




Review

Advanced Technology in Agriculture Industry by Implementing Image Annotation Technique and Deep Learning Approach: A Review

Normaisharah Mamat¹, Mohd Fauzi Othman¹, Rawad Abdoulghafor^{2,*}, Samir Brahim Belhaouari^{3,*}, Normahira Mamat⁴ and Shamsul Faisal Mohd Hussein¹

- ¹ Department of Electronic System Engineering, Malaysia-Japan International Institute of Technology, University Teknologi Malaysia, Jalan Sultan Yahya Petra, Kuala Lumpur 54100, Malaysia; normaisharah@utm.my (N.M.); mdfauzi@utm.my (M.F.O.); foxfaisal@yahoo.co.uk (S.F.M.H.)
- ² Computational Intelligence Group Research, Faculty of Information and Communication Technology, International Islamic University Malaysia, Kuala Lumpur 53100, Malaysia
- ³ Division of Information and Computing Technology, College of Science and Engineering, Hamad Bin Khalifa University, Education City, Doha P.O. Box 34110, Qatar
- ⁴ Faculty of Electronic Engineering Technology, Universiti Malaysia Perlis, Kampus Pauh Putra, Arau 02600, Malaysia; normahira@unimap.edu.my
- * Correspondence: rawad@iiu.edu.my (R.A.); sbelhaouari@hbku.edu.qa (S.B.B.)

Abstract: The implementation of intelligent technology in agriculture is seriously investigated as a way to increase agriculture production while reducing the amount of human labor. In agriculture, recent technology has seen image annotation utilizing deep learning techniques. Due to the rapid development of image data, image annotation has gained a lot of attention. The use of deep learning in image annotation can extract features from images and has been shown to analyze enormous amounts of data successfully. Deep learning is a type of machine learning method inspired by the structure of the human brain and based on artificial neural network concepts. Through training phases that can label a massive amount of data and connect them up with their corresponding characteristics, deep learning can conclude unlabeled data in image processing. For complicated and ambiguous situations, deep learning technology provides accurate predictions. This technology strives to improve productivity, quality and economy and minimize deficiency rates in the agriculture industry. As a result, this article discusses the application of image annotation in the agriculture industry utilizing several deep learning approaches. Various types of annotations that were used to train the images are presented. Recent publications have been reviewed on the basis of their application of deep learning with current advancement technology. Plant recognition, disease detection, counting, classification and yield estimation are among the many advancements of deep learning architecture employed in many applications in agriculture that are thoroughly investigated. Furthermore, this review helps to assist researchers to gain a deeper understanding and future application of deep learning in agriculture. According to all of the articles, the deep learning technique has successfully created significant accuracy and prediction in the model utilized. Finally, the existing challenges and future promises of deep learning in agriculture are discussed.

Keywords: image annotation; deep learning; agriculture; plant recognition; disease detection; counting; classification; yield estimation



Citation: Mamat, N.; Othman, M.F.; Abdoulghafor, R.; Belhaouari, S.B.; Mamat, N.; Mohd Hussein, S.F. Advanced Technology in Agriculture Industry by Implementing Image Annotation Technique and Deep Learning Approach: A Review. *Agriculture* **2022**, *12*, 1033. <https://doi.org/10.3390/agriculture12071033>

Academic Editor: Yanbo Huang

Received: 10 June 2022

Accepted: 4 July 2022

Published: 15 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The agriculture sector is the backbone of most countries, providing enormous employment opportunities to the community as well as goods manufacturing and food supply. Fruit plantation is one of the most important agricultural activities. The production and protection of fruit per capita has recently been considered an essential indicator of a country's growth and quality of life [1]. Population growth of 7.2 to 9.6 billion people is expected

by 2100. The advanced approach of smart agriculture must be used to meet the demand for food from agriculture [2]. Several studies have recommended that the critical issue of improving management and production in the agriculture industry is addressed [3,4]. Agriculture production has challenges in terms of productivity, environmental impact and sustainability. Agriculture ecosystems necessitate constant monitoring of several variables, resulting in a large amount of data. The data could be in the form of images that can be processed with various image processing algorithms to identify plants, diseases and other cases in varied agricultural situations [5]. Advanced technology improvements have been made in agriculture with limited resources to ensure production, quality, processing, storage and distribution [6]. The technology used in this field involves various scientific disciplines covering sensors, big data, artificial intelligence and robotics [7]. Apart from using sensor technology to advance the agriculture industry [8], the use of image annotation techniques to improve agriculture production is a relatively new invention in technology.

Image annotation has attracted widespread attention in the past few years due to the rapid growth of image data [9–11]. This method is used to analyze big data images and predict labels for the images [12]. Image annotation is the technique of labeling an image with keywords which reflect the character of the image and assist in the intelligent retrieval of relevant images using a simple query representation [13]. Image annotation in the agriculture sector can annotate images according to the user's requirement. Everything from plants and fruits to soil can be annotated to be recognized and classified. Moreover, it helps in plant detection, classification and segmentation based on the plant species, type, health condition or maturity. It can predict the label of a given image and can correspond well to the image content [12]. Image annotation can describe images at the semantic level and has many applications that are not only focused on image analysis but also on urban management and biomedical engineering. Basically, image annotation algorithms are divided into traditional and deep neural network-based methods [14]. However, traditional or manual image annotation has inherent weaknesses. Therefore, automatic image annotation (AIA) was introduced in the late 1990s by Mori et al. [15].

The objective of automatic image annotation is to predict several textual labels for an unseen image representing its content, which is a labeling problem. This technology automatically annotates the image using its semantic tags and has been applied in image retrieval classification and the medical domain. The training data attempt to teach a model to assign semantic labels to the new image automatically. One or more tags will be transferred to the image based on image metadata or visual features. For instance, the technology has been proposed in many areas and shows outstanding achievement [13,16]. Large amounts of data are required to improve the accuracy of annotating images of plants or diseases. To assist researchers in overcoming these severe challenges, Deng et al. [17] introduced ImageNet, a publicly available collection of existing plants extensively used in computer vision. It has been frequently used as a benchmark for various visualization types of computer vision issues. Another public dataset is PlantVillage [18], an open-access platform for disease plant leaf images by Penn State University. Moreover, the datasets that are dedicated to fruit detection are MinneApple [19], Date Fruit [20] and MangoYOLO [21], weed control datasets are DeepWeeds [22] and Open Plant Phenotype Dataset [23] and a dataset of plant seedlings at different growth stages is V2 Plant Seedling Dataset [24].

AIA can be classified into many categories. The difference in the classes is based on the contribution, computational complexity, computational time and annotation accuracy. One of the categories is deep learning-based image annotation [25,26]. Deep learning in research on AIA has attracted extensive attention in the theoretical study and various image processing and computer vision task applications. It shows high potential in image processing capabilities for the future needs of agriculture [27,28]. Deep learning, which is a subset of machine learning, was firstly introduced by Dechter [29] in 1986 to machine learning and by Aizerberg et al. [30] in 2000 to the artificial neural network. It can transform the data using various functions that allow data representation in a hierarchical way and defined as a simpler concept. It learns to perform any task directly from the images and

produce high-accuracy responses [31,32]. Several AIA techniques have been proposed other than the deep learning approach such as support vector machines, Bayesian, texture resemblance and instance-based method. Deep learning techniques, on the other hand, have succeeded in image processing throughout the last decade [33]. The high accuracy of deep learning is generated by high computational and storage requirements during the training and inference phase. This is because the training process is both space consuming and computationally intensive, as millions of parameters are needed to refine over multiple periods of time [34]. Due to complexity of the data models, training is quite expensive. Furthermore, deep learning necessitates the use of costly graphic user interfaces (GPUs) and many machines. This raises the cost to the users. The image annotation training set based on deep learning can be classified into supervised, unsupervised and semi-supervised categories.

Supervised deep learning involves training a data sample from a data source that has been classified correctly. Its algorithm is trained on input data that has been labeled for a certain output until it is able to discern the underlying links between the inputs and output findings. The system is supplied with labeled datasets during the training phase, which will inform it which outputs are associated with certain input values. Supervised learning provides a significant challenge due to the requirement of a huge amount of labeled data [35,36] and at least hundreds of annotated images are required during the supervised training [37]. The training approach consists of providing a large number of annotated images to the algorithm to assist the model to learn, then testing the trained model on unannotated images. To determine the accuracy of this method, annotated images with hidden labels are often employed in the algorithm's testing stage. Thus, annotated images for training supervised deep learning models achieve acceptable performance levels. Most of the studies applied supervised learning, as this method promises high accuracy as proposed in [38–40]. Another attractive annotation method is based on unsupervised learning. Unsupervised learning, in contrast to supervised learning, deals with unlabeled data. In addition, labels for these cases are frequently difficult to obtain due to insufficient knowledge data or the labeling is prohibitively expensive. Furthermore, the lack of labels makes setting goals for the trained model problematic. Consequently, determining whether or not the results are accurate is difficult. The study by [41] employed unsupervised learning in two real weed datasets using a recent unsupervised deep clustering technique. These datasets' results signal a potential direction in the use of unsupervised learning and clustering in agricultural challenges. For circumstances where cluster and class numbers vary, the suggested modified unsupervised clustering accuracy has proven to be a robust and easier to interpret evaluation clustering measure. It is also feasible to demonstrate how data augmentation and transfer learning can significantly improve unsupervised learning.

Semi-supervised learning, like supervised and unsupervised learning, involves working with a dataset. However, the dataset is separated into labeled and unlabeled parts. When the labeling of acquired data is too difficult or expensive, this technique is frequently used. In fact, it is also possible to use it if the labeled data are poor quality [42]. The fundamental issue in large-scale image annotation approaches based on semi-supervised learning is dealing with a large, noisy dataset in which the number of images expands faster. The ability to identify unwanted plants has improved because of the advancement in farm image analysis. However, the majority of these systems rely on supervised learning, which necessitates a large number of manually annotated images. As a result, due to the huge variety of plant species being cultivated, supervised learning is economically infeasible for the individual farmer. Therefore, [43–45] proposed an unsupervised image annotation technique to solve weed detection in farms using deep learning approaches.

Deep learning has significant potential in the agriculture sector in increasing the amount and quality of the produce by image-based classification. Consequently, many researchers have employed the technology and method of deep learning to improve and automate tasks [3]. Its role in this sector gives excellent results in plant counting, leaf counting, leaf segmentation and yield prediction [46]. Noon et al. [47] have reviewed

the application of deep learning in the agriculture sector by identifying plant leaf stress in early detection to enable farmers to apply the suitable treatment. Deep learning is effective in detecting leaf stress for various plants. However, implementing deep learning in agriculture requires a large amount of data regarding the plants, in terms of collecting and processing. The necessary data are basically collected using wireless sensors, drones, robots and satellites [48]. The more data used to train the deep learning model, the more robust and pervasive the model becomes [49].

Unmanned aerial vehicles (UAVs) and unmanned ground vehicles (UGVs) are examples of robotics systems that provide a cost-effective, adaptable and scalable solution for product management and crop quality [50]. Weeds are able to reduce crop production and their growth must be monitored regularly to keep them under control. Additionally, applying the same amount of herbicide to the entire field results in waste, pollution and a higher cost for farmers. The combination of image analytics from UAV footage and precision agriculture is able to assist agronomists in advising farmers on where to focus herbicides in particular regions in the field [51,52]. As stated in [53], the first stage in site-specific weed management is to detect weed patches in the field quickly and accurately. Therefore, the authors proposed object detection implemented with Faster RCNN in training and evaluating weed detection in soybean fields using a low-altitude UAV. The proposed technique was the best model in detecting weeds by obtaining an intersection over union (IoU) performance of 0.85. Franco et al. [54] have captured a thistle weed species, *Cirsium arvense*, in cereal crops by utilizing a UAV. This tool is used to gather a view of an agriculture site with detailed exploration and is attractive due to its low operational costs and flexible driving. A UAV captured RGB images of thistles at 50 m above the ground, annotated weed and cereal classes and grouped them under a unique label of pixels. According to [51], labeling plants in a field image consumes a lot of time and there is very little attention paid to annotating the data by training a deep learning model. Therefore, the authors proposed a deep learning technique to detect weeds using UAV images by applying overlapping windows for weed detection [51]. Deep learning techniques will provide the probability of the plant being a weed or crop for each window location. Deep learning can make harvesting robots more effective when generating robust and reliable computer vision algorithms to detect fruit [55]. The usage of UAVs in dataset collection has also been applied in palm oil tree detection [56], rice phenology [57], detection and classification of soybean pests [58], potato plant detection [59], paddy field yield assessment [60] and corn classification [61].

Over the last few decades, UGVs have been used to achieve efficiency, particularly by reducing manpower requirements. UGVs have been employed for soil analysis [62], precision spraying [63], controlled weeding [64] and crop harvesting [65]. Mazzia et al. [66] employed a UGV for path planning using deep learning as an estimator. Row-based crops are ideal for testing and deploying UGVs that can monitor and manage to harvest the crops. The study proposed by the authors proved the feasibility of the deep learning technique by demonstrating the viability of a complete autonomous global path planner. In [67], a robot harvester with the implementation of a deep learning algorithm is used to detect an obstacle and observe the surrounding environment for rice. The image cascade network's employment successfully detects obstacles and avoids collision with an average success rate of 96.6%. Besides UAVs and UGVs, deep learning provides a practical solution in the agriculture field from satellite imagery. A vital component of agricultural monitoring systems is having accurate maps of crop types and acreage. Therefore, the application of satellites is able to determine the boundary of smallholder farms since their boundaries are hazy, in irregular shapes and frequently mixed with other land uses. Persello et al. [68] presented a deep learning technique to automatically delineate smallholder farms using a convolutional network in combination with a globalization and grouping algorithm. The proposed solution outperforms alternative strategies by autonomously delineating field boundaries with F scores greater than 0.7 and 0.6 for the proposed test regions, respectively. Furthermore, satellites are implemented to capture images in identifying crops as presented in [69]. The authors utilized multiexposure satellite imagery of agricultural

land using image analysis and deep learning techniques for edge segmentation in an image. The implementation of a CNN for image edge smoothing achieves accuracy of 98.17%. According to [70], enough data should be collected for training in order to predict crop yields and forecast crop prices reliably. Data availability is a significant limitation that can be overcome using satellite imagery that can cover huge geographic areas. The combination of utilizing deep learning using satellite imagery applications gives a significant advantage results in extracting field boundaries [71], monitoring agricultural areas [72], weather prediction [73], crop classification [74] and soil moisture forecast [75].

Various implementations of deep learning in agriculture approaches have been extensively reviewed in recent years as proposed in [5,37,76–79]. Among those, Koirala et al. [77] reviewed the application of deep learning in fruit detection and yield estimation, Zhang et al. [80] explore dense scene analysis of the application deep learning in agriculture and Moazzam et al. [79] emphasized the challenges of weed and crop classification using deep learning. Based on the great attention on the implementation of deep learning in the agriculture sector in recent years, and contrary to existing surveys, this article concisely reviews the use of deep learning techniques in image annotation, focusing on plants and crop areas. This review article presents the most recent five years of research on this method in agriculture, covering the new technology and trends. The presentation covers the techniques of annotating images, the learning techniques, the various architectures proposed, the tools used and, finally, the applications. The application issues are basically in plant detection, disease detection, counting, yield estimation, segmentation and classification in the agriculture sector. These tasks are difficult to perform manually, time consuming and require workforce involvement. The lack of people's ability to identify objects for these tasks is finally compensated for by using current technology and trends, particularly image annotation and deep learning techniques, which also boost process efficiency. There are many different types of plants. To identify plants, especially rare ones, knowledge is required. Additionally, a systematic and disciplined approach to classifying various plants is crucial for recognizing and categorizing the vast amount of data acquired on the many known plants. To solve this problem, plant detection and classification are crucial tasks. Since segmentation helps to extract features from an image, it will improve classification accuracy. A crucial concern in agriculture is disease detection. Disease control procedures can waste time and resources and result in additional plant losses without accurate identification of the disease and its causative agent. Furthermore, in the agriculture industry, counting is essential in managing orchards, yet it can be difficult because of various issues, including overlapping. In particular, counting leaves provides a clear image of the plant's condition and stage of development. Especially in the age of global climate change, agricultural output assessment is essential for solving new concerns in food security. Accurate yield estimation benefits famine prevention efforts in addition to assisting farmers in making appropriate economic and management decisions. Therefore, this manuscript emphasizes these efforts to boost agriculture production by summarizing these tasks using deep learning, which improves prediction and accuracy. Various architecture structures of CNNs are well described as a reference for researchers to better understand the implementation of deep learning in the agriculture sector to illustrate how they work. This article also proposes the future trends and technology that could be implemented to improve the quality and productivity in the agriculture field.

2. Deep Learning for Image Annotation

Image annotation using deep learning is the most informative method that requires more complex training data. It is essential for functional datasets because it informs the training model about the crucial parts of the image and may use those details to recognize the classes in test images. The majority of automatic image annotation methods perform by extracting features from training and testing images at the first step. Secondly, based on the training data, the annotation model is developed. Finally, annotations are developed based on the characteristics of the test images [81]. Figure 1 illustrates the detail of the image

annotation process. Feature extraction is a technique for indexing and extracting visual content from images. Color, texture, shape and domain-specific features are examples of primitive or low-level image features [82].

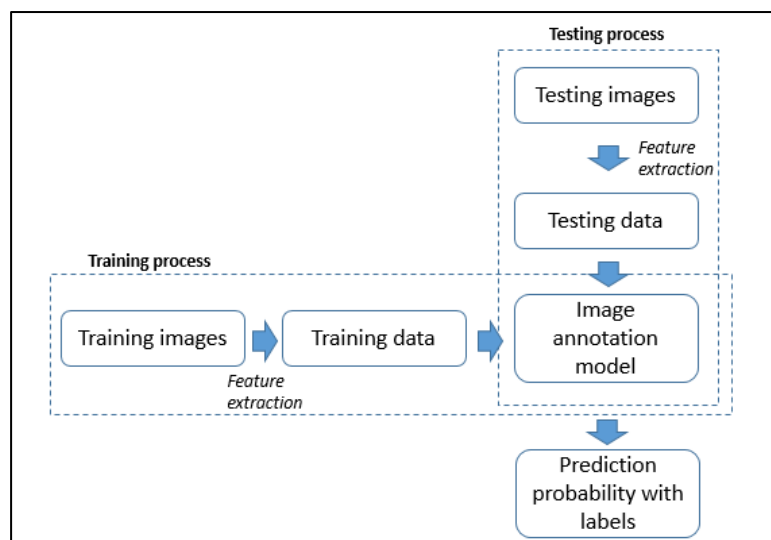


Figure 1. The process of image annotation algorithm.

Depending on the approach utilized, various annotation types are used to annotate images. The popular image annotation techniques employed in agriculture based on deep learning are bounding box [83–86] and segmentation [87–90]. The study in [91] proposed the tools to boost the efficiency of identifying agriculture images, which frequently have more various objects and more detailed shapes than those in many general datasets. Feature extraction in the architecture of deep learning can be found in imaging applications. Different types of this architecture in deep learning that have frequently been applied in recent years are unsupervised pre-trained networks (UPNs), recurrent neural networks (RNNs) and convolutional neural networks (CNNs) [92]. An RNN has the advantage of processing time-series data and making decisions about the future based on historical data. An RNN has been proposed by Alibabaei et al. [93] to predict tomato yield according to the date, climate, irrigation amount and soil water content. RNN architecture consists of long-shot term memory (LSTM), gated recurrent units (GRUs), bidirectional LSTM (BLSTM) and bidirectional GRU (BGRU). The study shows that BLSTM is able to capture the relationship of the past and new observations and accurately predict the yield. However, the BLSTM model has a longer training time compared to implemented models. The authors also conclude that deep learning has the ability to estimate the yield at the end of the seasons.

A CNN is mainly used among deep learning architecture due to its high detection accuracy, reliability and feasibility [94]. CNNs or convNets are designed to learn the spatial features, for example edges, textures, corners or more abstract shapes. The core of learning these characteristics is the diverse and successive transformation of the input object, which is convolution at different spatial scales such as pooling operation. This operation identifies and combines both high-level concepts and low-level features [95]. This method has been proven to be good in extracting abstract features from a raw image through convolutional and pooling layers [96]. The architecture of CNNs was introduced by Fukushima [97] who proposed the algorithm of supervised and unsupervised training of the parameter that learns from the incoming data. In general, a CNN receives the image data that form input layers and generates a vector of different characteristics assigned to object classes in the form of an output layer. There are hidden layers between the input and output layers consisting of a series of convolution and pooling layers and ending with a fully connected layer [98]. CNNs are widely used as a powerful class of models to classify images in a multiple problems in agriculture such as fruit classification, plant disease detection, weed

identification and pest classification [99]. In addition, they can also detect and count the number of crops. Huang et al. [100] chose a CNN to classify green coffee beans because CNN characteristics are good at extracting image color and shape.

Two categories of object detection in deep learning are defined by drawing bounding boxes around the images and classifying the object's pixels. From a label perspective, drawing rectangular bounding boxes around the object is much easier compared to labeling the object's pixels by drawing outlines. However, from a mapping perspective, pixel-level object detection is more accurate compared to the bounding box technique [101]. According to Hamidinekoo et al. [102], it is challenging to segment and compute the detection of individual fruits from images. Therefore, the authors applied a CNN to classify various parts of the plant inflorescence and estimate fruit numbers from the images. CNNs are also used in detecting fruit and disease. Onishi et al. [103] proposed a high-speed and accurate method to detect the position of fruit and automated harvesting using a robot arm. The authors utilized a shot multibox detector (SSD) based on the CNN method to detect objects in an image using a single deep neural network. To achieve a high level of recognition accuracy, the SSD creates multiscale predictions from multiscale feature maps and explicitly separates the predictions based on ratio aspect. The image of fruit detection utilized in this method is shown in Figure 2. Other fruits and leaves occlude some apples, but the method can still detect the apples. The result of the study showed that the fruit detection using the SSD is 90% and this accuracy was achieved in only 2 s.

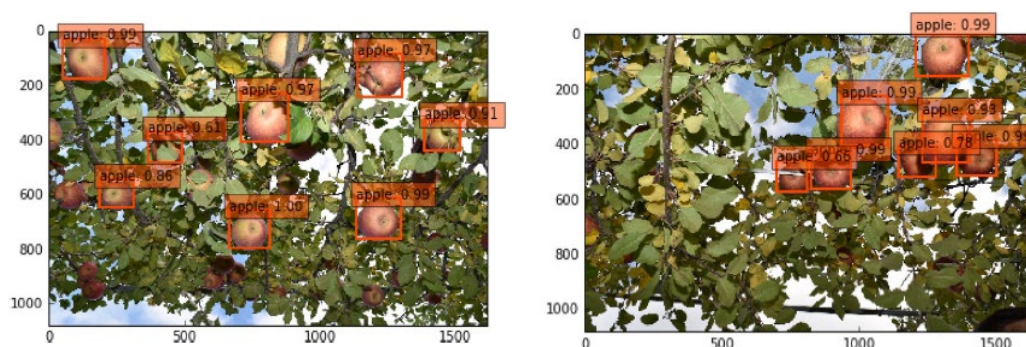


Figure 2. Fruit detection using CNN [103].

Another major concern in the agriculture sector nowadays is that many pathogens and insects threaten many farms. Since deep learning can dive into deep analysis and computation, this technique is one of the prominent methods for plant disease detection [104]. Many approaches help to monitor the health of the crop, from semantic segmentation to other popular image annotation techniques. When compared to labeling data for classification, segmentation data are more challenging. Several image annotations based on supervised learning for object segmentation methods have been presented in recent years for this reason. Sharma et al. [105] used image segmentation to detect disease by employing the CNN method. In order to obtain maximum data on disease symptoms, the image is segmented by extracting the affected parts of leaves rather than the whole images. The quantifying result for each type of disease shows that the data are trained very well and achieved that excellent result even under real conditions. Kang and Chen [106] performed detection and segmentation of apple fruit and branches as shown in Figure 3. As shown in Figure 3a–f, apples are drawn in distinct colors, and branches are drawn in blue. These detections and segmentations are recognized by utilizing a CNN. The experiment achieved 0.873 accuracy of instance segmentation of apple fruits and 0.794 accuracy of branch segmentation.

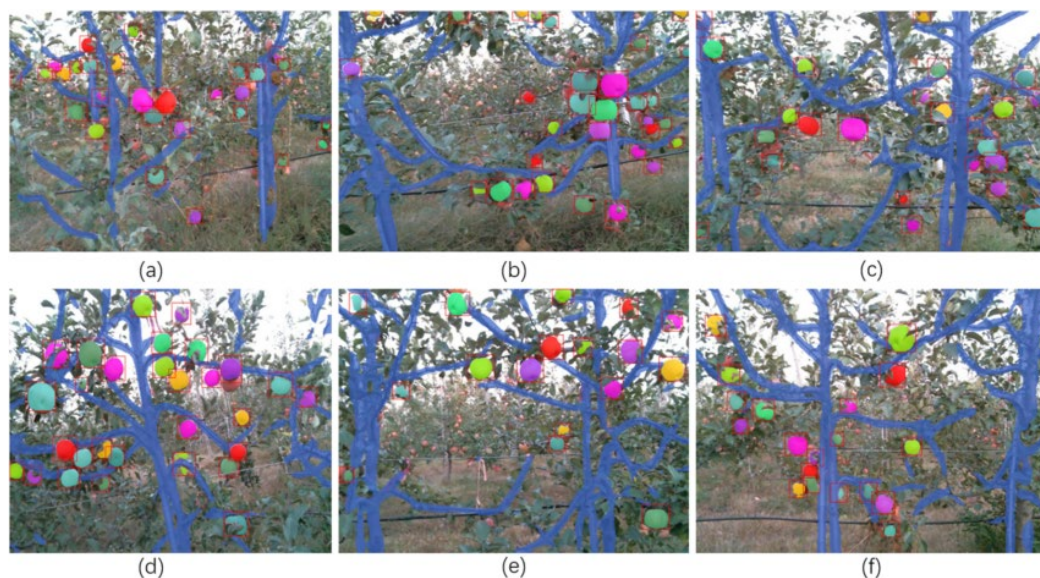


Figure 3. Detection and segmentation of fruit and branch [106].

Khattak et al. [107] proposed a CNN to identify fruits and leaves in healthy and diseased conditions. The result shows that the CNN has a test accuracy of 94.55 percent, making it a suggested support tool for farmers in classifying citrus fruit/leaf condition as either healthy or diseased. In yield estimation, Yang et al. [108] trained a CNN to estimate corn grain yield. The experiment conducted by the authors produced 75.50% classification accuracy of spectral and color images. Fuentes [109] successfully proved that the implementation of a deep learning technique can detect disease and pests in tomato plants. In addition, the technique is able to deal with a complex scenario from the surrounding area of the plant. The result obtained is shown in Figure 4a–d, where the deep learning generates high accuracy in detecting disease and pests. The image from left to right for each sub-figure is the input image, annotated image and predicted results.

The architectures of CNNs have been classified gradually with the increasing number of convolutional layers, namely LeNet, AlexNet, Visual Geometri Group 16 (VGG16), VGG19, ResNet, GoogLeNet ResNext, DenseNet and You Only Look Once (YOLO). The differences between these architectures are the number of layers, non-linearity function and the pooling type used [110]. Mu et al. [111] applied VggNet to detect the quality of blueberry through the skin pigments during the seven stages of its maturity. The technique was used to solve the difficulty and identify the maturity and quality grade of the blueberry fruit measured by the human eye. In fact, the method has improved the accuracy and efficiency of detection of the quality of blueberry. Lee et al. [112] proposed three types of CNN architecture with different layers, namely, VGG16 with 16 layers, InceptionV3 with 48 layers and GoogLeNetBN with 34 layers. The InceptionV2 inspired GoogLeNetBN and InceptionV3 architecture and has the capability of improving the accuracy and reducing the complexity of computation. Batch normalization (BN) has been proven to be able to limit overfitting and speed up convergence. In a study by [113], three CNN architectures, AlexNet, InceptionV3 and SqueezeNet, were compared to assess their accuracy in evaluating tomato late blight disease. Among these architectures, AlexNet generates the highest accuracy in feature extraction with 93.4%. Gehlot and Saini [114] also compared the performance of CNN architectures in classifying diseases in tomato leaves. The architectures assessed in the study are AlexNet, GoogLeNet, VGG-16, ResNet-101 and DenseNet-121. The accuracy of all these architectures are almost equal. However, the size of DenseNet-121 is much smaller, at 89.6MB, and the largest size is 504.33 MB, obtained by ResNet-101.

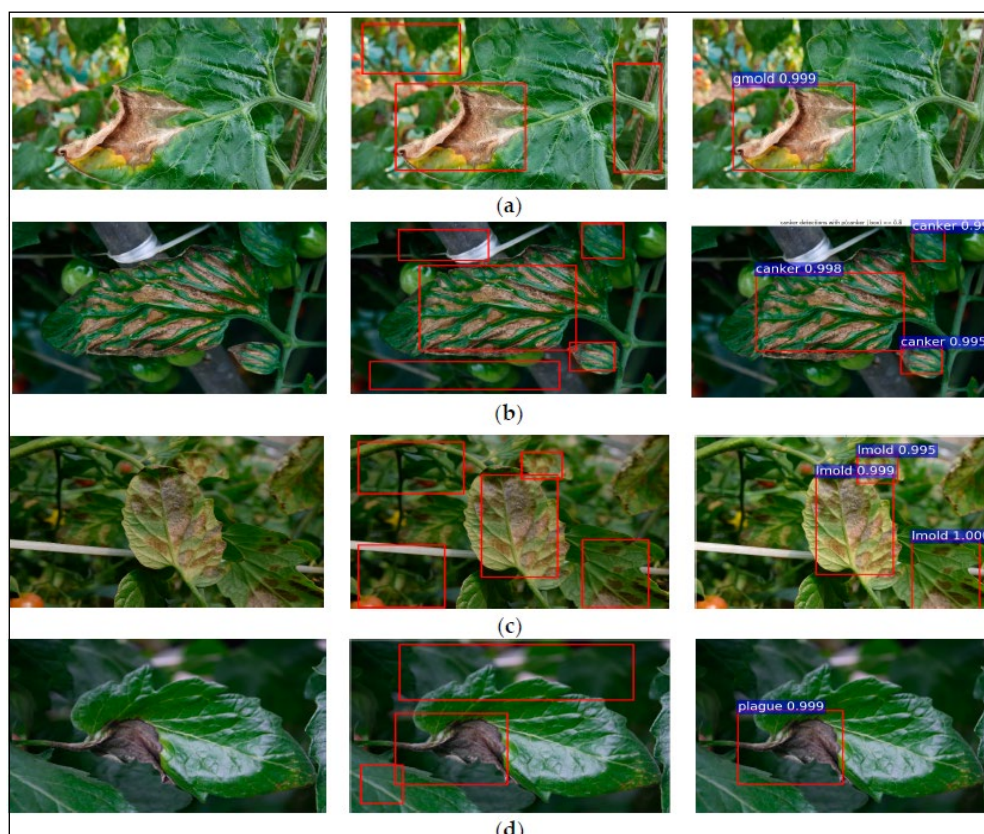


Figure 4. Detection result of disease and pests that affected tomato plants. (a) Gray mold (b) Canker (c) Leaf mold (d) Plague [109].

Figure 5 presents the details on the image annotation and its deep learning approach technique. Low-level features are used to represent images in image classification and retrieval. The initial stage in semantic comprehension is to extract efficient and effective visual features from an image's unstructured array of pixels. The performance of semantic learning approaches is considerably improved by appropriate feature representation. Numerous feature extraction techniques, including image segmentation, color features, texture characteristics, shape features and spatial relationships, have been proposed [115]. There are five categories of image annotation methods, which are generative model-based image annotation, nearest neighbor-based image annotation, discriminative model-based image annotation, tag completion-based image annotation and deep learning-based image annotation [25,26]. In the past decade, tremendous progress has been made in deep learning techniques, allowing image annotation tasks to be solved using deep learning-based feature representation. The most recent advancements in deep learning enable a number of deep models for large-scale image annotation. A CNN is commonly used by deep learning-based approaches to extract robust visual characteristics. Several versions of CNN architecture, such as LeNet, VGG, GooLeNet, etc., have been proposed. The following section describes the most commonly employed CNN architectures. The four types of image annotation are image classification, object detection or recognition, segmentation and boundary recognition. All of these task types can be annotated using deep learning techniques. The training process of deep learning can be supervised, unsupervised or semi-supervised, depending on how the neural network is used. In most cases, supervised learning is used to predict a label or a number. Commonly used benchmarks for evaluating image annotation techniques are based on the performance metrics. Section 4.8 provides the specifics on performance evaluation metrics.

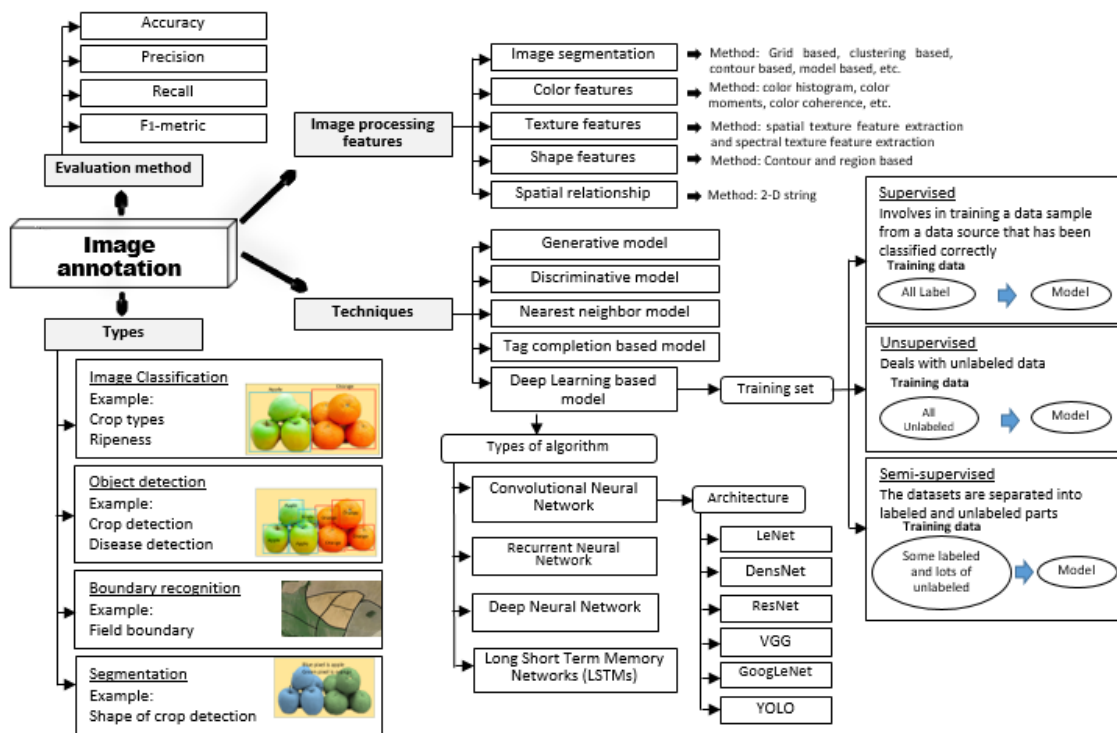


Figure 5. Image annotation for deep learning-based technique.

3. Deep Learning Architecture

A CNN is a special type of multilayer neural network used to recognize visual patterns directly from pixel images with minimal processing. The computer views an image as an array of numbers representing each pixel. Therefore, it is important that the relationship between the pixels persists even after the network has processed the image. To store the spatial relationship between pixels, a CNN is used, in which various mathematical operations are stacked on top of each other to create layers of the network [38].

The CNN architecture consists of convolutional layers, pooling layers and fully connected layers [116]. The basic architecture of a CNN is displayed in Figure 6.

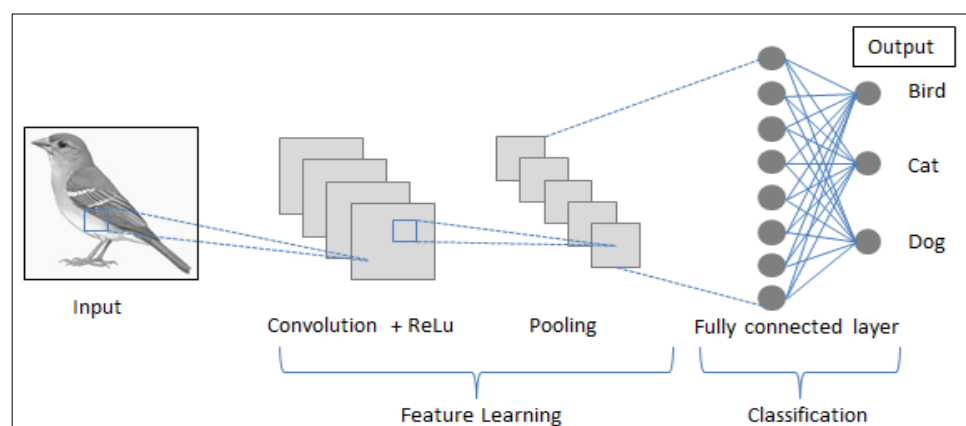


Figure 6. Basic architecture of CNN.

3.1. Convolutional Layer

In the feature learning process, the input image implemented with convolutional operation transfers the input matrices with convolutional kernels or can be understood as filters. These convolutional kernel operations, namely channels, kernel size, strides, padding and activation function, are used in a conventional image processing technique where the parameter needs to be set manually. These operations should be determined and

optimized based on the practical problem [117]. Each kernel slides over the input images and extracts features from the images. The sliding filter or kernel that slides horizontally and vertically is known as convolutional operation [118]. Das et al. [119] have explained the convolution process of strides and padding, where the strides act to reduce the data size by slides in each step in feature maps. The dimensions for the feature map can be maintained through the padding process. Padding will add zeros to the input matrix symmetrically. The process of strides and padding are shown in Figure 7.

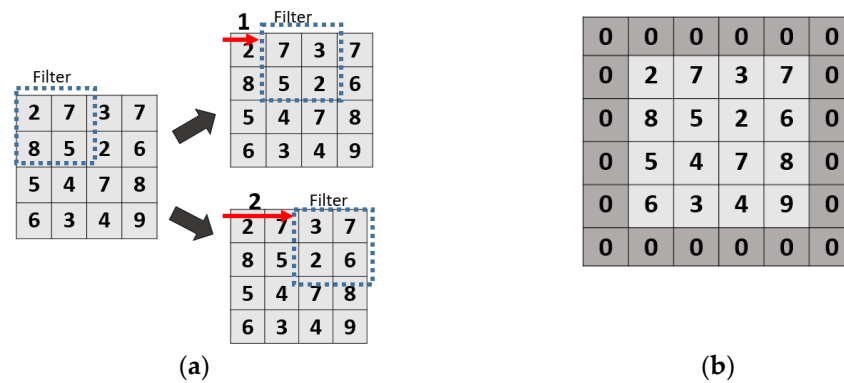


Figure 7. Convolution process: (a) stride, (b) padding.

The sliding process connects each neuron after the shift and provides a complete tiling for the input image. All the weights and biases for all neurons are combined to detect the same feature for all locations of the input image [120]. The output for the next layer, $a_{i,j}$ for convolutional operation, is computed as follows:

$$a_{ij} = \sigma((W * X)_{ij} + b) \tag{1}$$

where σ is non-linearity introduced in the network, W is the filter or kernel that slides over the input image, X is the input that is provided to the layer and b is the bias term of the filter [121].

3.2. Activation Function

Rectified linear unit (ReLU) is a most notable non-saturated activation function used to enhance the performance of a CNN. The operation of ReLU is shown in Figure 8. It is defined as in (2), where $z_{i,j,k}$ is the activation function input at (i, j) on the k th channel. *Max* operation in the equation allows the computation to be faster than the activation function of the sigmoid of tanh and does not face a gradient vanishing problem like tanh and sigmoid functions. Moreover, it allows the network to easily achieve sparse representation while inducing sparsity in the hidden units [116,122].

$$a_{i,j,k} = \max(z_{1,j,k}, 0) \tag{2}$$

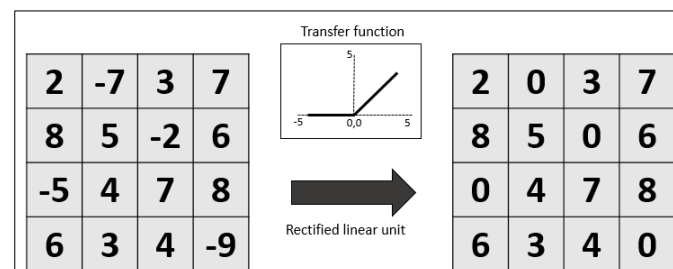


Figure 8. ReLU operation.

3.3. Pooling Layer

The pooling layer was firstly introduced in [123] in order to minimize the processing of the data. Pooling layers, also known as downsampling, generate smaller feature maps by reducing the parameter number and dimensionality in the input images. Even larger images are shrunk down, and the most important features in the images are preserved. The maximum values from each patch are kept by preserving the best fit in the feature [124]. There are two commonly used pooling functions, which are average pooling and maximum pooling. The average pooling calculates the average value at each patch on the feature map, and the maximum pooling calculates the maximum value on the feature map. The example of these pooling operations are shown in Figure 9.

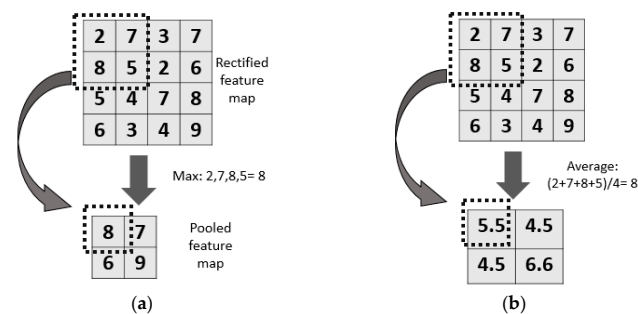


Figure 9. Example of pooling operation: (a) maximum pooling, (b) average pooling.

3.4. Fully Connected Layer

The fully connected layer is the final layer after the convolutional and the pooling layers. Here, the data are transformed to a one-dimensional layer and each neuron is connected directly to a neuron in the previous layer. The structure for this layer may consist of one or more hidden layers. The softmax activation function is usually applied in a fully connected layer to classify the input by generating a probability between 0 and 1. A softmax activation function is defined as in Equation (3) [125].

$$f_{c1} = f\left(b + \sum_{q=1}^M w_{1,q} * o_q\right) \quad (3)$$

3.5. Loss Function

In every CNN architecture, the last layer is called the output layer. The final classification occurs by calculating the prediction error produced by the CNN over the training data using a loss function. The loss function is the crucial component of the CNN to predict error through gradient calculation. Most of the studies on CNNs employ softmax or cross-entropy loss as the encoded output [126,127].

4. Improvement of CNN Architecture

CNNs received proper attention after the success of the AlexNet architecture in 2012 and this achievement was the start of the other CNN architectures [128]. The others CNN architectures are described in the next subsection.

4.1. LeNet

LeNet was the earliest CNN architecture, introduced by LeCun [129] in 1998. The structure consists of three convolutional layers and two fully connected layers. The architecture of LeNet is shown in Figure 10. The network contains five layers with learnable parameters, and combines average pooling and three sets of convolutions layers. There are two fully connected layers after the convolution and pooling process. At the end, a softmax classifier sorts the images into their appropriate categories. The study presented in [130] employed this architecture to detect and identify plant disease of potato and tomato.

A batch size of 150 epochs was used to train the model and resulted in accuracy of detection and recognition of 99%.

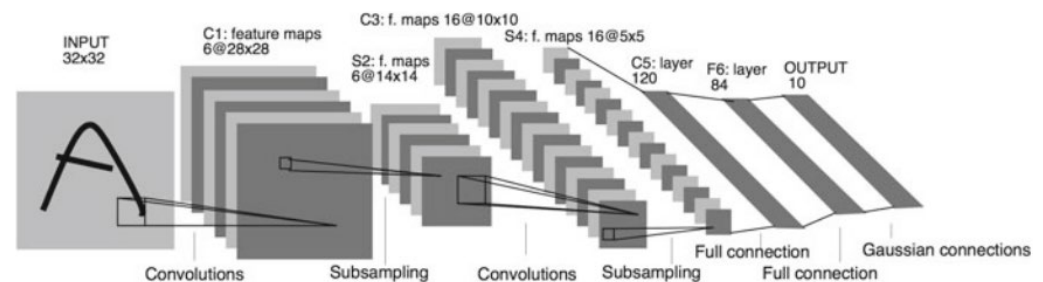


Figure 10. Architecture of LeNet [129].

4.2. AlexNet

AlexNet was proposed by Alex Krizhevsky [131] in 2012 during the ImageNet Large Scale Recognition Challenge and won the competition. The proposed architecture reduced error from 26% to 15.3% by utilizing the convolutional layers, max pooling layers, data augmentation, dropout, ReLU activations and SGD. AlexNet with 60 million parameters has eight layers, five convolutional layers and three fully connected layers. Every convolutional and fully connected layer used non-saturated ReLU gives the training response over tanh and sigmoid is improved [132]. Figure 11 shows the architecture of the AlexNet convolutional network that was proposed by Patino et al. [133] in classification of tropical fruits with 2633 images of fruits divided into 15 categories, including high variability and complexity. The authors of [134] employed AlexNet to train different datasets consisting of vegetables images. According to the experiment, the accuracy rate reached 92.1% compared to the SVM method with 80.5%.

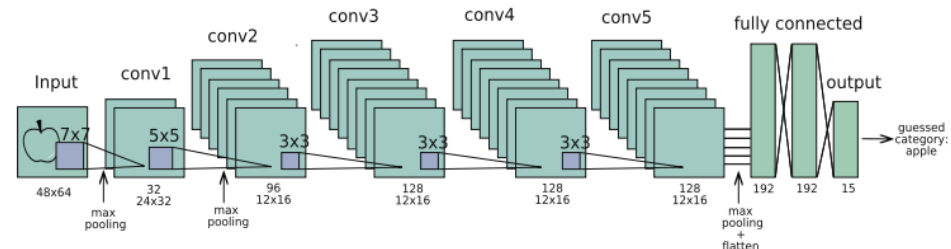


Figure 11. Architecture of AlexNet [133].

4.3. VGG

VGG architecture was first proposed by Simonyan and Zisserman [135] in 2014 by improving AlexNet by changing the kernel filter's size. At the same time, the generation of VGG aimed to improve the training time and reduce the number of parameters. It has been applied in various image classification tasks and was trained on more than 14 million images consisting of 1000 classes. It improved the AlexNet model that was considered the most popular image classifier and carried with it the ReLU tradition of AlexNet. There are many variants of VGGNet, including VGG-16, VGG-19, etc. The architecture of VGG-16 consists of a block of five convolutional layers and three fully connected layers containing 138 M parameters [136]. Figure 12 shows the architecture of VGG-16 as proposed by [137] in classification of jujube. Contrasting with AlexNet, VGG-16 has a deeper network and uniform structure consisting of 16 trainable layers containing 13 convolutional layers and three fully connected layers.

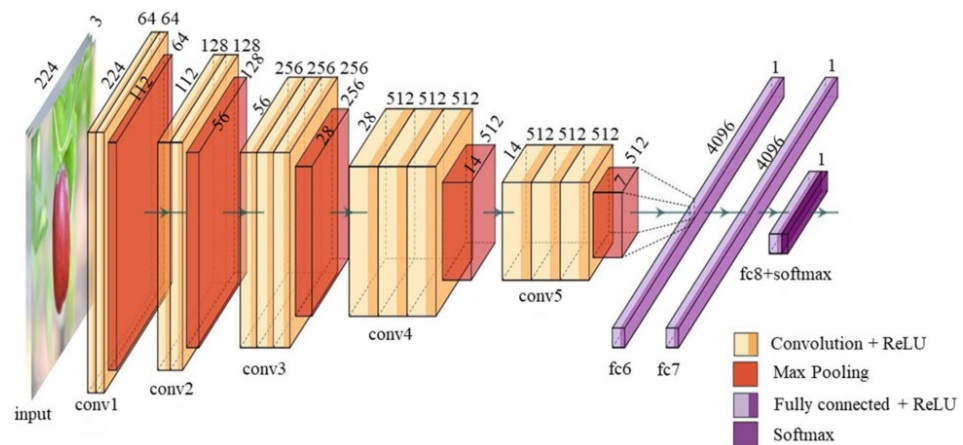


Figure 12. Architecture of VGG-16 [137].

4.4. GoogLeNet/Inception

GoogLeNet is based on the architecture of Inception and uses the module that allows the network to choose between multiple convolution filter sizes in each block. It was proposed by research at Google in 2014 and won the ILSVRC 2014 image classification challenge. The error rate generated by GoogLeNet showed a significant decrease compared to AlexNet. The architecture consists of a 22-layer deep network assessing the quality in detection and classification [138]. Then, the authors of [139] improved the architecture to InceptionV3 by updating the ImageNet classification accuracy. The updated Inception is referred to as InceptionVN. Then, in 2016, the architecture of Inception was updated to InceptionV4 by combining the architecture of Inception together with residual connection in research by Ni et al. [140].

Ni et al. [141] implemented GoogLeNet due to its superior performance in identification of fruit and vegetables. This architecture was used to monitor the change process of banana. The model was trained for 4320 iterations to recognize the freshness of banana. The model obtained recognition accuracy of 98.92%. Its architecture is illustrated in Figure 13.

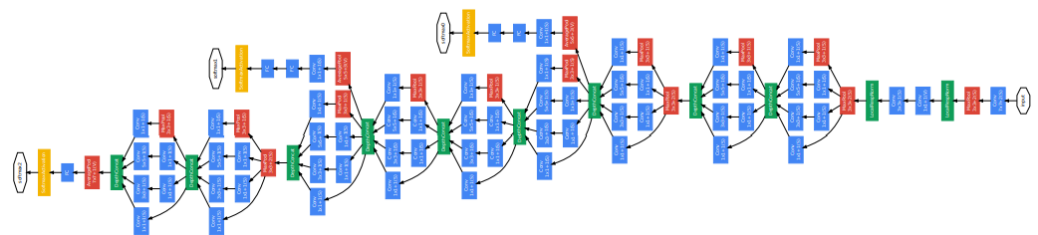


Figure 13. GoogLeNet architecture [138].

4.5. Residual Network (ResNet)

ResNet is a specific type of neural network that was introduced by He et al. [142] in 2015 and won 1st place in the ILSVRC 2015 competition by achieving an error rate of 3.5%. It has the ability to train a network with 100 layers and 1000 layers. The layer in ResNet receives the input from the previous layer and its residual units. The architecture consists of 34 layers, starting with one additional maxpooling layer, and ends with one average pooling layer [143]. The architecture of ResNet is shown in Figure 14.

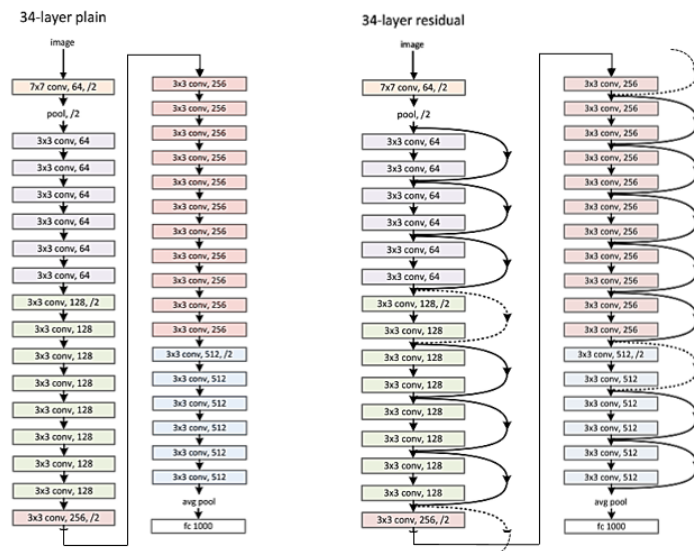


Figure 14. Network architecture for plain network with 34 parameter layers and residual network with 34 parameter layers [142].

4.6. DenseNet

DenseNet refers to a densely connected convolutional network introduced by Huang et al. [144] and it has an interesting pattern of connections, in which each layer is connected to the others within a dense block. All previous layers are used as input and its own feature maps are used as the input for all subsequent layers. This means all layers are able to access the feature maps. DenseNet can alleviate the vanishing gradient problem, promote feature reuse, strengthen feature propagation and significantly reduce the number of parameters. The structure of DenseNet with five layers and expansion of four is shown in Figure 15. The limitation of DenseNet is the large memory consumption. Therefore, Huang et al. [145] suggested CondenseNet to reduce the memory and speed it up by learning groups of convolution operations and pruning while training [146].

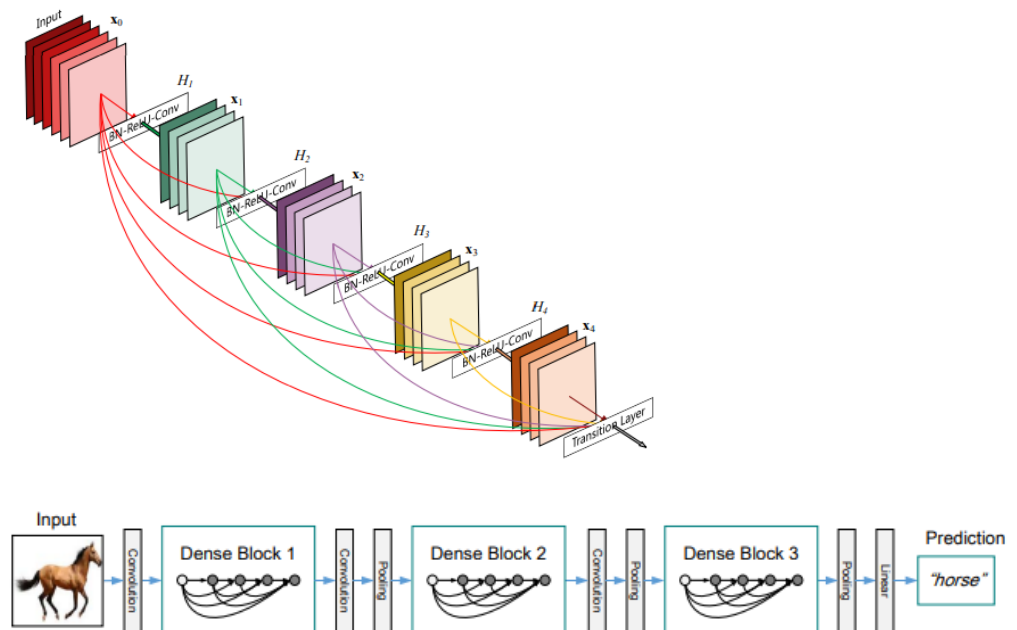


Figure 15. Dense block and layers in DenseNet [144].

4.7. You Only Look Once (YOLO)

YOLO was developed by Redmon [147] in 2015 to reframe cognitive problems as regression problems rather than classification problems. YOLO uses a single neural network to predict the bounding box and assign class probabilities. The model of YOLO is simple and able to train directly from full images. The loss function that is trained by YOLO corresponds directly to detection performance and the entire model is trained together. The architecture of YOLO is shown in Figure 16. It has 24 convolutional layers used to extract features from an image and ends with two fully connected layers that are used to predict the probabilities and coordinates of the output. There are many variants of YOLO that have been developed as an improvement of the previous version, namely YOLOv2 [148], YOLOv3 [149] and YOLOv4 [150]. Basically, the enhancement of the version is based on the framework design where the usual YOLO uses DarkNet that is trained on ImageNet. Then, the framework for YOLOv2 was improved to DarkNet-19, YOLOv3 with Darknet53 and YOLOv4 with DarkNet with CSPDarkNet53. Lippi et al. [151] preferred YOLO in early detection of pests as this model represented the fastest and most effective solution. Among these various versions of YOLO, the authors implemented YOLOv4 as it has been proven to outperform the previous ones in terms of accuracy and speed on the assorted standard dataset. YOLOv3 with Darknet53 framework has been employed by Chang et al. [152] to achieve real-time plant species recognition. The experiment’s findings demonstrate that the deep classifier was able to identify three different plants. Gai et al. [153] improved YOLOv4 in a cherry fruit detection application. The model of YOLOv4 was improved by replacing its backbone network, CSPDarkNet53, with DenseNet. The improvement generated advanced feature extraction, deepened the network structure and provided higher speed detection than the previous YOLOv4. The average accuracy given by the improved YOLOv4 is 0.15 higher than YOLOv4. In 2020, Jocher [154] released YOLOv5, which has fast, accurate and easy to train characteristics. It is well known for successful real-time object detection trained on the COCO dataset. The backbone for YOLOv5 is cross stage partial network (CSPNet) that is used to extract rich informative features from an input image and, by utilizing deeper networks, the processing time has been improved. YOLOv5 is implemented in detecting wheat spikes using UAV [155], detecting maturity of strawberry fruit [156], detecting defects of kiwi fruit [157] and detecting apple fruit [158].

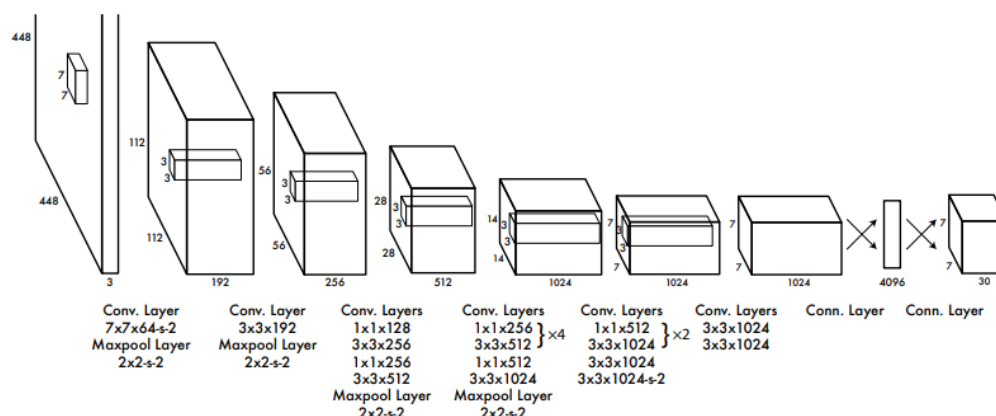


Figure 16. The architecture of YOLO [147].

There are many more ConvNets that have been proposed. ConvNets have improved greatly over time, owing primarily to increased processing power, new concepts, experiments and worldwide interest in deep learning. Those ConvNets are summarized in Table 1. The accuracy values were taken from image classification on ImageNet, the database platform consisting of a large visual database intended for use in the development of visual object recognition software.

Table 1. ConvNet details.

CNN Architecture	Year	Developed by	Characteristics	ImageNet Top Accuracy	ImageNet Top-5 Accuracy	Number of Parameters
LeNet	1998	Yann LeCun et al. [159]	Small and easy to understand	98.35%	-	60 thousand
AlexNet	2012	Alex Krizhevsky et al. [131]	First major CNN model that used GPU for training	63.3%	84.60%	60 million
VGG-16	2014	Simonyan, and Zisserman [135]	-Good architecture for particular task benchmark -Attractive feature of architectural simplicity comes at high cost -Progressing network requires a lot of computation -Has 16 layers	74.4%	91.90%	138 million
VGG-19	2014	Simonyan and Zisserman [135]	-Has 19 layers	74.5%	90.9%	144 million
GoogLeNet/InceptionV1	2014	Google [138]	-Designed to work well under strict constraint of memory and computational budget -Trains faster than VGG	74.80%	92.2%	4 million
InceptionV3	2014	Szegedy et al. [139]	Has higher efficiency and deeper network compared than InceptionV1 and -V2	78.8%	94.4%	24 million
InceptionV4	2014	Szegedy et al. [140]	Has more Inception modules than InceptionV3 and uniform, more simplified architecture	80.0%	95.0%	48 million
Inception-ResNetV2	2014	Szegedy et al. [140]	-Hybrid Inception version with enhancement of recognition performance -Has computational cost similar to InceptionV4	80.1%	95.1%	56 million
YOLO	2015	Joseph Redmon [147]	Superb speed (45 frames per second)	76.5%	93.3%	60 million
ResNet-50	2015	Kaiming He [142]	-Introduces a skip connection to adapt the input from the previous layer to the next layer by maintaining the input -Deep network with 50 layers	76.0%	93.0%	26 million
ResNet-152	2015	Kaiming He [142]	Very deep network of 152 layers	77.8%	93.8%	60 million
DenseNet-121	2016	Gao Huang et al. [144]	-Has 120 convolutions and 4 AvgPool -Each layer has connection to every other layer	74.98%	92.29%	8 million
DenseNet-264	2016	Gao Huang et al. [145]	-Has 264-layer DenseNet	77.85%	93.88%	34 million
YOLOv2/YOLO9000	2016	Joseph Redmon and Ali Farhadi [148]	-Improvement from YOLOv1 in variety of ways -Uses Darknet-19 as a backbone -Real-time object detection model with a single stage	86%	-	59 million

Table 1. Cont.

CNN Architecture	Year	Developed by	Characteristics	ImageNet Top Accuracy	ImageNet Top-5 Accuracy	Number of Parameters
YOLOv3	2018	Joseph Redmon and Ali Farhadi [149]	-Improved version of YOLOv1 and v2 -Significant differences between the previous versions in terms of speed, precision and class specifications	86.3%	-	86 million
Big Transfer (BiT-L)	2019	Kolesnikov et al. [160]	Pre-trains on large supervised source datasets and fine-tunes the model on a target task	87.54%	98.5%	928 million
YOLOv4	2020	Alexey [150]	-Most recent YOLO series version for fast object detection in a single image -Uses CSPDarknet53 as a backbone	86.8%	-	193 million
YOLOv5	2020	Glenn Jocher [154]	-Has three important parts: backbone, neck and head model -CSPNet is used as a backbone	87.1%	-	296 million
Noisy Student Training EfficientNet-L2	2020	Xie et al. [161]	-A semi-supervised learning method that performs well even when labeled data are plentiful	88.4%	98.7%	480 million
Meta Pseudo Labels	2021	Pham et al. [162]	Has a teacher network that generates pseudo labels from unlabeled data in order to teach a student network	90.2%	98.8%	480 million

4.8. Performance Metric

Deep learning methods have overcome the agricultural issues in crop detection [163], counting [98,164], classification [165], segmentation [166], disease diagnosis [118,167], etc.

After creating a model based on a deep learning technique and receiving some output in the form of a class, the next step is to use test datasets to determine how effective the technique is. The most crucial aspect in data science research is to evaluate the model, which determines how accurate the prediction is. Deep learning algorithms are evaluated using a variety of performance metrics. Each deep learning result generates accuracy based on the percentage of accuracy, intersection of union (IoU), performance score (F_1), mean average precision (mAP) and correlation coefficient (R^2). This study is focused on these performance metrics that were employed in the previous studies in this article. It is crucial to pick the right metrics to evaluate the deep learning technique used. The metrics are used to determine how deep learning algorithm performance is evaluated and compared. The simplest intuitive performance metric is accuracy, which is just the ratio of properly predicted observations to the total observations. A high accuracy percentage shows which model is the best and how good the model has performed. IoU, also known as the Jaccard index, is the computation of the ratio of the intersection and the union of two sets. These region-based measures do not examine the accuracy of the segmented region boundaries, which is significant during automated tree training operations. In addition, IoU measurements are strict because they penalize false positives and favor regional uniformity over border accuracy [168,169]. The F_1 score computes the performance of detection by using recall and precision. Recall and precision measure the fraction of true-positive objects that are successfully detected and objects in the prediction [106]. The F_1 score is greatest at 1 (perfect precision and recall) and lowest at 0. In other words, recall is the number of well-predicted positives divided by the total number of positives. It shows the percentage of positives that are well predicted. Precision is similar to recall as it shows the number of positive predictions generated. It divides the number of predicted positives by all the positives predicted. If the value of precision is high, this means the majority of the positive predictions for the objects are correctly predicted as positive. The calculations of accuracy, precision, recall, F_1 score, *IoU* and class accuracy are shown in (4)–(8). True positive is the annotation that is correctly drawn with an *IoU* of > 0.5 , true negative is every part of an image that does not predict an object, false positive is a missing annotation and false negative is an annotation that has an *IoU* score of < 0.5 .

$$Accuracy = \frac{True\ Positive + True\ Negative}{True\ Positive + False\ Positive + True\ Negative + False\ Negative} \quad (4)$$

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (5)$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (6)$$

$$F_1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (7)$$

$$IoU = \frac{True\ Positive}{True\ Positive + False\ Positive + False\ Negative} \quad (8)$$

5. Results and Discussion

Deep learning architecture is adaptable, which means it may be used to solve new challenges in the future. Moreover, this method can be applied to a wide range of tasks and data types. Therefore, this study summarizes the previous studies implementing a deep learning algorithm in the agriculture sector, shown in Table 2. The application of deep learning in agriculture is to detect and classify crops, disease, yield estimation, border extraction, etc. The first step in those studies involves gathering a correctly annotated dataset

that is large enough for a complex model to produce satisfactory results when trained on it. In order to perform successfully, a CNN requires a lot of training data. If the dataset is not particularly large, though, image augmentation can be employed to make a small dataset appear larger. It has been observed that augmentation of existing data, rather than collecting new data, enhances the classification accuracy of a deep learning model. In fact, the data augmentation technique is able to avoid overfitting problems and achieve high accuracy.

Basically, all the studies proposed smartphones to capture the object images due to the advancement of the technology's resolution, while also being simple and cost-effective. The smartphone's rapid evolution has elevated it to the foremost choice in the area of the agriculture industry. Another ability provided by the smartphone is to detect the object in real-time with the advancement of the deep learning employment method in the application of object detection. Most methods to extract the border are based on the utilization of satellite images due to the imagery captured, guaranteeing a powerful method without physical contact, wide views and consisting of a big data revolution. The image that is processed through a satellite is computationally intensive, therefore, deep learning is very helpful in analyzing the image provided. In image processing, the images most commonly used are red, green and blue (RGB). These images will generate a color image on the screen when RGB images are stacked on top of each other.

According to the previous studies, many of the datasets were utilized to train the images and annotated images of multiple regions have been proposed. Images are initially segmented or bounded into multiple regions such as color, shape and texture features and are extracted. Automatic image annotation aims to learn a semantic concept model from a large number of datasets or image samples and apply it to new images. After the training, images are labeled with semantic labels, and test images can be assessed using keywords, similar to how the training of image features. As presented in the summary table, many studies on automatic image annotation have applied deep learning-based approaches. Convolution procedures extract image features and a deep neural network training model establishes the connection between the images and labels. A deep learning algorithm is generated based on the image features primarily using the supervised technique. The algorithm is tested with sample images and the prediction performance is analyzed. The annotation result generates either a bounding box or a segmentation with the predicted performance value. Deep learning has made tremendous progress in image labeling by automatically annotating images to recognize the plant species, maturity and health. The proposed deep learning technique is tested with other techniques to test its performance. The best performance or higher accuracy results indicate that the proposed technique is a better model to employ when detecting or classifying an object. In addition, the capability of the proposed deep learning technique is tested on a variety of datasets. Even when using the same technique, various datasets generate different performances due to differences in dataset properties. However, during the training process, the model's performance can be improved by increasing the number of epochs and iteration value. As proposed in [170], different datasets are used to detect various diseases in different plants. The disease is successfully detected by implementing a VGG-16-based deep learning method, which is compared to InceptionV3 and GoogLeNetBN. VGG-16 outperforms the other techniques used with 98.8% accuracy. According to the study, pre-training with plant-specific tasks reduced the impact of overfitting for a deeper Inception model, but the VGG-16 model demonstrated better generalization when adapting to new data. In addition, the fact that VGG-16 outperforms the Inception technique is because of a lack of variability in the dataset, which limits the implementation of deeper architectures. The summaries of the various techniques from the previous studies provide a wide overview of the performance metric result when deep learning algorithms are used for the task. This critical information about performance metric can help researchers choose a suitable deep learning algorithm for their studies. In some circumstances, the model does not produce particularly accurate predictions. Training the method also takes a long time. As a result, it is crucial to improve accuracy while decreasing training time. The number of epochs, input size, network depth and width or slow weight update can contribute to training time taken.

Table 2. Summary of previous studies on deep learning implementation.

Authors	Research Problem	Dataset Collection Method	Dataset Pre-Processing Method	DL Method	Datasets Used	Result (Accuracy, Score, Detection Time)
[171]	Disease detection	Digital camera and smartphone used to collect tomato images in planting greenhouse	Bounding box	Yolo V3	Tomato	Accuracy = 92.39% (20.39 ms)
[172]	Disease detection	Collected RGB color images using Nikon 7200d camera with image resolution 4000 × 6000 pixels	-	CNN (SVM classifier)	4447 Turkey-PlantDataset	Accuracy score: 97.56% (PlantDiseaseNet-MV) 96.83% (PlantDiseaseNet-EF)
[167]	Disease detection	Collected RGB images using Canon Rebel T5i DSLR and smartphones	-	CNN	3651 apple leaves	Accuracy = 97%
[173]	Disease detection	Image collected by two digital cameras and five smartphone cameras	-	GoogLeNet, Xception and Inception-ResNet-v2	4727 tea leaves	Accuracy = 89.64%
[112]	Disease detection	Used initial version of the PV dataset that is publicly available	-	VGG-16 InceptionV3 GoogLeNetBN	14 crop plants, 38 crop–disease pairs, and 26 crop–disease categories	Accuracy = 98.8% (VGG-16)
[174]	Disease detection	Images are captured by hand	Augmentation	Ensemble of pre-trained DenseNet121 EfficientNetB7 and EfficientNet NoisyStudent	3651 high-grade images of apple leaves with various foliar diseases	Ensemble of pre-trained DenseNet121, EfficientNetB7 and EfficientNet NoisyStudent: Accuracy = 96.25% DenseNet121: Accuracy = 95.26% EfficientNwtB7: Accuracy = 95.62% NoisyStudent: Accuracy = 91.24%
[136]	Disease classification	Image source is online dataset from Plant Pathology 2020-FGVC7	Data augmentation	ResNetV2	1821 images of apple tree leaves	Accuracy = 94.7%
[175]	Plant disease detection	Mobile phone	-	Baseline model of ResNet18 ResNet34 ResNet50	54,305 leaf images, PlantVillage and 1747 coffee leaves	Accuracy = 99%

Table 2. Cont.

Authors	Research Problem	Dataset Collection Method	Dataset Pre-Processing Method	DL Method	Datasets Used	Result (Accuracy, Score, Detection Time)
[170]	Crop detection	Multicopter DJI Phantom 4 drone using RGB camera with 4000 × 3000 pixels	Bounding box	EfficientDet-D1 SSD MobileNetv2 SSD ResNet50 Faster R-CNN ResNet50	197 images of paddy seedlings	EfficientDet-D1: Precision = 0.83 Recall = 0.71 F ₁ = 0.77 Faster R-CNN: Precision = 0.82 Recall = 0.64 F ₁ = 0.72
[176]	Fruit detection	Smartphone (Galaxy S9, Samsung Electronics) with 5312 × 2988 pixels	Bounding box	Canopy-attention-YOLOv4	480 raw apple tree images	Precision = 94.89% Recall = 90.08% F ₁ = 92.52% (0.19 s)
[177]	Fruit detection	QG Raspberry Pi_Sony IMX477 and OAK-D color camera	Bounding box	SSD MobileNet-V1	1929 images of grape bunches	mAP = 66.96%
[178]	Fruit detection	Canon Powershot G16 camera	Bounding box	Improved YOLOv5	1214 apple images	Recall = 91.48% Precision = 83.83% mAP = 86.75% F ₁ = 87.49%
[179]	Fruit detection	Dataset obtained from GrapeCS-ML and Open Image Dataset v6	Bounding box	YOLOv3 YOLOv4 YOLOv5	2985 images of grapes	YOLOv5: F ₁ =0.76 YOLOv4: F ₁ =0.77
[49]	Fruit detection	Image collected with 4032 × 3024-pixel smartphone camera (iPhone 7 Plus, Apple)	Bounding box	AlexNet ResNet101 DarkNet53, Improve YOLOv3	849 apple images	Improve YOLOv3: F ₁ = 95.0% DarkNet53+YOLOv3: F ₁ = 94.6% AlexNet+Faster R-CNN: F ₁ = 91.2%
[153]	Fruit detection	Image captured by 3000 × 4000 pixel Sony DSC-HX400 Camera and 40 million-pixel Huawei mobile phone	Bounding box	Improved YOLO-V4	400 images of cherries fruit	F ₁ score = 0.947 Iou = 0.856 (0.467 s)

Table 2. Cont.

Authors	Research Problem	Dataset Collection Method	Dataset Pre-Processing Method	DL Method	Datasets Used	Result (Accuracy, Score, Detection Time)
[103]	Fruit position detection and harvesting robot	Harvesting robot equipped with stereo camera and robot arm	-	SSD (VGGNet) R-CNN YOLO	169 images of apples	SSD (VGGNet): Accuracy = 90% (2 s)
[180]	Fruit detection and counting	Captured images by DJI MAVIC Air2 drones, SLR cameras (Panasonic DMC-G7) and Honor 20 mobile phone	Bounding box	YOLOv5-CS (citrus sort)	More than 3000 original images of green citrus	mAP = 98.23% Precision = 86.97% Recall = 97.66%
[98]	Fruit counting	Collected 128 × 128-pixel images from Google Images	Generate synthetic image	Inception-ResNet	24,000 tomato images	Average accuracy = 91%
[94]	Olive fruit fly detection and counting	Collected images from Dacus Image Recognition Toolkit (DIRT)	Bounding box	Modified YOLOv4	848 images of olive fruit fly	mAP = 96.68% (52.46 h) Precision = 0.84 Recall = 0.97 F ₁ score = 0.90
[181]	Leaf counting	Captured image by Cannon Rebel XS camera	Bounding box	Faster R-CNN Tiny YOLOv3	1000 images of Arabidopsis plants	F ₁ score = 0.94
[102]	Fruit count	Flatbed scanner (Plustek, OpticPro A320) with 3600 × 5200 pixels	Patch classifier	LeNet DenseNet	2552 images of mature inflorescences	DenseNet: Precision = 91.8% Recall = 92% LeNet: Precision = 77.8% Recall = 76.2%
[182]	Plant counting	RGB images taken using UAV with 256 × 256 pixels	Segmentation	Mask R-CNN	Potato and lettuce plants	Potato: Precision = 0.997 Recall = 0.825
[183]	Crop classification	Images captured by Landsat-8 satellites	Semantic segmentation	DNN	Corn, soybean, barley, spring wheat, dry bean, sugar beet and alfalfa area	F1=0.8476 Precision = 0.8463 Recall = 0.8536
[184]	Crop classification	Dataset taken from previous study in which the size of images is 1280 × 1024 pixels	Augmentation (crop)	AlexNet	13,200 white cabbage seedlings	Accuracy = 94%

Table 2. Cont.

Authors	Research Problem	Dataset Collection Method	Dataset Pre-Processing Method	DL Method	Datasets Used	Result (Accuracy, Score, Detection Time)
[185]	Fruit classification	RGB images captured using smartphone camera (LG-V20)	Data augmentation technique (flipping, horizontal, vertical)	VGG-16	1300 four classes of dates	Accuracy = 98.49% Precision = 96.63% Recall = 97.33%
[137]	Fruit classification	Images captured using Nikon D7500 camera with 3024 × 4032 and 6000 × 4000 pixels	Image augmentation (rotation, flip, brightness, adjustment, contrast and saturation enhancement)	Jujube classification network based on DL technique SVM AlexNet VGG-16 ResNet	1700 images of 20 jujube varieties	Jujube classification network-based DL technique: Accuracy = 84.16% ResNet-18: Accuracy = 78.25% VGG-16: Accuracy = 71.42% AlexNet: Accuracy = 65.36% SVM: Accuracy = 60.84%
[186]	Classify fruit images	Image taken by digital camera (Nikon D7100)	-	VGG-16	440 images of litchi and lychee	Accuracy = 98.33%
[187]	Freshness and fruit classification	Real-world images from the internet	Augmentation (rotation and horizontal flipping)	RestNet-50 + RestNet-101	Fresh and rotten fruits, e.g., apple, banana, orange, lemon, pear, strawberry and others	RestNet-50+RestNet-101: Average accuracy = 98.50% for freshness 97.43% for classification VGG-16: 94.79% for freshness 94.90% for classification
[188]	Classification of seedless and seeded fruit	Images taken by a digital camera (COOLPIX P520, Nikon) with image size of 1600 × 1200 pixels	Augmentation (brightness changes, rotation and horizontal and vertical flip)	VGG16 RestNet-50 InceptionV3 InceptionResNetV2	599 images of seeded and seedless fruit	VGG-16: Accuracy = 89% ResNet50: Accuracy = 86% InceptionV3: Accuracy = 91% InceptionResNetV2: Accuracy = 85%
[189]	Detection and classification	Recorded using video camera with full HD definition (1920 × 1080)	Bounding box	Tiny YOLOv3	Unripe, ripe and overripe coffee fruit at density of 5000 trees hectare ⁻¹	mAP = 84.0% F ₁ score = 82.0% Precision = 82.0% Recall = 83%

Table 2. Cont.

Authors	Research Problem	Dataset Collection Method	Dataset Pre-Processing Method	DL Method	Datasets Used	Result (Accuracy, Score, Detection Time)
[190]	Detection, segmentation and classification	Dataset taken from a previous study [191]	Segmentation	R-CNN	1036 reproductive structures (flower buds, flowers, immature fruits and mature fruit)	Average counting precision = 77.9%
[40]	Detection and segmentation	Images captured using camera	Bounding box	CNN	300 images of grapes	F ₁ score= 0.91
[106]	Detection and segmentation	Intel RealSense D-435 RGB-camera and Logitech C615 webcam	Semantic segmentation	DaSNet-v2	1277 images of apple trees (fruit and branches)	Recall = 0.868 Precision = 0.88 Accuracy = 0.873
[168]	Image segmentation	RGB images and 3D point cloud data (Kinect V2 sensor)	Semantic segmentation	SegNet	Apple trees	F ₁ score = 0.93
[93]	Yield estimation	Images from Google Images	-	LSTM GRU BLSTM BGRU	Tomato, potato	R ² = 0.97 to 0.99 (BLSTM)
[192]	Yield estimation	Image taken from UAV (DJI Phantom 4 Pro)	Bounding box	Region convolutional neural network	592 trees of apples	R ² = 0.86
[72]	Monitoring agricultural area	Images collected by optical satellite sensors of SPOT, Landsat-8 and Sentinel-1A	Aumentation (patch normalization)	Spatio-temporal-spectral deep learning	Paddy field area	F ₁ score = 0.93
[193]	Border extraction	Images collected by Sentinel-2 and Landsat-8 satellites	Semantic segmentation	ResUNet	Field border	Accuracy = 85.60%
[194]	Border extraction	RGB image and near-infrared-2 bands taken using WorldView-3 satellite image	Polygon	FCNN (UNet, SegNet and DenseNet)	Smallholder farm in Bangladesh	Precision for all proposed FCNNs is up to 0.8

Based on the result, most of the previous studies presented deep learning implemented for annotated image results as an accuracy percentage. This is because accuracy simply gives a ratio of correctly drawn annotations to total predicted annotations. Therefore, it is the most intuitive approach to evaluate a task's performance. Accuracy is not just a simple measurement, but it is also the least insightful when it comes to evaluating an annotation task's performance. In fact, most real-world annotation ignores false negatives and false positives which could lead to prejudice and inaccurate conclusions about the quality of the mission. In some cases, the authors prefer the F_1 score because it elegantly summarizes a model's prediction effectiveness by merging two competing metrics: recall and precision. The different usages of the performance for accuracy are employed when true positives and true negatives are more important, while when false negatives and false positives are critical, then the F_1 score is used. Another case is if the class distribution is similar, then the accuracy can be employed. Meanwhile, if there are imbalanced classes, then F_1 is a good choice. Overall, all these evaluation metrics help present the quality of the annotated images proposed using the deep learning technique.

Based on the previous studies, the image annotation implemented in agriculture is summarized in the pie chart shown in Figure 17. Most of the deep learning techniques are applied to detect and classify plants and their diseases. Crop detection and classification are difficult tasks due to the wide range of interclass forms, colors and textures. As a result of these limitations, there is a shortage of automated fruit classification systems for various groups. Crop detection and classification using an advanced information system could effectively identify the right fruit with the right nutrition. On the other hand, plant detection and classification are applied in harvesting robots to pick fruit and vegetables. Robotic harvesting has a high potential to be used in crop detection by reducing the cost of labor while improving fruit quality. In fact, the problems of plant diseases are a worldwide issue related to food production. Plant diseases adversely affect the economy and incur losses to farmers. Therefore, utilizing deep learning for annotating images in agriculture helps earlier disease detection and prevents the plants from becoming worse.

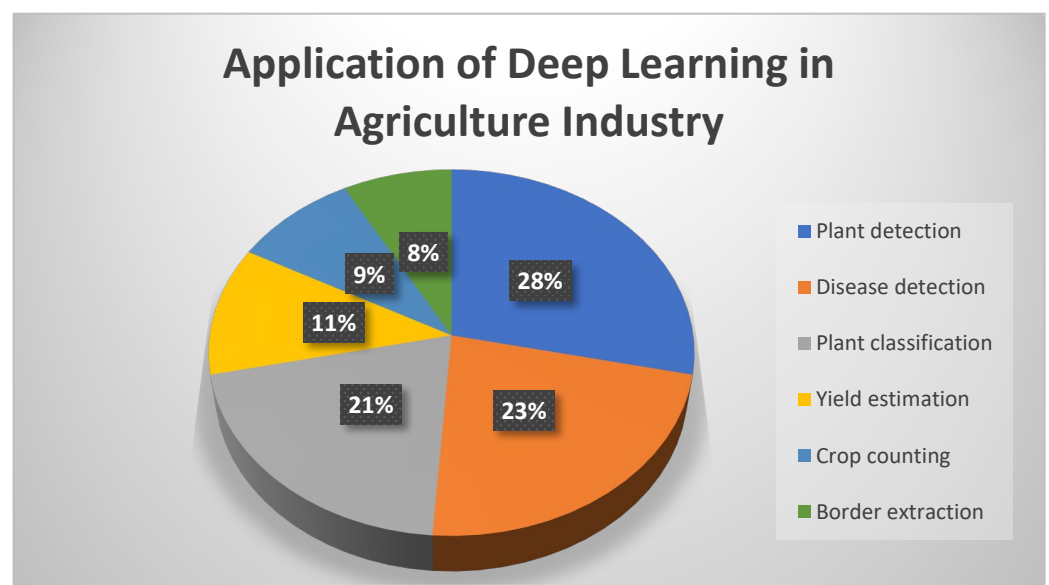


Figure 17. Image annotation applications in agriculture industry.

6. Conclusions

This study presented a comprehensive review of the application of image annotation using the deep learning technique in the agriculture field. Image annotation is extremely useful in the agriculture industry in increasing the crop production. It assists in recognizing and classifying the plants and their diseases. The employment of deep learning in image annotation generates a high performance on the dataset or image based on the accurate pre-

diction. In previous studies, bounding boxes were one of the most popular and recognized image annotation methods. Annotators are expected to outline the object in bounding boxes in accordance with the specified deep learning requirements. In addition, bounding boxes are also one of the least expensive and time-consuming annotation methods available.

The application of deep learning in agriculture industries, as well as the technologies presented to improve agriculture productivity, are studied in this article. The CNN type of deep learning has been widely used in the agriculture sector since it promises to recognize features from images without the need for human interaction. Moreover, the technique provided superior classification and precision accuracy when compared to other techniques. In this article, the implementation of various architectures of deep learning is discussed in order to evaluate the method's effectiveness in terms of plant and disease detection, classification, counting and segmentation of plants. This review has found that deep learning has high performance in image processing techniques.

Despite the efforts of numerous researchers, the task of developing a rapid and reliable fruit detection system remains under investigation. This is due to a wide range of colors, shape, sizes, textures and reflectance qualities of the fruit in field settings. Many advancements of the architecture network were created to improve image detection, classification and segmentation accuracy. It is crucial to find a suitable deep learning architecture in order to produce high accuracy, low error rate and shorter training time. The architectures of deep learning that are commonly used are summarized in this study. The results reveal that ConvNet's accuracy has been gradually improving over time in most circumstances. Each learning-based algorithm evaluation is an important aspect of any project. The model will generate satisfactory results when tested using a performance metric. Performance metrics assist in identifying how effectively the model generalizes to new data. The majority of the previous studies in this article employed the accuracy metric to evaluate model performance. However, accuracy is inappropriate when interacting with imbalanced data, in which the number of points in one class is significantly more than the number of points in another. Other performance metrics, including the F_1 score, recall and precision, can help to overcome this accuracy issue. For recommendations, the training time and inaccuracy in future studies on deep learning in agriculture should be reduced by employing advancement of deep learning architecture.

The technique of deep learning can be used more widely in agriculture industries to improve plant productivity and quality. To encourage the usage of greater intelligence, deep learning can be integrated with other technologies such as robotics and the Internet of Things (IoT). The use of deep learning in robotic harvesting, planting and logistics could be beneficial. Using IoT technologies, farmers can boost yields by controlling every variable in crop production, such as moisture levels, soil conditions, pest stress and microclimates. Precision agriculture allows farmers to enhance efficiency and minimize expenses by providing more precise strategies for planting and growing crops. However, these technologies may include all levels of security. It also addresses new particular security concerns such as accuracy, device and data integrity. Security issues could consist of hijacking autonomous devices such as UAVs and robots. If a malicious agent hijacks an autonomous system, the hijacker can control and direct the device remotely without authorization. This type of violence could have several consequences, including the inability of the system to fulfill a task. The malfunction could result in significant losses due to incorrect crop management, equipment damage and the autonomous system itself. Security concerns must be integrated into the system to maximize their effectiveness. It is critical to create security schemes to detect incidents and avoid corrupt or inconsistent data. As a result, advancing to the next levels of robots and IoT technologies necessitates solutions with security measures that provide dependability and accuracy in implementing these systems.

In addition, the requirement for massive amounts of labeled data remains a major obstacle for supervised deep learning methods. This problem is particularly apparent in the agriculture industry, where hundreds of images must first be annotated by humans before training. In addition, the labeling process frequently needs the participation of some

field experts who are in short supply. If there are not enough labeled data for supervised learning algorithms to work, semi-supervised learning can be utilized in the agriculture sector to solve real-world problems. Semi-supervised learning is an excellent compromise between supervised and unsupervised learning. Researchers have spent the majority of their time on organizing the data, therefore, semi-supervised learning allows working with limited data. This learning is typically employed when labeling or acquiring data is too complex or expensive. Additionally, it is also feasible to use it if the quality of labeled data is poor.

Author Contributions: Writing—original draft preparation, N.M. (Normaisharah Mamat); supervision and formal analysis, M.F.O.; writing—review and editing, R.A., S.B.B. and S.F.M.H.; methodology, N.M. (Normahira Mamat); and funding acquisition, R.A. and S.B.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Universiti Teknologi Malaysia (Profesional Development Research University research grant: Q.K130000.21A2.05E38) and the APC was funded by Qatar National Library.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Acknowledgments: The publication of this article was funded by the Qatar National Library. The authors would like to acknowledge the Qatar National Library for supporting the publication of this article and Universiti Teknologi Malaysia (Professional Development Research University research grant: Q.K130000.21A2.05E38) for financially supporting this research.

Conflicts of Interest: All authors declare that they have no conflicts of interest.

References

1. Khan, T.; Sherazi, H.; Ali, M.; Letchmunan, S.; Butt, U. Deep Learning-Based Growth Prediction System: A Use Case of China Agriculture. *Agronomy* **2021**, *11*, 1551. [[CrossRef](#)]
2. Ahmad, N.; Singh, S. Comparative study of disease detection in plants using machine learning and deep learning. In Proceedings of the 2nd International Conference on Electrical, Computer and Energy Technologies (ICECET), Prague, Czech Republic, 20–22 July 2021; pp. 54–59. [[CrossRef](#)]
3. Zheng, Y.-Y.; Kong, J.-L.; Jin, X.-B.; Wang, X.-Y.; Su, T.-L.; Zuo, M. CropDeep: The Crop Vision Dataset for Deep-Learning-Based Classification and Detection in Precision Agriculture. *Sensors* **2019**, *19*, 1058. [[CrossRef](#)] [[PubMed](#)]
4. Velumani, K.; Madec, S.; de Solan, B.; Lopez-Lozano, R.; Gillet, J.; Labrosse, J.; Jezequel, S.; Comar, A.; Baret, F. An automatic method based on daily in situ images and deep learning to date wheat heading stage. *Field Crop. Res.* **2020**, *252*, 107793. [[CrossRef](#)]
5. Santos, L.; Santos, F.N.; Oliveira, P.M.; Shinde, P. Deep learning applications in agriculture: A short review. In Proceedings of the Iberian Robotics Conference, Proceedings of the Robot 2019: Fourth Iberian Robotics Conference, Porto, Portugal, 20–22 November 2019; Springer: Cham, Switzerland, 2019; pp. 139–151.
6. Khan, N.; Ray, R.; Sargani, G.; Ihtisham, M.; Khayyam, M.; Ismail, S. Current Progress and Future Prospects of Agriculture Technology: Gateway to Sustainable Agriculture. *Sustainability* **2021**, *13*, 4883. [[CrossRef](#)]
7. Cecotti, H.; Rivera, A.; Farhadloo, M.; Pedroza, M.A. Grape detection with convolutional neural networks. *Expert Syst. Appl.* **2020**, *159*, 113588. [[CrossRef](#)]
8. Kayad, A.; Paraforos, D.; Marinello, F.; Fountas, S. Latest Advances in Sensor Applications in Agriculture. *Agriculture* **2020**, *10*, 362. [[CrossRef](#)]
9. Cheng, C.Y.; Liu, L.; Tao, J.; Chen, X.; Xia, R.; Zhang, Q.; Xiong, J.; Yang, K.; Xie, J. The image annotation algorithm using convolutional features from intermediate layer of deep learning. *Multimed. Tools Appl.* **2021**, *80*, 4237–4261. [[CrossRef](#)]
10. Niu, Y.; Lu, Z.; Wen, J.-R.; Xiang, T.; Chang, S.-F. Multi-Modal Multi-Scale Deep Learning for Large-Scale Image Annotation. *IEEE Trans. Image Process.* **2018**, *28*, 1720–1731. [[CrossRef](#)] [[PubMed](#)]
11. Blok, P.M.; van Henten, E.J.; van Evert, F.K.; Kootstra, G. Image-based size estimation of broccoli heads under varying degrees of occlusion. *Biosyst. Eng.* **2021**, *208*, 213–233. [[CrossRef](#)]
12. Chen, S.; Wang, M.; Chen, X. Image Annotation via Reconstitution Graph Learning Model. *Wirel. Commun. Mob. Comput.* **2020**, *2020*, 8818616. [[CrossRef](#)]
13. Bhagat, P.; Choudhary, P. Image annotation: Then and now. *Image Vis. Comput.* **2018**, *80*, 1–23. [[CrossRef](#)]
14. Wang, R.; Xie, Y.; Yang, J.; Xue, L.; Hu, M.; Zhang, Q. Large scale automatic image annotation based on convolutional neural network. *J. Vis. Commun. Image Represent.* **2017**, *49*, 213–224. [[CrossRef](#)]

15. Mori, Y.; Takahashi, H.; Oka, R. Image-to-word transformation based on dividing and vector quantizing images with words. In Proceedings of the First International Workshop on Multimedia Intelligent Storage and Retrieval Management, Orlando, FL, USA, October 1999; pp. 1–9.
16. Ma, Y.; Liu, Y.; Xie, Q.; Li, L. CNN-feature based automatic image annotation method. *Multimed. Tools Appl.* **2019**, *78*, 3767–3780. [[CrossRef](#)]
17. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.
18. Mohanty, S.P.; Hughes, D.P.; Salathé, M. Using Deep Learning for Image-Based Plant Disease Detection. *Front. Plant Sci.* **2016**, *7*, 1419. [[CrossRef](#)] [[PubMed](#)]
19. Hani, N.; Roy, P.; Isler, V. MinneApple: A Benchmark Dataset for Apple Detection and Segmentation. *IEEE Robot. Autom. Lett.* **2020**, *5*, 852–858. [[CrossRef](#)]
20. Altaheri, H.; Alsulaiman, M.; Muhammad, G.; Amin, S.U.; Bencherif, M.; Mekhtiche, M. Date fruit dataset for intelligent harvesting. *Data Brief* **2019**, *26*, 104514. [[CrossRef](#)]
21. Koirala, A.; Walsh, K.B.; Wang, Z.; McCarthy, C. Deep learning for real-time fruit detection and orchard fruit load estimation: Benchmarking of ‘MangoYOLO’. *Precis. Agric.* **2019**, *20*, 1107–1135. [[CrossRef](#)]
22. Olsen, A.; Konovalov, D.A.; Philippa, B.; Ridd, P.; Wood, J.C.; Johns, J.; Banks, W.; Girgenti, B.; Kenny, O.; Whinney, J.; et al. DeepWeeds: A Multiclass Weed Species Image Dataset for Deep Learning. *Sci. Rep.* **2019**, *9*, 2058. [[CrossRef](#)]
23. Madsen, S.L.; Mathiassen, S.K.; Dyrmann, M.; Laursen, M.S.; Paz, L.-C.; Jørgensen, R.N. Open Plant Phenotype Database of Common Weeds in Denmark. *Remote Sens.* **2020**, *12*, 1246. [[CrossRef](#)]
24. Giselsson, T.M.; Jørgensen, R.N.; Jensen, P.K.; Dyrmann, M.; Midtiby, H.S. Midtiby, A public image database for benchmark of plant seedling classification algorithms. *arXiv* **2017**, arXiv:1711.05458.
25. Cheng, Q.; Zhang, Q.; Fu, P.; Tu, C.; Li, S. A survey and analysis on automatic image annotation. *Pattern Recognit.* **2018**, *79*, 242–259. [[CrossRef](#)]
26. Randive, K.; Mohan, R. A State-of-Art Review on Automatic Video Annotation Techniques. In Proceedings of the International Conference on Intelligent Systems Design and Applications, Vellore, India, 6–8 December 2018; Springer: Cham, Switzerland, 2018; pp. 1060–1069. [[CrossRef](#)]
27. Sudars, K.; Jasko, J.; Namatevs, I.; Ozola, L.; Badaukis, N. Dataset of annotated food crops and weed images for robotic computer vision control. *Data Brief* **2020**, *31*, 105833. [[CrossRef](#)]
28. Cao, J.; Zhao, A.; Zhang, Z. Automatic image annotation method based on a convolutional neural network with threshold optimization. *PLoS ONE* **2020**, *15*, e0238956. [[CrossRef](#)]
29. Dechter, R. Learning while searching in constraint-satisfaction problems. In Proceedings of the Fifth National Conference on Artificial Intelligence (AAAI-86), Philadelphia, PA, USA, 11–15 August 1986.
30. Aizenberg, I.; Aizenberg, N.N.; Vandewalle, J.P. *Multi-Valued and Universal Binary Neurons: Theory, Learning and Applications*; Springer Science & Business Media: Cham, Switzerland, 2000.
31. Schmidhuber, J. Deep learning. *Scholarpedia* **2015**, *10*, 32832. [[CrossRef](#)]
32. Schmidhuber, J. Deep Learning in Neural Networks: An Overview. *Neural Netw.* **2015**, *61*, 85–117. [[CrossRef](#)] [[PubMed](#)]
33. Adnan, M.M.; Rahim, M.S.M.; Rehman, A.; Mehmood, Z.; Saba, T.; Naqvi, R.A. Automatic Image Annotation Based on Deep Learning Models: A Systematic Review and Future Challenges. *IEEE Access* **2021**, *9*, 50253–50264. [[CrossRef](#)]
34. Chen, J.; Ran, X. Deep Learning with Edge Computing: A Review. *Proc. IEEE* **2019**, *107*, 1655–1674. [[CrossRef](#)]
35. Caron, M.; Bojanowski, P.; Joulin, A.; Douze, M. Deep clustering for unsupervised learning of visual features. In Proceedings of the ECCV: European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; pp. 139–156. [[CrossRef](#)]
36. Bakhshipour, A.; Jafari, A. Evaluation of support vector machine and artificial neural networks in weed detection using shape features. *Comput. Electron. Agric.* **2018**, *145*, 153–160. [[CrossRef](#)]
37. Kamilaris, A.; Prenafeta-Boldú, F.X. Deep learning in agriculture: A survey. *Comput. Electron. Agric.* **2018**, *147*, 70–90. [[CrossRef](#)]
38. Bresilla, K.; Perulli, G.D.; Boini, A.; Morandi, B.; Corelli Grappadelli, L.; Manfrini, L. Single-Shot Convolution Neural Networks for Real-Time Fruit Detection Within the Tree. *Front. Plant Sci.* **2019**, *10*, 611. [[CrossRef](#)]
39. Tsironis, V.; Bourou, S.; Stentoumis, C. Evaluation of Object Detection Algorithms on A New Real-World Tomato Dataset. *ISPRS Arch.* **2020**, *43*, 1077–1084. [[CrossRef](#)]
40. Santos, T.T.; de Souza, L.L.; dos Santos, A.A.; Avila, S. Grape detection, segmentation, and tracking using deep neural networks and three-dimensional association. *Comput. Electron. Agric.* **2020**, *170*, 105247. [[CrossRef](#)]
41. Dos Santos Ferreira, A.; Freitas, D.M.; da Silva, G.G.; Pistori, H.; Folhes, M.T. Unsupervised deep learning and semi-automatic data labeling in weed discrimination. *Comput. Electron. Agric.* **2019**, *165*, 104963. [[CrossRef](#)]
42. Van Engelen, J.E.; Hoos, H.H. A survey on semi-supervised learning. *Mach. Learn.* **2020**, *109*, 373–440. [[CrossRef](#)]
43. Shorewala, S.; Ashfaq, A.; Sidharth, R.; Verma, U. Weed Density and Distribution Estimation for Precision Agriculture Using Semi-Supervised Learning. *IEEE Access* **2021**, *9*, 27971–27986. [[CrossRef](#)]
44. Hu, C.; Thomasson, J.A.; Bagavathiannan, M.V. A powerful image synthesis and semi-supervised learning pipeline for site-specific weed detection. *Comput. Electron. Agric.* **2021**, *190*, 106423. [[CrossRef](#)]
45. Khan, S.; Tufail, M.; Khan, M.T.; Khan, Z.A.; Iqbal, J.; Alam, M. A novel semi-supervised framework for UAV based crop/weed classification. *PLoS ONE* **2021**, *16*, e0251008. [[CrossRef](#)]

46. Karami, A.; Crawford, M.; Delp, E.J. Automatic Plant Counting and Location Based on a Few-Shot Learning Technique. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 5872–5886. [[CrossRef](#)]
47. Noon, S.K.; Amjad, M.; Qureshi, M.A.; Mannan, A. Use of deep learning techniques for identification of plant leaf stresses: A review. *Sustain. Comput. Inform. Syst.* **2020**, *28*, 100443. [[CrossRef](#)]
48. Fountsop, A.N.; Fendji, J.L.E.K.; Atemkeng, M. Deep Learning Models Compression for Agricultural Plants. *Appl. Sci.* **2020**, *10*, 6866. [[CrossRef](#)]
49. Xuan, G.; Gao, C.; Shao, Y.; Zhang, M.; Wang, Y.; Zhong, J.; Li, Q.; Peng, H. Apple Detection in Natural Environment Using Deep Learning Algorithms. *IEEE Access* **2020**, *8*, 216772–216780. [[CrossRef](#)]
50. Rahnemoonfar, M.; Sheppard, C. Real-time yield estimation based on deep learning. Autonomous Air and Ground Sensing Systems for Agricultural Optimization and Phenotyping II. In Proceedings of the SPIE Commercial + Scientific Sensing and Imaging, Anaheim, CA, USA, 8 May 2017; p. 1021809. [[CrossRef](#)]
51. Bah, M.D.; Hafiane, A.; Canals, R. Deep Learning with Unsupervised Data Labeling for Weed Detection in Line Crops in UAV Images. *Remote Sens.* **2018**, *10*, 1690. [[CrossRef](#)]
52. Huang, H.; Lan, Y.; Yang, A.; Zhang, Y.; Wen, S.; Deng, J. Deep learning versus Object-based Image Analysis (OBIA) in weed mapping of UAV imagery. *Int. J. Remote Sens.* **2020**, *41*, 3446–3479. [[CrossRef](#)]
53. Veeranampalayam Sivakumar, A.N.; Li, J.; Scott, S.; Psota, E.; Jhala, A.J.; Luck, J.D.; Shi, Y. Comparison of Object Detection and Patch-Based Classification Deep Learning Models on Mid- to Late-Season Weed Detection in UAV Imagery. *Remote Sens.* **2020**, *12*, 2136. [[CrossRef](#)]
54. Franco, C.; Guada, C.; Rodríguez, J.T.; Nielsen, J.; Rasmussen, J.; Gómez, D.; Montero, J. Automatic detection of thistle-weeds in cereal crops from aerial RGB images. In Proceedings of the International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, Cádiz, Spain, 11–15 June 2018; Springer: Cham, Switzerland, 2018; pp. 441–452. [[CrossRef](#)]
55. Kalampokas, T.; Vrochidou, E.; Papakostas, G.A.; Pachidis, T.; Kaburlasos, V.G. Grape stem detection using regression convolutional neural networks. *Comput. Electron. Agric.* **2021**, *186*, 106220. [[CrossRef](#)]
56. Liu, X.; Ghazali, K.H.; Han, F.; Mohamed, I.I. Automatic Detection of Oil Palm Tree from UAV Images Based on the Deep Learning Method. *Appl. Artif. Intell.* **2021**, *35*, 13–24. [[CrossRef](#)]
57. Yang, Q.; Shi, L.; Han, J.; Yu, J.; Huang, K. A near real-time deep learning approach for detecting rice phenology based on UAV images. *Agric. For. Meteorol.* **2020**, *287*, 107938. [[CrossRef](#)]
58. Tetila, E.C.; Machado, B.B.; Astolfi, G.; Belete, N.A.D.S.; Amorim, W.P.; Roel, A.R.; Pistori, H. Detection and classification of soybean pests using deep learning with UAV images. *Comput. Electron. Agric.* **2020**, *179*, 105836. [[CrossRef](#)]
59. Mhango, J.; Harris, E.; Green, R.; Monaghan, J. Mapping Potato Plant Density Variation Using Aerial Imagery and Deep Learning Techniques for Precision Agriculture. *Remote Sens.* **2021**, *13*, 2705. [[CrossRef](#)]
60. Tri, N.C.; Duong, H.N.; Van Hoai, T.; Van Hoa, T.; Nguyen, V.H.; Toan, N.T.; Snasel, V. A novel approach based on deep learning techniques and UAVs to yield assessment of paddy fields. In Proceedings of the 2017 9th International Conference on Knowledge and Systems Engineering (KSE), Hue, Vietnam, 19–21 October 2017; pp. 257–262.
61. Trujillano, F.; Flores, A.; Saito, C.; Balcazar, M.; Racoceanu, D. Corn classification using Deep Learning with UAV imagery. An operational proof of concept. In Proceedings of the IEEE 1st Colombian Conference on Applications in Computational Intelligence (ColCACI), Medellin, Colombia, 16–18 May 2018; pp. 1–4. [[CrossRef](#)]
62. Vaeljaots, E.; Lehiste, H.; Kiik, M.; Leemet, T. Soil sampling automation case-study using unmanned ground vehicle. *Eng. Rural Dev.* **2018**, *17*, 982–987. [[CrossRef](#)]
63. Cantelli, L.; Bonaccorso, F.; Longo, D.; Melita, C.D.; Schillaci, G.; Muscato, G. A Small Versatile Electrical Robot for Autonomous Spraying in Agriculture. *AgriEngineering* **2019**, *1*, 29. [[CrossRef](#)]
64. Cutulle, M.A.; Maja, J.M. Determining the utility of an unmanned ground vehicle for weed control in specialty crop system. *Ital. J. Agron.* **2021**, *16*, 1426–1435. [[CrossRef](#)]
65. Jun, J.; Kim, J.; Seol, J.; Kim, J.; Son, H.I. Towards an Efficient Tomato Harvesting Robot: 3D Perception, Manipulation, and End-Effector. *IEEE Access* **2021**, *9*, 17631–17640. [[CrossRef](#)]
66. Mazzia, V.; Salvetti, F.; Aghi, D.; Chiaberge, M. Deepway: A deep learning estimator for unmanned ground vehicle global path planning. *arXiv* **2020**, arXiv:2010.16322.
67. Li, Y.; Iida, M.; Suyama, T.; Suguri, M.; Masuda, R. Implementation of deep-learning algorithm for obstacle detection and collision avoidance for robotic harvester. *Comput. Electron. Agric.* **2020**, *174*, 105499. [[CrossRef](#)]
68. Persello, C.; Tolpekin, V.; Bergado, J.; de By, R. Delineation of agricultural fields in smallholder farms from satellite images using fully convolutional networks and combinatorial grouping. *Remote Sens. Environ.* **2019**, *231*, 111253. [[CrossRef](#)]
69. Mounir, A.J.; Mallat, S.; Zrigui, M. Analyzing satellite images by apply deep learning instance segmentation of agricultural fields. *Period. Eng. Nat. Sci.* **2021**, *9*, 1056–1069. [[CrossRef](#)]
70. Gastli, M.S.; Nassar, L.; Karray, F. Satellite images and deep learning tools for crop yield prediction and price forecasting. In Proceedings of the 2021 International Joint Conference on Neural Networks (IJCNN), Shenzhen, China, 18–22 July 2021; pp. 1–8. [[CrossRef](#)]
71. Waldner, F.; Diakogiannis, F.I. Deep learning on edge: Extracting field boundaries from satellite images with a convolutional neural network. *Remote Sens. Environ.* **2020**, *245*, 111741. [[CrossRef](#)]

72. Nguyen, T.T.; Hoang, T.D.; Pham, M.T.; Vu, T.T.; Huynh, Q.-T.; Jo, J. Monitoring agriculture areas with satellite images and deep learning. *Appl. Soft Comput.* **2020**, *95*, 106565. [[CrossRef](#)]
73. Dhyani, Y.; Pandya, R.J. Deep learning oriented satellite remote sensing for drought and prediction in agriculture. In Proceedings of the 2021 IEEE 18th India Council International Conference (INDICON), Guwahati, India, 19–21 December 2021; pp. 1–5. [[CrossRef](#)]
74. Gadiraju, K.K.; Ramachandra, B.; Chen, Z.; Vatsavai, R.R. Multimodal deep learning based crop classification using multispectral and multitemporal satellite imagery. In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Virtual Event, 6–10 July 2020; pp. 3234–3242. [[CrossRef](#)]
75. Ahmed, A.; Deo, R.; Raj, N.; Ghahramani, A.; Feng, Q.; Yin, Z.; Yang, L. Deep Learning Forecasts of Soil Moisture: Convolutional Neural Network and Gated Recurrent Unit Models Coupled with Satellite-Derived MODIS, Observations and Synoptic-Scale Climate Index Data. *Remote Sens.* **2021**, *13*, 554. [[CrossRef](#)]
76. Wang, C.; Liu, B.; Liu, L.; Zhu, Y.; Hou, J.; Liu, P.; Li, X. A review of deep learning used in the hyperspectral image analysis for agriculture. *Artif. Intell. Rev.* **2021**, *54*, 5205–5253. [[CrossRef](#)]
77. Koirala, A.; Walsh, K.B.; Wang, Z.; McCarthy, C. Deep learning—Method overview and review of use for fruit detection and yield estimation. *Comput. Electron. Agric.* **2019**, *162*, 219–234. [[CrossRef](#)]
78. Darwin, B.; Dharmaraj, P.; Prince, S.; Popescu, D.; Hemanth, D. Recognition of Bloom/Yield in Crop Images Using Deep Learning Models for Smart Agriculture: A Review. *Agronomy* **2021**, *11*, 646. [[CrossRef](#)]
79. Moazzam, S.I.; Khan, U.S.; Tiwana, M.I.; Iqbal, J.; Qureshi, W.S.; Shah, S.I. A Review of application of deep learning for weeds and crops classification in agriculture. In Proceedings of the 2019 International Conference on Robotics and Automation in Industry (ICRAI), Rawalpindi, Pakistan, 21–22 October 2019; pp. 1–6. [[CrossRef](#)]
80. Zhang, Q.; Liu, Y.; Gong, C.; Chen, Y.; Yu, H. Applications of Deep Learning for Dense Scenes Analysis in Agriculture: A Review. *Sensors* **2020**, *20*, 1520. [[CrossRef](#)]
81. Chen, Y.; Zeng, X.; Chen, X.; Guo, W. A survey on automatic image annotation. *Appl. Intell.* **2020**, *50*, 3412–3428. [[CrossRef](#)]
82. Bouchakwa, M.; Ayadi, Y.; Amous, I. A review on visual content-based and users' tags-based image annotation: Methods and techniques. *Multimedia Tools Appl.* **2020**, *79*, 21679–21741. [[CrossRef](#)]
83. Dananjayan, S.; Tang, Y.; Zhuang, J.; Hou, C.; Luo, S. Assessment of state-of-the-art deep learning based citrus disease detection techniques using annotated optical leaf images. *Comput. Electron. Agric.* **2022**, *193*, 106658. [[CrossRef](#)]
84. He, Z.; Xiong, J.; Chen, S.; Li, Z.; Chen, S.; Zhong, Z.; Yang, Z. A method of green citrus detection based on a deep bounding box regression forest. *Biosyst. Eng.* **2020**, *193*, 206–215. [[CrossRef](#)]
85. Morbekar, A.; Parihar, A.; Jadhav, R. Crop disease detection using YOLO. In Proceedings of the 2020 International Conference for Emerging Technology (INCET), Belgaum, Karnataka, India, 5–7 June 2020; pp. 1–5. [[CrossRef](#)]
86. Lamb, N.; Chuah, M.C. A strawberry detection system using convolutional neural networks. In Proceedings of the 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 10–13 December 2018; pp. 2515–2520. [[CrossRef](#)]
87. Tassis, L.M.; de Souza, J.E.T.; Krohling, R.A. A deep learning approach combining instance and semantic segmentation to identify diseases and pests of coffee leaves from in-field images. *Comput. Electron. Agric.* **2021**, *186*, 106191. [[CrossRef](#)]
88. Fawakherji, M.; Youssef, A.; Bloisi, D.; Pretto, A.; Nardi, D. Crop and weeds classification for precision agriculture using context-independent pixel-wise segmentation. In Proceedings of the Third IEEE International Conference on Robotic Computing (IRC), Naples, Italy, 25–27 February 2019; pp. 146–152. [[CrossRef](#)]
89. Bosilj, P.; Aptoula, E.; Duckett, T.; Cielniak, G. Transfer learning between crop types for semantic segmentation of crops versus weeds in precision agriculture. *J. Field Robot.* **2019**, *37*, 7–19. [[CrossRef](#)]
90. Storey, G.; Meng, Q.; Li, B. Leaf Disease Segmentation and Detection in Apple Orchards for Precise Smart Spraying in Sustainable Agriculture. *Sustainability* **2022**, *14*, 1458. [[CrossRef](#)]
91. Wspanialy, P.; Brooks, J.; Moussa, M. An image labeling tool and agricultural dataset for deep learning. *arXiv* **2021**, arXiv:2004.03351.
92. Biffi, L.; Mitishita, E.; Liesenberg, V.; Santos, A.; Gonçalves, D.; Estrabis, N.; Silva, J.; Osco, L.P.; Ramos, A.; Centeno, J.; et al. ATSS Deep Learning-Based Approach to Detect Apple Fruits. *Remote Sens.* **2020**, *13*, 54. [[CrossRef](#)]
93. Alibabaei, K.; Gaspar, P.D.; Lima, T.M. Crop Yield Estimation Using Deep Learning Based on Climate Big Data and Irrigation Scheduling. *Energies* **2021**, *14*, 3004. [[CrossRef](#)]
94. Mamdouh, N.; Khattab, A. YOLO-Based Deep Learning Framework for Olive Fruit Fly Detection and Counting. *IEEE Access* **2021**, *9*, 84252–84262. [[CrossRef](#)]
95. Kattenborn, T.; Leitloff, J.; Schiefer, F.; Hinz, S. Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. *ISPRS J. Photogramm. Remote Sens.* **2021**, *173*, 24–49. [[CrossRef](#)]
96. Zhu, X.X.; Tuia, D.; Mou, L.; Xia, G.-S.; Zhang, L.; Xu, F.; Fraundorfer, F. Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 8–36. [[CrossRef](#)]
97. Fukushima, K. Neocognitron: A hierarchical neural network capable of visual pattern recognition. *Neural Netw.* **1988**, *1*, 119–130. [[CrossRef](#)]
98. Rahnemoonfar, M.; Sheppard, C. Deep Count: Fruit Counting Based on Deep Simulated Learning. *Sensors* **2017**, *17*, 905. [[CrossRef](#)]
99. Thenmozhi, K.; Reddy, U.S. Crop pest classification based on deep convolutional neural network and transfer learning. *Comput. Electron. Agric.* **2019**, *164*, 104906. [[CrossRef](#)]

100. Huang, N.; Chou, D.; Lee, C.; Wu, F.; Chuang, A.; Chen, Y.; Tsai, Y. Smart agriculture: Real-time classification of green coffee beans by using a convolutional neural network. *IET Smart Cities* **2020**, *2*, 167–172. [\[CrossRef\]](#)
101. Asad, M.H.; Bais, A. Weed detection in canola fields using maximum likelihood classification and deep convolutional neural network. *Inf. Process. Agric.* **2020**, *7*, 535–545. [\[CrossRef\]](#)
102. Hamidinekoo, A.; Martínez, G.A.G.; Ghahremani, M.; Corke, F.; Zwiggelaar, R.; Doonan, J.H.; Lu, C. DeepPod: A convolutional neural network based quantification of fruit number in Arabidopsis. *GigaScience* **2020**, *9*, g1aa012. [\[CrossRef\]](#)
103. Onishi, Y.; Yoshida, T.; Kurita, H.; Fukao, T.; Arihara, H.; Iwai, A. An automated fruit harvesting robot by using deep learning. *ROBOMECH J.* **2019**, *6*, 13. [\[CrossRef\]](#)
104. Adi, M.; Singh, A.K.; Reddy, H.; Kumar, Y.; Challa, V.R.; Rana, P.; Mittal, U. An overview on plant disease detection algorithm using deep learning. In Proceedings of the 2021 2nd International Conference on Intelligent Engineering and Management (ICIEM), London, UK, 28–30 April 2021; pp. 305–309. [\[CrossRef\]](#)
105. Sharma, P.; Berwal, Y.P.S.; Ghai, W. Performance analysis of deep learning CNN models for disease detection in plants using image segmentation. *Inf. Process. Agric.* **2020**, *7*, 566–574. [\[CrossRef\]](#)
106. Kang, H.; Chen, C. Fruit detection, segmentation and 3D visualisation of environments in apple orchards. *Comput. Electron. Agric.* **2020**, *171*, 105302. [\[CrossRef\]](#)
107. Khattak, A.; Asghar, M.U.; Batool, U.; Ullah, H.; Al-Rakhami, M.; Gumaei, A. Automatic Detection of Citrus Fruit and Leaves Diseases Using Deep Neural Network Model. *IEEE Access* **2021**, *9*, 112942–112954. [\[CrossRef\]](#)
108. Yang, W.; Nigon, T.; Hao, Z.; Paiao, G.D.; Fernández, F.G.; Mulla, D.; Yang, C. Estimation of corn yield based on hyperspectral imagery and convolutional neural network. *Comput. Electron. Agric.* **2021**, *184*, 106092. [\[CrossRef\]](#)
109. Fuentes, A.; Yoon, S.; Kim, S.C.; Park, D.S. A Robust Deep-Learning-Based Detector for Real-Time Tomato Plant Diseases and Pests Recognition. *Sensors* **2017**, *17*, 2022. [\[CrossRef\]](#)
110. Maheswari, P.; Raja, P.; Apolo-Apolo, O.E.; Pérez-Ruiz, M. Intelligent Fruit Yield Estimation for Orchards Using Deep Learning Based Semantic Segmentation Techniques—A Review. *Front. Plant Sci.* **2021**, *12*, 684328. [\[CrossRef\]](#)
111. Mu, C.; Yuan, Z.; Ouyang, X.; Sun, P.; Wang, B. Non-destructive detection of blueberry skin pigments and intrinsic fruit qualities based on deep learning. *J. Sci. Food Agric.* **2021**, *101*, 3165–3175. [\[CrossRef\]](#) [\[PubMed\]](#)
112. Lee, S.H.; Goëau, H.; Bonnet, P.; Joly, A. New perspectives on plant disease characterization based on deep learning. *Comput. Electron. Agric.* **2020**, *170*, 105220. [\[CrossRef\]](#)
113. Verma, S.; Chug, A.; Singh, A.P. Application of convolutional neural networks for evaluation of disease severity in tomato plant. *J. Discret. Math. Sci. Cryptogr.* **2020**, *23*, 273–282. [\[CrossRef\]](#)
114. Gehlot, M.; Saini, M.L. Analysis of different CNN architectures for tomato leaf disease classification. In Proceedings of the 2020 5th IEEE International Conference on Recent Advances and Innovations in Engineering (ICRAIE), Jaipur, India, 1–3 December 2020; pp. 1–6. [\[CrossRef\]](#)
115. Zhang, D.; Islam, M.; Lu, G. A review on automatic image annotation techniques. *Pattern Recognit.* **2012**, *45*, 346–362. [\[CrossRef\]](#)
116. Jmour, N.; Zayen, S.; Abdelkrim, A. Convolutional neural networks for image classification. In *2018 International Conference on Advanced Systems and Electric Technologies (IC_ASET)*; IEEE: New York, NY, USA, 2018; pp. 397–402. [\[CrossRef\]](#)
117. Zhou, L.; Zhang, C.; Liu, F.; Qiu, Z.; He, Y. Application of Deep Learning in Food: A Review. *Compr. Rev. Food Sci. Food Saf.* **2019**, *18*, 1793–1811. [\[CrossRef\]](#)
118. Lee, K.B.; Cheon, S.; Kim, C.O. A Convolutional Neural Network for Fault Classification and Diagnosis in Semiconductor Manufacturing Processes. *IEEE Trans. Semicond. Manuf.* **2017**, *30*, 135–142. [\[CrossRef\]](#)
119. Das, P.; Yadav, J.K.P.S.; Yadav, A.K. An Automated Tomato Maturity Grading System Using Transfer Learning Based AlexNet. *Ing. Des Syst. Inf.* **2021**, *26*, 191–200. [\[CrossRef\]](#)
120. Palsson, F.; Sveinsson, J.R.; Ulfarsson, M.O. Multispectral and Hyperspectral Image Fusion Using a 3-D-Convolutional Neural Network. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 639–643. [\[CrossRef\]](#)
121. Indolia, S.; Goswami, A.K.; Mishra, S.; Asopa, P. Conceptual Understanding of Convolutional Neural Network—A Deep Learning Approach. *Procedia Comput. Sci.* **2018**, *132*, 679–688. [\[CrossRef\]](#)
122. Gu, J.; Wang, Z.; Kuen, J.; Ma, L.; Shahroudy, A.; Shuai, B.; Liu, T.; Wang, X.; Wang, G.; Cai, J.; et al. Recent advances in convolutional neural networks. *Pattern Recognit.* **2018**, *77*, 354–377. [\[CrossRef\]](#)
123. LeCun, Y.; Boser, B.; Denker, J.; Henderson, D.; Howard, R.; Hubbard, W.; Jackel, L. Handwritten digit recognition with a back-propagation network. *Adv. Neural Inf. Process. Syst.* **1989**, *2*, 396–404.
124. Sharma, N.; Jain, V.; Mishra, A. An Analysis of Convolutional Neural Networks for Image Classification. *Procedia Comput. Sci.* **2018**, *132*, 377–384. [\[CrossRef\]](#)
125. Sargül, M.; Ozyildirim, B.; Avci, M. Differential convolutional neural network. *Neural Netw.* **2019**, *116*, 279–287. [\[CrossRef\]](#)
126. Zeng, W.; Li, M.; Zhang, J.; Chen, L.; Fang, S.; Wang, J. High-order residual convolutional neural network for robust crop disease recognition. In Proceedings of the 2nd International Conference on Computer Science and Application Engineering, Hohhot, China, 22–24 October 2018; pp. 1–5. [\[CrossRef\]](#)
127. Mohammadi, S.; Belgiu, M.; Stein, A. 3D fully convolutional neural networks with intersection over union loss for crop mapping from multi-temporal satellite images. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 11–16 July 2021; pp. 5834–5837. [\[CrossRef\]](#)

128. Prilianti, K.R.; Brotosudarmo, T.H.P.; Anam, S.; Suryanto, A. Performance comparison of the convolutional neural network optimizer for photosynthetic pigments prediction on plant digital image. *AIP Conf. Proc.* **2019**, *2084*, 020020. [[CrossRef](#)]
129. Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [[CrossRef](#)]
130. Dubey, A. Agricultural plant disease detection and identification. *Int. J. Electr. Eng. Technol.* **2020**, *11*, 354–363.
131. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Processing Syst.* **2012**, *60*, 84–90. [[CrossRef](#)]
132. Liu, X.; Han, F.; Ghazali, K.H.; Mohamed, I.I.; Zhao, Y. A review of convolutional neural networks in remote sensing image. In Proceedings of the 2019 8th International Conference on Software and Computer Applications, Penang, Malaysia, 19–21 February 2019; pp. 263–267. [[CrossRef](#)]
133. Cheng, L.; Leung, A.C.S.; Ozawa, S. (Eds.) In Proceedings of the Neural Information Processing: 25th International Conference, ICONIP 2018, Siem Reap, Cambodia, 13–16 December 2018; Springer: Cham, Switzerland, 2018.
134. Zhu, L.; Li, Z.; Li, C.; Wu, J.; Yue, J. High performance vegetable classification from images based on AlexNet deep learning model. *Int. J. Agric. Biol. Eng.* **2018**, *11*, 190–196. [[CrossRef](#)]
135. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. In Proceedings of the 3rd International Conference on Learning Representations (ICLR2015), San Diego, CA, USA, 7–9 May 2015; pp. 1–14.
136. Alsayed, A.; Alsabei, A.; Arif, M. Classification of Apple Tree Leaves Diseases using Deep Learning Methods. *Int. J. Comput. Sci. Netw. Secur.* **2021**, *21*, 324–330.
137. Meng, X.; Yuan, Y.; Teng, G.; Liu, T. Deep learning for fine-grained classification of jujube fruit in the natural environment. *J. Food Meas. Charact.* **2021**, *15*, 4150–4165. [[CrossRef](#)]
138. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
139. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826. [[CrossRef](#)]
140. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A.A. Inception-v4, inception-resnet and the impact of residual connections on learning. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017.
141. Ni, J.; Gao, J.; Deng, L.; Han, Z. Monitoring the Change Process of Banana Freshness by GoogLeNet. *IEEE Access* **2020**, *8*, 228369–228376. [[CrossRef](#)]
142. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [[CrossRef](#)]
143. Deeba, K.; Amutha, B. WITHDRAWN: ResNet—Deep neural network architecture for leaf disease classification. *Microprocess. Microsyst.* **2020**, 103364. [[CrossRef](#)]
144. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2261–2269. [[CrossRef](#)]
145. Huang, G.; Liu, S.; van der Maaten, L.; Weinberger, K.Q. CondenseNet: An efficient densenet using learned group convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 2752–2761. [[CrossRef](#)]
146. Zhang, K.; Guo, Y.; Wang, X.; Yuan, J.; Ding, Q. Multiple feature reweight DenseNet for image classification. *IEEE Access* **2019**, *7*, 9872–9880. [[CrossRef](#)]
147. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788. [[CrossRef](#)]
148. Redmon, J.; Farhadi, A. YOLO9000: Better, faster, stronger. In Proceedings of the IEEE Conference on Computer Vision And Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 7263–7271. Available online: http://www.worldscientific.com/doi/abs/10.1142/9789812771728_0012 (accessed on 15 March 2022).
149. Redmon, J.; Farhadi, A. YOLOv3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767. Available online: <http://arxiv.org/abs/1804.02767> (accessed on 17 March 2022).
150. Bochkovskiy, A.; Wang, C.; Liao, H. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934.
151. Lippi, M.; Bonucci, N.; Carpio, R.F.; Contarini, M.; Speranza, S.; Gasparri, A. A YOLO-based pest detection system for precision agriculture. In Proceedings of the 2021 29th Mediterranean Conference on Control and Automation (MED), Puglia, Italy, 22–25 June 2021; pp. 342–347. [[CrossRef](#)]
152. Chang, C.-L.; Chung, S.-C. Improved deep learning-based approach for real-time plant species recognition on the farm. In Proceedings of the 12th International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDS), Porto, Portugal, 20–22 July 2020; pp. 1–5. [[CrossRef](#)]
153. Gai, R.; Chen, N.; Yuan, H. A detection algorithm for cherry fruits based on the improved YOLO-v4 model. *Neural Comput. Appl.* **2021**, 1–12. [[CrossRef](#)]

154. Jocher, G.; Stoken, A.; Borovec, J.; Chaurasia, A.; Changyu, L. Yolov5, Code Repos. 2020. Available online: <https://github.com/ultralytics/yolov5> (accessed on 5 April 2022).
155. Zhao, J.; Zhang, X.; Yan, J.; Qiu, X.; Yao, X.; Tian, Y.; Zhu, Y.; Cao, W. A Wheat Spike Detection Method in UAV Images Based on Improved YOLOv. *Remote Sens.* **2021**, *13*, 3095. [[CrossRef](#)]
156. Fan, Y.; Zhang, S.; Feng, K.; Qian, K.; Wang, Y.; Qin, S. Enhancement, Strawberry Maturity Recognition Algorithm Combining Dark Channel Enhancement and YOLOv5. *Sensors* **2022**, *22*, 419. [[CrossRef](#)]
157. Yao, J.; Qi, J.; Zhang, J.; Shao, H.; Yang, J.; Li, X. A Real-Time Detection Algorithm for Kiwifruit Defects Based on YOLOv. *Electronics* **2021**, *10*, 1711. [[CrossRef](#)]
158. Kuznetsova, A.; Maleva, T.; Soloviev, V. Detecting apples in orchards using YOLOv3 and YOLOv5 in general and close-up images. In Proceedings of the International Symposium on Neural Networks, Cairo, Egypt, 4–6 December 2020; pp. 233–243.
159. LeCun, Y.; Haffner, P.; Bottou, L.; Bengio, Y. Object recognition with gradient-based learning. In *Shape, Contour and Grouping in Computer Vision*; Springer: Berlin/Heidelberg, Germany, 1999; pp. 319–345.
160. Kolesnikov, A.; Beyler, L.; Zhai, X.; Puigcerver, J.; Yung, J.; Gelly, S.; Houlsby, N. Big Transfer (BiT): General visual representation learning. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; Springer: Cham, Switzerland, 2020; pp. 491–507. [[CrossRef](#)]
161. Xie, Q.; Luong, M.-T.; Hovy, E.; Le, Q.V. Self-Training with noisy student improves imagenet classification. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 10684–10695. [[CrossRef](#)]
162. Pham, H.; Dai, Z.; Xie, Q.; Le, Q.V. Meta Pseudo Labels. 2020. Available online: <http://arxiv.org/abs/2003.10580> (accessed on 6 April 2022).
163. Mureşan, H.; Oltean, M. Fruit recognition from images using deep learning. *Acta Univ. Sapientiae Inform.* **2018**, *10*, 26–42. [[CrossRef](#)]
164. Chen, S.W.; Shivakumar, S.S.; Dcunha, S.; Das, J.; Okon, E.; Qu, C.; Taylor, C.J.; Kumar, V. Counting Apples and Oranges with Deep Learning: A Data-Driven Approach. *IEEE Robot. Autom. Lett.* **2017**, *2*, 781–788. [[CrossRef](#)]
165. Zhong, L.; Hu, L.; Zhou, H. Deep learning based multi-temporal crop classification. *Remote Sens. Environ.* **2019**, *221*, 430–443. [[CrossRef](#)]
166. Marani, R.; Milella, A.; Petitti, A.; Reina, G. Deep neural networks for grape bunch segmentation in natural images from a consumer-grade camera. *Precis. Agric.* **2021**, *22*, 387–413. [[CrossRef](#)]
167. Thapa, R.; Zhang, K.; Snavely, N.; Belongie, S.; Khan, A. The Plant Pathology Challenge 2020 data set to classify foliar disease of apples. *Appl. Plant Sci.* **2020**, *8*, e11390. [[CrossRef](#)]
168. Majeed, Y.; Zhang, J.; Zhang, X.; Fu, L.; Karkee, M.; Zhang, Q.; Whiting, M.D. Deep learning based segmentation for automated training of apple trees on trellis wires. *Comput. Electron. Agric.* **2020**, *170*, 105277. [[CrossRef](#)]
169. Garcia-Garcia, A.; Orts-Escolano, S.; Oprea, S.; Villena-Martinez, V.; Garcia-Rodriguez, J. A review on deep learning techniques applied to semantic segmentation. *arXiv* **2017**, arXiv:1704.06857.
170. Anuar, M.M.; Halin, A.A.; Perumal, T.; Kalantar, B. Aerial Imagery Paddy Seedlings Inspection Using Deep Learning. *Remote Sens.* **2022**, *14*, 274. [[CrossRef](#)]
171. Liu, J.; Wang, X. Tomato Diseases and Pests Detection Based on Improved Yolo V3 Convolutional Neural Network. *Front. Plant Sci.* **2020**, *11*, 898. [[CrossRef](#)]
172. Turkoglu, M.; Yanikoğlu, B.; Hanbay, D. PlantDiseaseNet: Convolutional neural network ensemble for plant disease and pest detection. *Signal Image Video Process.* **2021**, *16*, 301–309. [[CrossRef](#)]
173. Krisnandi, D.; Pardede, H.F.; Yuwana, R.S.; Zilvan, V.; Heryana, A.; Fauziah, F.; Rahadi, V.P. Diseases Classification for Tea Plant Using Concatenated Convolution Neural Network. *CommIT J.* **2019**, *13*, 67–77. [[CrossRef](#)]
174. Bansal, P.; Kumar, R.; Kumar, S. Disease Detection in Apple Leaves Using Deep Convolutional Neural Network. *Agriculture* **2021**, *11*, 617. [[CrossRef](#)]
175. Afifi, A.; Alhumam, A.; Abdelwahab, A. Convolutional Neural Network for Automatic Identification of Plant Diseases with Limited Data. *Plants* **2021**, *10*, 28. [[CrossRef](#)] [[PubMed](#)]
176. Lu, S.; Chen, W.; Zhang, X.; Karkee, M. Canopy-attention-YOLOv4-based immature/mature apple fruit detection on dense-foliage tree architectures for early crop load estimation. *Comput. Electron. Agric.* **2022**, *193*, 106696. [[CrossRef](#)]
177. Aguiar, A.S.; Magalhães, S.A.; dos Santos, F.N.; Castro, L.; Pinho, T.; Valente, J.; Martins, R.; Boaventura-Cunha, J. Grape Bunch Detection at Different Growth Stages Using Deep Learning Quantized Models. *Agronomy* **2021**, *11*, 1890. [[CrossRef](#)]
178. Yan, B.; Fan, P.; Lei, X.; Liu, Z.; Yang, F. A Real-Time Apple Targets Detection Method for Picking Robot Based on Improved YOLOv. *Remote Sens.* **2021**, *13*, 1619. [[CrossRef](#)]
179. Sozzi, M.; Cantalamessa, S.; Cogato, A.; Kayad, A.; Marinello, F. Automatic Bunch Detection in White Grape Varieties Using YOLOv3, YOLOv4, and YOLOv5 Deep Learning Algorithms. *Agronomy* **2022**, *12*, 319. [[CrossRef](#)]
180. Lyu, S.; Li, R.; Zhao, Y.; Li, Z.; Fan, R.; Liu, S. Green Citrus Detection and Counting in Orchards Based on YOLOv5-CS and AI Edge System. *Sensors* **2022**, *22*, 576. [[CrossRef](#)] [[PubMed](#)]
181. Buzzy, M.; Thesma, V.; Davoodi, M.; Velni, J.M. Real-Time Plant Leaf Counting Using Deep Object Detection Networks. *Sensors* **2020**, *20*, 6896. [[CrossRef](#)] [[PubMed](#)]

182. Machefer, M.; Lemarchand, F.; Bonnefond, V.; Hitchins, A.; Sidiropoulos, P. Mask R-CNN Refitting Strategy for Plant Counting and Sizing in UAV Imagery. *Remote Sens.* **2020**, *12*, 3015. [[CrossRef](#)]
183. Sun, Z.; Di, L.; Fang, H.; Burgess, A. Deep Learning Classification for Crop Types in North Dakota. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 2200–2213. [[CrossRef](#)]
184. Perugachi-Diaz, Y.; Tomczak, J.M.; Bhulai, S. Deep learning for white cabbage seedling prediction. *Comput. Electron. Agric.* **2021**, *184*, 106059. [[CrossRef](#)]
185. Nasiri, A.; Taheri-Garavand, A.; Zhang, Y.-D. Image-based deep learning automated sorting of date fruit. *Postharvest Biol. Technol.* **2019**, *153*, 133–141. [[CrossRef](#)]
186. Osako, Y.; Yamane, H.; Lin, S.-Y.; Chen, P.-A.; Tao, R. Cultivar discrimination of litchi fruit images using deep learning. *Sci. Hortic.* **2020**, *269*, 109360. [[CrossRef](#)]
187. Kang, J.; Gwak, J. Ensemble of multi-task deep convolutional neural networks using transfer learning for fruit freshness classification. *Multimedia Tools Appl.* **2021**, *81*, 22355–22377. [[CrossRef](#)]
188. Masuda, K.; Suzuki, M.; Baba, K.; Takeshita, K.; Suzuki, T.; Sugiura, M.; Niikawa, T.; Uchida, S.; Akagi, T. Noninvasive Diagnosis of Seedless Fruit Using Deep Learning in Persimmon. *Hortic. J.* **2021**, *90*, 172–180. [[CrossRef](#)]
189. Bazame, H.C.; Molin, J.P.; Althoff, D.; Martello, M. Detection, classification, and mapping of coffee fruits during harvest with computer vision. *Comput. Electron. Agric.* **2021**, *183*, 106066. [[CrossRef](#)]
190. Goëau, H.; Mora-Fallas, A.; Champ, J.; Love, N.L.R.; Mazer, S.J.; Mata-Montero, E.; Joly, A.; Bonnet, P. A new fine-grained method for automated visual analysis of herbarium specimens: A case study for phenological data extraction. *Appl. Plant Sci.* **2020**, *8*, e11368. [[CrossRef](#)]
191. Goëau, H.; Mora-Fallas, A.; Champ, J.; Love, N.L.R.; Mazer, S.J.; Mata-Montero, E.; Joly, A.; Bonnet, P. Fine-grained automated visual analysis of herbarium specimens for phenological data extraction: An annotated dataset of reproductive organs in *Strepanthus* herbarium specimens. *Zenodo Repos.* **2020**, *10*. [[CrossRef](#)]
192. Apolo-Apolo, O.E.; Pérez-Ruiz, M.; Guanter, J.M.; Valente, J. A Cloud-Based Environment for Generating Yield Estimation Maps from Apple Orchards Using UAV Imagery and a Deep Learning Technique. *Front. Plant Sci.* **2020**, *11*, 1086. [[CrossRef](#)]
193. Sharifi, A.; Mahdipour, H.; Moradi, E.; Tariq, A. Agricultural Field Extraction with Deep Learning Algorithm and Satellite Imagery. *J. Indian Soc. Remote Sens.* **2022**, *50*, 417–423. [[CrossRef](#)]
194. Yang, R.; Ahmed, Z.U.; Schulthess, U.C.; Kamal, M.; Rai, R. Detecting functional field units from satellite images in smallholder farming systems using a deep learning based computer vision approach: A case study from Bangladesh. *Remote Sens. Appl. Soc. Environ.* **2020**, *20*, 100413. [[CrossRef](#)] [[PubMed](#)]