International Journal of Engineering and Technology Innovation, vol. 12, no. 3, 2022, pp. 225-246

Deep Learning for Image-Based Plant Growth Monitoring: A Review

Yin-Syuen Tong[1,2], Tou-Hong Lee[2], Kin-Sam Yen[1,*]

[1]School of Mechanical Engineering, Universiti Sains Malaysia, Nibong Tebal, Malaysia

[2]TMS LITE Sendirian Berhad, Sungai Ara, Malaysia

Received 09 November 2021; received in revised form 16 February 2022; accepted 17 February 2022

DOI: https://doi.org/10.46604/ijeti.2022.8865

Abstract

Deep learning (DL) approaches have received extensive attention in plant growth monitoring due to their ground-breaking performance in image classification; however, the approaches have yet to be fully explored. This review article, therefore, aims to provide a comprehensive overview of the work and the DL developments accomplished over the years. This work includes a brief introduction on plant growth monitoring and the image-based techniques used for phenotyping. The bottleneck in image analysis is discussed and the need of DL methods in plant growth monitoring is highlighted. A number of research works focused on DL based plant growth monitoring-related applications published since 2017 have been identified and included in this work for review. The results show that the advancement in DL approaches has driven plant growth monitoring towards more complicated schemes, from simple growth stages identification towards temporal growth information extraction. The challenges, such as resource-demanding data annotation, data-hungriness for training, and extraction of both spatial and temporal features simultaneously for accurate plant growth prediction, however, remain unsolved.

Keywords: deep learning, CNN, phenotyping, plant growth monitoring

1. Introduction

Climate change has resulted in a decline in the number of pollinators and water levels. This has presented new challenges for crop growers in producing enough crops to feed the world population, which has been projected to reach 8.6 billion by 2030 [1-4]. Moving toward food security and sustainability, there is a pressing need for growers to optimally utilize resources to maximize the yield and quality of crops produced, thus making plant growth monitoring a cornerstone in modern precision farming. The development of a plant is the ultimate result of the complex interaction between its genotypes and the environment. Therefore, a deep understanding of a particular plant is necessary to assist in plantation management, particularly in making decisions related to fertilization, harvesting, and early pest and disease prevention plans [5-9]. Information collected from growth monitoring can be used to reveal the relationship between a plant's genes and its traits, and hence serves as a reference or indicator in plant breeding programs [10-12].

Traditional machine learning (ML) methods such as decision trees, naïve Bayes algorithm, fuzzy logic, support vector machine, and gradient boosting algorithm usually require human involvement in feature extraction and preprocessing steps prior to model use [13-15]. Manual hand-crafted feature extraction and non-standardized preprocessing steps not only limit the model scalability but also renders the analytics work time-consuming and challenging. Experts with adequate knowledge are always necessary and deemed critical [16]. In the field of plant phenotyping, minute variations between plants of different species and cultivars are always expected. One would normally require long hours of study before fully grasping the key knowledge to plants of interest for recognition and classification.

* Corresponding author. E-mail address: meyks@usm.my

Tel.: +6045996387

In addition, a number of features such as color, texture, and shapes identified and extracted under a specific circumstance for a particular ML method are always application-specific [17-21]. A similar set of features extracted under different circumstances might not be applicable and has no guarantee to the performance of the same ML method. In contrast, deep learning (DL) methods that are based on multi-layer neural networks generally offer superior adaptability [22-24]. In DL methods, feature extraction or knowledge representation can be achieved automatically during the model training. Human intervention is not a must in this process. The extracted features by the model could be further generalized to similar domains and applications using a transfer learning approach. As a result, automatic feature extraction has become a great advantage of DL methods over traditional ML methods. Furthermore, the automatic feature extraction-enabled DL models offer an end-to-end solution to generalize a direct mapping from input images to the expected outputs [25-26].

To date, there is yet to be a review focusing on plant growth monitoring as a computer vision task despite the imminent need for faster and more robust analysis. This article therefore aims to address the gap of knowledge on DL approaches, models, and techniques that have been successfully developed and deployed for plant growth monitoring. To provide a clear picture of the advancement and challenges in plant growth monitoring programs, this review discusses the approaches that have been used to formulate monitoring tasks, as well as the DL models and techniques used to process various plant data from available resources. This review aims to uncover the latest state of DL development in the related fields, to identify the challenges as well as the need and direction of future research for plant growth monitoring, which will eventually benefit the prospective researchers in the fields.

A four-step process, as outlined in Snyder's work [27], was employed to produce this review. The process starts with the design phase, wherein research questions, search strategy, and selection criteria were defined. The search for related work from various sources was then carried out. The selected work was analyzed and reviewed to produce useful findings. Finally, the findings were structured and summarized. In the early design phase, the following research questions were formulated, as follows.

(1) What are the state-of-the-art DL approaches that have been reported for plant growth monitoring?

(2) What is the performance of these DL models?

(3) What are the challenges posed in relation to the application of DL in plant growth monitoring?

In the search for related works, several keywords and combinations of "deep learning", "CNN", "plant growth monitoring", "plant development", and "plant phenotyping" were used. The studies collected in this review were sourced from several prominent online platforms: Elsevier ScienceDirect, SpringerLink, and Google Scholar. In addition, the search was also carried out on websites of renowned plant-related journals such as Plant Physiology, Plants, Plant Methods, and Nature Methods. Only articles that are related to the use of DL network for quantification of plant growth, monitoring of plant growth, or harvestability events such as emergence and flowering (i.e., as observed in 2-D images), were selected for this review. The early parts of this review focus on the definition of plant growth, the overview of image-based phenotyping, the challenges in traditional image processing, and a brief introduction to DL. The remainder of this review is organized into three sections. It begins with a presentation of a list of growth monitoring-related publications and a summary of the studies conducted. The subsequent sections discuss the findings that include the generic DL frameworks and models used for plant growth monitoring. The final section provides an overview of the challenges and future prospects for image-based plant growth monitoring.

2. Growth Definition and Phenotyping

A plant undergoes a series of events in its lifetime. The dynamic development of a plant, or its "growth stage", serves as key information for critical decision-making related to breed selection and other day-to-day farming activities, such as fertilization and harvesting. The term "growth" has been used to describe a broad range of plant features, including the increase

in plant or organ size, cell architecture, and structural biomass [28]. While there is currently no universal definition for plant growth, the description of growth for various plant species is an interesting subject to be studied, particularly in areas related to plant phenotyping. Conventionally, the growth of a plant is characterized by literal descriptions or growth scales such as Biologische Bundesanstalt, Bundessortenamt and Chemical industry (BBCH) scale [29], based on visual observations. The introduction of uniform coding for each growth stage that provides ordinal measures of plant growth has enabled exchanges in work related to phenology among the scientific community, which include those for agricultural purposes and even extend to climate change studies [30-31].

Besides, researchers have attempted to represent the plant development stage via quantitative approaches using plant parts and organs such as leaves [32], stems [33], flowers [34], and fruits [35]. Changes in these phenotypes over a period of time were taken as the features reflecting plant development. In fact, a wide range of spatial and temporal scales have been used by researchers to represent the dynamic growth process. For example, in USA National Phenology Network (USA-NPN) data collection protocols, the observable growth stage of a plant (phenophase) was recorded alongside its frequency at a particular time point (i.e., the proportion of bloom on a plant), while the collective growth of a plant of similar species is expressed in terms of the magnitude observed across specified a time interval (i.e., the proportion of individual plant flowering in a month) [36].

Das Choudhury [37] presented a taxonomy to divide plant phenotypes into three primary types, namely structural, physiological, and temporal properties. The authors further categorized the plant structural and physiological phenotypes: (1) holistic, which describes the properties derived from whole-plant geometrics; and (2) component, which describes the properties derived from the measurement of individual parts and organs to address the spatial scale in phenotypes. For time-related phenotypes, trajectory-based measures that reflect the quantifiable changes over time were used in techniques to address the temporal properties in plant growth alongside the event-based phenotypes that indicate the distinct salient stages in a plant's life cycle [37].

2.1. Image-based phenotyping

A successful growth monitoring program implementation for agricultural activities requires more than a mere understanding and quantification of plant growth. An appropriate automated pipeline is vital. For centuries, plant growth characterization or phenotyping has heavily relied on manual inspection by experienced experts. The tedious and expertise-demanding process might thus inevitably involve subjective judgment. In addition, phenotyping accuracy is also constrained by human visual and cognitive limitations in detecting minute changes in plants. Nowadays, as a replacement for manual observation processes, plant growth measurements and status are extracted automatically from images of the plant, which are captured by cameras and light sensors. With the growing availability of inexpensive optical sensors such as digital RGB cameras, image-based techniques have become a popular subject in current plant phenotyping studies [38]. Sensing techniques used to collect pixelated representations from visible light, infrared, fluorescence response, and temperature have been applied in plant phenotyping. A comprehensive review of imaging techniques was presented in the work of Li et al. [39].

Apart from the common RGB cameras, multi-spectral or hyperspectral imaging systems that utilize spectrometers have been designed to capture plant images at wavelengths ranging from visible light (400 nm to 700 nm) to short-wave infrared (1000 nm to 2000 nm) [40-41]. These systems enable the phenotyping of traits that are not visible to humans, expanding the knowledge boundaries beyond the understanding gained from human visual perception. Various morphological and structural features such as leaf area, plant height, and stem width can be directly measured from 2-D digital images, while a number of derived measures such as color indices and normalized difference vegetation index (NDVI) have been applied in quantifying plant growth [42-44]. In addition to the image content, the metadata of an optical image such as its F-stop and exposure time were also reported to provide meaningful information in estimating a plant's physical properties [45].

Another type of phenotyping technique is the fluorescence imaging system. When fluorophores in a plant are brought to a brief excitation by exposure to a lamp or laser beam, light is emitted at a characteristic wavelength. This technique is commonly used to monitor chlorophyll content, plant photosynthesis activity, and other metabolites, whereby changes reflect the biotic and abiotic stresses in the plant [46]. However, fluorescence imaging systems have limited applications in the field of plant phenotyping, primarily due to its requirements for a dark-adapted reference, as well as for sufficient closeness to a plant canopy [47]. Besides, the use of thermal imaging that operates in infrared wavelengths for the study of the surface temperature of plants opens up a window to its applications for monitoring changes in plant transpiration and photosynthesis based on plant temperature, which are the two features related to stomatal conductance [48-50].

Similar to other applications, the difficulty in image segmentation due to the effect of surrounding temperature and changing weather presents a challenge in the implementation of outdoor monitoring systems [39]. In recent years, the ability to overcome cloud cover has rendered more advanced techniques such as synthetic aperture radar (SAR), which are particularly favorable for remote sensing under all weather conditions. Most importantly, an SAR radar demonstrates good potential in the ability to penetrate through plant canopy and ground, enabling the phenotyping of underground crops [51-52].

In phenotyping studies, particularly for plantations, the ambiguity between organ expansion and movement are often presented in 2-D images, especially when the plant images are captured under different lighting or at different sun azimuth angle. Thus, there is a need for better sampling to the third dimension (3-D) for better quantification of growth [28]. Among the 3-D imaging techniques, light detection and ranging (LiDAR), which involves laser beam steering for measurement reading, has been used to form cloud points that correspond to a plant's geometrics, allowing for accurate representation in 3-D space. Besides, researchers have proposed and successfully implemented cost-effective geometric-based approaches using 3-D reconstruction by one or more optical cameras. These reconstruction techniques include depth of focus (DoF), time of flight (ToF), and structured light [53].

In general, imaging techniques in plant phenotyping differ from each other not only in sensing principles, but also in terms of the spatial scale and the corresponding image acquisition platform used. Fig. 1 illustrates a collection of imaging techniques and scales in plant phenotyping. Generally, the aforementioned imaging techniques can also fall into organ-level [54], canopy-level [55], and remote sensing categories [56].

In remote sensing phenotyping, plants in open fields are captured with terrestrial satellite instruments [57] and unmanned aerial vehicles (UAVs) [58]. Generally, the latter has the advantage of providing higher spatial resolution images at a lower cost. Small-sized UAVs can also be employed for active farming operations such as dispensing pesticides and fertilizers in addition to remote monitoring use [59]. In contrast, the imaging at the canopy-level or organ-level such as plant leaves and roots is more prevalent in laboratories and greenhouses where phenotyping samples are provided in individual pots or trays.

Despite the superior spatial resolution, canopy-level imaging methods appear to have limited commercial applicability that demands image acquisition in a large quantity and fine temporal resolution for prompt informed crop management decisions. However, high-throughput phenotyping can be achieved at the canopy-level imaging when coupled with automatic or mobilized acquisition platforms. For instance, Tisné et al. [60] implemented a sequential movement and rotation table to enable high-throughput phenotyping of 735 Arabidopsis thaliana in pots. The authors also highlighted the resultant high spatial homogeneity of the plants, which is an important feature for its potential application.

Nonetheless, the canopy-level image acquisition carried out in plantation fields is not limited to the use of stationary platforms [61], but is also commonly supported by ground-based vehicles mounted with various sensors [62-63]. While there is an inevitable trade-off between spatial and temporal resolution among the platforms at different scales, the information obtained from images captured using canopy-level sensing technologies was reported to link well with observations acquired by airborne or satellite remote sensing sensors [64].

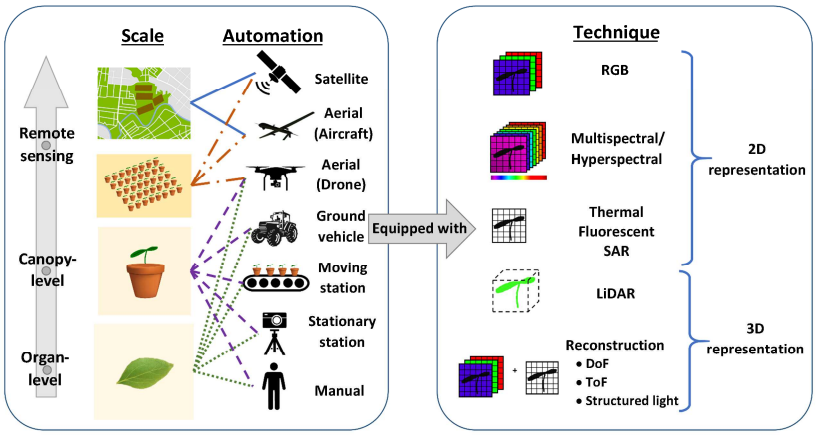


Fig. 1 Scale, automation, and techniques used in plant phenotyping imaging

2.2. The bottleneck in image analysis

In the wake of the Internet of Things (IoT) and Big Data, plant phenotyping work may now be facilitated by the possible access to a huge amount of multimodal and complex data. Sensing technologies have been developed to deliver data at multiple dimensions and scales. Despite the advances in imaging that open up opportunities for the study of more complex phenotypes, limitations in image analysis were identified as a bottleneck in plant phenotyping [65]. Even four years later, the need for an accurate image analysis technique in plant phenotyping studies to cater to the blooming data and information persists [66].

Traditional image analysis usually involves heavy reliance on human experience to design the preprocessing operations and feature extraction steps to derive useful traits from images. Adjustment of the image intensity and segmentation of the plant of interest from the background are among the early preprocessing steps that are commonly applied in image analysis. In common rule-based algorithms that lack the learning ability to adapt to potential variations, it is an enormous challenge to increase the flexibility of the heavily hand-engineered algorithms to cater to the many variations that may be encountered during plant deployment.

For instance, the inconsistent illumination due to uncontrollable weather at plantation fields brings about a major challenge in plant phenotyping image analysis. The presence of shadows or insufficient exposure in an image may hinder successful segmentation. Besides, challenges in segmentation may also come from cluttered backgrounds, where the overlapping parts between plants make the analysis unfeasible at the individual plant level. Even in the absence of other plants, self-occlusion that occurs when a plant develops past its early seedling stage also presents obstacles in the image segmentation of distinct plant parts [37]. Additionally, plant movement due to the wind, which consequently results in random shifts in images, is also an inevitable challenge for close-up field images.

Moreover, changes in the scale of the object due to variation in size and position of plants often cause difficulty in image analysis. For instance, the direction of the leave expansion may be different from one plant to another. Also, the growth of different parts of a plant may not be uniform at different stages [37], while plants of similar species may appear differently due to different growing environments [47]. Thus, apart from hand-crafted solutions, there is a need to stretch the realms of image analysis to include more flexible alternatives such as DL algorithms.

3. Deep Learning Applications in Plant Growth Monitoring

DL is essentially a subset of ML that consists of multi-layer artificial neural networks (ANNs) [67]. They are data-driven models whereby their outcomes and parameters are determined by the examples used in the model training process. The representation of data can be learned in several ways. The learning approaches include (1) supervised learning, where ground truth is used as a reference, and (2) unsupervised learning which focuses on discovering data patterns among samples. The

combination of both approaches (i.e., semi-supervised learning) offers a middle ground between the advantages and pitfalls of each learning. Another type of learning is reinforcement learning, which utilizes trial-and-error techniques that are structured by a reward and penalty mechanism.

Although changes in biological matter are generally influenced by genetic compositions and physiological factors, it is difficult to generalize such influences on plant growth. Due to this reason, modeling plant growth using conventional ML models such as support vector machines (SVMs), k-nearest neighbors (k-NNs), and decision tree becomes challenging. In fact, better performance of DL models has been reported in comparison to that of conventional ML models, particularly within the scope of image-based plant phenotyping studies [68-69].

Since the successful demonstration of "deeper" convolutional neural layers in the ImageNet large scale visual recognition challenge (ILSVRC) in 2012 [70], which displayed a ground-breaking performance in image classification, DL has gained the attention of researchers from various disciplines of computer vision, as well as in plant phenotyping studies. Over the years, variants of DL have evolved and extended to models that target complex computer vision tasks such as classification, regression, object detection, object semantic and instance segmentation, etc. [71]. Among the DL networks, the convolutional neural network (CNN), a subclass that utilizes shared weights in convolutional kernels, has been found to be more effective at encoding the hierarchical pattern primarily in image-based and video-based data, than the fully-connected multilayer perceptron. In the agriculture field, CNN has also been explored for image analysis for different applications such as disease detection, land cover classification, fruit counting, and weed recognition [72].

To date, there are a total of 23 studies which are relevant to DL applications for plant growth monitoring. These studies were published between the years 2017 and 2021, with the majority of them being published in 2020, as summarized in Fig. 2. This finding indicates that the exploration of DL for plant growth monitoring applications is a relatively new and emerging field.

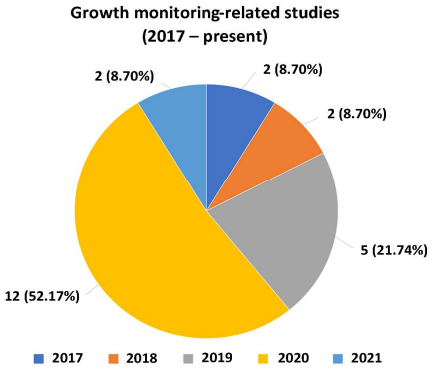


Fig. 2 Year of publication of plant growth monitoring-related studies

3.1. Generic deep learning framework for plant growth monitoring

In general, the DL networks that have been explored for plant growth monitoring applications can be divided into two groups: the pure CNN network and the hybrid CNN-LSTM network, as illustrated in Fig. 3. One important determinant in selecting between the two is the form of the input data. In the pure CNN network group, the input data is commonly in the form of grayscale images or color 2-D images, providing only the spatial information. On the other hand, the hybrid CNN-LSTM network group takes a series of images across a period of time, which provides both the spatial and temporal information.

For an image captured at any time point, pure CNNs may be used for a range of tasks in the plant monitoring process. With properly scaled imaging scope and preprocessing techniques, image patches covering plant organs or whole plants can be assigned to class labels related to maturity or growth stages using classification models. On the other hand, meaningful measures such as plant dimensions and NDVI can be directly extracted from the images via a regression approach. In cases of large-scale images, localization of an individual plant or its organ in an image may be carried out using two methods: by the combination of a sliding window mechanism and a classifier, or directly by object detection or instance segmentation CNNs.

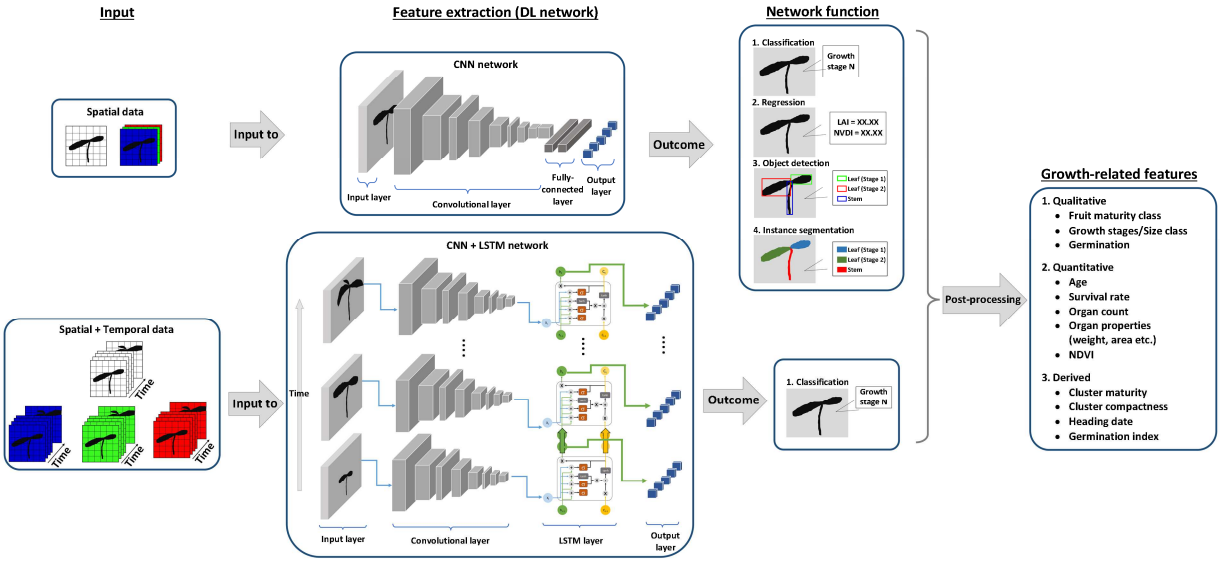


Fig. 3 Illustration of a workflow using DL framework for morphological plant growth monitoring

In addition, the hybrid of CNNs and recurrent neural networks (RNNs) such as long short-term memory (LSTM) may be used, considering the availability of time-series images that are taken within specific time frames. By incorporating sequential elements in addition to spatial information, monitoring using the CNN-LSTM hybrid architecture may be a good option to better utilize additional information. The output of the CNN feature extraction layer of plant images at each time frame is fed into the LSTM model for the classification of the output within a particular time frame, making time-series classification attainable. In comparison to the pure CNN network formulated for different functions, the use of the hybrid network for plant growth monitoring is relatively new [73]. However, the incorporation of temporal elements into spatial-based CNNs has been identified as one of the promising directions for future work related to plant growth monitoring studies, highlighting the potentially broad application of hybrid networks [74-75].

Next, the later part of the plant monitoring framework consists of the post-processing process. Based on pre-defined assumptions, the output of the DL network will be translated into phenotypes for monitoring purposes. For example, with the object detection output, the count of the flowers in a plant can be obtained and later used to characterize cotton flowering patterns [76]. In fact, depending on the post-processing method, the phenotypes used as the final output in monitoring systems using DL can be quantitative or qualitative, or a combination of both. For instance, the qualitative phenotype at a particular growth stage that is directly obtained from the classification output may be further processed into a quantitative phenotype that measures the collective growth stage of multiple plants, such as the count of a growth stage present.

From the reviewed accounts, plants and their corresponding growth stages can be categorized into four groups. The first is the seedling during vegetative growth (e.g., Arabidopsis [75], red clover, Alfalfa [73], white cabbage [77], lettuce [78-80], etc.), which represents the majority. The second group comprises crops in the reproductive phase (e.g., paddy [81-82], cotton [76], etc.), fruits at the harvest stage (e.g., date [83], blueberry [84], tomato [85], apple [86], etc.), and seed germination (e.g., maize, rye, pearl millet [74], etc.). The works related to the classification are summarized in Table 1, which also provides an overview of the plant growth monitoring approaches reported in the reviewed studies. Information such as plants under study, growth tasks, and corresponding phenotypes are presented to provide an overview of the studies conducted. Besides, in the implementation of the data-driven DL models, data preparation is an important preliminary step that directly affects model performance and applicability. Thus, details including the imaging system, data labeling, and preprocessing in the studies are also outlined in Table 1.

Details regarding the DL approach used to carry out the growth task in each study are listed in Table 2. Specifically, the optimal DL architecture in each study has been highlighted with the corresponding quantitative measures of model performance. Besides, additional details such as training settings and post-processing steps are also included to reflect the training process and strategies used to adapt the model output for different growth tasks.

Table 1 Overview of previous studies and data preparation

No.	Year	Plant	Growth category	Growth-related phenotype	Growth-related task	Imaging	Data	Labelling	Preprocessing
1 [87]	2017	Pineapple	Harvest phase (fruit)	Qualitative phenotype (fruit maturity)	Classification of pineapples based on maturity level	• RGB images (side views of pineapples)	• Train: 243 images • Validation: 27 images	• 3 classes (unripe, partially ripe, and fully ripe)	• Background and crown removal • Resizing to 200 × 200 px
2 [88]	2018	Weed (18 species)	Vegetative growth	Event-based qualitative phenotype (growth stages)	Classify weeds into 9 classes based on leaf counts	• RGB images (from cell phones, consumer cameras, and industrial cameras)	• 9649 images collected (18 weed species/family from 3 growing seasons) • Train: 11907 images • Validation: 2516 images	• 9 classes (1, 2, 3, 4, 5, 6, 7, 8, and >8 leaves)	• Data augmentation (horizontal flip, rotation, zoom, width shift, and Gaussian smoothing filters)
3 [83]	2019	Date	Harvest phase (fruit)	Qualitative phenotype (fruit maturity)	Classification of date fruits based on maturity stages and defect detection	• RGB images (top views with simple backgrounds)	• 1357 images augmented to 37050 images • Train: 30688 images • Validation: 6368 images • Test: 199 images	• 4 classes (Khalal, Rutab, Tamar, and defect) • Refer to Codex standard for dates (CXS 143-1985)	• 37056 augmented images • Data augmentation (rotation, height shift, width shift, zoom, horizontal-flip, and shear intensity)
4 [73]	2020	Red clover and Alfalfa	Vegetative growth (seedling)	Event-based qualitative phenotype (growth stages)	Monitoring of seedling emergence timing	• RGB time-lapse images (top views) at 3280 × 2464 px	• 42000 image sequences from 200 plants • Clover set for training and validation • Alfalfa set for testing	• 4 classes (growth events) • Some annotations by experts	• Registration and cropping to 89 × 89 px • Raw/HSV green channel (backgrounds removed)
5 [89]	2020	Gynura bicolor DC	Vegetative growth to harvest phase	Event-based qualitative phenotype (growth stages)	Classification of growth based on period/stages for harvesting (targeting the 2nd class for harvesting)	• RGB images (top views with simple backgrounds, e.g., black cloth)	• Two-scale inputs (intact leaves and interpolation for leaf patches) from 764 leaves	• 3 classes (early, middle, and late) • Performed with help from the experts	• Data augmentation (mirror rotation, image sharpening, and color jittering) • Resizing to 224 × 244 px
6 [90]	2020	Wheat and barley	Vegetative growth	Qualitative phenotype (growth stages)	Classification of crops based on growth stages	• RGB images (top views and 45° views) • Manual imaging at a constant 2 m height	• 138,000 images extracted from the videos taken in 7 fields	• 12 classes (wheat) • 11 classes (barley) • Labelling by experts according to Zadoks scale	• Resizing to 256 × 256 px
7 [77]	2021	White cabbage	Vegetative growth (seedling)	Qualitative phenotype (survival)	Prediction of seedling survival	• Grayscale images (top views of multiple plants in a tray) at 1280 × 1024 px	• 13200 seedlings • Train: 7920 images • Validation: 2640 images • Test: 2640 images • Day 4 to 7 data (visible above soil & no overlapping)	• 2 classes (successful / not) • 6 classes on day 14, of which class 5 and 6 are regarded as successful growth according to the experts	• Cropping to 64 - 75 px and resizing to 64 × 64 px • Normalization to [-0.5, 0.5] • Data augmentation (90° rotation, horizontal, vertical flip, and affine transformations)
8 [80]	2019	Lettuce	Vegetative growth	Trajectory quantitative phenotype (3-D size and fresh weight)	Monitoring of the growth rate of lettuce	• Time-lapse RGB images at 3280 × 2464 px • Stations in greenhouses • Acquisition every half an hour from 6 am to 6 pm, daily	• 1218 images from 5 lettuces • Train: 850 images • Test: 200 images • Set for monitoring: 168 images	• Manual area label	• Resizing to 800 × 600 px
9 [91]	2020	Little gem romaine lettuce	Vegetative growth	Trajectory quantitative phenotype (3-D size and fresh weight)	Monitoring of the growth rate (crop size) based on captured images and estimation of fresh weight based on crop size	• Time-lapse RGB images (top and side views) at 1920 × 1080 px • Each image contains 3 plants • Acquisition every 30 min from 6 am to 6 pm	• 1350 images (obtained from imaging technique and Google search engine)	• 2 classes (background/leafy) • Manual measurements (width, depth, height, and weight) as ground truth	-
10 [84]	2020	Blueberry	Harvest phase (fruit)	Quantitative phenotype (counts, maturity, and compactness)	Characterization of yield-related traits based on grape phenology for harvesting strategy (compactness, maturity, and counts)	• RGB images (complex backgrounds in the field, simple backgrounds in the field, and laboratory backgrounds) at 4896 × 2760 px	• Multi-view of grape clusters • Train: 524 images • Validation: 145 images • Test: 55 images	• 2 classes (maturity) • Hue value for maturity annotation	• Data augmentation (horizontal rotation of 50% and vertical rotation at 90°, 180°, and 270°) • Multiplication with a random value between 0.5 and 1.5 • Blurring with a Gaussian kernel with a sigma of 5.0 • Resizing to 1024 × 1024 px
11 [85]	2020	Tomato	Harvest phase (fruit)	Qualitative phenotype (growth stages)	Detection of tomato fruit ripeness	• RGB and depth images (multi-height views) at 720 × 1280 px	• 123 images of 1612 fruits • 2:1 train-test split	• 2 classes (ripeness) • Manual labelling based on detection of ripeness by empirical chromaticity values (red channel > 1.4 green channel)	-
12 [86]	2019	Apple	Harvest phase (fruit)	Qualitative phenotype (growth stages)	Detection of apple at different growth stages	• RGB images (multi-views) • Complex scene (orchard) at 3000 × 3000 px	• 480 images	• 3 classes (young, expanding, and ripe)	• Data augmentation from 480 to 4800 images • Resizing to 512 × 512 px
13 [76]	2020	Cotton	Reproductive phase (flower)	Trajectory quantitative phenotype (flower counts)	Characterization based on flowering patterns for genotype classification	• RGB images (multi-views and complex backgrounds in a single plant field) • Ground vehicle (average of 2-3 days of scanning interval)	• 8666 images (475 used for training object detection network) from 23 genotypes of 116 plants in the field	• 3 classes (target plant, emerging bloom, and non-bloom) • 5 classes (target plant, emerging bloom, region with specular reflectance, opened boll, and others)	-

Table 1 Overview of previous studies and data preparation (continued)

No.	Year	Plant	Growth category	Growth-related phenotype	Growth-related task	Imaging	Data	Labelling	Preprocessing
14 [74]	2020	• Maize (Zea mays) • Rye (Secale cereale) • Pearl millet (Pennisetum glaucum)	Germination phase (seed)	Event-based qualitative phenotype (germination) and derived quantitative phenotype (germination index)	Germination detection of different crops to assess the quality of seeds	• RGB time-lapse images (top views) • Seeds in petri dishes (multiple dish images) • Simple background images (black cloth) • Acquisition within 30 min of timeframe	• 23797 images from 2449 seeds • Petri dish-based stratification (randomly split to train, validate, and test at 8-1-1 ratio based on each petri dish to avoid overlapping)	• Bounding box labeling of each seed into 2 classes (germinating and not germinating)	• Cropping to 624 × 624 px for each petri dish
15 [92]	2020	Tomato	Harvest phase (fruit)	Qualitative phenotype (growth stages)	Detection and classification of tomatoes into different maturity grades	• RGB images • Different lighting conditions by capturing images at 9 am, 12 pm, and 5 pm • 680 × 480 px, 96 dpi, 24-bit JPEG images	• Train: 231 images (1193 fruits and 421 flowers) • Validation: 46 images	• 3 classes of object (leaf, flower, and fruit) • 3 classes of maturity (green, turning, and red) based on RGB and HSV values	-
16 [93]	2020	Mango (Mangifera indica)	Reproductive phase	Trajectory quantitative phenotype (panicle counts)	Classification of flowering stages and identification of flowering events	• RGB images (dual views) • Acquisition by agricultural vehicles • Orchard A (994 trees; 1988 images per week at 2464 × 2048 px) • Orchard B (24 trees; 6000 × 4000 px)	• 2 cultivars in orchards • Train: 3178 panicles (54 images) • Validation: 635 panicles (6 images) • Test: 2853 panicles (48 images from Orchard B)	• 3 classes (panicle stages) • Two types of bounding boxes: upright and rotated	• Random rotation at 40°
17 [94]	2021	Coconut	Harvest phase (fruit)	Qualitative phenotype (fruit maturity)	Classification of coconuts based on different maturity stages	• RGB images (multiple views) with complex backgrounds	• 2000 images captured in the field and retrieved via Google Images	• 2 classes (tender and mature)	• 20000 augmented images • Data augmentation (horizontal flip, image rotation by 90°, and color transformation) • Rescaling of image length to 516 px while maintaining the aspect ratio
18 [78]	2019	Lettuce	Vegetative growth	Qualitative phenotype (emergence of lettuce and classes)	Monitoring of lettuce growth (detection and classification of lettuce into different sizes)	• NDVI top-view images from aerial imaging using sensors from very light aircraft (VLA), with GSD = 3 cm	• Images of 100,000 lettuce heads for training (50% lettuce and 50% others) • Images of 60 patches (11330 × 6600 px) in a 7-hectare field consisting of ~ million lettuces (each patch >300 and <1000 lettuce heads)	• 20 × 20 px bounding boxes to enclose each lettuce head	• Contrast limited adaptive histogram equalization (CLAHE) for normalized NDVI images • Cropping to 250 × 250 px for local analysis
19 [81]	2019	Paddy rice	Reproductive phase	Event-based quantitative phenotype (heading date at 50% flowering panicle counts)	Heading date estimation by characterizing flowering patterns based on flowering panicle region counts	• RGB images acquired from the field at ground level (5184 × 3456 px) • Acquisition within 5-min time frame between 8 am and 4 pm	• 6000 images (equal number of positives and negatives)	• 2 classes (flowering region and non-flowering region)	• Resizing to 224 × 224 px
20 [75]	2017	• Arabidopsis thaliana rosette • Nicotiana tabacum rosette	Vegetative growth (seedling)	Quantitative phenotype (leaf counts and age)	Estimation of rosette phenomics (leaf counts and age measured in hours after germination) and mutant classification	• RGB images (top views with pot backgrounds)	• 2 species and mutant • IPPN dataset	• Age refers to the time after germination • 5 classes of Arabidopsis mutants	-
21 [42]	2018	Wheat	Vegetative growth	Quantitative phenotype (NDVI)	Estimation of vegetative index from images and prediction of vegetative index at unseen growth stages	• Aerial multispectral imaging	• Train: 8064 images • Test: 4023 images	• Labelling of NDVI based on calculated values from the NIR and red image	-
22 [79]	2020	Lettuce	Vegetative growth (seedling)	Qualitative phenotype (leaf fresh weight, leaf dry weight, and leaf area)	Estimation of growth-related traits	• RGB and depth images (top views), i.e., 1920 × 1080 px (RGB) and 512 × 424 px (depth) by Kinects. • Daily image acquisition	• 3 cultivars in greenhouse based on 286 RGB and depth images • Train-test split of 8(20% val): 2	• 3 traits (leaf fresh weight, leaf dry weight, and leaf area) based on on-field measurements	• 5954 augmented images • Data augmentation (rotation at 90°, 180°, and 270°; vertical and horizontal flipping; brightness adjustment) • Resizing to 128 × 128 × 3 px
23 [82]	2020	Rice	Vegetative growth	Quantitative phenotype (leaf area index)	Estimation of leaf area index based on captured images	• RGB images at 5472 × 3648 px • Image acquisition by UAV	• 4 varieties in paddy fields	• Ground measurement by automatic area meter, i.e., divided into 60 × 60 cm area • 9 types of color indices extracted from multispectral data	• Cropping to 100 × 100 px

Table 2 DL approaches in previous studies and the model performance

No.	Year	DL network function	Selected DL architecture	Reason	Training setting	Hardware	Performance	Comparative measures	Post-processing procedures
1 [87]	2017	Classification	Custom LeNet-like (2 conv + 1 FC)	-	• Adadelta back-propagation algorithm • Learning rate = 1.0 • Trained for 100 epochs • Early-stopping	-	• Classification accuracy = 92.6% • Average precision = 94%	-	-
2 [88]	2018	Classification	ImageNet-pretrained Inception-v3	• Good performance, ease of implementation, and relatively low computational cost	• Adam optimizer • Mini-batch size = 32 • Dropout = 0.4 • Model is trained 20 times	-	• Average classification accuracy = 70% • Average bias of predictions = 0.07 leaves • Absolute average bias of predictions = 0.51 leaves	-	• Outputs from 20 models are combined for prediction
3 [83]	2019	Classification	ImageNet-pretrained modified VGG-16	-	• Transfer learning, fine-tuning at stages with decreasing frozen layers in blocks • RMSProp optimizer • Mini-batch size = 32 • Dropout = 0.5	-	• Average accuracy = 96.98%. • Average AUC = 98.19%	-	-
4 [73]	2020	Classification	CNN-LSTM	-	-	-	• Average F1 score = 88.5%	• Multiclass CNN (chosen from VGG-16, ResNet50, and DenseNet121), 2-class CNN, and ConvLSTM	• Smoothing of filter (median of classes by majority voting) at 4 time frames (1 hour)
5 [89]	2020	Classification	G-net and L-net similar 6 layers deep fusion network + 2 FC layers + output custom GL-CNN (fusion CNN based on global and local multi-features)	-	• Stochastic gradient descent optimizer • Batch size = 20 • Initial learning rate = 0.005 (decreased by 1/5 every 20 epochs) momentum = 0.9 • Dropout = 0.5 • Training epoch = 100	• Intel Xeon® 2.20 GHz, 32 RAM, two GTX1080Ti	• Accuracy = 95.63% • F1 score = 95.25%	• AlexNet, VGG-16, GoogleNet, ResNet50, DenseNet, and early and late fusion	-
6 [90]	2020	Classification	ImageNet-pretrained VGG-19	-	-	-	• 99.7%-100% classification accuracy for both crops in both views	• Custom 5-layer CNN and SVM with conventional feature extraction	-
7 [77]	2021	Classification	Pretrained AlexNet	• Limited data	• Stochastic gradient descent optimizer • Training epoch = 500 • Batch size = 128 • Learning rate = 1e-4 • Weight decay = 1e-5	• Nvidia GTX1080	• Testing accuracy = 94% • Testing AUC = 95%	• Linear regression, Multi-layered perceptron (MLP), DenseNet, ResNet, and VGG-16 • All CNNs outperformed logistic regression (LR) and MLP	-
8 [80]	2019	Instance segmentation	COCO-pretrained mask R-CNN	-	• Initial learning rate = 0.0001 (decay by factor of 10 at 12000 and 16000 iterations) • 8 images for each GPU in each epoch	• Two GTX 1080 Ti GPU	Leaf area estimation: • Mean accuracy = 97.63% • Maximum error = 0.25 cm^2	-	• Determination of growth rates based on the leaf area
9 [91]	2020	Instance segmentation	Mask R-CNN	-	-	-	Instance segmentation: • Training loss = 0.21; validation loss = 0.31 Predictions: • Error within 30mm (18.7%) for length and width • Error < 0.5 g for fresh weight	-	• Determination of side view area, height, width, centroid side, top view area, centroid top and depth extracted from the masks, and bonding boxes using image moments and Green's theorem • Growth rates calculated using variations in area of each plant and linear regression based on fresh weight
10 [84]	2020	Instance segmentation	COCO-pretrained mask R-CNN (ResNet101 + FPN)	• Superior performance	• Iterative annotation strategy plus manual correction • Learning rate = 0.001 • Momentum = 0.9 • Transfer learning (MSCOCO pre-trained weights)	• Nvidia Tesla V100	Instance segmentation: • At 0.5 IoU Validation: • mAP = 78.3% • mask mIoU = 90.6% Testing: • mAP = 71.6% • mask mIoU = 90.4% Linear regression for traits: • R^2 = 0.886 • RMSE = 1.484	-	• Extract count, maturity ratio determination (percent of mature/total), and compactness by segments
11 [85]	2020	Instance segmentation	ImageNet pretrained mask R-CNN with ResNeXt101	-	• Stochastic gradient descent optimizer • L_2 regularization • Weight decay = 0.01 • Batch size = 1 • No. of Iteration = 200,000	-	At 0.5 IoU: • Average precision = 92.5% • Average recall = 90.5% • Average F1 score = 91.5%	• Large improvement as compared to the hand-crafted algorithm • Surpasses the performance reported in previous studies • Compares feature extractors (ResNet50 and ResNet101) • Compares post-processing (yes/no)	• Discard fruit from background in the segmentation output by empirical threshold using depth image
12 [86]	2019	Object detection	YOLOv3 with DenseNet	-	• Batch size = 8 • Momentum = 0.9 • Initial learning rate = 0.001 • Decay = 0.0005 • Training steps = 70000	-	• Average F1 score = 81.7% • Average IoU = 89.6% • Average detection time = 0.304 s per frame	• YOLO-v2, YOLO-v3, and FasterR-CNN with VGG-16	-

Table 2 DL approaches in previous studies and the model performance (continued)

No.	Year	DL network function	Selected DL architecture	Reason	Training setting	Hardware	Performance	Comparative measures	Post-processing procedures
13 [76]	2020	Object detection	COCO-pretrained faster R-CNN with Inception-ResNet v2 as the feature extractor	-	• Use 5-class labelling strategy • Classification confidence score = 0.7 for bloom counting	-	Object detection: • Overall mAP = 86% Regression analysis as compared to manual counting: • R^2 = 0.88 • RMSE = 0.80	-	• Plant-based counting • Take maximum count for each plant from 4 images
14 [74]	2020	Object detection	Faster R-CNN with pretrained Inception-ResNet v2	• Real-time detection is not required • Higher accuracy than YOLO and SSD	• Transfer learning • Hyperparameter search by internal random search	• 28 Intel CPU cores, 768 GB, four RTX 2080Ti	• Testing mAP of 97.9%, 94.2%, and 94.3% for ZM, SC, and PG, respectively	• Compares different feature extractors (ResNet50, ResNet101, and Inceptionv2) • Compare results of germination count vs manual	-
15 [92]	2020	Object detection	SSD with Pascal VOC-pretrained VGG-16	-	• Transfer learning and fine-tuning with replaced output	-	• Overall detection accuracy of 100% (flower) and 95.99% (fruit)	• R-CNN	• The results from fruit detection are used for maturity grading by SVM, kNN, and ANN
16 [93]	2020	Object detection	R^2CNN-Upright (the highest F1-score); MangoYOLO-Upright (the highest mAP)	-	R^2CNN-Upright: • Train to 146k iterations MangoYOLO-Upright: • Train to 77.5k iterations • Batch size = 32	• Intel Xeon Gold 1626 CPU, Tesla P100 GPU	R^2CNN-Upright: • mAP = 70.9% • F1-score = 82.0% MangoYOLO-Upright: • mAP = 72.2% • F1-score = 76.5%	• R^2CNN-Rotated, MangoYOLO-Rotated, and YOLOv3-Rotated	• The peak of the flowering is identified based on the weekly panicle counts
17 [94]	2021	Object detection	Faster R-CNN with ResNet-50	-	-	• i7-CPU@ 4.5 GHz, 4 GB GPU NVIDIA GeForce GTX 1650 + Google Colab with Tesla K80 GPU	• mAP = 89.4% at 0.5 IoU • Detection speed = 3.124 s/image	• Faster R-CNN with ResNet 101, ResNet 152, Inception V2, Inception ResNet V2, and NASNet • SSD, YOLO V3, and R-FCN	-
18 [78]	2019	Object detection (CNN + sliding window); Clustering (size categorization)	20 × 20 px sliding window CNN (4 conv + 1 FC)	• Small dataset size & target only binary classification • Smaller NN; less training time; faster execution for promptly use	• Early-stopping	-	• Classification accuracy > 98%	-	• Non-maximum suppression due to sliding window
19 [81]	2019	Object detection (CNN classifier + sliding window)	Sliding window ImageNet pretrained ResNet50	-	• Transfer learning • Stochastic gradient descent optimizer • Learning rate = 0.001 • Momentum = 0.9 • Trained on one cultivar • Tested on other five cultivars for generalization	-	Detection: • 70%-80% F1-score for different cultivars Heading date estimation: • MAE < 1 day	• Scale-invariant feature transform (SIFT) + SVM	• The results from object detection are used for counting • Take 50% flowering count date as a heading date
20 [75]	2017	Regression	Custom multi-layer CNN (3 conv + 1FC)	-	-	-	Age regression: • Mean MAE = 20.8 h • Standard deviation MAE = 14.4 h	-	-
21 [42]	2018	Regression	Modified AlexNet	• Suitable for content and resolution of input images	• Stochastic gradient descent optimizer • Mini-batch size = 72 • Base learning rate = 0.01 (with decay function) • Momentum = 0.9 • Weight decay = 0.0005	• Xeon 128 GB RAM, GTX TITAN X	Correlation between the aerial and ground measurement: • R^2 = 0.99 • RMSE = 0.019	-	-
22 [79]	2020	Regression	Custom multi-layer CNN (5 conv + 1 FC + 3 outputs)	-	• Stochastic gradient descent optimizer • Dropout = 0.5 • Learning rate = 0.001 (dropped by factor of 0.1 every 20 epochs) • Mini-batch size = 128 • Training epoch = 300	• Intel i7 CPU 3.2 GHz, 8 GB RAM, GTX1060	• R^2 > 0.89 • NRMSE ≤ 27.6 %	• SVM, random forest (RF), and linear regression from the hand-extracted features	-
23 [82]	2020	Regression	Modified ResNeXt	-	• Adam Optimizer • No. of epoch = 100 • Batch size = 16 • Weight decay = 0	-	• R^2 = 0.963 • RMSE= 0.334	• Plant canopy analyzer, color indices in regression models, and ML algorithms (ANN, partial least squares regression (PLSR), RF, and support vector regression (SVR))	-

Despite the different plants and monitoring tasks discussed in the reviewed studies, all of them shared similarities of DL applications in their experimental protocols. In general, the DL approaches employed in the reviewed studies can be divided into three categories, namely classification approach, instance segmentation or object detection approach, and regression. Notably, the regression approach was used in all studies that involved the determination of vegetative growth and leaf-based phenotypes.

3.1.1. Plant growth classification

In the plant growth classification, the chronological development of plants over time is categorized into multiple stages, which make up a set of class labels. In general, the ground truth references used in the classification are usually based on qualitative definitions formulated by field experts. This is despite some studies relating class definitions to image color features such as hue and chromaticity values [87, 92]. A classification-based DL model is used to predict the class label for the representation of the entire input image. By examining the DL model predictions at a particular time point within a specific time frame, the temporal growth characteristics of a plant such as growth rate can be monitored.

From 30688 constant background training images, Nasiri et al. [83] achieved an overall classification accuracy of 96.98% to simultaneously distinguish healthy dates from defects, and categorize them into three maturity levels using a modified VGG16 architecture. Notably, the authors employed gradient-weighted class activation mapping (Grad-CAM) to visualize and justify the classification results. Using leaf images from different scales (whole leaf and patches), Hao et al. [89] investigated classification based on three stages of growth for harvesting using a fusion strategy. The concatenation of global-scaled and local-scaled features prior to the fully-connected layers demonstrated the advantage over fusion at earlier or later layers, with a 95.63% classification accuracy recorded during testing.

In a large-scale experiment carried out by Bauer et al. [78], a CNN classifier was trained to classify iceberg lettuce into three head sizes from airborne normalized NDVI images. A shallow network of five hidden layers was employed for binary classification based on the presence of the lettuce head, with an accuracy of over 98%, which was followed by classification based on the size, depending on the results obtained from k-Means clustering. To detect seedling emergence timing, Samiei et al. [73] used a combination of CNN and LSTM architecture to classify the three sequential development stages of red clover seedlings from time-lapse images. The proposed CNN-LSTM model outperformed other architectures, namely, multi-class, 2-class, and convolutional LSTM (ConvLSTM) models, attaining an average of 91% accuracy. The model retained a high performance with a 90% accuracy when tested on alfalfa seedling images, demonstrating the ability of the model to generalize across different plant species. Besides, in an experiment by Perugachi-Diaz et al. [77], AlexNet [70] was employed to predict the survival of white cabbage seedlings based on grayscale images. The model performed well with an impressive overall classification accuracy of 94%, which was supported by the output from manual inspection carried out by field experts.

3.1.2. Instance segmentation and object detection for growth stage identification

Apart from classification, DL architectures were designed for more advanced instance segmentation and object detection tasks which provide not only object class information, but also its location in images. These architectures have also been implemented in growth monitoring. In these approaches, the DL model is trained to localize all possible objects that are present in an input image and assign a class label to each of the identified objects. The position of objects in the image is represented in two ways: by bounding boxes or groups of pixels. Depending on the context used in the image and annotation, these approaches enable plant growth monitoring with more information than the simple distinct growth stages that are obtained from the pure image classification approach. Using the object location or region, more complex growth traits can be derived for growth monitoring. For instance, growth monitoring by object counting (e.g., number of fruits, leaves, etc.) and object morphological analysis (e.g., determining area, convex hull, etc.) can be carried out based on the detection or segmentation results.

In a study [76] to characterize cotton flowering patterns, the COCO-pretrained Faster R-CNN developed for bloom counting scored an 86% mean average precision (mAP). The authors reported a 3% increase in the average precision of emerging bloom detection, with an increased number of classes used to differentiate between other non-blooming objects. The plant-based counting strategy employed based on the results from object detection achieved an $R^2 = 0.88$ at RMSE 0.8 when examined against manual counting ground truth data. A large improvement was observed in the average differences as compared to manual ground truth when the finer temporal resolution was allowed.

Desai et al. [81] performed object detection using a combination of sliding window and a CNN (pretrained ResNet50) to detect flowering regions in time-series images. The detection achieved an average F1 score of 77%, which was later used to characterize the flowering pattern in paddy to estimate the rice heading date. Considering the heading date as the day when 50% panicle exsertion was observed, an average error of less than 1 day was reported for the resultant heading date estimation.

In another study by Ni et al. [84], a Mask R-CNN was employed to detect and segment individual fruit from images of blueberry clusters. A mAP of 78.3% and 71.6% was attained for validation and testing, respectively. The inconsistency in maturity annotation due to differences in illumination was addressed by using a hue value of the fruit in the image as the quantitative reference to determine its maturity level. The authors also employed a semi-automatic strategy in data annotation, whereby models trained on a relatively small amount of manually labeled data were used to generate annotation for a larger amount of other unseen data. Albeit the requirement for manual correction, this strategy was proven to save time. Besides, a comparison with the classical color and geometry-based computer vision algorithms revealed the superior performance of Mask R-CNN in catering to the variability in the dataset, as reported by Afonso et al. [85].

3.1.3. Regression for plant growth analysis

In contrast to the classification approach, growth monitoring using the regression approach involves the use of a set of growth measurements, such as plant height and number of leaves as references. These references may be extracted from input images using an image processing pipeline, or obtained from external non-imaging sources such as a separate handheld device. DL networks are employed to carry out the regression of these quantitative phenotypes directly from the images.

In a study by Ubbens and Stavness [75], plant phenomics were extracted for several purposes that include leaf counting, mutant classification, and age regression. The work was carried out separately using multiple CNN architectures for similar data. The CNN was used to directly estimate the plant age, which is an important growth indicator measured as the hour after germination. As a result, the authors reported an absolute difference of 20.8 hours with a standard deviation of 14.4 hours in the data labeled with the age range between 392 hours and 620 hours. By taking ground-based NDVI measurement as ground truth, Khan et al. [42] also demonstrated the successful estimation (MSE of 0.019) of NDVI from aerial RGB images using a modified version of AlexNet.

In an investigation by Zhang et al. [79], a custom CNN was used to carry out regression of growth-related traits such as leaf fresh weight (LFW), leaf dry weight (LDW), and leaf area (LA) with actual measurements as ground truth to monitor lettuce growth. The prediction of each trait was linked separately to each hidden unit at the last fully-connected layer to enable separate predictions. Collectively, the trained models showed good agreement with a reference marking R^2 value of more than 0.89 and a normalized root mean square error of less than 27%.

3.1.4. Performance evaluation

Notably, all the approaches reviewed in the previous sections are of the supervised learning type. In classification, confusion matrix metrics such as accuracy, precision, recall, and F1 score that examine the correctness of every prediction made by a model are commonly used to reflect model performance. Besides, the performance for object detection and instance segmentation

network can be more accurately represented by average precision (AP) and intersection over union (IoU) [95]. In contrast, correlation and distance measures such as root mean square error (RMSE) and mean absolute error (MAE) that represent the proximity in value relative to a reference are calculated to characterize the performance in the regression approach.

In addition to being a determinant in the algorithms for many applications, a supervised DL network is essentially a tool for non-linear mappings between a set of inputs and outputs [96]. The outputs of the DL network are usually further processed to adapt for a particular application. To better reflect the actual performance in the growth monitoring task, overall evaluation other than those that have been addressed by DL networks is also sought. This may be done via a direct comparison of the final output with the manually obtained output. For example, the germination indices obtained from the object detection task in the work of Genze et al. [74] were used for cross-checking with manual measurements besides the performance in germination detection. As another example, the estimated heading date output from a model was compared against the heading date determined via observation [81]. Despite the expensive cost, these justifications are required to ensure the applicability of the selected DL network.

On the other hand, there is a lack of information regarding model inference time despite its vitality to facilitate prompt farming decisions. However, this may not be of imminent concern considering the relatively long data collection and processing time in the plant domain industry. The transition from one growth stage to another may occur gradually in plants, especially for perennial plants which can span over few years. Nonetheless, object detection in plant monitoring may be carried out within seconds as reported by Tian et al. [86] (0.304 s/frame using YOLOv3 on 512 × 512 px input images) and Parvathi et al. [94] (3.142 s/image using Faster R-CNN on images with 516 px length). Note that the speed performances from different studies may not be directly comparable due to variations in factors such as input image resolution [97], preprocessing methods [98], and hardware used.

In fact, the advancement of DL networks and their increasing availability as research tools can be inferred by examining the date of publication of the work reviewed. All three reviewed papers which dated before 2019 demonstrated the application of classification and regression approaches. Since 2019, a noticeable shift in DL network functions for plant monitoring has been observed—from simple image classification and regression approaches, toward more complicated object detection or segmentation computer vision tasks. Two studies involving object detection using a sliding window have been published in 2019, while the majority of the studies conducted in the following year employed more sophisticated networks such as region-based convolutional neural networks (R-CNNs) and you only look once (YOLO) for the same task. Besides, the transition may also be attributed to the rapid development of energy-efficient hardware architectures in processing units, computing memory, and data compression techniques in recent years [99]. Coupled with the availability of optimized DL architectures, hardware advancement brings about few obstacles for researchers to explore more complex DL solutions in their work.

3.2. The hybrid CNN-LSTM for plant growth monitoring

In image-based plant growth monitoring that constitutes mainly the decision-making based on information obtained from images, CNNs with the convolutional layers have been commonly used in DL architecture applications to generate meaningful spatial representation over multidimensional data. However, considering that the plant growth can also be viewed as a series of events unfolded over time in a plant life cycle, spatial information should not be the sole consideration in the monitoring process. Specifically, a hybrid DL architecture that consists of CNN and LSTM brings about the possibility of better deciphering the spatial-temporal nature of the plants.

Since 2012, CNN architectures for image classification and regression have rapidly evolved over the short course of 5 years [100]. Early CNN architectures such as AlexNet [70], ZFNet [101], and VGG [102] share the common generic CNN structure that involves feeding a fixed size input into the stacks of convolutional layers that are end-to-end connected, prior to further being input into the fully-connected layers and the final output layer. The plain convolutional layers in the generic structure have later been extended to variants with modularized convolutional layers (InceptionNet [103]) and variants with

skip connections in between them (ResNet [104] and DenseNet [105]). In 2015, advances against the vanishing or exploding gradient problem [106] in CNNs has reached a milestone when the ResNet variant with parameters up to 152-layer deep was proven applicable.

LSTM [107] is another type of RNN that can retain long-term dependencies. It is commonly employed in supervised sequence learning, where at least one of the inputs or target outputs is presented in the sequential or time-related form. Since its introduction in 1997, LSTM has been applied in weather forecasting [108], speech recognition [109], and language modeling [110]. However, designed to only process 1-D data, conventional LSTM offers limited solutions to problems with data in high dimensionality. Particularly in image analysis with 2-D spatial data, the application of LSTM would require an intermediate apparatus to bridge the gap within data representations. In this case, CNNs which function as feature extractors, are deemed fit for the task.

In CNN-LSTM architecture, the input is fed to a series of CNN layers which are followed by LSTM layers before the generation of the output. The combination of feature extraction by CNN with LSTM enables sequence learning using higher-order data. By using the hybrid architecture, the temporal component in time-series data can be exploited, which translates into a sequence instead of another spatial dimension to be fed into 2-D CNN. CNN-LSTM reportedly displayed superior performance in time-series prediction in various fields as compared to the conventional LSTM or pure CNN models [111-113]. Besides, the hybrid architecture has also been employed in applications including video activity recognition, image captioning, and video description [114].

Interestingly, the reviewed studies on plant growth monitoring using a pure classification or regression approach share the tendency to employ CNN with simpler architectures. Except for [73], the majority of these studies, which date from 2017 or later, have chosen AlexNet and VGG networks, which are the exemplary representations of the plain generic CNN structure. On the other hand, several studies have come up with their customized architectures, which were also developed based on the plain generic CNN structure. These findings indicate that the simple architectures may remain competitive as a network of choice for straightforward image classification and regression tasks. In addition, this choice has also been attributed to the limited data availability [77] and suitability in terms of image resolution and content [42].

In the object detection or instance segmentation approach, neural network architectures may be divided according to their detection paradigms, namely single-stage or two-stage detection. In general, the former is typically represented by YOLO networks [115-118], which has an advantage in terms of detection speed. On the other hand, the latter is represented by R-CNN networks [119-121], which achieve relatively high detection accuracy [122]. Among the reviewed studies, R-CNN variants have been found dominant in terms of choice of network. This may be due to the nature of growth monitoring which has to cater to the gradual manifestation of plant growth that may take several days or even months. Specifically, the authors in [74] selected the Faster R-CNN since real-time detection was not required in their study. Also, the Mask R-CNN [123] network, an extension to the R-CNN network that was introduced in 2017, has appeared as the go-to for growth monitoring involving instance segmentation.

4. Challenges and Future Prospects

Even though DL frameworks have shown great potential in plant growth monitoring, a few challenges remain. These come from the data preparation phase, whereby the natural complexities of plants take a toll on the imaging and annotation process. Similarly, the post-processing steps that involve the interpretation of DL network output have been found to be largely reliant on the preceding plant growth assumptions, indicating an obstacle to automation.

While limitations in spatial resolution for the detection of small object or discrimination of gradual changes in appearance may be resolved with the use of suitable imaging, other challenges such as occlusion in plant organs or between plants are particularly critical in detection-based monitoring, as reported in previous studies [78, 84, 86]. In this case, a viable solution is

to project information from 2-D images to a higher 3-D point cloud, despite the complications that may follow [124]. Non-invasive imaging of underground plants has long been a challenge in phenotyping imaging. The fact that none of the reviewed studies have reported underground crops reflects this particular challenge in this area of study.

In supervised learning-based DL models, knowledge occurs in an iterative training process by tuning the parameters of the model. Each example is fed into the DL model as a representation of ground truth. The model is trained to output its current predictions, and the deviations from the predefined ground truth are computed using the cost function. Iteratively, the parameters in the model are updated to minimize the loss, targeting closer predictions to the annotated ground truth. Thus, it is evident that accurate annotation is undeniably crucial in the preliminary step to build a useful DL model. However, data annotation in plant monitoring is a resource-demanding task.

In comparison with the imaging technology which has advanced tremendously since automation, annotating plant growth based on images is a laborious process that often demands the assistance of field experts. This process, which is yet to be automated, has limited the exploitation of the power of data-driven DL. It would require a large number of examples to sufficiently train a DL network for practical use in plant growth monitoring tasks. The more the trainable parameters, the larger the dataset required to prevent the model from overfitting. Due to the close visual appearance of crops in between the continuous growth transitions, a precise annotation of the different growth stages is a shared challenge in the growth monitoring of different crops [73-74, 85, 89]. In some cases, only experts with specific knowledge of the problem domain are qualified to carry out the manual annotation. While annotation strategies such as semi-automatic labeling as in [84] may offer a potential solution, there is definitely no guarantee that the labels introduced in the annotation system, such as the number of classes, will always be optimum for the task [76].

Finally, it is evident that the DL has been found to be more commonly used as a feature extraction tool for growth monitoring applications at the early stages of growth. Dependent on the task formulation, the results obtained using the DL networks are later subjected to hand-crafted processing to form phenotypes that are believed to indicate the dynamic plant growth. While the DL enables impressive performance in feature extraction, there is a plain reliance on the manual monitoring process. In short, there is a need for proactive research to explore the incorporation of learning algorithms apart from the supervised learning in growth monitoring. In addition, another possible future development is to employ ensemble learning. Instead of selecting a trained model and relying solely on it, predictions from multiple models can be used for better generalization [125].

The exploitation of temporal information and the selection of time points may be a vital consideration in realizing practical growth monitoring in the future. Since plant growth is a continuous dynamic measure that spans over a plant lifetime, it is only intuitive to consider time points apart from the visual changes in growth monitoring. It is undeniable that the irreversible plant development has a sequential relationship with the growth stage. In this case, the combination of CNN with RNN architecture, such as LSTM that was designed to retain and discard cell states from one time frame to another, can potentially be a better choice over the pure CNN architecture, which lacks the temporal information for model development.

In particular, the use of temporal information in the form of time-series images has been highlighted with its advantage over the naïve classification approach [73]. Also, a large improvement in average difference when compared to the manual ground truth was observed when imaging with the finer temporal resolution was allowed by Jiang et al. [76]. Perugachi-Diaz et al. [77] also noticed that the model trained with data from the early growth stage may generalize well when fed with data obtained from the later growth stage, but not vice versa, indicating the importance of the time component in growth monitoring. In addition, considering the future time points as proactive plant monitoring, forecasting plant growth may be another area worth venturing into [126-127].

5. Conclusions

To the best of authors' knowledge, this is the first article which provides a comprehensive review of DL approaches for image-based plant growth monitoring covering the research background, the efforts, the developments as well as the achievements that have been accomplished over the years. Notably, this review has witnessed a paradigm shift in plant growth monitoring applications. From traditional image processing to the employment of CNN networks since 2017, current applications have arrived at the exploration of hybrid networks that incorporate the temporal information of plant growth. At the same time, the plant monitoring frameworks that have been presented in previous research, have also progressed from straightforward classification and regression approaches toward more complicated object detection and segmentation, since 2019.

Despite advances in imaging techniques, annotation of plant growth remains a challenging aspect in plant monitoring, even with the help of field experts. It is also hard to be overlooked that the utilization of DL models in the past is almost exclusively limited to features extraction. Further exploration to benefit the use of DL throughout the monitoring process is needed to reduce human intervention. The exploitation of temporal information for plant growth monitoring using DL architectures, such as hybrid CNN-LSTM models, is called for further investigations. Lastly, model generalization could be another interesting prospect to be explored for future study, for example, a generic model which is capable of assessing growth patterns of various plant species and types.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] S. L. Althaus, M. R. Berenbaum, J. Jordan, and D. A. Shalmon, "No Buzz for Bees: Media Coverage of Pollinator Decline," Proceedings of the National Academy of Sciences, vol. 118, no. 2, Article no. e2002552117, January 2021.

[2] E. Fereres, F. Orgaz, and V. Gonzalez-Dugo, "Reflections on Food Security under Water Scarcity," Journal of Experimental Botany, vol. 62, no. 12, pp. 4079-4086, August 2011.

[3] P. J. Gregory, J. S. Ingram, and M. Brklacich, "Climate Change and Food Security," Philosophical Transactions of the Royal Society B: Biological Sciences, vol. 360, no. 1463, pp. 2139-2148, November 2005.

[4] "World Population Prospects 2019: Data Booklet", https://population.un.org/wpp/Publications/Files/WPP2019_Highlights.pdf, June 21, 2019.

[5] M. F. Dreccer, G. Molero, C. Rivera-Amado, C. John-Bejai, and Z. Wilson, "Yielding to the Image: How Phenotyping Reproductive Growth Can Assist Crop Improvement and Production," Plant Science, vol. 282, pp. 73-82, May 2019.

[6] T. T. Tran, J. W. Choi, T. T. H. Le, and J. W. Kim, "A Comparative Study of Deep CNN in Forecasting and Classifying the Macronutrient Deficiencies on Development of Tomato Plant," Applied Sciences, vol. 9, no. 8, Article no. 1601, April 2019.

[7] G. Xing, K. Liu, and J. Gai, "A High-Throughput Phenotyping Procedure for Evaluation of Antixenosis against Common Cutworm at Early Seedling Stage in Soybean," Plant Methods, vol. 13, no. 1, Article no. 66, August 2017.

[8] R. Sujatha, J. M. Chatterjee, N. Z. Jhanjhi, and S. N. Brohi, "Performance of Deep Learning vs Machine Learning in Plant Leaf Disease Detection," Microprocessors and Microsystems, vol. 80, Article no. 103615, February 2021.

[9] R. Bhagwat and Y. Dandawate, "A Review on Advances in Automated Plant Disease Detection," International Journal of Engineering and Technology Innovation, vol. 11, no. 4, pp. 251-264, September 2021.

[10] K. Mochida, D. Saisho, and T. Hirayama, "Crop Improvement Using Life Cycle Datasets Acquired under Field Conditions," Frontiers in Plant Science, vol. 6, Article no. 740, September 2015.

[11] S. A. Prado, L. Cabrera-Bosquet, A. Grau, A. Coupel-Ledru, E. J. Millet, C. Welcker, et al., "Phenomics Allows Identification of Genomic Regions Affecting Maize Stomatal Conductance with Conditional Effects of Water Deficit and Evaporative Demand," Plant, Cell, and Environment, vol. 41, no. 2, pp. 314-326, February 2018.

[12] X. Zhang, C. Huang, D. Wu, F. Qiao, W. Li, L. Duan, et al., "High-Throughput Phenotyping and QTL Mapping Reveals the Genetic Architecture of Maize Plant Growth," Plant Physiology, vol. 173, no. 3, pp. 1554-1564, January 2017.

[13] J. Heaton, "An Empirical Analysis of Feature Engineering for Predictive Modeling," SoutheastCon, pp. 1-6, April 2016.
[14] U. Shruthi, V. Nagaveni, and B. K. Raghavendra, "A Review on Machine Learning Classification Techniques for Plant Disease Detection," 5th International Conference on Advanced Computing and Communication Systems, pp. 281-284, March 2019.
[15] W. Yi, S. Dai, Y. Jiang, C. Yuan, and L. Yang, "Computer-Aided Visual Modeling of Rice Leaf Growth Based on Machine Learning," 23rd International Conference on Soft Computing and Measurements, pp. 226-229, May 2020.
[16] A. Paturkar, G. S. Gupta, and D. Bailey, "Plant Trait Segmentation for Plant Growth Monitoring," 35th International Conference on Image and Vision Computing New Zealand, pp. 1-6, November 2020.
[17] A. H. B. A. Wahab, R. Zahari, and T. H. Lim, "Detecting Diseases in Chilli Plants Using K-Means Segmented Support Vector Machine," 3rd International Conference on Imaging, Signal Processing, and Communication, pp. 57-61, July 2019.
[18] B. Patel and A. Sharaff, "Feature Fusion Based Growth Analysis of Chhattisgarh Rice Plants Using Machine Learning Technique," 7th International Conference on Signal Processing and Integrated Networks, pp. 814-818, February 2020.
[19] B. Bose, J. Priya, S. Welekar, and Z. Gao, "Hemp Disease Detection and Classification Using Machine Learning and Deep Learning," International Conference on Parallel and Distributed Processing with Applications, Big Data and Cloud Computing, Sustainable Computing and Communications, Social Computing and Networking (ISPA/BDCloud/SocialCom/SustainCom), pp. 762-769, November 2020.
[20] M. Hesami and A. M. P. Jones, "Modeling and Optimizing Callus Growth and Development in Cannabis Sativa Using Random Forest and Support Vector Machine in Combination with a Genetic Algorithm," Applied Microbiology Biotechnology, vol. 105, no. 12, pp. 5201-5212, June 2021.
[21] N. Nandhini and J. G. Shankar, "Prediction of Crop Growth Using Machine Learning Based on Seed Features," ICTACT Journal on Soft Computing, vol. 11, no. 1, pp. 2232-2236, October 2020.
[22] J. Chai, H. Zeng, A. Li, and E. W. T. Ngai, "Deep Learning in Computer Vision: A Critical Review of Emerging Techniques and Application Scenarios," Machine Learning with Applications, vol. 6, Article no. 100134, December 2021.
[23] N. O'Mahony, S. Campbell, A. Carvalho, S. Harapanahalli, G. V. Hernandez, L. Krpalkova, et al., "Deep Learning vs. Traditional Computer Vision," https://arxiv.org/ftp/arxiv/papers/1910/1910.13796.pdf, October 10, 2015.
[24] D. Radovanović and S. Đukanović, "Image-Based Plant Disease Detection: A Comparison of Deep Learning and Classical Machine Learning Algorithms," 24th International Conference on Information Technology, pp. 1-4, February 2020.
[25] S. Srinivas, R. K. Sarvadevabhatla, K. R. Mopuri, N. Prabhu, S. S. S. Kruthiventi, and R. V. Babu, "A Taxonomy of Deep Convolutional Neural Nets for Computer Vision," Frontiers in Robotics and AI, vol. 2, Article no. 36, January 2016.
[26] J. Long, E. Shelhamer, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431-3440, June 2015.
[27] H. Snyder, "Literature Review as a Research Methodology: An Overview and Guidelines," Journal of Business Research, vol. 104, pp. 333-339, November 2019.
[28] J. J. Olas, F. Fichtner, and F. Apelt, "All Roads Lead to Growth: Imaging-Based and Biochemical Methods to Measure Plant Growth," Journal of Experimental Botany, vol. 71, no. 1, pp. 11-21, January 2020.
[29] O. Baddour, H. Kontongomde, E. Koch, E. Bruns, F. M. Chmielewski, C. Defila, et al., Guidelines for Plant Phenological Observations, Geneva: World Meteorological Organization, 2009.
[30] U. Meier, H. Bleiholder, L. Buhr, C. Feller, H. Hack, M. Heß, et al., "The BBCH System to Coding the Phenological Growth Stages of Plants—History and Publications," Plant, vol. 61, no. 2, pp. 41-52, February 2009.
[31] G. Zhao, Y. Gao, S. Gao, Y. Xu, J. Liu, C. Sun, et al., "The Phenological Growth Stages of Sapindus Mukorossi According to BBCH Scale," Forests, vol. 10, no. 6, Article no. 462, May 2019.
[32] C. Campillo, M. I. García, C. Daza, and M. H. Prieto, "Study of a Non-Destructive Method for Estimating the Leaf Area Index in Vegetable Crops Using Digital Images," HortScience Horts, vol. 45, no. 10, pp. 1459-1463, October 2010.
[33] S. D. Choudhury, S. Goswami, S. Bashyam, T. Awada, and A. Samal, "Automated Stem Angle Determination for Temporal Plant Phenotyping Analysis," Proceedings of the IEEE International Conference on Computer Vision Workshops, pp. 2022-2029, October 2017.
[34] B. Chacón, R. Ballester, V. Birlanga, A. G. Rolland-Lagan, and J. M. Pérez-Pérez, "A Quantitative Framework for Flower Phenotyping in Cultivated Carnation (Dianthus Caryophyllus L.)," PLOS ONE, vol. 8, Article no. e82165, December 2013.
[35] D. F. M. Cortes, R. S. Catarina, G. B. D. A. Barros, F. A. S. Arêdes, S. F. d. Silveira, G. A. Ferreguetti, et al., "Model-Assisted Phenotyping by Digital Images in Papaya Breeding Program," Scientia Agricola, vol. 74, pp. 294-302, August 2017.
[36] A. H. Rosemartin, E. G. Denny, K. L. Gerst, R. L. Marsh, E. E. Posthumus, T. M. Crimmins, et al., "USA National Phenology Network Observational Data Documentation," U.S. Department of the Interior and U.S. Geological Survey, Report 2018-1060, April 25, 2018.

[37] S. Das Choudhury, A. Samal, and T. Awada, "Leveraging Image Analysis for High-Throughput Plant Phenotyping," Frontiers in Plant Science, vol. 10, Article no. 508, April 2019.
[38] Z. Li, R. Guo, M. Li, Y. Chen, and G. Li, "A Review of Computer Vision Technologies for Plant Phenotyping," Computers and Electronics in Agriculture, vol. 176, Article no. 105672, September 2020.
[39] L. Li, Q. Zhang, and D. Huang, "A Review of Imaging Techniques for Plant Phenotyping," Sensors, vol. 14, no. 11, pp. 20078-20111, November 2014.
[40] A. Brugger, J. Behmann, S. Paulus, H. G. Luigs, M. T. Kuska, P. Schramowski, et al., "Extending Hyperspectral Imaging for Plant Phenotyping to the UV-Range," Remote Sensing, vol. 11, no. 12, Article no. 1401, June 2019.
[41] P. Mishra, S. Lohumi, H. A. Khan, and A. Nordon, "Close-Range Hyperspectral Imaging of Whole Plants for Digital Phenotyping: Recent Applications and Illumination Correction Approaches," Computers and Electronics in Agriculture, vol. 178, Article no. 105780, November 2020.
[42] Z. Khan, V. Rahimi-Eichi, S. Haefele, T. Garnett, and S. J. Miklavcic, "Estimation of Vegetation Indices for High-Throughput Phenotyping of Wheat Using Aerial Imaging," Plant Methods, vol. 14, no. 1, Article no. 20, March 2018.
[43] M. Sancho-Adamson, M. I. Trillas, J. Bort, J. A. Fernandez-Gallego, and J. Romanyà, "Use of RGB Vegetation Indexes in Assessing Early Effects of Verticillium Wilt of Olive in Asymptomatic Plants in High and Low Fertility Scenarios," Remote Sensing, vol. 11, no. 6, Article no. 607, March 2019.
[44] S. C. Kefauver, G. El-Haddad, O. Vergara-Diaz, and J. L. Araus, "RGB Picture Vegetation Indexes for High-Throughput Phenotyping Platforms (HTPPs)," Remote Sensing for Agriculture, Ecosystems, and Hydrology XVII, vol. 9637, Article no. 96370J, October 2015.
[45] B. T. W. Putra, P. Soni, B. Marhaenanto, S. S. Harsono, and S. Fountas, "Using Information from Images for Plantation Monitoring: A Review of Solutions for Smallholders," Information Processing in Agriculture, vol. 7, no. 1, pp. 109-119, March 2020.
[46] M. L. Pérez-Bueno, M. Pineda, and M. Barón, "Phenotyping Plant Responses to Biotic Stress by Chlorophyll Fluorescence Imaging," Frontiers in Plant Science, vol. 10, Article no. 1135, September 2019.
[47] J. Yao, D. Sun, H. Cen, H. Xu, H. Weng, F. Yuan, et al., "Phenotyping of Arabidopsis Drought Stress Response Using Kinetic Chlorophyll Fluorescence and Multicolor Fluorescence Imaging," Frontiers in Plant Science, vol. 9, Article no. 603, May 2018.
[48] I. Leinonen, O. M. Grant, C. P. P. Tagliavia, M. M. Chaves, and H. G. Jones, "Estimating Stomatal Conductance with Thermal Imagery," Plant, Cell, and Environment, vol. 29, pp. 1508-1518, August 2006.
[49] V. Sagan, M. Maimaitijiang, P. Sidike, K. Eblimit, K. T. Peterson, S. Hartling, et al., "UAV-Based High Resolution Thermal Imaging for Vegetation Monitoring, and Plant Phenotyping Using ICI 8640 P, FLIR Vue Pro R 640, and thermoMap Cameras," Remote Sensing, vol. 11, no. 3, Article no. 330, February 2019.
[50] J. Urban, M. Ingwers, M. A. McGuire, and R. O. Teskey, "Stomatal Conductance Increases with Rising Temperature," Plant Signaling and Behavior, vol. 12, no. 8, Article no. e1356534, August 2017.
[51] C. Baer, S. Gutierrez, J. Jebramcik, J. Barowski, F. Vega, and I. Rolfes, "Ground Penetrating Synthetic Aperture Radar Imaging Providing Soil Permittivity Estimation," IEEE MTT-S International Microwave Symposium, pp. 1367-1370, October 2017.
[52] T. Roitsch, L. Cabrera-Bosquet, A. Fournier, K. Ghamkhar, J. Jiménez-Berni, F. Pinto, and E. S. Ober, "Review: New Sensors and Data-Driven Approaches—A Path to Next Generation Phenomics," Plant Science, vol. 282, pp. 2-10, May 2019.
[53] Y. Lin, "LiDAR: An Important Tool for Next-Generation Phenotyping Technology of High Potential for Plant Phenomics?" Computers and Electronics in Agriculture, vol. 119, pp. 61-73, November 2015.
[54] C. S. Bekkering, J. Huang, and L. Tian, "Image-Based, Organ-Level Plant Phenotyping for Wheat Improvement," Agronomy, vol. 10, no. 9, Article no. 1287, August 2020.
[55] F. Baret, S. Madec, K. Irfan, J. Lopez, A. Comar, M. Hemmerlé, et al., "Leaf-Rolling in Maize Crops: From Leaf Scoring to Canopy-Level Measurements for Phenotyping," Journal of Experimental Botany, vol. 69, no. 10, pp. 2705-2716, April 2018.
[56] M. Tattaris, M. P. Reynolds, and S. C. Chapman, "A Direct Comparison of Remote Sensing Approaches for High-Throughput Phenotyping in Plant Breeding," Frontiers in Plant Science, vol. 7, Article no. 1131, August 2016.
[57] O. N. Lungu, L. M. Chabala, and C. Shepande, "Satellite-Based Crop Monitoring and Yield Estimation—A Review," Journal of Agricultural Science, vol. 13, no. 1, pp. 180-194, December 2020.
[58] C. Xie and C. Yang, "A Review on Plant High-Throughput Phenotyping Traits Using UAV-Based Sensors," Computers and Electronics in Agriculture, vol. 178, Article no. 105731, November 2020.
[59] D. C. Tsouros, S. Bibi, and P. G. Sarigiannidis, "A Review on UAV-Based Applications for Precision Agriculture," Information, vol. 10, no. 11, Article no. 349, November 2019.

[60] S. Tisné, Y. Serrand, L. Bach, E. Gilbault, R. Ben Ameur, H. Balasse, et al., "Phenoscope: An Automated Large-Scale Phenotyping Platform Offering High Spatial Homogeneity," The Plant Journal, vol. 74, no. 3, pp. 534-544, May 2013.

[61] S. Shajahan, I. Cannayen, and J. Hendrickson, "Monitoring Plant Phenology Using Phenocam: A Review," ASABE Annual International Meeting, Article no. 162461829, July 2016.

[62] J. Underwood, A. Wendel, B. Schofield, L. McMurray, and R. Kimber, "Efficient In-Field Plant Phenomics for Row-Crops with an Autonomous Ground Vehicle," Journal of Field Robotics, vol. 34, no. 6, pp. 1061-1083, September 2017.

[63] A. Gebremedhin, P. Badenhorst, J. Wang, K. Giri, G. Spangenberg, and K. Smith, "Development and Validation of a Model to Combine NDVI and Plant Height for High-Throughput Phenotyping of Herbage Yield in a Perennial Ryegrass Breeding Program," Remote Sensing, vol. 11, no. 21, Article no. 2494, October 2019.

[64] J. Zhang, Y. Huang, R. Pu, P. Gonzalez-Moreno, L. Yuan, K. Wu, et al., "Monitoring Plant Diseases and Pests through Remote Sensing Technology: A Review," Computers and Electronics in Agriculture, vol. 165, Article no. 104943, October 2019.

[65] M. Minervini, H. Scharr, and S. A. Tsaftaris, "Image Analysis: The New Bottleneck in Plant Phenotyping [Applications Corner]," IEEE Signal Processing Magazine, vol. 32, no. 4, pp. 126-131, July 2015.

[66] C. Zhao, Y. Zhang, J. Du, X. Guo, W. Wen, S. Gu, et al., "Crop Phenomics: Current Status and Perspectives," Frontiers in Plant Science, vol. 10, Article no. 714, June 2019.

[67] A. Shrestha and A. Mahmood, "Review of Deep Learning Algorithms and Architectures," IEEE Access, vol. 7, pp. 53040-53065, April 2019.

[68] X. Yang and M. Sun, "A Survey on Deep Learning in Crop Planting," IOP Conference Series: Materials Science and Engineering, vol. 490, Article no. 062053, April 2019.

[69] M. P. Pound, J. A. Atkinson, A. J. Townsend, M. H. Wilson, M. Griffiths, A. S. Jackson, et al., "Deep Machine Learning Provides State-of-the-Art Performance in Image-Based Plant Phenotyping," GigaScience, vol. 6, no. 10, pp. 1-10, October 2017.

[70] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," Communication of the ACM, vol. 60, pp. 84-90, June 2017.

[71] Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," Nature, vol. 521, pp. 436-444, May 2015.

[72] A. Kamilaris and F. X. Prenafeta-Boldú, "A Review of the Use of Convolutional Neural Networks in Agriculture," The Journal of Agricultural Science, vol. 156, no. 3, pp. 312-322, June 2018.

[73] S. Samiei, P. Rasti, J. Ly Vu, J. Buitink, and D. Rousseau, "Deep Learning-Based Detection of Seedling Development," Plant Methods, vol. 16, no. 1, Article no. 103, July 2020.

[74] N. Genze, R. Bharti, M. Grieb, S. J. Schultheiss, and D. G. Grimm, "Accurate Machine Learning-Based Germination Detection, Prediction and Quality Assessment of Three Grain Crops," Plant Methods, vol. 16, no. 1, Article no. 157, December 2020.

[75] J. R. Ubbens and I. Stavness, "Deep Plant Phenomics: A Deep Learning Platform for Complex Plant Phenotyping Tasks," Frontiers in Plant Science, vol. 8, Article no. 1190, July 2017.

[76] Y. Jiang, C. Li, R. Xu, S. Sun, J. S. Robertson, and A. H. Paterson, "DeepFlower: A Deep Learning-Based Approach to Characterize Flowering Patterns of Cotton Plants in the Field," Plant Methods, vol. 16, no. 1, Article no. 156, July 2020.

[77] Y. Perugachi-Diaz, J. M. Tomczak, and S. Bhulai, "Deep Learning for White Cabbage Seedling Prediction," Computers and Electronics in Agriculture, vol. 184, Article no. 106059, May 2021.

[78] A. Bauer, A. G. Bostrom, J. Ball, C. Applegate, T. Cheng, S. Laycock, et al., "Combining Computer Vision and Deep Learning to Enable Ultra-Scale Aerial Phenotyping and Precision Agriculture: A Case Study of Lettuce Production," Horticulture Research, vol. 6, Article no. 70, June 2019.

[79] L. Zhang, Z. Xu, D. Xu, J. Ma, Y. Chen, and Z. Fu, "Growth Monitoring of Greenhouse Lettuce Based on a Convolutional Neural Network," Horticulture Research, vol. 7, Article no. 124, August 2020.

[80] J. Y. Lu, C. L. Chang, and Y. F. Kuo, "Monitoring Growth Rate of Lettuce Using Deep Convolutional Neural Networks," ASABE Annual International Meeting, Article no. 1900341, July 2019.

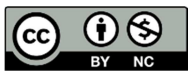
[81] S. V. Desai, V. N. Balasubramanian, T. Fukatsu, S. Ninomiya, and W. Guo, "Automatic Estimation of Heading Date of Paddy Rice Using Deep Learning," Plant Methods, vol. 15, no. 1, Article no. 76, July 2019.

[82] T. Yamaguchi, Y. Tanaka, Y. Imachi, M. Yamashita, and K. Katsura, "Feasibility of Combining Deep Learning and RGB Images Obtained by Unmanned Aerial Vehicle for Leaf Area Index Estimation in Rice," Remote Sensing, vol. 13, no. 1, Article no. 84, December 2021.

[83] A. Nasiri, A. Taheri-Garavand, and Y. D. Zhang, "Image-Based Deep Learning Automated Sorting of Date Fruit," Postharvest Biology and Technology, vol. 153, pp. 133-141, July 2019.

[84] X. Ni, C. Li, H. Jiang, and F. Takeda, "Deep Learning Image Segmentation and Extraction of Blueberry Fruit Traits Associated with Harvestability and Yield," Horticulture Research, vol. 7, Article no. 110, July 2020.
[85] M. Afonso, H. Fonteijn, F. S. Fiorentin, D. Lensink, M. Mooij, N. Faber, et al., "Tomato Fruit Detection and Counting in Greenhouses Using Deep Learning," Frontiers in Plant Science, vol. 11, Article no. 1759, November 2020.
[86] Y. Tian, G. Yang, Z. Wang, H. Wang, E. Li, and Z. Liang, "Apple Detection during Different Growth Stages in Orchards Using the Improved YOLO-V3 Model," Computers and Electronics in Agriculture, vol. 157, pp. 417-426, February 2019.
[87] A. A. Azman and F. S. Ismail, "Convolutional Neural Network for Optimal Pineapple Harvesting," ELEKTRIKA—Journal of Electrical Engineering, vol. 16, no.2, pp. 1-4, August 2017.
[88] N. Teimouri, M. Dyrmann, P. R. Nielsen, S. K. Mathiassen, G. J. Somerville, and R. N. Jørgensen, "Weed Growth Stage Estimator Using Deep Convolutional Neural Networks," Sensors, vol. 18, no. 5, Article no. 1580, May 2018.
[89] X. Hao, J. Jia, A. M. Khattak, L. Zhang, X. Guo, W. Gao, et al., "Growing Period Classification of Gynura Bicolor DC Using GL-CNN," Computers and Electronics in Agriculture, vol. 174, Article no. 105497, July 2020.
[90] S. Rasti, C. J. Bleakley, G. C. M. Silvestre, N. M. Holden, D. Langton, and G. M. P. O'Hare, "Crop Growth Stage Estimation Prior to Canopy Closure Using Deep Learning Algorithms," Neural Computing and Applications, vol. 33, no. 5, pp. 1733-1743, March 2021.
[91] A. Reyes-Yanes, P. Martinez, and R. Ahmad, "Real-Time Growth Rate and Fresh Weight Estimation for Little Gem Romaine Lettuce in Aquaponic Grow Beds," Computers and Electronics in Agriculture, vol. 179, Article no. 105827, December 2020.
[92] R. G. D. Luna, E. P. Dadios, A. A. Bandala, and R. R. P. Vicerra, "Tomato Growth Stage Monitoring for Smart Farm Using Deep Transfer Learning with Machine Learning-Based Maturity Grading," Journal of Agricultural Science, vol. 42, no. 1, pp. 24-36, January 2020.
[93] A. Koirala, K. B. Walsh, Z. Wang, and N. Anderson, "Deep Learning for Mango (Mangifera indica) Panicle Stage Classification," Agronomy, vol. 10, no. 1, Article no. 143, January 2020.
[94] S. Parvathi and S. Tamil Selvi, "Detection of Maturity Stages of Coconuts in Complex Background Using Faster R-CNN Model," Biosystems Engineering, vol. 202, pp. 119-132, February 2021.
[95] P. Shan, "Image Segmentation Method Based on K-Mean Algorithm," EURASIP Journal on Image and Video Processing, vol. 2018, Article no. 81, March 2018.
[96] G. Bebis and M. Georgiopoulos, "Feed-Forward Neural Networks," IEEE Potentials, vol. 13, no. 4, pp. 27-31, October 1994.
[97] S. Srivastava, A. V. Divekar, C. Anilkumar, I. Naik, V. Kulkarni, and V. Pattabiraman, "Comparative Analysis of Deep Learning Image Detection Algorithms," Journal of Big Data, vol. 8, no. 1, Article no. 66, May 2021.
[98] M. Hashemi, "Enlarging Smaller Images before Inputting into Convolutional Neural Network: Zero-Padding vs. Interpolation," Journal of Big Data, vol. 6, no. 1, Article no. 98, November 2019.
[99] M. Capra, B. Bussolino, A. Marchisio, M. Shafique, G. Masera, and M. Martina, "An Updated Survey of Efficient Hardware Architectures for Accelerating Deep Convolutional Neural Networks," Future Internet, vol. 12, no. 7, Article no. 113, July 2020.
[100] W. Wang and Y. Yang, "Development of Convolutional Neural Network and Its Application in Image Classification: A Survey," Optical Engineering, vol. 58, no. 4, Article no. 040901, April 2019.
[101] M. D. Zeiler and R. Fergus, "Visualizing and Understanding Convolutional Networks," European Conference of Computer Vision, pp. 818-833, September 2014.
[102] K. Simonyan and A. J. C. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," https://arxiv.org/pdf/1409.1556.pdf, April 10, 2015.
[103] C. Szegedy, L. Wei, J. Yangqing, P. Sermanet, S. Reed, D. Anguelov, et al., "Going Deeper with Convolutions," IEEE Conference on Computer Vision and Pattern Recognition, pp. 1-9, June 2015.
[104] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," IEEE Conference on Computer Vision and Pattern Recognition, pp. 770-778, June 2016.
[105] G. Huang, Z. Liu, L. Van Der Maaten, and K. Weinberger, "Densely Connected Convolutional Networks," IEEE Conference on Computer Vision and Pattern Recognition, pp. 2261-2269, July 2017.
[106] X. Glorot and Y. Bengio, "Understanding the Difficulty of Training Deep Feedforward Neural Networks," 13th International Conference on Artificial Intelligence and Statistics, pp. 249-256, May 2010.
[107] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, pp. 1735-1780, November 1997.
[108] Z. Karevan and J. A. K. Suykens, "Transductive LSTM for Time-Series Prediction: An Application to Weather Forecasting," Neural Networks, vol. 125, pp. 1-9, May 2020.

[109] H. Sak, A. Senior, and F. J. A. Beaufays, "Long Short-Term Memory Based Recurrent Neural Network Architectures for Large Vocabulary Speech Recognition," https://arxiv.org/pdf/1402.1128.pdf, February 05, 2014.
[110] V. Frinken, F. Zamora-Martínez, S. España-Boquera, M. J. Castro-Bleda, A. Fischer, and H. Bunke, "Long-Short Term Memory Neural Networks Language Modeling for Handwriting Recognition," 21st International Conference on Pattern Recognition, pp. 701-704, November 2012.
[111] J. Zhu, H. Chen, and W. Ye, "A Hybrid CNN-LSTM Network for the Classification of Human Activities Based on Micro-Doppler Radar," IEEE Access, vol. 8, pp. 24713-24720, February 2020.
[112] C. Uyulan, "Development of LSTM & CNN Based Hybrid Deep Learning Model to Classify Motor Imagery Tasks," https://www.biorxiv.org/content/biorxiv/early/2020/12/28/2020.09.20.305300.full.pdf, November 23, 2020.
[113] M. Alhussein, K. Aurangzeb, and S. I. Haider, "Hybrid CNN-LSTM Model for Short-Term Individual Household Load Forecasting," IEEE Access, vol. 8, pp. 180544-180557, October 2020.
[114] J. Donahue, L. A. Hendricks, M. Rohrbach, S. Venugopalan, S. Guadarrama, K. Saenko, et al., "Long-Term Recurrent Convolutional Networks for Visual Recognition and Description," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, pp. 677-691, April 2017.
[115] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," IEEE Conference on Computer Vision and Pattern Recognition, pp. 779-788, June 2016.
[116] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," IEEE Conference on Computer Vision and Pattern Recognition, pp. 6517-6525, July 2017.
[117] J. Redmon and A. J. A. Farhadi, "YOLOv3: An Incremental Improvement," https://arxiv.org/pdf/1804.02767.pdf, April 08, 2018.
[118] A. Bochkovskiy, C. Y. Wang, and H. J. A. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," https://arxiv.org/pdf/2004.10934.pdf, April 23, 2020.
[119] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," IEEE Conference on Computer Vision and Pattern Recognition, pp. 580-587, June 2014.
[120] R. Girshick, "Fast R-CNN," IEEE International Conference on Computer Vision, pp. 1440-1448, December 2015.
[121] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137-1149, June 2017.
[122] S. A. Sanchez, H. J. Romero, and A. D. Morales, "A Review: Comparison of Performance Metrics of Pretrained Models for Object Detection Using the TensorFlow Framework," IOP Conference Series: Materials Science and Engineering, vol. 844, Article no. 012024, June 2020.
[123] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," IEEE International Conference on Computer Vision, pp. 2980-2988, October 2017.
[124] S. Paulus, "Measuring Crops in 3D: Using Geometry for Plant Phenotyping," Plant Methods, vol. 15, no. 1, Article no. 103, September 2019.
[125] M. A. Ganaie, M. Hu, M. Tanveer, and P. J. A. Suganthan, "Ensemble Deep Learning: A Review," https://arxiv.org/pdf/2104.02395.pdf, April 06, 2021.
[126] S. Sakurai, H. Uchiyama, A. Shimada, and R. I. Taniguchi, "Plant Growth Prediction Using Convolutional LSTM," 14th International Conference on Computer Vision Theory and Applications, pp. 105-113, February 2019.
[127] R. Yasrab, J. Zhang, P. Smyth, and M. P. Pound, "Predicting Plant Growth from Time-Series Data Using Deep Learning," Remote Sensing, vol. 13, no. 3, Article no. 331, January 2021.



Copyright© by the authors. Licensee TAETI, Taiwan. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-NC) license (https://creativecommons.org/licenses/by-nc/4.0/).