fruits, and prediction of harvest date within an error range of $\pm 2.03$ days
[94]	Greenhouse lettuce	Greenhouse hydroponics system	Estimation of growth-related traits (leaf fresh weight (LFW), leaf dry weight (LDW), and leaf area (LA))	CNN (compared with other methods like SVR, RF, and LR)	286 digital images and 286 depth images	Kinect 2.0 depth sensor	Structural features derived from depth images	R2 values for CNN: LFW (0.8938), LDW (0.8910), and LA (0.9156)

scenarios, such as varying lighting conditions, pepper overlaps, and leaf occlusions. Achieving an average FPS value of 5, their approach holds promise for real-time monitoring of sweet pepper growth. Nevertheless, there is still room for improvement, especially in terms of real-time performance (e.g., inference speed), to make it more optimized and viable for large-scale practical applications, particularly in automated fruit-picking systems.

In recent years, the automation of crop harvesting, specifically tomatoes, has gained considerable attention in agricultural robotics. Su et al. [101] targeted tomato maturity detection, an essential aspect for determining post-harvest logistics such as transportation and storage. Their SE-YOLOv3-MobileNetV1 model excelled in classifying tomatoes into four distinct maturity levels with an average precision value of 97.5%. The incorporation of the Squeeze-and-Excitation attention mechanism ensured accurate detection while keeping the model lightweight, an attribute essential for embedded development in robotic applications. However, despite the robustness of their model, additional progress is required to address the challenges posed by real-world circumstances, such as the mutual occlusion caused by leaves and fruits. Meanwhile, Yuan et al.'s [82] study focuses on detecting cherry tomatoes in a complex greenhouse setting. Considering the operational environment and the precision offered by deep learning, they chose the SSD model and further experimented with varying base networks and input sizes. Their results demonstrate significant improvements in automatic cherry tomato detection, with a fascinating 98.85% precision achieved using the Inception V2 network. However, the challenge of detecting side-grown tomatoes underscores the need to fine-tune the model for complex scenarios. Moreira et al. [102] highlighted the crucial step towards achieving fully automated robotic harvesting: the development of an accurate fruit detection system. They proposed a deep learning-based system using SSD MobileNet v2 and YOLOv4 models to detect tomatoes and introduced an innovative histogram-based HSV colour space model for classifying their ripening stage. Notably, YOLOv4 displayed an impressive performance in both detection and classification, with an impressive F1-Score of 85.81% in the detection task. Yet, challenges persisted in identifying the middle stages of ripening due to subtle colour variations. This underpins the importance of continual model refinement, especially when differentiating closely related classes. In another similar study, Mao et al. [103] strived to enhance the accuracy and practicality of cucumber detection in complex environments. They introduced a multi-path convolutional neural network (MPCNN) with colour component selection and a support vector machine (SVM). The methodology effectively identified the cucumber region by reducing the background interference and emphasizing the colour differences between the cucumber and its surroundings. This approach yielded satisfactory results, with over 90% pixels in cucumber images being accurately classified. However, it remains uncertain how this model would perform under

different environmental conditions or with other cucumber varieties.

These applications of deep learning-based computer vision for automated recognition and classification of greenhouse crops shows promising results, but also persisting challenges. Studies have utilized techniques like YOLO, Mask RCNN, Faster R-CNN, MobileNets, and Swin Transformers to accurately detect and classify various crop species, cultivars, and growth stages. The complexities of distinguishing between highly similar classes, handling occlusion, and adapting models across diverse environments and lighting conditions remain active research problems. However, deep learning methodologies have achieved remarkable accuracies surpassing traditional techniques. Continual refinement of models and architectures tailored to specific crops, growth conditions, and agricultural tasks is still needed for robust real-world performance. Table 4 summarizes the technical details of the studies presented in this subsection.

### C. CROP DISEASE MONITORING

In agriculture, plant diseases always remain a major issue that causes significant losses in the world's food production, particularly in controlled settings like greenhouses. It is severely affecting the yield and quality of agricultural products and has become a key concern in the development of global agriculture. Traditional identification methods, largely manual and guided by pathologists, often lack the desired speed and precision, making them unsuitable for the fast-paced requirements of modern agriculture. Therefore, the development of deep learning and computer vision presents a promising path toward fast and accurate disease recognition; however, environmental factors such as different lighting and leaf occlusion continue to be ongoing challenges that researchers work to overcome in the area of greenhouse disease monitoring. At present, advances in imaging technology have led to the creation of several open-source image datasets featuring various horticultural crops. Notable datasets include ImageNet, PlantVillage, and OUFV. These collections have significantly improved the accuracy of image classification and recognition. Such large-scale image datasets are extensively used, offering plenty of feature information for training deep neural network models in horticultural research. With PlantVillage([https://www.tensorflow.org/datasets/catalog/plant\\_village](https://www.tensorflow.org/datasets/catalog/plant_village)), Wspanaly et al. [104] explored diseases affecting tomato leaves and highlighted the importance of automated disease detection because of its cost-effective implications. Their system can identify various tomato leaf diseases and evaluate their severity. While the method showed the potential to detect previously unseen diseases and yielded severity estimations comparable to human assessments, biases in the dataset, especially concerning the background, could limit its real-world applicability. Future efforts should concentrate on diversifying datasets to ensure the development of more generalized and robust models.

TABLE 4. Summary of recognition and classification of crops.

Reference	Plant Type	Growing Medium	Task	Method Applied	Dataset	Imaging Sensor	Feature Type	Performance
[102]	Tomato	Soil-based	Detection and Classification of tomatoes according to ripening stage	YOLOv4, SSD MobileNet v2 and HSV color space model	Two public dataset has used: 1)AgRobTomato 2)RpiTomato 1029 images in total	ZED Camera, Raspberry Pi High Quality camera	HSV color space	YOLOv4 for Detection: F1-Score of 85.81% YOLOv4 for Classification: Macro F1-Score of 74.16% HSV color space model: Balanced Accuracy of 68.10%
[101]	Tomato	Soil-based	Tomato maturity classification	SE-YOLOv3-MobileNetV1	Training set: 369 Test set: 93 Total 462 images	vivo S5 rear camera with a resolution of 48 megapixels	Color and shape	mAP: 97.5% for SE-YOLOv3-MobileNetV1, Detection speed: faster than YOLOv3 and YOLOv5
[99]	Strawberry	Soil-based	Detection and Plumppness Grading of Strawberries	Improved Faster-RCNN	400 images (panoramic strawberry images)	MV-SUF1200M-T industrial cameras with a resolution of 12.0 megapixels	Morphological features and RGB color space	Improved Faster-RCNN achieved an Accuracy of 0.860 and IOU of 0.892
[82]	Cherry Tomato	Soil-based	Detection under various conditions	SSD300, SSD512 SSD-MobileNet, SSD_Inception V2	Training set: 2768 Test set: 346 Validation set: 346. Total 3460 images	Intel RealSense D435 depth camera	Color, shape, and morphological features	Precision of Separated: 94%, Occlusion: 92%, Sunny or Uneven Illumination: 96%, Shadow: 98%, Side-grown: 74% The Average precision of SSD_Inception V2 reaches 98.85%



TABLE 4. (Continued.) Summary of recognition and classification of crops.

Reference	Plant Type	Growing Medium	Task	Method Applied	Dataset	Imaging Sensor	Feature Type	Performance
[98]	Tomato (Solanum lycopersicum)	Rockwool system	Image Classification/Segmentation	TheLNet270v1	Total 13,766 thermal images	Compact long-wave thermal camera	Colorful pixels	91% accuracy (BF score), 95.75% training, 95.23% validation
[97]	Cucumber	Soil-based	Image recognition of cucumber canopy vine tops	YOLOv5-CA-LS (Improved version of YOLOv5 using CA mechanism and modified loss functions)	Total 814 images were used as raw data	Canon G7Xmark II	Shape, color, and spatial features	Average recognition accuracy of YOLOv5-CA-LS model: 97.1% under different angles and cloudy conditions
[100]	Sweet Pepper	Soil-based	Instance segmentation and Detection of green, red, and yellow sweet peppers	Improved Mask RCNN with Swin Transformer and UNet3+	Kaggle and Veer platforms (2286 images of green, red, and yellow sweet peppers)	Not specified	Geometric, depth, and color features	Detection AP: 98.1%, AR: 99.4%, Segmentation AP: 94.8%, F1 score: 98.8%, FPS: 5
[103]	Cucumber	Soil-based	Cucumber Detection for Mechanical Harvesting	1) I-RELIEF color component selection 2) Multi-path convolutional neural network(MPCNN) 3) SVM feature classification	Images collected in cucumber planting base in Shouguang, China (225 images taken, 218 effective samples used)	Sony IMX386 with a resolution of 4032 x 3016	Color components from 5 color spaces: RGB, HSI, YCbCr, Lab, and YIQ	More than 90% pixels of cucumber images correctly classified, misidentified pixels less than 22%

Similarly, using the PlantVillage dataset, Restrepo-Arias et al. [105] introduced a novel diagnostic approach that highlights the impact of genotypic and phenotypic characteristics on how plants respond to pathogens. Their method, which emphasizes texture-based features and uses Bayesian Optimization to train artificial neural networks, achieved an impressive accuracy of up to 96.31% with MobileNet. This method's emphasis on textural features shows promise, potentially reducing biases arising from leaf morphology. However, exploring more plant datasets and, experimenting with different image sizes, not relying solely on texture, and considering various features might enhance its classification accuracy and broaden its applicability.

Despite the widespread use of open-source datasets like PlantVillage, crop recognition systems are still in the development stage and have not been established on a large scale. As a result, most researchers prefer to experiment with their own collected image sets. Zhao et al. [106] collected images of healthy and diseased strawberry varieties to build their dataset. They introduced a modified Faster RCNN architecture, emphasizing multiscale feature fusion. Achieving a commendable mAP of 92.18%, their method stands out for its efficiency and accuracy. Still, continuous refinement of their model is required for its adaptability across a broader spectrum of strawberry diseases. Xu et al. [107] addressed the challenge of melon leaf disease detection using an innovatively pruned version of YOLO v5s combined with ShufNet v2. Their strategy achieved an impressive 95.7% mAP@0.5. By focusing on smaller disease features, they achieved real-time detection in intricate greenhouse environments. The model's speed and efficiency, with an inference time of just 13.8 ms, are noteworthy. Their work underscores the power of leveraging refined neural networks for specific tasks, though the expansion to other crops will determine its broader relevance. Zhang et al. [108] employed the EfficientNet-B4 model to identify diseases in cucumber leaves, achieving an impressive 97% accuracy. However, external factors like lighting introduced challenges in distinguishing similar diseases. While their approach offers promise for real-time greenhouse monitoring, addressing environmental variables and optimizing for different devices, remain crucial for broader applicability. Zhang et al. [109] utilized color and color-infrared (CIR) images to diagnose wheat diseases like leaf rust and tan spots. Employing deep features extracted through the ResNet101 model, their approach achieved notable accuracies up to 84%. While their approach maximizes the advantages of automated image analysis over manual observation, the deep-learning model's supremacy in capturing finer features stands out, making it a promising direction for further advancements in disease detection.

These investigations show that deep-learning models, particularly when trained on comprehensive image datasets, offer substantial promise in timely and accurately detecting plant diseases in controlled environments like greenhouses. The fusion of computer vision and advanced imaging technology has shown the potential to revolutionize traditional

agricultural monitoring practices. A detailed summary of the techniques and their respective performances in the discussed research can be found in table 5.

#### D. AUTOMATIC HARVESTING

Historically, the agricultural landscape was dominated by manual and labor-intensive processes, leading to increased costs and limitations in efficiency. The recent advent and integration of computer vision technology have marked a significant agricultural transformation. Today, the use of advanced technology in the form of intelligent harvesters equipped with vision-based robotics is on the rise. Groundbreaking research in this area has reshaped the paradigms of contemporary agricultural production, both investigating and implementing these technological advancements. For instance, Rong et al. [110]'s exploration into robotic harvesting of greenhouse tomatoes sheds light on the current technological challenges in replicating the efficiency and accuracy of manual harvesting. They proposed an advanced system to accurately identify tomato positions and determine the best grasping technique. By integrating a YOLOv5m-based detection mechanism, they achieved impressive recognition accuracies of 90.2% for tomato bunches and 97.3% for individual fruits. However, while their optimized strategies reduced collision impacts on the manipulator's grasp, yielding a promising harvesting success rate of 72.1%, their average harvesting time of 14.6 seconds per fruit still underscores the need for speed improvements.

Similarly, Benavides et al. [111] attempted to automate tomato crop harvesting using a Computer Vision System (CVS), with a primary focus on the detection and localization of ripe tomatoes. By employing a myriad of digital image processing tools and basic trigonometry, the system successfully classified around 80.8% of beef tomatoes and 87.5% of cluster tomatoes as "collectible." An outstanding achievement was the sub-millisecond processing time, a significant leap from previous methodologies. Despite the advancements, challenges persist in terms of ambient lighting conditions and the variability of the working environment, emphasizing the need for more adaptable and robust systems.

A different study by Rong et al. [112] highlights the need for a fully automated mushroom harvesting robot. Their innovative robot system, employing Intel RealSense D435i for imaging and an improved SSD algorithm for detection, offers a promising mushroom recognition success rate of 95%. The system's robustness is further evidenced by an admirable harvesting success rate of 86.8% within an average time of 8.85 seconds per mushroom. However, their initial approach to mushroom recognition faced challenges with varying illumination, adhesion intricacies, and posture identification, underscoring the need for deep learning models with more sophisticated recognition algorithms to effectively negotiate the intricacies involved in mushroom identification across diverse circumstances, hence enhancing the process. Nonetheless, the fragile nature of oyster mushrooms highlights

**TABLE 5.** Summary of greenhouse crop disease monitoring.

Reference	Plant Type	Growing Medium	Task	Method Applied	Dataset	Imaging Sensor	Feature Type	Performance
[106]	Strawberry	Soil-based	Detection of healthy strawberries and seven strawberry diseases under natural conditions	Faster R_CNN with multiscale feature fusion network (ResNet, FPN, and CBAM blocks)	Self-collected and from Jeonbuk National University shared open-source dataset, resulting in a total of 3600 images	Sony digital camera	Spatial and color features	mAP of 92.18% with an average detection time of 229 ms
[108]	Cucumber	Soil-based	Identification and classification of cucumber leaf diseases	EfficientNet-B4 improved with Ranger optimizer	Dataset consist of four classes where total images was 2816	Digital camera	Color, texture, and other morphological features	Model accuracy with EfficientNet-B4: 97%
[109]	Wheat	Pots (Deepot D40: 6.4 × 25.4 cm)	Diagnosis of leaf rust, tan spot, and leaf rust + tan spot diseases	Image processing, handcrafted features (HF) extraction, Custom CNN, feature fusion, SVM classifier	In total 160 pots of plants images	Canon EOS Rebel T7i (color images) and multi-band camera (CIR images)	Color and texture	Accuracy of 75%, 84%, 71% for leaf rust, tan spot, and leaf rust + tan spot, respectively
[152]	Tomato and Corn	Not Specified	Disease Prediction	1) SSD Mobile Net 2) EfficientDet D0 3) Faster RCNN	Dataset contains 1000 images, where 80% used as training set and rest 20% as test set	Not Specified	Color and texture	Faster_RCNN had best performance with 80.4% accuracy

TABLE 5. (Continued.) Summary of greenhouse crop disease monitoring.

Reference	Plant Type	Growing Medium	Task	Method Applied	Dataset	Imaging Sensor	Feature Type	Performance
[104]	Tomato	Not Specified	Recognize diseases, detect unseen diseases, estimate per-leaf severity.	ResNet-50	PlantVillage Dataset. (tomato crops) The subset of tomato images contains 16415 diseased leaf images exhibiting nine different conditions and 1590 healthy leaves	20-megapixel camera	Color, leaf surface, shape, and background	Classification accuracy of four variants: Color 0.970 Color-masked 0.940 Background-only 0.635 Silhouette 0.523
[105]	Tomato	Not Specified	Disease diagnosis based on image texture	Small CNNs with the Bayesian Optimization technique	PlantVillage dataset (tomato crops) 26,148 images after balance	20-megapixel camera	Color and texture features	Accuracy: MobileNet: 96.31%, SqueezeNet: 95.05%
[107]	Melon (Cucumis melon. L.)	Soil-based	Detection of leaf disease (powdery mildew and real leaves)	Pruned-YOLO v5s+Shufe (PYSS)	Dataset contains 2162 images in total	Vivo IQOO Neo5 mobile phone	Color, textures, and spots	mAP@0.5 for melon powdery mildew: 93.2%, and for melon real leaves: 98.2%

the critical need for advancements in end-effector designs, ensuring minimal damage during the harvesting process.

Recent innovations by Liu et al. [113] explored the application of the DA-Mask RCNN model for detecting green asparagus. The aim was to enhance detection precision during the autonomous harvesting of green asparagus by integrating MASK RCNN with depth information. The addition of the depth filter showed significant promise, especially under varying illumination conditions, achieving precision values of up to 0.993. While the model demonstrated resilience against false positives in bright lighting, further optimization, particularly for nighttime scenarios, is crucial for broader applications.

These studies underscore the transformative potential of combining computer vision with deep-learning models in reshaping harvesting practices. While significant advancements have been achieved, the interplay between accuracy and operational efficiency emerges as a central concern. Table 6 summarizes the technical details of the studies discussed in this subsection.

#### E. YIELD ESTIMATION

Yield estimation in agricultural production is pivotal for stakeholders across the spectrum—from farmers to agricultural enterprises. It aids in strategically navigating post-harvest operations, driving marketing initiatives, and optimizing resource allocation. While traditional techniques have dominated this space, there is an emergent reliance on cutting-edge computational methodologies. The intersection of computer vision and agriculture has seen innovative developments, particularly with the employment of deep learning models to address challenges like overlapping crops, dense vegetation, and varying light conditions specially in greenhouse setup. This synthesis of technology and agriculture brings new methodologies, delivering improved precision in predictions, even as complexities arise. Wang et al. [114] employed an improved version of the YOLOv3 deep learning model to estimate tomato yields in artificially lit plant factories (PFAL). By refining the traditional YOLO algorithm, they achieved a notable mean average precision (mAP) of 99.3%—a 2.7% improvement over the original YOLOv3. Notably, their approach excelled in distinguishing densely packed and obscured fruits, paving the way for real-time crop monitoring and dynamic yield estimation. However, challenges persist due to the complex lighting conditions in PFAL environment and similarities between green fruit and their surrounding vegetation. In another study, Maji et al. [115] investigated wheat yield estimation through SlynNet, a hybrid deep learning approach that combines Mask R-CNN and U-Net. This approach effectively captures wheat morphological features, attaining a high mAP of 97.57% in spike detection. The study underscored the technique's resilience to natural field constraints like overlapping and varying resolution. Yet, while SlynNet proved excellence for its detection capabilities, it acknowledges the challenge in

precisely estimating grain yield from spikelet counts. Further studies focused on investigating the intricate anatomical details of spikelets could potentially enhance the precision of yield prediction. The growth and development of generative organs in greenhouse plants are essential for both yield estimation and higher productivity. Given the challenges in the greenhouse environment, such as leaf and branch obstructions and the risk of duplicate counts, there is a pressing need for more efficient methods. Manual counting approaches are time-consuming and often marred by inaccuracies, highlighting the necessity for rapid and automated solutions. Egi et al. [116] addressed this need by innovatively incorporating a drone-based AI system to detect and count greenhouse tomatoes. Their emphasis was not limited to the fruits; they also targeted the flowers. Using the YOLO V5 and Deep Sort algorithms, their method showcased remarkable accuracies of 99% for green tomatoes and 85% for red tomatoes. Nonetheless, their approach faced challenges with flower detection, securing only a 50% accuracy. While their achievements are commendable, the potential inaccuracies from drone movements, combined with the shortcomings in flower detection, hint at areas for improvement. An expanded dataset could offer a potential solution. Zhou et al. [117] utilized an “Improved ResNet” deep learning model to segment accurately and grade broccoli heads in greenhouse conditions. Their method achieved an impressive accuracy of 0.896 for broccoli head segmentation, even in varied lighting conditions. While the model's performance is notable, its dependency on controlled environments and data processing pipeline reliance on manual settings poses challenges for real-world applications. However, the potential integration of semi-supervised learning in future work indicates promising strides towards refining and optimizing crop yield estimation in greenhouse farming.

From the discussed studies centered on yield estimation in agricultural settings, particularly in greenhouses, it becomes evident that the scientific community is making strides in leveraging advanced computational techniques. These pioneering methodologies, while exhibiting remarkable promise, also underscore the complexities of the greenhouse environment. Further investigation and advancement in these methods could potentially bring about a paradigm shift in the agricultural sector—including farmers and enterprises—strategize planting, harvesting, and marketing initiatives. Table 7 provides more details about the complex technical aspects of the studies discussed here.

#### F. CROP HEALTH ANALYSIS

In the realm of greenhouse plant health analysis, advanced computer vision techniques have provided innovative solutions for the prompt detection of crop biotic and abiotic stresses, nutrient deficiencies, and water stress, ensuring superior plant quality. The ability to promptly identify biotic and abiotic stresses is imperative for effective greenhouse crop management and optimal plant health. While visual inspection is tedious and subjective, computer vision and spectral

TABLE 6. Summary of greenhouse automatic harvesting task.

Reference	Plant Type	Growing Medium	Task	Method Applied	Dataset	Imaging Sensor	Feature Type	Performance
[111]	Tomato (Beef and Cluster types)	Mediterranean greenhouse (Parral-type)	Automate the detection of ripe Tomatoes and localization of Tomato peduncles for robotic harvesting.	Digital image processing tools, Trigonometry, and numerical geometrical descriptors	175 images in total	A Genie Nano with an OPT-M26014MCN optics	Color and shape	Beef Tomatoes: Accuracy: 80.8%, Processing Time: ~27ms; Cluster Tomatoes: Accuracy: 87.5%, Processing Time: ~29ms
[113]	Green Asparagus	Soil-based	Detection of Green Asparagus and 3D reconstruction for harvesting	DA-Mask RCNN model	13,500 images under different weather condition	Tuyang FM810 camera	RGB and Depth Info	Precision: 0.993, Recall: 0.971, F1-Score: 0.982; Accuracy of 3D reconstruction was 99.72%
[110]	Tomato	Nijawan Water Field Greenhouse, Xiangcheng District, Suzhou, China	Robotic Harvesting system (Identifying the position of the tomato fruit and estimating the grasping pose)	YOLOv5m, optimal sorting algorithm and fruit nearest neighbour positioning algorithm	Tomato fruit dataset with 3369 images and tomato cluster dataset with 358 images	Intel Realsense D435 RGB-D camera	Color and shape	Fruit bunch recognition accuracy: 90.2%; Fruit recognition accuracy: 97.3%; Harvesting success rate: 72.1%; Average harvesting time: 14.6 s for a single fruit
[112]	Oyster Mushroom	Soil-based (Cuboid bags being used for cultivating oyster mushrooms)	Design and development of a Mushroom harvesting robot (Mushroom recognition)	Improved SSD	4600 labeled images were collected for building dataset	Intel RealSense D435i RGB-D camera	RGB Point cloud	Mushroom recognition: 95%, Harvesting: 86.8%, Single harvest time: 8.85s

imaging offer great promise for automated, non-invasive assessment of crop condition. As Taha et al. [118] employed spectral analysis and machine learning models to estimate nutrient contents in aquaponically grown lettuce. Using the selected optimal wavelengths, they achieved commendable predictive accuracies ( $R^2 \geq 0.94$ ), suggesting a potential automated solution for nutrient estimation in aquaponics. However, the real challenge lies in translating these lab-based experiments to real-world greenhouses where external factors can significantly impact spectral readings. Eshkabilov et al. [119] also employed hyperspectral imaging within the 400-1000 nm range to predict nutrient concentrations in lettuce cultivars using PLSR and PCA models, achieving high accuracy ( $R^2 = 0.784-0.987$ ) with laboratory measurements. However, while hyperspectral imaging holds promise, its implementation might be restrictive due to cost implications, especially for small-scale greenhouse farmers.

RGB imaging has also been utilized for abiotic stress detection. Lak et al. [120] developed a classification algorithm using visible light images and MLP neural network modeling to detect water stress in greenhouse tomatoes. After optimizing with PCA, the technique achieved 83.3% accuracy in distinguishing normal and water-stressed plants using only RGB image features. Surprisingly, adding thermal imagery did not improve results. Similarly, Levanon et al. [121] utilized RGB and thermal imaging along with neural networks to predict water and nutrient stress in banana plantlets. The multi-modal data fusion approach enabled models to achieve high prediction accuracy of over 90% for four stress classes. However, the small sample size of 16 plants may limit generalization.

Beyond abiotic factors, computer vision shows promise for biotic disease screening. Najafian et al. [122] introduced a large dataset of over 40,000 wheat kernel images to detect Fusarium damage using deep learning models like EfficientNet and ResNet. The semi-supervised approach reached F1 scores up to 84.29% for binary classification but was limited for multi-class tasks. Overall, the dataset provides a valuable benchmark, but more samples are needed. Janani et al. [123] proposed a unique method to identify nitrogen nutrient levels in groundnut leaves using the CNN-based HVN model, achieving a training accuracy of 95% and validation accuracy of 92%. Although their method highlighted the significance of leaf color in determining nitrogen concentration, they acknowledged potential inaccuracies stemming from external factors unrelated to nitrogen levels. Such nuances emphasize the importance of considering all potential variables in crop health prediction.

Overall, these studies demonstrate that computer vision and spectroscopy can enable automated, non-destructive monitoring of multiple crop health indicators. However, translating these technologies into commercial systems will require more robust models that fuse multi-modal data to overcome environmental noise, larger datasets to improve generalizability, and model optimization to account for subtleties between stress symptoms and normal variations.

Table 8 provides more details about the complex technical aspects of the studies discussed here.

### G. PEST AND INSECT MONITORING

Monitoring pests and insects in greenhouses has always been a primary concern for sustainable agriculture, impacting both the quality and quantity of yield. It has now become an exciting field in the agricultural domain, emphasizing the critical need for precise, real-time solutions. Traditional methods have shown limitations in accuracy and response time. Consequently, computer vision and deep learning, particularly convolutional neural networks (CNNs), are being increasingly explored to identify pests like whiteflies and thrips, aiming to streamline and improve the detection process in greenhouse environments. Traditionally, farmers manually sample, count, and identify pests, a time-consuming and error-prone process. The integration of computer vision presents a more effective and accurate alternative, and a wide range of research is exploring its possibilities in the context of integrated pest management (IPM) [124], [125], [126], [127], [128], [129], [130], [131]. In a recent study, Liu et al. [129] investigated real-time pest detection on crops using advanced computer vision and deep learning. By adopting convolutional neural networks (CNNs) and generating a virtual database for training, they achieved an impressive 97.8% detection accuracy for various invertebrate pests on crops. In the future, utilizing multispectral or hyperspectral imaging may make it possible to identify well-camouflaged pests. Furthermore, the development of ground-based robotic system capable of performing real-time proximal detection of invertebrate pests could be the future of pest management. Yang et al. [128] focused their research on identifying greenhouse pests, especially whiteflies and thrips, using image processing. Their novel method employed dual color spaces (HIS and Lab) combined with advanced ensemble learning classifiers, resulting in a commendable 95.73% recognition accuracy. While this approach substantially minimized manual intervention, the study underscores the challenge of uniformly illuminating complex greenhouse environments, possibly impacting accuracy in varied scenarios. Despite this success, there is still room for improvement. The system could be more optimized by reducing the number of false positive detections. Moreover, extending the system's capability to a broader range of environments beyond greenhouses is on the horizon. On a similar note, Li et al. [131] also conducted research on detecting whitefly and thrips from sticky trap images using a deep learning model, 'TPest-RCNN', based on the Faster R-CNN. This model, fine-tuned for small insect detection, showed excellent precision with a mAP of 0.95. While the results were promising, challenges like diverse pest densities and lighting conditions remain. However, their approach sets a robust groundwork for real-time monitoring, which can significantly aid timely interventions. In a distinctive approach, Zhao et al. [130] tackled pest detection in *Brassica chinensis* through images

TABLE 7. Summary of greenhouse yield estimation.

Reference	Plant Type	Growing Medium	Task	Method Applied	Dataset	Imaging Sensor	Feature Type	Performance
[115]	Wheat	Soil (Pots filled with 12.5 Kg uniformly mixed soil)	Spike and skeleton detection for high-throughput yield estimation	SlypNet (utilizing Mask R-CNN and U-Net)	3600 wheat plants image	Prosilica GT 6600 (6576x4384 pixels)	Spike and skeleton morphological features	Mask R-CNN for spike detection: mAP of 97.57%, F1 of 0.67, MCC of 0.91; Spikelet detection accuracy ~99%
[114]	Dwarf Micro-Tom tomato	Soilless (substrate potted and artificial-lighting-hydroponic)	Intelligent recognition and yield estimation of tomato	Improved YOLOv3 deep learning model with generalized intersection over union (GIoU)	3,680 original images and 18,400 enhanced images	Intel Real sense D455 RGB-D camera and iDS-TCV441-CF industrial camera	Multi-scale features, color, and shape	Mean Average Precision (mAP) of improved model: 99.3%. Processing time for a single image: 15ms
[116]	Tomato	Soil-based	Detect and count tomato flowers and fruits for yield estimation	YOLO V5 & Deep Sort deep learning algorithms	Dataset contains 2923 images and 6957 annotations (3 classes: red tomato, green tomato, flower)	DJI Spark drone (12 MP camera)	RGB color feature, shape and texture	Best model: accuracy of 0.618 at mAP 0.5; Average F1 score: 0.63
[117]	Broccoli	Soil-based	Segmentation and Grading for yield estimation	Improved ResNet, PSOA, and Otsu method	Private image dataset with 506 images (Three classes: Zheqing95, Tailv1, and Tailv2)	Canon EOS 90D camera with 32.5 megapixels resolution and a 22.3 mm x 14.8 mm CMOS sensor	Color and shape	Accuracy of 0.896 on test set for broccoli head segmentation; Grading accuracy for Zheqing95, Tailv1, and Tailv2 were 0.879, 0.853, and 0.841 respectively; Average running time of 0.18 s for segmenting an image



**TABLE 8.** Summary of greenhouse crop health analysis.

Reference	Plant Type	Growing Medium	Task	Method Applied	Dataset	Imaging Sensor	Feature Type	Performance
[119]	Four lettuce cultivars	Rockwool cubes with hydroponic nutrient solution	Rapid detection of nutrient content of hydroponically grown lettuce cultivars	Partial squares regression (PLSR) and principal component analysis (PCA)	The size of a hyperspectral image of one leaf was in the range of [1900x1024] - [2200x1024] pixels on 448 bands.	Specim FX-10 Hyperspectral Camera (wavelength range of 400–1,000 nm)	HIS color features	Both PLSR and PCA models produced well correlated outcome in comparison with the laboratory measured freshly cut lettuce leaves in predicting nutrient contents with $R^2 = 0.784-0.987$
[123]	Groundnut	Soil-based	Detection and classification of nitrogen nutrient level	Median filter and CNN-based HVN model	600 leaves	Not Specified	Texture and spatial features	Developed CNN-based HVN model achieved 95% training accuracy and 92% validation accuracy
[121]	Banana plantlets	Soil-based	Abiotic stress prediction	Neural networks (Resnet-50, GoogLeNet, MobileNet, VGG16, etc.)	1020 RGB and Thermal images	RGB: Samsung SM-G930F, 4032x3024 Thermal: Opgal-ThermApp, 384 x 288	Not-Specified	Neural networks exhibited prediction rates over 90% amongst the four classes.
[118]	Romaine lettuce (var. longifolia)	Aquaponics (Various Nutrient Treatments)	Estimation of nutrient contents in aquaponically grown lettuce	PLSR, BPNN, and RF	128 lettuce seedlings with four NPK treatments	FieldSpec4, Pro FR portable spectroradiometer	Leaf reflectance spectra (optimal wavebands)	BPNN obtained highest predictive accuracy $R^2-p = 0.97$
[120]	Tomato	Loamy sand in greenhouse	Detection of water stress	PCA and MLPNN	Not Specified	Thermal and Visible (RGB) via infrared camera (ITI-P400)	Thermal and visible features	83.3% accuracy

captured by unmanned aerial vehicles (UAV). Given the challenges of aerial imagery such as image blur and small object sizes, their strategy leveraged deep learning combined with improved CenterNet algorithms, realizing a detection performance of up to 94.7% R-squared. Despite its potential, aerial imagery's inherent issues, like blurring, could affect its widespread adaptability unless integrated with superior quality cameras or advanced image restoration techniques. In a different study, Lins et al. [127] addressed the task of automating the counting and classification of aphids, particularly *Rhopalosiphum padi*, using a software named 'AphidCV'. This software, rooted in computer vision and machine learning, not only expedited the counting process but also introduced morphometry data. While its prowess is undoubted, some limitations in classifying winged aphids were evident. Diversifying the training dataset, perhaps with data augmentation techniques, could be beneficial. This solution holds promise for its extension to other aphid species and direct application in field-based pest management, reducing reliance on labor-intensive manual methods.

Given the advancements detailed above, computer vision, particularly when coupled with deep learning methodologies like CNNs, showcases transformative potential in the realm of greenhouse pest detection and management. Nevertheless, as research progresses, it is imperative to address inherent challenges, such as diverse pest densities and varied lighting conditions, to ensure consistent, real-time accuracy. Embracing multi-modal data fusion, hyperspectral imaging, and more extensive datasets can further amplify the precision and reliability of these systems within greenhouse environments. Table 9 details the technical aspects of the discussed studies.

#### H. GREENHOUSE CROP QUALITY INSPECTION

The quality control and grading of agricultural commodities is essential for determining their market value, safety, and appeal to consumers [132]. Manual inspection, however, can be inconsistent, labor-intensive, and unscalable for large-scale production. The integration of computer vision and artificial intelligence has unlocked exciting new possibilities for automated, non-destructive, and real-time assessment of horticultural produce quality [133], [134], [135], [136], [137]. As explored by Tan et al. [134], deep convolutional neural networks demonstrated high accuracy (mean average precision of 95.52%) in classifying the maturity of tomatoes based on color features extracted from images. The use of task-specific architectures like Mask-RCNN allowed precise segmentation and localization of the produce within complex backgrounds. However, the model training time of 6 hours indicates room for optimization in computational efficiency for real-time adoption.

Beyond maturity, quality parameters like pest/disease damage, ripeness, and shelf-life have also been automated using imaging techniques. For instance, Hendrawan et al. [135] developed a convolutional neural network model to categorize large green chili peppers into three maturity classes

with 91.27% accuracy. The model shows promise for rapid, objective maturity grading to standardize quality. However, classification accuracy was lower for immature peppers, warranting further research into data augmentation and transfer learning to improve model robustness. In another study, Shi et al. [136] combined deep learning and causal analysis to predict maturity dates of leafy greens in greenhouses, achieving a root mean squared error of only 2.49 days. While novel, the approach struggled with crops in late static growth stages, suggesting the need for adaptive models that emphasize historical data over static phenotypes in late stages.

Wei et al. [137] developed a model using grape skin color analysis and a backpropagation neural network to predict the maturity of greenhouse-grown grapes, achieving up to 79.4% accuracy. A two-factor color model performed better than single-color predictors. However, prediction accuracy varied between grape varieties depending on color changes during ripening. Custom varietal models or adaptive techniques may further improve prediction.

Zhu et al. [133] proposed a computer vision approach using YOLOv5 and OpenCV to grade mushrooms in greenhouses based on size features automatically. The model achieved 96% accuracy in identifying and measuring occluded mushrooms under varying illumination. However, further work is needed to optimize mushroom recognition speed and expand functionality for tasks like robotic spraying. Nyalala et al. [138] presented a technique using machine learning and image processing to estimate the weight and volume of tomatoes on a simulated greenhouse conveyor system. Occluded tomatoes were segmented using polygon approximation before extraction of shape features. The best models achieved high correlation with reference measurements, demonstrating feasibility for in-line, non-destructive quality screening. Nonetheless, additional testing is required across diverse tomato varieties and shape features.

The studies presented highlight the immense potential of computer vision and AI techniques to automate the quality assessment of greenhouse-grown horticulture produce. Advanced models have enabled non-destructive evaluation of multiple quality traits, from external features like pest/disease damage and maturity to internal parameters like texture and shelf-life. However, enhancements in model versatility, accuracy, and computational efficiency are needed to account for the diversity of greenhouse varieties and environments. Current systems also have limited real-world testing beyond controlled settings. Expanding datasets, integrating multi-modal sensor inputs, and optimizing deep learning architectures tailored for greenhouse conditions will be critical next steps. While research is still progressing, computer vision and AI solutions promise to transform quality control practices for greenhouse horticulture. Automated, real-time quality grading and defect detection could provide invaluable objective data to support selective harvesting, packing, pricing, and sales for greenhouse produce. Overall, these emerging technologies are poised to enhance productivity, reduce waste, and add value to the competitive greenhouse industry. Table 10

TABLE 9. Summary of greenhouse pest and insect management.

Reference	Plant Type	Growing Medium	Task	Method Applied	Dataset	Imaging Sensor	Feature Type	Performance
[131]	Green pepper plants	Polyethylene film covered greenhouse	Detection of tiny pest (whitefly and thrips)	TPest-RCNN (based on Faster R-CNN)	84 images from 2 weeks for dataset 1 and 42 images from another week for dataset 2	Image-acquisition device made by the National Engineering Research Center for Information Technology in Agriculture	Handcrafted features (color, shape, texture)	Mean F1 score: 0.944, mAP: 0.952
[130]	Brassica chinensis	Greenhouse	Pest disease detection of Brassica chinensis	Improved Center-Net with attention mechanism and DIOU loss	Aerial images of Brassica chinensis bitten by flea beetle: 42,452 wormhole annotations	SG906 MAX1 UAV with RGB imaging system; Resolution: 4096 x 3072	Wormhole morphology (texture and shape)	Detection accuracy: AP50: 87.2%, R-squared: 94.7%
[128]	Cucumber	Cucumber greenhouse at the National Experiment Station for Precision Agriculture in Beijing	Recognition and counting of greenhouse pests	Segmentation in HSI & Lab color spaces using Pre-witt & Canny operator. Ensemble learning for recognition.	Images from National Experiment Station for Precision Agriculture in Beijing	T500 industrial mobile phone camera	Color (HSI and RGB) and shape features.	Segmentation accuracy: 91.57%, Recognition accuracy: 95.73%
[129]	Multiple crops and fruit trees	Controlled field environments	Detection of invertebrate pests on crops	Different CNN architecture along with ResNet152-V2	Total of 10C-database:6073 and 5C-database:29507	Smartphones	Morphological features, color variations, textural features	ResNet152-V2 achieved an accuracy of 97.8%

presents the technical details of several studies conducted in this area.

### I. SEED QUALITY ANALYSIS

Seeds are crucial to modern agriculture, determining both food supply and crop yield [118]. Traditional manual assessments of seed quality, while essential, are laborious and prone to inaccuracies. The commercial seed industry is now leaning into computer vision technology, tapping into its potential to extract seed features with precision. With the rapid advancement of various imaging techniques, enhanced by deep learning, this technology sets new benchmarks for seed quality evaluation in greenhouse settings [140], [141], [142], [143], [144], [145]. For example, Medeiros et al. [141] investigated the application of Convolutional Neural Networks (CNN) with X-ray imagery to determine crambe seed quality. Their deep learning models robustly categorized seeds based on their tissue integrity, germination, and vigor with impressive accuracies of 91%, 95%, and 82%, respectively. This substantiates the vast potential of X-ray imagery in furnishing critical insights into the physical and physiological attributes of seeds. However, a lingering concern is the reliance on digital radiographic images, which, while powerful, might not capture the complete essence of seed vitality in diverse scenarios. On a similar note, Hong et al. [142] employed a combination of hyperspectral and X-ray imaging techniques for a nondestructive viability prediction of pepper seeds. The ensemble-based fusion model, integrating both hyperspectral and X-ray data, stood out with an accuracy of 92.51%. This approach demonstrates that combining different imaging modalities can produce more accurate categorization results. Nonetheless, the research's dependency on just two pepper-seed cultivars may demand broader experimental validations for holistic reliability across varied conditions. In a different study, Lube et al. [143] introduced the MultipleXLab system: a flexible platform for monitoring seed germination and root growth. Using deep learning methodologies, they showcased the system's capability to screen seed vigor and evaluate seedling responses under varied conditions. The system, however, presents two key challenges: its restricted experimental duration due to agar dehydration and a potential light deprivation issue for seeds placed in specific positions. Overcoming these limitations could elevate its utility. Gao et al. [144] presented an innovative end-to-end platform named HyperSeed, which is adept at providing hyperspectral information specifically for seeds. Their application on rice seeds using a 3D convolutional neural network (3D CNN) outperformed traditional methods like the support vector machine (SVM) model, boasting an impressive 97.5% accuracy. Nevertheless, the system is constrained by its single-threaded software nature and demands exploration of global spatial traits, hinting towards potential future advancements. On the other hand, Sabanci et al. [140] conducted a study to distinguish between tomato seed cultivars, employing a multi-tier deep learning approach. They

initially utilized convolutional neural network (CNN) models for seed image classification, with MobileNetv2 showcasing the highest efficacy. Furthermore, they leveraged deep features from this model to feed a Bidirectional Long Short-Term Memory (BiLSTM) network, pushing the classification accuracy to a notable 96.09%. Despite the promising results, the study underscores the importance of diverse datasets for model robustness and suggests hyperparameter optimization to enhance performance.

To synthesize, while computer vision technologies are profoundly redefining seed quality assessments in greenhouse farming, there remains an imperative for further refinements to harness their full potential. Such tools, when impeccably optimized, can provide real-time and nondestructive solutions to the complexities of seed quality evaluation in the modern agricultural landscape. Table 11 provides a concise summary of several studies conducted in this field.

### J. WEED MANAGEMENT

Weed management poses a significant challenge in contemporary agriculture, as weeds compete with crops for vital resources such as light, water, nutrients, and space. Weeds are a primary factor behind agricultural yield losses. Study [146] highlighted that weeds account for about 34% of yield losses, substantially more than losses from pests (18%) or pathogens (16%). Recently, the agriculture sector has seen a growing interest in integrating modern weed management techniques with computer vision. Extensive research has been carried out using these cutting-edge technologies to control and manage weed growth in greenhouses. In one study, Koparan et al. [147] emphasized the role of image background in deep learning models for weed detection. The study applied advanced architectures like VGG16 and ResNet50, highlighting a decrease in model accuracy when transitioning from a uniform to a non-uniform background and vice versa. However, when combined data sets from both backgrounds were used, the performance surged to nearly 99%. Despite their rigorous approach, the limitations lie in the model's dependency on the image backgrounds, underscoring the need for more diverse training data. This could enhance the model's adaptability across varying environmental conditions. In another study, Wang et al. [148] underscored the challenges posed by the limited availability of weed datasets in the field. The study introduced "Weed25", a dataset encapsulating images of 25 weed species and utilized state-of-the-art models like YOLOv3, YOLOv5, and Faster R-CNN to achieve impressive accuracy rates hovering around 92%. However, the research confines its scope to only 25 species, leaving room for the inclusion of more diverse weed types, which could lead to further refinement of models tailored for precision in greenhouse weed management. Oda et al. [149] developed a multispectral camera system specifically for recognizing weeds within crops. Their findings revealed that the infrared band was more precise than other bands, highlighting its importance in plant detection. This affirms

TABLE 10. Summary of greenhouse crop quality inspection.

Reference	Plant Type	Growing Medium	Task	Method Applied	Dataset	Imaging Sensor	Feature Type	Performance
[153]	Little Gem Romaine Lettuce	Aquaponic grow beds	Monitor growth rate and estimate fresh weight	MASK-RCNN, and regression analysis for image segmentation	1350 images from various sources	2 ELP 1080P Webcam (2.8–12 mm HD Varifocal Lens)	height, depth, width, area	Size error: 30 mm (18.7%), Weight error: 0.5 g (8.3%)
[137]	Grapes (Drunk Incense, Muscat Hamburg, Xiang Yue)	Soil-based (clay loam soil)	Prediction of grape maturity	Back-propagation neural network (BPNN)	RGB images of grapes	Not Specified	RGB, HSI values from grape skin	Prediction accuracy: Drunk Incense (78.2%), Muscat Hamburg (77.0%), Xiang Yue (78.0%)
[136]	Leafy greens	Soil-based	Maturity prediction	Deep convolutional neural network combined with a crop growth model based on causal inference	362 images (3:1:1 ratio for training, verification, and testing)	Not Specified	Color and texture	Classification accuracy of SE-ResNet34 (+cutout) = 0.75, RMSE = 2.49 days
[134]	Tomato	Soil-based	Tomato grading based on maturity color	Computer Vision (CV) and DNN using Mask-RCNN Inception Resnet-V2	2,100 images (Unripe, Partially Ripe, Totally Ripe)	Logitech Webcam C922 Pro Stream	RGB color	mAP: 95.52%

TABLE 11. Summary of greenhouse seed quality analysis.

Reference	Plant Type	Growing Medium	Task	Method Applied	Dataset	Imaging Sensor	Feature Type	Performance
[143]	Arabidopsis and Tomato	Half Murashige and Skoog (MS) plates, and nylon mesh along with soil and polystyrene plates	Phenotyping of seed germination and early root growth	SeedNet and RootNet based on the U-Net architecture	A total of 3200 annotated plants	Canon 5DSr SLR, Canon MP-E 65 mm f/2.8 1-5x lens, Canon EF 40 mm f/2.8 STM lens	Structural and morphological features	F1 score for SeedNet: 0.8048; RootNet: 0.7395
[142]	Pepper	Soil-based	Prediction of pepper seed Viability	Hyperspectral and X-ray imaging techniques along with CNN	X-ray images of 1152 pepper seeds	Hyperspectral camera (HSI-200), & X-ray imaging device (X-ray SF160NCT)	Spectral features	Accuracies: Hyperspectral model: 88.99%, X-ray model: 75.33%, Fusion model: 92.51%
[141]	Crambe abyssinica	blotting paper moistened with distilled water	Seed quality Classification based on X-ray images	Custom CNN model with hyperparameter tuning	Radiographic images of 600 seeds labeled, based on their physical integrity, viability and vigor	Faxitron MX-20 device	Morphological features from X-ray imaging	Classification accuracy: 91% (integrity), 95% (germination), 82% (vigor)
[144]	Rice	Pasteurized soil (with some germination process)	Hyperspectral analysis and seed classification	3D convolutional neural network (3D CNN) and SVM	Pixel-based: 274,641; Seed-based: 200	High-performance line-scan image spectrograph	hyperspectral reflectance	Best accuracy: 3D CNN: 97.5% (seed-level) and 94.21% (pixel-level)

the potential of combining computer vision with multispectral imaging in enhancing post-emergence herbicide applications. Nonetheless, challenges persist, including the influence of leaf overlap and varied light intensities on detection accuracy. Addressing these challenges head-on could pave the way for more robust and accurate weed detection models. Koparan et al. [150] examined the role of site-specific weed management using RGB image texture features. Their methodology compared the Support Vector Machine (SVM) and the deep learning-based VGG16 models. Remarkably, their deep learning approach, specifically the VGG16 model, showcased an impressive f1-score value of 100% for corn classification, which is a breakthrough for the corn crop production system. While their method holds potential, one must consider the vast variability of crops and regions. A significant highlight of their work is the effectiveness of VGG16 in identifying weeds in the presence of various crops, elucidating the intricacies of weed-crop dynamics in precision agriculture. In a different study, Rai et al. [151] deployed deep learning models on edge devices, aiming to detect weeds through aerial imagery. By comparing both heavyweight and lightweight deep learning models, they showed that lightweight models, specifically CSPMobileNet-v2 and YOLOv4-lite, achieved a mean average precision(mAP) of 83.2% and 82.2%, respectively. Their approach offers real-time detection with commendable accuracy but, critically, must be scrutinized for potential hardware biases and storage challenges, especially with high-resolution aerial data. The flexibility offered by edge computing cannot be ignored; however, securing these devices remains a pertinent concern. Drawing insights from those studies, it is evident that integrating computer vision techniques holds immense potential for refining greenhouse weed management. However, achieving precision remains a formidable challenge. Diverse data sources, enhanced models, and advanced imaging techniques may collectively steer the future of this realm, allowing for more sustainable and efficient weed control in greenhouses. A brief overview of different studies in this area, along with their technical details, can be found in Table 12.

## V. CRITICAL CHALLENGES FACED BY COMPUTER VISION TECHNOLOGY IN SMART GREENHOUSE OPERATIONS AND THE WAY OUT

While computer vision has shown immense potential across various greenhouse applications, as highlighted in this review, significant challenges must be addressed for effective real-world deployment.

### A. LACK OF LARGE-SCALE STANDARDIZED DATASETS

A major limitation identified is the shortage of large-scale standardized image datasets for the greenhouse domain (as noted in [66], [102], [105]). Most studies rely on small proprietary datasets collected by the researchers themselves, often just a few hundred images, which restricts generalization of techniques ([93], [98], [107]). There is a need to establish

extensive public databases encapsulating the diversity of greenhouse environments, with variability in factors like lighting, humidity, crop types, growth stages, and imaging angles represented ([92], [95], [106]). Centralized repositories like PlantVillage offer a valuable start but have limited coverage and annotation complexity. Constructing large-scale greenhouse image datasets with standard formatting and annotations will be critical to benchmark performance of computer vision techniques and fuel advancements ([94], [114], [115]).

### B. NEED FOR SPECIALIZED AND OPTIMIZED MODELS

While deep neural networks have driven progress in many visual recognition tasks, off-the-shelf models pre-trained on open-field images fail to account for the unique intricacies of greenhouse environments ([96], [97], [105]). Networks designed for generic datasets do not transfer effectively to greenhouses, as the specialized appearance and growth patterns of crops in controlled settings are not represented ([99], [108], [109]). The frequent occlusion from dense foliage, glass surface reflections, condensation, extreme illumination fluctuations, and evolution of visual features across plant growth requires specialized model architectures and training strategies ([95], [100], [114]). For instance, Li et al. [97] found that directly applying YOLOv3 for cucumber canopy recognition resulted in insufficient accuracy, requiring an enhanced model with additional coordinate attention modules. Xu et al. [107] showed that extensive pruning and adaptation of YOLOv5s was needed for effective real-time melon disease detection in greenhouses. Rong et al. [110] demonstrated that off-the-shelf YOLOv5 had difficulties recognizing occluded tomato bunches, needing optimization of the loss function and model architecture. These studies underscore the need to tailor base models to address nuances like occlusion, illumination variation, and growth patterns unique to greenhouse environments. Beyond architectures, the training process must also account for greenhouse-specific factors. Models like those proposed by Cong et al. [100], Gang et al. [93], and Liu et al. [113] illustrate the benefits of customized training regimes using specialized greenhouse datasets over pre-trained networks. However, collecting exhaustive labelled greenhouse data can be challenging. Recent work by Moysiadis et al. [95] demonstrates the potential of simulated or synthetically augmented data for greenhouse model training. Such emerging data-centric solutions must be explored in combination with adapted model architectures.

Overall, while deep learning has shown immense potential, realizing robust computer vision for greenhouses will require specialized model architectures and training techniques tailored to the unique intricacies of controlled environments ([92], [94], [98]). Greenhouse-specific networks, re-training regimes, augmented data, and continual optimization will be instrumental in developing computer vision solutions adept at the nuances of greenhouse conditions.

TABLE 12. Summary of greenhouse weed management.

Reference	Plant Type	Growing Medium	Task	Method Applied	Dataset	Imaging Sensor	Feature Type	Performance
[150]	Four weeds types and six crops type	Waldron green-house at North Dakota State University	Classification of weeds and crops	SVM with ReliefF, VGG16 with transfer learning	3792 RGB images (2271 weeds, 1521 crops) from greenhouse	Google Pixel 5 (12.2MP main & 16MP ultra-wide camera)	RGB texture features, LBP & GLCM	VGG16 f1-scores: 93%-97.5%, And for Corn class: 100%
[151]	Weeds (specifically kochia, ragweed, horse-weed)	Outdoor (Green-house at NDSU-main campus)	Weed detection in aerial-image context and videos	Eight DL models (CSPDarkNet-53, DarkNet-53, DenseNet-201, ResNet-50, CSPMobileNet-v2, YOLOv4-lite, EfficientNet-B0, DarkNet-Ref)	SpotWeeds dataset Total images used for training and validation: 1,096	DJI Phantom 4 Pro (RGB camera)	Color and texture features	CSPMobileNet-v2: mAP of 83.2%; YOLOv4-lite: mAP of 82.2%; ResNet-50: mAP of 79.6%
[147]	Weeds and Crops	Scenario-1: Potting mix; Scenario-2: Black pebbled gravel	Weed identification	CNN (VGG16 & ResNet50)	Scenario-1: 2868 images Scenario-2: 3488 images	Canon EOS T7 digital camera	Color and texture features	Non-uniform background f1-score: VGG16: 82.75% and, ResNet50: 75%; Uniform background f1-score: VGG16: 77.5% and, ResNet50: 68.4%
[148]	25 different weed species	Fields and lawns in Chongqing, China	Weed identification and detection	YOLOv3, YOLOv5, and Faster R-CNN	Weed25; containing 14035 images of 25 different weed species	Nikon D5300 SLR and a smartphone (Huawei Enjoy 9S)	Morphological features, Color variations, Leaf textures	Detection accuracy: YOLOv3 - 91.8%, YOLOv5 - 92.4%, Faster R-CNN - 92.15%



### C. CONSTRAINTS IN COMPUTATIONAL RESOURCES

Many state-of-the-art techniques like Mask R-CNN are computationally intensive, making real-time deployment difficult given the constrained resources of embedded greenhouse systems ([111], [114]). Hardware devices in agricultural settings often have tight power budgets, restricting the complexity of models that can be run ([93], [116]). While cloud-based solutions can provide greater parallel processing power, sole reliance on the cloud introduces drawbacks like network latency, connectivity dependencies, and data privacy concerns. A balanced approach could be edge-cloud co-design, with lightweight models handling core functionality on-device while leveraging the cloud for more intensive computations. Still, optimizing models and inference pipelines for efficient execution on low-power devices remains a key challenge ([112], [117]). Research on extremely lightweight yet accurate architectures specialized for greenhouse conditions is imperative.

### D. LIMITED INTEGRATION OF DIVERSE SENSING MODALITIES

While most research relies solely on RGB images, other sensing modalities like hyperspectral imaging can provide valuable crop insights ([118], [119], [120]). However, techniques to integrate and fuse multi-modal data sources are still emergent. Capturing relationships between RGB, spectral, depth, thermal, and other data could significantly enhance model robustness and performance ([121], [122]). Developing sensor fusion methods and tailored multi-modal networks is an open research frontier.

### E. SHORTAGE OF SPECIALIZED AND INTERDISCIPLINARY TALENT

The effective development and implementation of computer vision technologies for greenhouses require skilled professionals spanning multiple disciplines ([14], [103], [104], [105]). Emerging deep learning solutions rely heavily on the parallel computing capabilities of GPUs ([64]). To fully leverage these tools requires both computer vision and agriculture expertise ([93], [114]). However, there is a talent shortage with competencies and experience across these domains ([97], [99]). Computer vision involves diverse fields like image processing, machine learning, and pattern recognition ([14]). Integrating techniques from these areas into the complexities of greenhouse farming demands specialized interdisciplinary knowledge ([95], [108]). From researchers advancing scientific innovations to technicians managing real-world deployment, skilled personnel able to bridge computer vision and agriculture are imperative but lacking ([100], [106]). Closing this talent gap across computer science and agricultural engineering is vital to drive progress. Initiatives to support education, training, and collaboration across disciplines are critical to develop professionals that can effectively apply computer vision to transform greenhouse agriculture.

In summary, major challenges like lack of sufficient labeled data, need for task-specific models, computational constraints,

and limited multi-modality demonstrate that significant work remains to develop computer vision techniques adept at the nuances of greenhouse environments. Substantial progress has been made, but overcoming these areas through data generation, model optimization, efficient designs, and sensor fusion integration will be essential to realize the full potential of computer vision for smart greenhouse agriculture.

### VI. CONCLUSION

This paper presented a focused analysis of recent advancements in applying computer vision and deep learning for greenhouse agriculture automation. Spanning diverse application areas like crop monitoring, disease detection, yield forecasting, and quality analysis, over 100 studies were reviewed in detail. The innovations showcased offer glimpses into the transformative potential of data-driven intelligent solutions for optimizing productivity and sustainability in controlled environments. However, significant challenges remain that must be addressed before widespread adoption. The lack of large-scale standardized datasets restricts model generalization and limits performance benchmarking. Meanwhile, computational constraints of embedded systems pose bottlenecks for real-time deployment. There is also a need for more skilled talent with expertise across computer vision, deep learning, and agriculture. Most importantly, the uniqueness of greenhouse environments demands specialized, optimized techniques - off-the-shelf solutions pre-trained on open field data often fail to transfer effectively. Nonetheless, the progress made indicates that an exciting future lies ahead. Expanded collaborations for standardized dataset development, computational advancements in low-power devices, interdisciplinary training programs, and research on tailored solutions for greenhouses will be critical to drive the field forward. Computer vision and deep learning have already shown initial success in automation tasks like robotic harvesting and quality assessment. With continued innovation, they are poised to transform greenhouse infrastructure worldwide - enabling autonomous, efficient, data-driven systems that enhance productivity, resilience, and sustainability. This timely review provided a holistic synthesis of the state-of-the-art, analyzed key challenges, and outlined prospective directions. By condensing current work and elucidating future needs, it aims to motivate and guide ongoing research to unlock the full potential of AI and computer vision for next-generation smart greenhouse agriculture globally. The possibilities are boundless, and the opportunities endless.

### ACKNOWLEDGMENT

The authors also sincerely acknowledge the AI tools ChatGPT and Anthropic's Claude for their assistance in brainstorming ideas, providing writing suggestions, formatting the manuscript, and other forms of support during the preparation of this paper. This research was supported in part by the Universiti Malaysia Pahang-AI Sultan Abdullah under the Product Development Grant UIC220803.

## REFERENCES

- [1] W. K. Mok, Y. X. Tan, and W. N. Chen, "Technology innovations for food security in Singapore: A case study of future food systems for an increasingly natural resource-scarce world," *Trends Food Sci. Technol.*, vol. 102, pp. 155–168, Aug. 2020, doi: [10.1016/J.TIFS.2020.06.013](https://doi.org/10.1016/J.TIFS.2020.06.013).
- [2] Y. Liu, X. Ma, L. Shu, G. P. Hancke, and A. M. Abu-Mahfouz, "From industry 4.0 to agriculture 4.0: Current status, enabling technologies, and research challenges," *IEEE Trans. Ind. Informat.*, vol. 17, no. 6, pp. 4322–4334, Jun. 2021, doi: [10.1109/TII.2020.3003910](https://doi.org/10.1109/TII.2020.3003910).
- [3] FAO, IFAD, UNICEF, WFP, and WHO, *The State of Food Security and Nutrition in the World 2022*, FAO, Rome, Italy, 2022, doi: [10.4060/cc0639en](https://doi.org/10.4060/cc0639en).
- [4] *The State of Food Security and Nutrition in the World 2023*, FAO, Rome, Italy, Jul. 2023, doi: [10.4060/cc3017en](https://doi.org/10.4060/cc3017en).
- [5] M. O. Ojo and A. Zahid, "Deep learning in controlled environment agriculture: A review of recent advancements, challenges and prospects," *Sensors*, vol. 22, no. 20, p. 7965, Oct. 2022, doi: [10.3390/S22207965](https://doi.org/10.3390/S22207965).
- [6] T. Karanisa, Y. Achour, A. Ouammi, and S. Sayadi, "Smart greenhouses as the path towards precision agriculture in the food-energy and water nexus: Case study of Qatar," *Environ. Syst. Decis.*, vol. 42, no. 4, pp. 521–546, Dec. 2022, doi: [10.1007/S10669-022-09862-2](https://doi.org/10.1007/S10669-022-09862-2).
- [7] R. Rayhana, G. Xiao, and Z. Liu, "Internet of Things empowered smart greenhouse farming," *IEEE J. Radio Freq. Identificat.*, vol. 4, no. 3, pp. 195–211, Sep. 2020, doi: [10.1109/JRFID.2020.2984391](https://doi.org/10.1109/JRFID.2020.2984391).
- [8] A. J. Castro, M. D. López-Rodríguez, C. Giagnocavo, M. Gimenez, L. Céspedes, A. La Calle, M. Gallardo, P. Pumares, J. Cabello, E. Rodríguez, D. Uclés, S. Parra, J. Casas, F. Rodríguez, J. S. Fernandez-Prados, D. Alba-Patiño, M. Expósito-Granados, B. E. Murillo-López, L. M. Vasquez, and D. L. Valera, "Six collective challenges for sustainability of Almería greenhouse horticulture," *Int. J. Environ. Res. Public Health*, vol. 16, no. 21, p. 4097, Oct. 2019, doi: [10.3390/IJERPH16214097](https://doi.org/10.3390/IJERPH16214097).
- [9] B. Yang and Y. Xu, "Applications of deep-learning approaches in horticultural research: A review," *Horticulture Res.*, vol. 8, no. 1, p. 123, Dec. 2021, doi: [10.1038/S41438-021-00560-9](https://doi.org/10.1038/S41438-021-00560-9).
- [10] V. G. Dhanya, A. Subeesh, N. L. Kushwaha, D. K. Vishwakarma, T. N. Kumar, G. Ritika, and A. N. Singh, "Deep learning based computer vision approaches for smart agricultural applications," *Artif. Intell. Agricult.*, vol. 6, pp. 211–229, Jan. 2022, doi: [10.1016/J.AIIA.2022.09.007](https://doi.org/10.1016/J.AIIA.2022.09.007).
- [11] Y.-H. Wang and W.-H. Su, "Convolutional neural networks in computer vision for grain crop phenotyping: A review," *Agronomy*, vol. 12, no. 11, p. 2659, Oct. 2022, doi: [10.3390/AGRONOMY12112659](https://doi.org/10.3390/AGRONOMY12112659).
- [12] H. Tian, T. Wang, Y. Liu, X. Qiao, and Y. Li, "Computer vision technology in agricultural automation—A review," *Inf. Process. Agricult.*, vol. 7, no. 1, pp. 1–19, Mar. 2020, doi: [10.1016/J.INPA.2019.09.006](https://doi.org/10.1016/J.INPA.2019.09.006).
- [13] K. Nemali, "History of controlled environment horticulture: Greenhouses," *HortScience*, vol. 57, no. 2, pp. 239–246, Feb. 2022, doi: [10.21273/HORTSCI16160-21](https://doi.org/10.21273/HORTSCI16160-21).
- [14] *Greenhouse Smart Application Project that Reduces Waste and Improves Workforce Efficiency at Farms in Spain*. Accessed: Jul. 26, 2023. [Online]. Available: <https://www.advantech.com/emt/resources/case-study/greenhouse-smart-application-project-that-reduces-waste-and-improves-workforce-efficiency-at-farms-in-spain>
- [15] C. Vastistas, D. D. Avgoustaki, and T. Bartzanas, "A systematic literature review on controlled-environment agriculture: How vertical farms and greenhouses can influence the sustainability and footprint of urban microclimate with local food production," *Atmosphere*, vol. 13, no. 8, p. 1258, Aug. 2022, doi: [10.3390/ATMOS13081258](https://doi.org/10.3390/ATMOS13081258).
- [16] M. S. Farooq, S. Riaz, M. A. Helou, F. S. Khan, A. Abid, and A. Alvi, "Internet of Things in greenhouse agriculture: A survey on enabling technologies, applications, and protocols," *IEEE Access*, vol. 10, pp. 53374–53397, 2022, doi: [10.1109/ACCESS.2022.3166634](https://doi.org/10.1109/ACCESS.2022.3166634).
- [17] D. K. Widyawati, A. Ambarwari, and A. Wahyudi, "Design and prototype development of Internet of Things for greenhouse monitoring system," in *Proc. 3rd Int. Seminar Res. Inf. Technol. Intell. Syst. (ISRITI)*, Dec. 2020, pp. 389–393, doi: [10.1109/ISRITI51436.2020.9315487](https://doi.org/10.1109/ISRITI51436.2020.9315487).
- [18] P. K. Tripathy, A. K. Tripathy, A. Agarwal, and S. P. Mohanty, "MyGreen: An IoT-enabled smart greenhouse for sustainable agriculture," *IEEE Consum. Electron. Mag.*, vol. 10, no. 4, pp. 57–62, Jul. 2021, doi: [10.1109/MCE.2021.3055930](https://doi.org/10.1109/MCE.2021.3055930).
- [19] V. Oguntosin, C. Okeke, E. Adetiba, A. Abdulkareem, and J. Olowoloni, "IoT-based greenhouse monitoring and control system," *Int. J. Comput. Digit. Syst.*, vol. 14, no. 1, pp. 469–483, Aug. 2023, doi: [10.12785/IJCDSD/140137](https://doi.org/10.12785/IJCDSD/140137).
- [20] M. Dhanaraju, P. Chenniappan, K. Ramalingam, S. Pazhanivelan, and R. Kaliaperumal, "Smart farming: Internet of Things (IoT)-based sustainable agriculture," *Agriculture*, vol. 12, no. 10, p. 1745, Oct. 2022, doi: [10.3390/AGRICULTURE12101745](https://doi.org/10.3390/AGRICULTURE12101745).
- [21] C. Maraveas, "Incorporating artificial intelligence technology in smart greenhouses: Current state of the art," *Appl. Sci.*, vol. 13, no. 1, p. 14, Dec. 2022, doi: [10.3390/APP13010014](https://doi.org/10.3390/APP13010014).
- [22] S. J. Soheli, N. Jahan, M. B. Hossain, A. Adhikary, A. R. Khan, and M. Wahiduzzaman, "Smart greenhouse monitoring system using Internet of Things and artificial intelligence," *Wireless Pers. Commun.*, vol. 124, no. 4, pp. 3603–3634, Jun. 2022, doi: [10.1007/S11277-022-09528-X](https://doi.org/10.1007/S11277-022-09528-X).
- [23] A. Escamilla-García, G. M. Soto-Zarazúa, M. Toledano-Ayala, E. Rivas-Araiza, and A. Gastélum-Barrios, "Applications of artificial neural networks in greenhouse technology and overview for smart agriculture development," *Appl. Sci.*, vol. 10, no. 11, p. 3835, May 2020, doi: [10.3390/APP10113835](https://doi.org/10.3390/APP10113835).
- [24] N. N. Misra, Y. Dixit, A. Al-Mallahi, M. S. Bhullar, R. Upadhyay, and A. Martynenko, "IoT, big data, and artificial intelligence in agriculture and food industry," *IEEE Internet Things J.*, vol. 9, no. 9, pp. 6305–6324, May 2022, doi: [10.1109/JIOT.2020.2998584](https://doi.org/10.1109/JIOT.2020.2998584).
- [25] X. Cao, Y. Yao, L. Li, W. Zhang, Z. An, Z. Zhang, S. Guo, L. Xiao, X. Cao, D. Luo. (2022). *iGrow: A Smart Agriculture Solution to Autonomous Greenhouse Control*. [Online]. Available: <https://github.com/holmescao/iGrow.git>
- [26] Z. Li, R. Guo, M. Li, Y. Chen, and G. Li, "A review of computer vision technologies for plant phenotyping," *Comput. Electron. Agricult.*, vol. 176, Sep. 2020, Art. no. 105672, doi: [10.1016/J.COMPAG.2020.105672](https://doi.org/10.1016/J.COMPAG.2020.105672).
- [27] N. T. Sinshaw, B. G. Assefa, S. K. Mohapatra, and A. M. Beyene, "Applications of computer vision on automatic potato plant disease detection: A systematic literature review," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–18, Nov. 2022, doi: [10.1155/2022/7186687](https://doi.org/10.1155/2022/7186687).
- [28] E. Mavridou, E. Vrochidou, G. A. Papakostas, T. Pachidis, and V. G. Kaburlasos, "Machine vision systems in precision agriculture for crop farming," *J. Imag.*, vol. 5, no. 12, p. 89, Dec. 2019, doi: [10.3390/JIMAGING5120089](https://doi.org/10.3390/JIMAGING5120089).
- [29] M. Ouhami, A. Hafiane, Y. Es-Saady, M. El Hajji, and R. Canals, "Computer vision, IoT and data fusion for crop disease detection using machine learning: A survey and ongoing research," *Remote Sens.*, vol. 13, no. 13, p. 2486, Jun. 2021, doi: [10.3390/RS13132486](https://doi.org/10.3390/RS13132486).
- [30] D. D. Cox and T. Dean, "Neural networks and neuroscience-inspired computer vision," *Current Biol.*, vol. 24, no. 18, pp. R921–R929, Sep. 2014, doi: [10.1016/J.CUB.2014.08.026](https://doi.org/10.1016/J.CUB.2014.08.026).
- [31] J. Memon, M. Sami, R. A. Khan, and M. Uddin, "Handwritten optical character recognition (OCR): A comprehensive systematic literature review (SLR)," *IEEE Access*, vol. 8, pp. 142642–142668, 2020, doi: [10.1109/ACCESS.2020.3012542](https://doi.org/10.1109/ACCESS.2020.3012542).
- [32] E. Dilek and M. Dener, "Computer vision applications in intelligent transportation systems: A survey," *Sensors*, vol. 23, no. 6, p. 2938, Mar. 2023, doi: [10.3390/s23062938](https://doi.org/10.3390/s23062938).
- [33] B. Santra and D. P. Mukherjee, "A comprehensive survey on computer vision based approaches for automatic identification of products in retail store," *Image Vis. Comput.*, vol. 86, pp. 45–63, Jun. 2019, doi: [10.1016/J.IMAVIS.2019.03.005](https://doi.org/10.1016/J.IMAVIS.2019.03.005).
- [34] A. Esteva, K. Chou, S. Yeung, N. Naik, A. Madani, A. Mottaghi, Y. Liu, E. Topol, J. Dean, and R. Socher, "Deep learning-enabled medical computer vision," *NPI Digit. Med.*, vol. 4, no. 1, p. 5, Jan. 2021, doi: [10.1038/S41746-020-00376-2](https://doi.org/10.1038/S41746-020-00376-2).
- [35] É. Zablocki, H. Ben-Younes, P. Pérez, and M. Cord, "Explainability of deep vision-based autonomous driving systems: Review and challenges," *Int. J. Comput. Vis.*, vol. 130, no. 10, pp. 2425–2452, Oct. 2022, doi: [10.1007/S11263-022-01657-X](https://doi.org/10.1007/S11263-022-01657-X).
- [36] L. Zhou, L. Zhang, and N. Konz, "Computer vision techniques in manufacturing," *IEEE Trans. Syst. Man, Cybern. Syst.*, vol. 53, no. 1, pp. 105–117, Jan. 2023, doi: [10.1109/TSMC.2022.3166397](https://doi.org/10.1109/TSMC.2022.3166397).
- [37] E. Bastug, M. Bennis, M. Medard, and M. Debbah, "Toward interconnected virtual reality: Opportunities, challenges, and enablers," *IEEE Commun. Mag.*, vol. 55, no. 6, pp. 110–117, Jun. 2017, doi: [10.1109/MCOM.2017.1601089](https://doi.org/10.1109/MCOM.2017.1601089).

- [38] Y. Kortli, M. Jridi, A. Al Falou, and M. Atri, "Face recognition systems: A survey," *Sensors*, vol. 20, no. 2, p. 342, Jan. 2020, doi: [10.3390/S20020342](https://doi.org/10.3390/S20020342).
- [39] M. S. Islam, M. M. Hasan, S. Abdullah, J. U. M. Akbar, N. H. M. Arafat, and S. A. Murad, "A deep spatio-temporal network for vision-based sexual harassment detection," in *Proc. Emerg. Technol. Comput., Commun. Electron. (ETCCE)*, Dec. 2021, pp. 1–6, doi: [10.1109/ETCCE54784.2021.9689891](https://doi.org/10.1109/ETCCE54784.2021.9689891).
- [40] J. F. S. Gomes and F. R. Leta, "Applications of computer vision techniques in the agriculture and food industry: A review," *Eur. Food Res. Technol.*, vol. 235, no. 6, pp. 989–1000, Dec. 2012, doi: [10.1007/S00217-012-1844-2](https://doi.org/10.1007/S00217-012-1844-2).
- [41] C. Chen, W. Zhu, and T. Norton, "Behaviour recognition of pigs and cattle: Journey from computer vision to deep learning," *Comput. Electron. Agricult.*, vol. 187, Aug. 2021, Art. no. 106255, doi: [10.1016/J.COMPAG.2021.106255](https://doi.org/10.1016/J.COMPAG.2021.106255).
- [42] J. U. M. Akbar, S. Fauzi Kamarulzaman, and E. H. Tusher, "Plant stem disease detection using machine learning approaches," in *Proc. 14th Int. Conf. Comput. Commun. Netw. Technol. (ICCCNT)*, New Delhi, India, Jul. 2023, pp. 1–8, doi: [10.1109/ICCCNT56998.2023.10307074](https://doi.org/10.1109/ICCCNT56998.2023.10307074).
- [43] W. Yang, L. Cai, and F. Wu, "Image segmentation based on gray level and local relative entropy two dimensional histogram," *PLoS One*, vol. 15, no. 3, Mar. 2020, Art. no. e0229651, doi: [10.1371/JOURNAL.PONE.0229651](https://doi.org/10.1371/JOURNAL.PONE.0229651).
- [44] G. Li, Y. Huang, Z. Chen, G. D. Chesser, J. L. Purswell, J. Linhoss, and Y. Zhao, "Practices and applications of convolutional neural network-based computer vision systems in animal farming: A review," *Sensors*, vol. 21, no. 4, p. 1492, Feb. 2021, doi: [10.3390/S21041492](https://doi.org/10.3390/S21041492).
- [45] S. Zheng, L. Qian, P. Li, C. He, X. Qin, and X. Li, "An introductory review of spiking neural network and artificial neural network: From biological intelligence to artificial intelligence," 2022, *arXiv:2204.07519*.
- [46] J. Chai, H. Zeng, A. Li, and E. W. T. Ngai, "Deep learning in computer vision: A critical review of emerging techniques and application scenarios," *Mach. Learn. Appl.*, vol. 6, Dec. 2021, Art. no. 100134, doi: [10.24433/CO.0411648.V1](https://doi.org/10.24433/CO.0411648.V1).
- [47] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: [10.1145/3065386](https://doi.org/10.1145/3065386).
- [48] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778, doi: [10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90).
- [49] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
- [50] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269, doi: [10.1109/CVPR.2017.243](https://doi.org/10.1109/CVPR.2017.243).
- [51] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Proc. 13th Eur. Conf. Comput. Vis. (ECCV)* Cham, Switzerland: Springer, Sep. 2014, pp. 818–833, doi: [10.1007/978-3-319-10590-1\\_53](https://doi.org/10.1007/978-3-319-10590-1_53).
- [52] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 580–587, doi: [10.1109/CVPR.2014.81](https://doi.org/10.1109/CVPR.2014.81).
- [53] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448, doi: [10.1109/ICCV.2015.169](https://doi.org/10.1109/ICCV.2015.169).
- [54] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017, doi: [10.1109/TPAMI.2016.2577031](https://doi.org/10.1109/TPAMI.2016.2577031).
- [55] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788, doi: [10.1109/CVPR.2016.91](https://doi.org/10.1109/CVPR.2016.91).
- [56] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Proc. 14th Eur. Conf.*, in Lecture Notes in Computer Science, vol. 9905, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Amsterdam, The Netherlands, Cham, Switzerland: Springer, Oct. 2016, pp. 21–37, doi: [10.1007/978-3-319-46448-0](https://doi.org/10.1007/978-3-319-46448-0).
- [57] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2999–3007, doi: [10.1109/ICCV.2017.324](https://doi.org/10.1109/ICCV.2017.324).
- [58] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988, doi: [10.1109/ICCV.2017.322](https://doi.org/10.1109/ICCV.2017.322).
- [59] J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang, and H. Lu, "Dual attention network for scene segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3141–3149, doi: [10.1109/CVPR.2019.00326](https://doi.org/10.1109/CVPR.2019.00326).
- [60] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer Verlag, 2015, pp. 234–241, doi: [10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28).
- [61] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-Net architecture for medical image segmentation," in *Proc. Int. Workshop Deep Learn. Med. Image Anal.* Cham, Switzerland: Springer, 2018, pp. 3–11, doi: [10.1007/978-3-030-00889-5\\_1](https://doi.org/10.1007/978-3-030-00889-5_1).
- [62] O. Oktay, J. Schlemper, L. Le Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention U-Net: Learning where to look for the pancreas," 2018, *arXiv:1804.03999*.
- [63] D. Jha, P. H. Smedsrud, M. A. Riegler, D. Johansen, T. D. Lange, P. Halvorsen, and H. D. Johansen, "ResUNet++: An advanced architecture for medical image segmentation," in *Proc. IEEE Int. Symp. Multimedia (ISM)*, Dec. 2019, pp. 225–2255, doi: [10.1109/ISM46123.2019.00049](https://doi.org/10.1109/ISM46123.2019.00049).
- [64] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, Y. Zhou. (2021). *TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation*. [Online]. Available: <https://github.com/Beckschen/>
- [65] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, 2015, pp. 3431–3440, doi: [10.1109/CVPR.2015.7298965](https://doi.org/10.1109/CVPR.2015.7298965).
- [66] H. Wu, J. Zhang, K. Huang, K. Liang, and Y. Yu, "FastFCN: Rethinking dilated convolution in the backbone for semantic segmentation," 2019, *arXiv:1903.11816*.
- [67] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," 2016, *arXiv:1606.00915*.
- [68] B. Hariharan, P. Arbeláez, R. Girshick, and J. Malik. (2014). *LNCS 8695-Simultaneous Detection and Segmentation*. [Online]. Available: <http://www.eecs.berkeley.edu/Research/Projects/CS/vision/shape/sds>
- [69] T. Chen, L. Lin, X. Wu, N. Xiao, and X. Luo, "Learning to segment object candidates via recursive neural networks," *IEEE Trans. Image Process.*, vol. 27, no. 12, pp. 5827–5839, Dec. 2018, doi: [10.1109/TIP.2018.2859025](https://doi.org/10.1109/TIP.2018.2859025).
- [70] L. Xiang and D. Wang, "A review of three-dimensional vision techniques in food and agriculture applications," *Smart Agricult. Technol.*, vol. 5, Oct. 2023, Art. no. 100259, doi: [10.1016/J.ATECH.2023.100259](https://doi.org/10.1016/J.ATECH.2023.100259).
- [71] Z. Kang, J. Yang, Z. Yang, and S. Cheng, "A review of techniques for 3D reconstruction of indoor environments," *ISPRS Int. J. Geo-Inf.*, vol. 9, no. 5, p. 330, May 2020, doi: [10.3390/IJGI9050330](https://doi.org/10.3390/IJGI9050330).
- [72] B. Cui, W. Tao, and H. Zhao, "High-precision 3D reconstruction for small-to-medium-sized objects utilizing line-structured light scanning: A review," *Remote Sens.*, vol. 13, no. 21, p. 4457, Nov. 2021, doi: [10.3390/RS13214457](https://doi.org/10.3390/RS13214457).
- [73] S. Yamamoto, M. Karkee, Y. Kobayashi, N. Nakayama, S. Tsubota, L. N. Thi Thanh, and T. Konya, "3D reconstruction of apple fruits using consumer-grade RGB-depth sensor," *Eng. Agricult., Environ. Food*, vol. 11, no. 4, pp. 159–168, Oct. 2018, doi: [10.1016/J.EAEF.2018.02.005](https://doi.org/10.1016/J.EAEF.2018.02.005).
- [74] J. Li, W. Gao, Y. Wu, Y. Liu, and Y. Shen, "High-quality indoor scene 3D reconstruction with RGB-D cameras: A brief review," *Comput. Vis. Media*, vol. 8, no. 3, pp. 369–393, Sep. 2022, doi: [10.1007/S41095-021-0250-8](https://doi.org/10.1007/S41095-021-0250-8).
- [75] V. Cantoni, S. Levialdi, and V. (Vito) Roberto, *Artificial Vision: Image Description, Recognition and Communication*. New York, NY, USA: Academic, 1997.
- [76] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: [10.1038/nature14539](https://doi.org/10.1038/nature14539).
- [77] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, Jan. 2015, doi: [10.1016/J.NEUNET.2014.09.003](https://doi.org/10.1016/J.NEUNET.2014.09.003).
- [78] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," *Comput. Electron. Agricult.*, vol. 147, pp. 70–90, Apr. 2018, doi: [10.1016/J.COMPAG.2018.02.016](https://doi.org/10.1016/J.COMPAG.2018.02.016).

- [79] X. Yu, H. Lu, and D. Wu, "Development of deep learning method for predicting firmness and soluble solid content of postharvest korla fragrant pear using vis/NIR hyperspectral reflectance imaging," *Postharvest Biol. Technol.*, vol. 141, pp. 39–49, Jul. 2018, doi: [10.1016/j.postharvbio.2018.02.013](https://doi.org/10.1016/j.postharvbio.2018.02.013).
- [80] S. Chen, Y. Cao, Y. Kang, P. Li, and B. Sun, "Deep feature representation based imitation learning for autonomous helicopter aerobatics," *IEEE Trans. Artif. Intell.*, vol. 2, no. 5, pp. 437–446, Oct. 2021, doi: [10.1109/TAI.2021.3053511](https://doi.org/10.1109/TAI.2021.3053511).
- [81] M. V. D. Prasad, B. JwalaLakshamma, A. Hari Chandana, K. Komali, M. V. N. Manoja, P. Rajesh Kumar, C. Raghava Prasad, S. Inthiyaz, and P. Sasi Kiran, "An efficient classification of flower images with convolutional neural networks," *Int. J. Eng. Technol.*, vol. 7, no. 1, p. 384, Dec. 2017, doi: [10.14419/IJET.V7I1.1.9857](https://doi.org/10.14419/IJET.V7I1.1.9857).
- [82] T. Yuan, L. Lv, F. Zhang, J. Fu, J. Gao, J. Zhang, W. Li, C. Zhang, and W. Zhang, "Robust cherry tomatoes detection algorithm in greenhouse scene based on SSD," *Agriculture*, vol. 10, no. 5, p. 160, May 2020, doi: [10.3390/agriculture10050160](https://doi.org/10.3390/agriculture10050160).
- [83] P. Kumari, V. Goswami, H. N., and R. S. Pundir, "Recurrent neural network architecture for forecasting banana prices in Gujarat, India," *PLoS One*, vol. 18, no. 6, Jun. 2023, Art. no. e0275702, doi: [10.1371/journal.pone.0275702](https://doi.org/10.1371/journal.pone.0275702).
- [84] J. A. Nasir, O. S. Khan, and I. Varlamis, "Fake news detection: A hybrid CNN-RNN based deep learning approach," *Int. J. Inf. Manage. Data Insights*, vol. 1, no. 1, Apr. 2021, Art. no. 100007, doi: [10.1016/j.jjimei.2020.100007](https://doi.org/10.1016/j.jjimei.2020.100007).
- [85] J. Oruh, S. Viriri, and A. Adegun, "Long short-term memory recurrent neural network for automatic speech recognition," *IEEE Access*, vol. 10, pp. 30069–30079, 2022, doi: [10.1109/ACCESS.2022.3159339](https://doi.org/10.1109/ACCESS.2022.3159339).
- [86] J. Ackerson, R. Dave, and N. Seliya, "Applications of recurrent neural network for biometric authentication & anomaly detection," *Information*, vol. 12, no. 7, p. 272, Jul. 2021, doi: [10.3390/INFO12070272](https://doi.org/10.3390/INFO12070272).
- [87] J. Yu, A. de Antonio, and E. Villalba-Mora, "Deep learning (CNN, RNN) applications for smart homes: A systematic review," *Computers*, vol. 11, no. 2, p. 26, Feb. 2022, doi: [10.3390/computers11020026](https://doi.org/10.3390/computers11020026).
- [88] B. Dhiman, Y. Kumar, and Y.-C. Hu, "A general purpose multi-fruit system for assessing the quality of fruits with the application of recurrent neural network," *Soft Comput.*, vol. 25, no. 14, pp. 9255–9272, Jul. 2021, doi: [10.1007/s00500-021-05867-2](https://doi.org/10.1007/s00500-021-05867-2).
- [89] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," 2014, *arXiv:1409.4842*.
- [90] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning transferable architectures for scalable image recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8697–8710, doi: [10.1109/CVPR.2018.00907](https://doi.org/10.1109/CVPR.2018.00907).
- [91] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," 2014, *arXiv:1411.4038*.
- [92] U. Lee, M. P. Islam, N. Kochi, K. Tokuda, Y. Nakano, H. Naito, Y. Kawasaki, T. Ota, T. Sugiyama, and D.-H. Ahn, "An automated, clip-type, small Internet of Things camera-based tomato flower and fruit monitoring and harvest prediction system," *Sensors*, vol. 22, no. 7, p. 2456, Mar. 2022, doi: [10.3390/s22072456](https://doi.org/10.3390/s22072456).
- [93] M.-S. Gang, H.-J. Kim, and D.-W. Kim, "Estimation of greenhouse lettuce growth indices based on a two-stage CNN using RGB-D images," *Sensors*, vol. 22, no. 15, p. 5499, Jul. 2022, doi: [10.3390/s22155499](https://doi.org/10.3390/s22155499).
- [94] L. Zhang, Z. Xu, D. Xu, J. Ma, Y. Chen, and Z. Fu, "Growth monitoring of greenhouse lettuce based on a convolutional neural network," *Horticulture Res.*, vol. 7, no. 1, p. 124, Dec. 2020, doi: [10.1038/s41438-020-00345-6](https://doi.org/10.1038/s41438-020-00345-6).
- [95] V. Moysiadis, G. Kokkonis, S. Bibi, I. Moscholios, N. Maropoulos, and P. Sarigiannidis, "Monitoring mushroom growth with machine learning," *Agriculture*, vol. 13, no. 1, p. 223, Jan. 2023, doi: [10.3390/agriculture13010223](https://doi.org/10.3390/agriculture13010223).
- [96] R. Shinoda, K. Motoki, K. Hara, H. Kataoka, R. Nakano, T. Nakazaki, and R. Noguchi, "RoseTracker: A system for automated Rose growth monitoring," *Smart Agricult. Technol.*, vol. 5, Oct. 2023, Art. no. 100271, doi: [10.1016/j.atech.2023.100271](https://doi.org/10.1016/j.atech.2023.100271).
- [97] M. Chen, X. Lang, X. Zhai, T. Li, and Y. Shi, "Intelligent recognition of greenhouse cucumber canopy vine top with deep learning model," *Comput. Electron. Agricult.*, vol. 213, Oct. 2023, Art. no. 108219, doi: [10.1016/j.compag.2023.108219](https://doi.org/10.1016/j.compag.2023.108219).
- [98] M. P. Islam, Y. Nakano, U. Lee, K. Tokuda, and N. Kochi, "TheLNet270v1—A novel deep-network architecture for the automatic classification of thermal images for greenhouse plants," *Frontiers Plant Sci.*, vol. 12, Jul. 2021, Art. no. 630425, doi: [10.3389/fpls.2021.630425](https://doi.org/10.3389/fpls.2021.630425).
- [99] C. Zhou, J. Hu, Z. Xu, J. Yue, H. Ye, and G. Yang, "A novel greenhouse-based system for the detection and plumpness assessment of strawberry using an improved deep learning technique," *Frontiers Plant Sci.*, vol. 11, p. 559, Jun. 2020, doi: [10.3389/fpls.2020.00559](https://doi.org/10.3389/fpls.2020.00559).
- [100] P. Cong, S. Li, J. Zhou, K. Lv, and H. Feng, "Research on instance segmentation algorithm of greenhouse sweet pepper detection based on improved mask RCNN," *Agronomy*, vol. 13, no. 1, p. 196, Jan. 2023, doi: [10.3390/agronomy13010196](https://doi.org/10.3390/agronomy13010196).
- [101] F. Su, Y. Zhao, G. Wang, P. Liu, Y. Yan, and L. Zu, "Tomato maturity classification based on SE-YOLOv3-MobileNetV1 network under nature greenhouse environment," *Agronomy*, vol. 12, no. 7, p. 1638, Jul. 2022, doi: [10.3390/agronomy12071638](https://doi.org/10.3390/agronomy12071638).
- [102] G. Moreira, S. A. Magalhães, T. Pinho, F. N. dos Santos, and M. Cunha, "Benchmark of deep learning and a proposed HSV colour space models for the detection and classification of greenhouse tomato," *Agronomy*, vol. 12, no. 2, p. 356, Jan. 2022, doi: [10.3390/agronomy12020356](https://doi.org/10.3390/agronomy12020356).
- [103] S. Mao, Y. Li, Y. Ma, B. Zhang, J. Zhou, and K. Wang, "Automatic cucumber recognition algorithm for harvesting robots in the natural environment using deep learning and multi-feature fusion," *Comput. Electron. Agricult.*, vol. 170, Mar. 2020, Art. no. 105254, doi: [10.1016/j.compag.2020.105254](https://doi.org/10.1016/j.compag.2020.105254).
- [104] P. Wspanialy and M. Moussa, "A detection and severity estimation system for generic diseases of tomato greenhouse plants," *Comput. Electron. Agricult.*, vol. 178, Nov. 2020, Art. no. 105701, doi: [10.1016/j.compag.2020.105701](https://doi.org/10.1016/j.compag.2020.105701).
- [105] J. F. Restrepo-Arias, J. W. Branch-Bedoya, and G. Awad, "Plant disease detection strategy based on image texture and Bayesian optimization with small neural networks," *Agriculture*, vol. 12, no. 11, p. 1964, Nov. 2022, doi: [10.3390/agriculture12111964](https://doi.org/10.3390/agriculture12111964).
- [106] S. Zhao, J. Liu, and S. Wu, "Multiple disease detection method for greenhouse-cultivated strawberry based on multiscale feature fusion faster R-CNN," *Comput. Electron. Agricult.*, vol. 199, Aug. 2022, Art. no. 107176, doi: [10.1016/j.compag.2022.107176](https://doi.org/10.1016/j.compag.2022.107176).
- [107] Y. Xu, Q. Chen, S. Kong, L. Xing, Q. Wang, X. Cong, and Y. Zhou, "Real-time object detection method of melon leaf diseases under complex background in greenhouse," *J. Real-Time Image Process.*, vol. 19, no. 5, pp. 985–995, Oct. 2022, doi: [10.1007/s11554-022-01239-7](https://doi.org/10.1007/s11554-022-01239-7).
- [108] P. Zhang, L. Yang, and D. Li, "EfficientNet-B4-ranger: A novel method for greenhouse cucumber disease recognition under natural complex environment," *Comput. Electron. Agricult.*, vol. 176, Sep. 2020, Art. no. 105652, doi: [10.1016/j.compag.2020.105652](https://doi.org/10.1016/j.compag.2020.105652).
- [109] Z. Zhang, P. Flores, A. Friskop, Z. Liu, C. Igathinathane, X. Han, H. J. Kim, N. Jahan, J. Mathew, and S. Shreya, "Enhancing wheat disease diagnosis in a greenhouse using image deep features and parallel feature fusion," *Frontiers Plant Sci.*, vol. 13, Mar. 2022, Art. no. 834447, doi: [10.3389/fpls.2022.834447](https://doi.org/10.3389/fpls.2022.834447).
- [110] J. Rong, P. Wang, T. Wang, L. Hu, and T. Yuan, "Fruit pose recognition and directional orderly grasping strategies for tomato harvesting robots," *Comput. Electron. Agricult.*, vol. 202, Nov. 2022, Art. no. 107430, doi: [10.1016/j.compag.2022.107430](https://doi.org/10.1016/j.compag.2022.107430).
- [111] M. Benavides, M. Cantón-Garbín, J. A. Sánchez-Molina, and F. Rodríguez, "Automatic tomato and peduncle location system based on computer vision for use in robotized harvesting," *Appl. Sci.*, vol. 10, no. 17, p. 5887, Aug. 2020, doi: [10.3390/app10175887](https://doi.org/10.3390/app10175887).
- [112] J. Rong, P. Wang, Q. Yang, and F. Huang, "A field-tested harvesting robot for oyster mushroom in greenhouse," *Agronomy*, vol. 11, no. 6, p. 1210, Jun. 2021, doi: [10.3390/agronomy11061210](https://doi.org/10.3390/agronomy11061210).
- [113] X. Liu, D. Wang, Y. Li, X. Guan, and C. Qin, "Detection of green asparagus using improved mask R-CNN for automatic harvesting," *Sensors*, vol. 22, no. 23, p. 9270, Nov. 2022, doi: [10.3390/s22239270](https://doi.org/10.3390/s22239270).
- [114] X. Wang, Z. Vladislav, O. Viktor, Z. Wu, and M. Zhao, "Online recognition and yield estimation of tomato in plant factory based on YOLOv3," *Sci. Rep.*, vol. 12, no. 1, p. 8686, May 2022, doi: [10.1038/s41598-022-12732-1](https://doi.org/10.1038/s41598-022-12732-1).
- [115] A. K. Maji, S. Marwaha, S. Kumar, A. Arora, V. Chinnusamy, and S. Islam, "SlypNet: Spikelet-based yield prediction of wheat using advanced plant phenotyping and computer vision techniques," *Frontiers Plant Sci.*, vol. 13, Aug. 2022, Art. no. 889853, doi: [10.3389/fpls.2022.889853](https://doi.org/10.3389/fpls.2022.889853).

- [116] Y. Egi, M. Hajyzadeh, and E. Eyceyurt, "Drone-computer communication based tomato generative organ counting model using YOLO v5 and deep-sort," *Agriculture*, vol. 12, no. 9, p. 1290, Aug. 2022, doi: [10.3390/agriculture12091290](https://doi.org/10.3390/agriculture12091290).
- [117] C. Zhou, J. Hu, Z. Xu, J. Yue, H. Ye, and G. Yang, "A monitoring system for the segmentation and grading of broccoli head based on deep learning and neural networks," *Frontiers Plant Sci.*, vol. 11, p. 402, Apr. 2020, doi: [10.3389/fpls.2020.00402](https://doi.org/10.3389/fpls.2020.00402).
- [118] M. F. Taha, A. I. ElManawy, K. S. Alshallash, G. ElMasry, K. Alharbi, L. Zhou, N. Liang, and Z. Qiu, "Using machine learning for nutrient content detection of aquaponics-grown plants based on spectral data," *Sustainability*, vol. 14, no. 19, p. 12318, Sep. 2022, doi: [10.3390/su141912318](https://doi.org/10.3390/su141912318).
- [119] S. Eshkabilov, A. Lee, X. Sun, C. W. Lee, and H. Simsek, "Hyperspectral imaging techniques for rapid detection of nutrient content of hydroponically grown lettuce cultivars," *Comput. Electron. Agricult.*, vol. 181, Feb. 2021, Art. no. 105968, doi: [10.1016/j.compag.2020.105968](https://doi.org/10.1016/j.compag.2020.105968).
- [120] M. Lak, S. Minaei, S. Soufizadeh, A. Banakar, and G. Polder, "Development of a smart machine vision based system to detect water stress in greenhouse tomato plants," *Agricult. Eng. Int. CIGR J.*, vol. 23, no. 2, pp. 141–151, 2021. [Online]. Available: <http://www.cigrjournal.org>
- [121] S. Levanon, O. Markovich, I. Gozlan, O. Bakhshian, A. Zvirin, Y. Honen, and R. Kimmel, "Abiotic stress prediction from RGB-T images of banana plantlets," 2020, *arXiv:2011.11597*.
- [122] K. Najafian, L. Jin, H. R. Kutcher, M. Hladun, S. Horovatin, M. A. Oviedo-Ludena, S. M. P. De Andrade, L. Wang, and I. Stavness, "Detection of fusarium damaged kernels in wheat using deep semi-supervised learning on a novel WheatSeedBelt dataset," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2023, pp. 660–669. [Online]. Available: <https://www.researchgate.net/publication/374662079>
- [123] M. Janani and R. Jebakumar, "Detection and classification of groundnut leaf nutrient level extraction in RGB images," *Adv. Eng. Softw.*, vol. 175, Jan. 2023, Art. no. 103320, doi: [10.1016/j.advengsoft.2022.103320](https://doi.org/10.1016/j.advengsoft.2022.103320).
- [124] X. Zhang, J. Bu, and X. Wang, "Automatic pest identification system based on deep learning and machine vision," *Frontiers Plant Sci.*, vol. 14, Sep. 2023, doi: [10.3389/fpls.2023.1255719](https://doi.org/10.3389/fpls.2023.1255719).
- [125] A. Albanese, M. Nardello, and D. Brunelli, "Automated pest detection with DNN on the edge for precision agriculture," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 11, no. 3, pp. 458–467, Sep. 2021, doi: [10.1109/JETCAS.2021.3101740](https://doi.org/10.1109/JETCAS.2021.3101740).
- [126] S. U. Rasheed, W. Muhammad, I. Qaiser, and M. J. Irshad, "A multispectral pest-detection algorithm for precision agriculture," *Eng. Proc.*, vol. 12, no. 1, p. 46, Jan. 2021, doi: [10.3390/engproc2021012046](https://doi.org/10.3390/engproc2021012046).
- [127] E. A. Lins, J. P. M. Rodriguez, S. I. Scoloski, J. Pivato, M. B. Lima, J. M. C. Fernandes, P. R. V. da Silva Pereira, D. Lau, and R. Rieder, "A method for counting and classifying aphids using computer vision," *Comput. Electron. Agricult.*, vol. 169, Feb. 2020, Art. no. 105200, doi: [10.1016/j.compag.2019.105200](https://doi.org/10.1016/j.compag.2019.105200).
- [128] Z. Yang, W. Li, M. Li, and X. Yang, "Automatic greenhouse pest recognition based on multiple color space features," *Int. J. Agricult. Biol. Eng.*, vol. 14, no. 2, pp. 188–195, 2021, doi: [10.25165/J.IJABE.20211402.5098](https://doi.org/10.25165/J.IJABE.20211402.5098).
- [129] H. Liu and J. S. Chahl, "Proximal detecting invertebrate pests on crops using a deep residual convolutional neural network trained by virtual images," *Artif. Intell. Agricult.*, vol. 5, pp. 13–23, Jan. 2021, doi: [10.1016/j.aiaa.2021.01.003](https://doi.org/10.1016/j.aiaa.2021.01.003).
- [130] R. Zhao and F. Shi, "A novel strategy for pest disease detection of brassica chinensis based on UAV imagery and deep learning," *Int. J. Remote Sens.*, vol. 43, nos. 19–24, pp. 7083–7103, Dec. 2022, doi: [10.1080/01431161.2022.2155082](https://doi.org/10.1080/01431161.2022.2155082).
- [131] W. Li, D. Wang, M. Li, Y. Gao, J. Wu, and X. Yang, "Field detection of tiny pests from sticky trap images using deep learning in agricultural greenhouse," *Comput. Electron. Agricult.*, vol. 183, Apr. 2021, Art. no. 106048, doi: [10.1016/j.compag.2021.106048](https://doi.org/10.1016/j.compag.2021.106048).
- [132] W. Nie, T. Li, and L. Zhu, "Market demand and government regulation for quality grading system of agricultural products in China," *J. Retailing Consum. Services*, vol. 56, Sep. 2020, Art. no. 102134, doi: [10.1016/j.jretconser.2020.102134](https://doi.org/10.1016/j.jretconser.2020.102134).
- [133] X. Zhu, K. Zhu, P. Liu, Y. Zhang, and H. Jiang, "A special robot for precise grading and metering of mushrooms based on YOLOv5," *Appl. Sci.*, vol. 13, no. 18, p. 10104, Sep. 2023, doi: [10.3390/app131810104](https://doi.org/10.3390/app131810104).
- [134] W. K. Tan, M. Amir Hakim Ismail, Z. Husin, and M. L. Yasruddin, "Automated tomato grading system using computer vision (CV) and deep neural network (DNN) algorithm," in *Proc. IEEE 12th Symp. Comput. Appl. Ind. Electron. (ISCAIE)*, May 2022, pp. 22–27, doi: [10.1109/ISCAIE54458.2022.9794557](https://doi.org/10.1109/ISCAIE54458.2022.9794557).
- [135] Y. Hendrawan, "Classification of large green chilli maturity using deep learning," *IOP Conf. Ser., Earth Environ. Sci.*, vol. 924, Dec. 2021, Art. no. 012009, doi: [10.1088/1755-1315/924/1/012009](https://doi.org/10.1088/1755-1315/924/1/012009).
- [136] J. Shi, F. Shi, and X. Huang, "Prediction of maturity date of leafy greens based on causal inference and convolutional neural network," *Agriculture*, vol. 13, no. 2, p. 403, Feb. 2023, doi: [10.3390/agriculture13020403](https://doi.org/10.3390/agriculture13020403).
- [137] X. Wei, L. Wu, D. Ge, M. Yao, and Y. Bai, "Prediction of the maturity of greenhouse grapes based on imaging technology," *Plant Phenomics*, vol. 2022, Jan. 2022, Art. no. 9753427, doi: [10.34133/2022/9753427](https://doi.org/10.34133/2022/9753427).
- [138] I. Nyalala, C. Okinda, Q. Chao, P. Mecha, T. Korohou, Z. Yi, S. Nyalala, Z. Jiayu, L. Chao, and C. Kunjie, "Weight and volume estimation of single and occluded tomatoes using machine vision," *Int. J. Food Properties*, vol. 24, no. 1, pp. 818–832, Jan. 2021, doi: [10.1080/10942912.2021.1933024](https://doi.org/10.1080/10942912.2021.1933024).
- [139] L. Zgripeacea, V. Putan, M. Ardelean, and C. B. Baneasa, "Technology of thin metal sheet cutting with fiber laser," in *Proc. J. Phys. Conf. Ser.*, vol. 1781, no. 1, Feb. 2021, Art. no. 012059, doi: [10.1088/1742-6596/1781/1/012059](https://doi.org/10.1088/1742-6596/1781/1/012059).
- [140] N. S. Narwade, P. D. Deshmukh, and D. V. Jadhav, "Early detection and grading of apple fruit diseases using deep learning," *J. Inst. Eng. India Ser. B*, vol. 101, no. 4, pp. 687–696, Aug. 2020, doi: [10.1007/s40031-020-00466-z](https://doi.org/10.1007/s40031-020-00466-z).
- [141] A. D. de Medeiros, R. C. Bernardes, L. J. da Silva, B. A. L. de Freitas, D. C. F. D. S. Dias, and C. B. da Silva, "Deep learning-based approach using X-ray images for classifying *Crambe abyssinica* seed quality," *Ind. Crops Products*, vol. 164, pp. 113378–113378, Jun. 2021. [Online]. Available: <https://doi.org/10.1016/j.indcrop.2021.113378>
- [142] S.-J. Hong, S. Park, A. Lee, S.-Y. Kim, E. Kim, C.-H. Lee, and G. Kim, "Nondestructive prediction of pepper seed viability using single and fusion information of hyperspectral and X-ray images," *Sens. Actuators A, Phys.*, vol. 350, Feb. 2023, Art. no. 114151, doi: [10.1016/j.sna.2022.114151](https://doi.org/10.1016/j.sna.2022.114151).
- [143] V. Lube, M. A. Noyan, A. Przybysz, K. Salama, and I. Blilou, "MultipleXLab: A high-throughput portable live-imaging root phenotyping platform using deep learning and computer vision," *Plant Methods*, vol. 18, no. 1, p. 38, Dec. 2022, doi: [10.1186/s13007-022-00864-4](https://doi.org/10.1186/s13007-022-00864-4).
- [144] T. Gao, A. K. N. Chandran, P. Paul, H. Walia, and H. Yu, "HyperSeed: An end-to-end method to process hyperspectral images of seeds," *Sensors*, vol. 21, no. 24, p. 8184, Dec. 2021, doi: [10.3390/s21248184](https://doi.org/10.3390/s21248184).
- [145] K. G. Falk, T. Z. Jubery, S. V. Mirnezami, K. A. Parnley, S. Sarkar, A. Singh, B. Ganapathysubramanian, and A. K. Singh, "Computer vision and machine learning enabled soybean root phenotyping pipeline," *Plant Methods*, vol. 16, no. 1, pp. 1–19, Jan. 2020, doi: [10.1186/s13007-019-0550-5](https://doi.org/10.1186/s13007-019-0550-5).
- [146] E.-C. Oerke, "Crop losses to pests," *J. Agricult. Sci.*, vol. 144, no. 1, pp. 31–43, Feb. 2006, doi: [10.1017/S0021859605005708](https://doi.org/10.1017/S0021859605005708).
- [147] S. G. C, C. Koparan, M. R. Ahmed, Y. Zhang, K. Howatt, and X. Sun, "A study on deep learning algorithm performance on weed and crop species identification under different image background," *Artif. Intell. Agricult.*, vol. 6, pp. 242–256, Jan. 2022, doi: [10.1016/j.aiaa.2022.11.001](https://doi.org/10.1016/j.aiaa.2022.11.001).
- [148] P. Wang, Y. Tang, F. Luo, L. Wang, C. Li, Q. Niu, and H. Li, "Weed25: A deep learning dataset for weed identification," *Frontiers Plant Sci.*, vol. 13, Nov. 2022, Art. no. 1053329, doi: [10.3389/fpls.2022.1053329](https://doi.org/10.3389/fpls.2022.1053329).
- [149] Y. S. Oda, L. O. de Oliveira, S. De Paula, A. O. de Oliveira, and J. C. de Castro Neto, "Weed detection among soybean plants in artificial lighting environment using multispectral images and computer vision," *Proc. SPIE*, vol. 12746, Aug. 2023, Art. no. 1274608, doi: [10.1117/12.2686194](https://doi.org/10.1117/12.2686194).
- [150] S. G. C, Y. Zhang, C. Koparan, M. R. Ahmed, K. Howatt, and X. Sun, "Weed and crop species classification using computer vision and deep learning technologies in greenhouse conditions," *J. Agricult. Food Res.*, vol. 9, Sep. 2022, Art. no. 100325, doi: [10.1016/j.jafr.2022.100325](https://doi.org/10.1016/j.jafr.2022.100325).
- [151] N. Rai, X. Sun, C. Igathinathane, K. Howatt, and M. Ostlie, "Aerial-based weed detection using low-cost and lightweight deep learning models on an edge platform," *J. ASABE*, vol. 66, no. 5, pp. 1041–1055, 2023, doi: [10.13031/ja.15413](https://doi.org/10.13031/ja.15413).

- [152] N. Fatima, S. A. Siddiqui, and A. Ahmad, (2021) *IoT-Based Smart Greenhouse with Disease Prediction Using Deep Learning*. [Online]. Available: [www.ijcacs.thesai.org](http://www.ijcacs.thesai.org)
- [153] A. Reyes-Yanes, P. Martinez, and R. Ahmad, "Real-time growth rate and fresh weight estimation for little gem romaine lettuce in aquaponic grow beds," *Comput. Electron. Agricult.*, vol. 179, Dec. 2020, Art. no. 105827, doi: [10.1016/j.compag.2020.105827](https://doi.org/10.1016/j.compag.2020.105827).



**JALAL UDDIN MD AKBAR** (Member, IEEE) was born in Chittagong, Bangladesh, in 1998. He received the B.S. degree (Hons.) in computer science engineering from International Islamic University Chittagong, in 2021. He is currently pursuing the master's degree by research in soft computing and intelligent systems with Universiti Malaysia Pahang Al-Sultan Abdullah (UMPSA), Pekan, Pahang, Malaysia. He has been a Research Assistant with the Soft Computing and Intelligent Systems Laboratory, UMPSA, since 2022. His research interests include machine learning, deep learning, and computer vision techniques for real-world applications. Based on his academic excellence, he received the prestigious Master's Research Scheme Scholarship from the Ministry of Higher Education, Malaysia. He has showcased his research work at various international conferences and forums. Through his dedication, he aims to advance innovations in artificial intelligence to create transformative solutions for challenging real-world problems.



**SYAFIQ FAUZI KAMARULZAMAN** (Senior Member, IEEE) received the B.Eng., master's, and Ph.D. degrees from the University of Tsukuba, Japan, in 2010, 2012, and 2015, respectively. He was a fellow of the Universiti Malaysia Pahang Al-Sultan Abdullah from 2012 to 2015. He is currently a Senior Lecturer with the Faculty of Computing, Universiti Malaysia Pahang Al-Sultan Abdullah. His research interests include control systems and artificial intelligence, which includes applications on autonomous vehicles, robotics, and the Internet of Things. He is a member of the Malaysian Robotics and Automation Society. He is a certified Cisco Instructor for CCNA Program and the Internet of Things.



**ABU JAFAR MD MUZAHID** received the B.S. (Hons.) and M.S. degrees in statistics from the Shahjalal University of Science and Technology, Bangladesh, and the master's degree in computing from Universiti Malaysia Pahang Al-Sultan Abdullah, Malaysia. He is currently pursuing the Ph.D. degree majoring in aerospace engineering with the Department of Mechanical, Aerospace, and Biomedical Engineering, The University of Tennessee, Knoxville. His research intersects include artificial intelligence, machine learning, and eco-friendly automotive innovations. He has an impressive international presence, having spearheaded AI and data-driven decision-making projects with institutional funding from Japan, Saudi Arabia, and Malaysia. His academic merit secured him the coveted Masters Research Scheme Scholarship from the Ministry of Higher Education, Malaysia. His trailblazing work has been consistently lauded at international forums, as evidenced by his Gold Medal wins at Citrex, the FYPro-COM Carnival, UMPSA, and the Special EURO BUSINESSHALLER Award at Malaysia Technology Expo 2022. His dedication to advancing green technology is not only recognized but also celebrated on the global stage.



**MD. ARAFATUR RAHMAN** (Senior Member, IEEE) received the Ph.D. degree in electronic and telecommunications engineering from the University of Naples Federico II, Naples, Italy, in 2013. He has around 15 years research and teaching experience in the domain of computer science and communications engineering. Currently, he is a Reader in cyber security with the School of Engineering, Computing and Mathematical Sciences, University of Wolverhampton, U.K. He was an Associate Professor with the Faculty of Computing, Universiti Malaysia Pahang (UMP). He was a Postdoctoral Research Fellow with the University of Naples Federico II, in 2014, and a Visiting Researcher with the Sapienza University of Rome, in 2016. His research interests include cyber security in particular focus on the Internet of Things (IoT), wireless communication networks, cognitive radio networks, 5G, vehicular communication, cyber physical systems, big data, cloud-fog-edge computing, and machine learning dependent applications. He has developed an excellent track record of academic leadership and management and execution of international ICT projects that are supported by agencies in the U.K., Italy, EU, and Malaysia. He has received number of prestigious international research awards, notably the Best Paper Award at ICNS-15 (Italy), IC0902 Grant (France), Italian Government Ph.D. Research Scholarship, and the IIUM Best Masters Student Award, the Best Supervisor Award at UMP, the Awards in International Exhibitions, including the Euro Business-HALLER Poland Special Award, the MTE 2022, Best Innovation Award, the MTE 2020, Malaysia, Diamond and Gold in BiS-17 U.K., the Best of the Best Innovation Award, and the Most Commercial IT Innovation Award, Malaysia, and Gold & Silver medals in iENA-17 Germany. He has coauthored of around 150 prestigious IEEE and Elsevier journal [e.g., IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, IEEE TRANSACTIONS ON GREEN COMMUNICATIONS AND NETWORKING, IEEE TRANSACTIONS ON SERVICES COMPUTING, *IEEE Communications Magazine*, *JNCA* (Elsevier), and *FGCS* (Elsevier)], and conference publications (e.g., IEEE GLOBECOM and IEEE DASC). He has served as a Specialty Chief Editor for *IoT Theory and Fundamental Research* (specialty section of *Frontiers in the Internet of Things*), an Advisory Board Member, an Editorial Board Member for *textitComputer Systems Science and Engineering and Tech Science Press & Computers* (MDPI), a Lead Guest Editor for IEEE ACCESS and *Computers*, an Associate Editor for IEEE ACCESS and *Patron*, and the General Chair, an Organizing Committee, the Publicity Chair, the Session Chair, a Programme Committee, and a member of Technical Programme Committee (TPC) in numerous leading conferences worldwide (e.g., IEEE GLOBECOM, IEEE DASC, IEEE ISCI, and IEEE ETCCCE) and journals. His name was enlisted inside the World top 2% scientists list released by Stanford University under the category of "Citation Impact in Single Calendar Year, in 2019, 2020, and 2021." He was awarded a Higher Education Academy (HEA) Fellowship from the U.K. He was a fellow of the IBM Center of Excellence and Earth Resources and Sustainability Center, Malaysia. He was endorsed by Royal Academy of Engineering, U.K., as a Global Talent under the category of "Exceptional Talent," in 2022.



**MUEEN UDDIN** (Senior Member, IEEE) received the Ph.D. degree from Universiti Teknologi Malaysia (UTM), in 2013. He is currently an Associate Professor in data and cybersecurity with the University of Doha for Science and Technology, Qatar. He has published more than 130 international journals and conference papers in highly reputed journals with a cumulative impact factor of over 300. His research interests include blockchain, cybersecurity, the IoT security, and network and cloud security.

...