

Survey paper

AI meets UAVs: A survey on AI empowered UAV perception systems for precision agriculture

Jinya Su^{a,b,*}, Xiaoyong Zhu^c, Shihua Li^{a,b}, Wen-Hua Chen^d^a School of Automation, Southeast University, Nanjing 210096, China^b Key Laboratory of Measurement and Control of Complex Systems of Engineering, Ministry of Education, Southeast University, 210096 Nanjing, China^c School of Electrical and Information Engineering, Jiangsu University, Zhenjiang 212013, China^d Department of Aeronautical and Automotive Engineering, Loughborough University, Loughborough LE11 3TU, UK

ARTICLE INFO

Article history:

Received 31 July 2022

Revised 14 October 2022

Accepted 3 November 2022

Available online 11 November 2022

Communicated by Zidong Wang

Keyword:

Deep learning

Machine learning

Precision agriculture

Remote sensing

Smart agriculture

Unmanned Aerial Vehicle (UAV)

ABSTRACT

Precision Agriculture (PA) promises to boost crop productivity while reducing agricultural costs and environmental footprints, and therefore is attracting ever-increasing interests in both academia and industry. This management strategy is underpinned by various advanced technologies including Unmanned Aerial Vehicle (UAV) sensing systems and Artificial Intelligence (AI) perception algorithms. In particular, due to their unique advantages such as a low cost, high spatio-temporal resolutions, flexibility, automation functions and minimized risk of operation, UAV sensing systems have been extensively applied in many civilian applications including PA since 2010. In parallel, AI algorithms (deep learning since 2012 in particular) are also drawing ever-increasing attention in different fields, since they are able to analyse an unprecedented volume/velocity/variety of data (semi-) automatically, which are also becoming computationally practical with the advancements of cloud computing, Graphics Processing Units and parallel computing. In this survey paper, therefore, a thorough review is performed on recent use of UAV sensing systems (e.g., UAV platforms, external sensing units) and AI algorithms (mainly supervised learning algorithms) in PA applications throughout the crop life-cycle, as well as the challenges and prospects for future development of UAVs and AI in agriculture sector. It is envisioned that this review is able to provide a timely technical reference, demystifying and promoting research, deployment and successful exploitation of AI empowered UAV perception systems for PA, and therefore contributing to addressing future agricultural and human nutrition challenges.

© 2022 Elsevier B.V. All rights reserved.

1. Introduction

1.1. Background

Agriculture: Agriculture plays an essential role worldwide (developing countries and regions in particular) and promises to provide the main source of food, income and employment (for 1 billion + people across the globe) to their rural populations. However, agriculture is now facing severe global grand challenges, e.g., 69% increase in food demand by 2050 to meet the unprecedented needs of 9.7 Billion people (by United Nation Food and Agriculture Organization) while with an increasing ageing-population structure (i.e., reduced and ageing farming workforce), decreasing ar-

able land and natural farming resources (e.g., land and water), increasingly unpredictable climatic conditions and production costs, and societal demand for shrinking agriculture's environmental footprints [1]. Recently, the outbreak of COVID-19 pandemic has also posed many new challenges in agriculture systems worldwide affecting major areas of production, dietary changes, environment, transportation, and overall supply chain [2]. Therefore, there is an urgent need to make agricultural systems being highly resource-efficient and sustainable, e.g., profitable, environmentally sustainable, and less manpower-dependent [3].

Precision Agriculture (PA) promises to boost crop productivity while reducing input costs and environmental footprints by focusing on the “four right” including the *right* management practice, at the *right* time, with the *right* amount of inputs and in the *right* place. It can also improve the quality of working environment and social aspects of farming [4]. This management strategy is underpinned by different advancements such as sensing technologies (i.e., analogue of the role of human eye), communication

* Corresponding author at: School of Automation, Southeast University, Nanjing 210096, China.

E-mail addresses: sucas@seu.edu.cn (J. Su), zyff@ujs.edu.cn (X. Zhu), lsh@seu.edu.cn (S. Li), W.Chen@lboro.ac.uk (W.-H. Chen).

systems (i.e., the role of mouth and ear for information transfer), Artificial Intelligence (AI) algorithms (i.e., the role of human brain), robotics and automation systems and technologies (e.g., variable rate technology) (i.e., the role of human foots and arms). Particularly, the advancement of sensing technologies makes it possible to acquire multi-modal data (e.g., spectral, spatial and semantic data) more efficiently while with unprecedented resolutions in terms of spectral, spatial and temporal for timely non-destructive monitoring, identifying in-field and temporal variability of crop and soil conditions for site-specific crop stress management. Moreover, AI algorithms (deep learning (DL) since 2012 in particular) are able to analyse an unprecedented volume/ velocity/ variety (3 V) of data, which are also becoming computationally practical with the popularity of cloud computing, edge intelligence (e.g., Central Processing Unit (CPU), Graphics Processing Units (GPU), Field-programmable gate array (FPGA), Application-specific integrated circuits (ASICs)), and parallel computing. In addition, robotics and automation technologies are also able to make precision and automated site-specific agriculture management (e.g., stress treatment) possible with the aid of navigation systems and variable-rate technologies. Among these advancements, Unmanned Aerial Vehicles (UAVs) represent one of the state-of-the-art (SOTA) robotic platforms for mobile data collection and target intervention (e.g., pesticide and fertilizer spraying), while AI algorithms are becoming the mainstream algorithms for data analysis and decision making, particularly in the era of “big data”.

Sensing systems: Timely obtaining plant status throughout their growth life-cycle is paramount for PA decision making and smart agriculture management. Conventional manual sensing by human field sampling, although being accurate, requires the instrument operators to perform intensive field data collection due to field spatial heterogeneity and therefore is destructive, sparse, labour-intensive, time-consuming and unsuitable for modern large-scale field applications. Deploying Internet of Things (IoTs) systems by distributing physical sensors in the field can partially alleviate the issues but is challenging as well, since they inevitably interfere with field operations such as tillage, planting, spraying and harvesting [5]. Therefore, different (semi-) autonomous platform-based mobile sensing systems have been developed in the past few decades to address these challenges such as ground-based (e.g., mobile device, robot, tractor) systems, UAV sensing systems, manned aircraft- and satellite-based systems [6]. In these mobile systems, various sensing units (e.g., optical sensors, LIDAR) are installed to acquire data reflecting chemical, biological, and physical characteristics of plants or soil of interest so that crop stress can be detected timely in a non-destructive manner and therefore treated precisely by agricultural robotic systems.

UAV sensing systems: Due to the increasing accessibility in terms of flexibility, automation functions (e.g., navigation and flight control), and cost, UAVs are drawing ever-increasing popularity in various civilian applications since 2010 (e.g., agriculture, forestry, environment, cargo transport, construction and mining, emergency management, industrial inspection, insurance, law enforcement, media and telecommunications, search and rescue, etc.) [7,8]. Major agricultural UAV companies include Da-Jiang Innovations (DJI), Parrot, Precisionhawk, AGEagle, Trimble Navigation, Microdrones, 3D Robotics, etc. [9], which are mainly focused on agricultural stress detection and drone spraying. Meanwhile, there are also many advances in the miniaturization of sensing units (e.g., spectral cameras, LIDAR) on-board UAVs for environmental perception. As a result, UAV (remote) sensing systems have become more modular, miniaturized and intelligent recently, particularly with the emergence of commercial UAV manufacturers such as DJI (Shenzhen, China). It is reported by Allied Market Research in 2021 that the global UAV market was valued at \$24.72 billion in

2020, and is projected to reach \$70.91 billion by 2030, witnessing a Compound Annual Growth Rate (CAGR) of 11.7% from 2021 to 2030, while the drone camera market garnered \$3.33 billion in 2019, and is expected to reach \$25.39 billion by 2027, witnessing a CAGR of 29.9% from 2020 to 2027.

UAV sensing systems have many unique advantages over ground-, aircraft- and satellite-based sensing systems. For example, ground-based manual approach [10], although being accurate, is time- and labor-intensive, costly, terrain-dependent and only provides insights on specific locations, while UAV sensing can avoid these problems due to its wider and denser coverage with minimal human involvement. Compared to manned aircraft- and satellite-based sensing systems, UAV sensing is more flexible, less reliant on cloud conditions, with a high cost-effectiveness and accuracy, with flexible multiple-sources data acquisition capacity due to various add-on sensing units, and possess user-defined spectral-spatial-temporal image resolutions (e.g., via user-defined sensors, flight height and revisit time) [11]. It is also noted that UAV sensing systems can also act as a bridge/relay between ground- and satellite-based sensing systems so that ground-truthing satellite data becomes much efficient for regional agricultural applications.

Artificial Intelligence (AI): Upon obtaining a large amount of aerial data, the next step would be data analysis and interpretation in support of decision making for agriculture management. Conventional manual data analysis based on human expert knowledge is important (i.e., indispensable in data labelling), but is generally time-consuming, labor-intensive, costly and therefore unsuitable for large-scale applications. Data-driven AI methods are becoming more and more popular due to their high effectiveness and efficiency, particularly with the advent of large-scale dataset and high-performance computing units. In particular, DL approaches [12,13], although first greeted with much skepticism, have proved to be effective in solving many practical problems and are therefore becoming an enabling tool in many areas such as transportation, healthcare, finance, industrial inspection among others [14]. These successes have been mainly attributed to three main factors including: (1) technical advancements in network architecture design (e.g., deeper, wider, smaller) [15], parameter initialization and training methods, (2) increasing availability of powerful computational hardware such as cloud computing, GPU and parallel processing, and (3) increasing availability of large-scale labelled datasets (e.g., ImageNet dataset with 1.2 million images classified in 1000 categories) for model training avoiding the problem of overfitting. AI has the potentials to transform agriculture at various agronomically relevant scales as well from plant and plot scales to field and regional scales. There are also many studies which have adopted AI algorithms in UAV based PA applications [16]. However, there is still a lack of survey paper systematically demonstrating how various AI algorithms are adopted to address the common challenges in various PA elements with UAV (remote) sensing images throughout crop growth lifecycle.

1.2. Aims and scope of this work

At present, due to the importance of this emerging field, there are already a few review papers on related topics (e.g., the use of UAV, sensing technologies, AI in agriculture), where some highly relevant ones are compared in Table 1 including review area, main topics and features and gaps against this study. [9] reviews the latest trends and applications of UAVs in agriculture with main emphasis on agricultural UAVs (e.g., platform, hardware, communication), control technologies (e.g., modelling, control, task allocation and swarm), and applications. However, AI algorithms and PA elements are not covered. [17] explores the types of sensors (e.g., location-based sensors, optical sensors) for smart farming,

Table 1

Summary of some existing survey papers on related topics: the use of UAVs, sensing technologies and AI in agriculture.

Ref	Review area	Main topics	Features and gaps
[9]	UAV in agriculture	• Agricultural UAV systems; • Control technologies of UAV; • UAV applications in agriculture.	• AI algorithms are not covered; • Different PA elements are not covered.
[17]	UAV in smart agriculture	• UAVs and agriculture sensors; • Applications and case studies; • Requirements, challenges and future directions.	• AI algorithms are not covered; • PA elements are not detailed.
[18]	UAV for PA applications	• Typical UAV PA applications (e.g. crop health, growth, weed, irrigation); • Typical data processing pipeline.	• AI algorithms are not covered; • PA elements are not systematic.
[19]	UAV RS for PA applications	• UAV RS in a number of PA applications; • Translating scientific results into common practice.	• AI algorithms are not covered; • PA elements and UAV sensors are not complete.
[20]	UAV RS for field phenotyping	• UAV sensing platforms for phenotyping; • Common methods for UAV based phenotyping; • Challenges of UAV based phenotyping.	• The focus is on phenotyping instead of PA elements; • AI algorithms are not systematic.
[22]	IoT and UAV in smart farming	• IoT in agriculture; • Some applications of UAVs in agriculture.	• UAVs in agriculture is not systematic; • AI algorithms are not covered.
[23]	UAV sensing for PA applications	• UAV sensing systems; • A number of UAV PA applications; • UAV deployment architectures in PA.	• AI algorithms are not covered; • Use of sensing in PA is not systematic.
[24]	UAVs and sensors in PA	• Technical characterization of UAVs and sensors in PA; • Practical challenges of UAV sensing in PA.	• AI algorithms are not covered; • PA elements are not systematic.
[25]	UAV for RS applications	• Generic UAV RS applications; • Improved UAV images over satellite RS.	• The focus is on generic RS instead of PA; • AI algorithms are not systematic.
[7]	DL for UAVs	• DL in UAV operations including perception, guidance, navigation and control.	• The focus is on UAV operations instead of PA.
[16]	DL for UAV agricultural images	• Object detection and semantic segmentation networks for UAV agricultural images; • UAV-based datasets.	• PA elements are not covered; • Only CNN approaches are considered.
[26]	DL in UAV RS	• DL based classification and regression for UAV RS; • Categorization was performed in architecture, evaluation, task, application and sensor.	• The main focus is on generic RS instead of PA; • Algorithms are limited to DL.

potential requirements and challenges for operating UAVs in smart agriculture, and also the future applications of UAVs in smart farming. However, the emphasis is on the role of bluetooth technology for UAV operations, where AI algorithms and PA elements are not covered or detailed. [18] presents a review on UAV applications for PA, where the focus is on different UAV PA applications, agricultural crops and typical data processing pipelines. [19] reviews the progress of UAV Remote Sensing (RS) in agriculture, where the focus is on several typical PA applications such as drought stress, weed and pathogen detection, nutrient status and growth vigor, and yield prediction, however, various methodologies are not detailed. [20] reviews the current status and perspectives on the use of UAV RS for field-based phenotyping (e.g., acquiring crop traits such as plant height, leaf area index, chlorophyll content, biomass and yield) [21], where the emphasis is on UAV sensing platforms (e.g., UAVs, sensors), commonly adopted methods (e.g., data pre-processing and typical data analysis) and the challenges for UAV based phenotyping. [22] reviews IoTs and UAVs in smart farming, however, the focus on IoTs in agriculture and the review for UAVs in smart farming is not systematic. [23] presents a survey on UAV sensing for PA, where the focus is on UAV sensing systems, various UAV PA applications and UAV deployment architectures. [24] provides a practical guide on technical characterization of common UAVs used in PA and helps select the most suitable UAVs and on-board sensors for various agricultural operations, however, the focus is on UAVs and UAV sensors for PA application without an detailed account of AI algorithms or PA elements. [25] performs a review on RS tasks that involve UAV data, however, this review is mainly on the generic UAV RS (e.g., land cover mapping, change detection) with a focus on improved UAV images (e.g., high resolution, better spectral data, multi-modal data) over conventional airborne and space-borne RS. [7] reviews DL methods and applications for UAV operations in terms of perception, guidance, navigation and control, rather than PA applications. [16] reviews recent Convolutional Neural Network (CNN) based methods for UAV based RS

images to classify agricultural crops, where the emphasis is on different CNN algorithms such as object detection networks, semantic segmentation networks, and UAV-based datasets. [26] presents a review of DL methods applied in UAV-based imagery, where the main focus is on classification and regression techniques instead of PA applications. There are also other review papers on the use of different UAV cameras for PA such as thermal camera [3,27], hyperspectral camera [28], which will not be detailed here due to their specific applications and the lack of space.

As evident in Table 1, although there are already a few studies on different aspects of PA (e.g., UAV for agriculture, DL for UAV RS), there is still a lack of comprehensive survey paper focusing on the interaction between AI and UAVs in the context of PA, where UAVs play the major role of data collection and AI plays a dominant role for data analysis for various PA elements in crop growth life-cycle. Therefore, in this survey paper, a non-exhaustive but systematic review is performed on recent use of UAV sensing systems and AI algorithms in PA applications, as well as challenges and prospects for future development of AI empowered UAV perception systems in agriculture sector. It is noted that the main scope of UAVs in this study is focused on UAV sensing systems, and therefore UAV spraying systems (e.g., decision implementation systems) [29] or seed sowing systems are not covered. In addition, the UAVs in this work mainly refer to small UAVs (less than 20 kg) due to the more strict regulations on large UAVs. It is also noted that the literature in this work, applications to different crops in particular, may not be complete as this work is mainly on highlighting application domains of AI empowered UAV perception systems instead of focusing on particular crops. To be more exact, the main contributions of this review paper are highlighted as below.

- An integrated PA framework is presented, highlighting the role of UAVs in data collection, AI algorithms in data analysis and decision making, and also various PA elements in crop growth life-cycle;

- UAV sensing systems are detailed including various platforms and sensing units for PA applications, where their pros and cons are compared in details along with some selection guidance;
- AI algorithms for UAV based PA applications are reviewed and categorized, where most studies are on supervised learning including regression, classification and object detection with a few studies on unsupervised learning, mainly addressing the issue of the requirement of large amounts of labelled dataset in supervised learning;
- Challenges of the use of AI and UAV sensing in PA applications are discussed along with possible solutions in support of their future rapid deployment and successful exploitation.

The remainder of this paper is organized as follows. In Section 2, we present PA framework, highlighting the role of AI, UAV sensing systems and also presenting various PA elements in roughly chronological order. UAV sensing systems including different UAV platforms and external sensing units for PA applications are discussed in Section 3. Section 4 discusses AI algorithms for UAV based PA applications, where the main focus and applications are supervised learning (e.g., regression, classification and object detection) with a few studies on unsupervised learning. In Section 5, challenges along with possible solutions are presented. Section 6 concludes this work.

2. UAV based PA framework

In this section, according to the closed-up agriculture field management process: data collection, analysis, decision and intervention, an integrated PA framework will be introduced (please refer to Fig. 1 for the integrated diagram) so that the roles of UAV sensing systems, AI algorithms and various PA elements can be illustrated. In traditional agriculture, fertilizers, irrigation, or pesticide control are usually applied at a same rate at prescribed times/frequencies (e.g., calendar-based strategy), in a uniform (or blanket) manner irrespective of crop status and spatial variability, and with a reactive (instead of proactive) approach [30]. This approach is becoming obsolete due to their inefficiency and un-sustainability in the use of agricultural resources, owing to the inherent spatiotemporal heterogeneity of crop stress. The concept of PA has emerged to make agriculture management efficient and sustainable. In particular, the International Society of Precision Agriculture

(<https://www.ispag.org/>) defines PA in 2019 as: “Precision agriculture is a management strategy that gathers, processes and analyzes temporal, spatial and individual data and combines it with other information to support management decisions according to estimated variability for improved resource use efficiency, productivity, quality, profitability and sustainability of agricultural production.”

Literature search is conducted in Web of Science (WoS) database to see the publication trend and also countries/regions in PA related areas. In particular, the number of publications by keyword “Precision Agriculture” (upper plots), and keywords “UAV” AND “Precision Agriculture”(lower plots) are displayed in Fig. 2. It follows from Fig. 2 that there are an exponentially increasing number of publications in “Precision Agriculture” since about 2000 (reaching 3000 publications in 2021), and the top two countries are China and USA. While in terms of “UAV” AND “Precision Agriculture”, the publications start to increase exponentially since about 2014 (reaching about 300 publications in 2021) and the top three countries are China, USA and Italy.

In the PA framework in Fig. 1, for fields of interest, UAVs, along with other sensing units (e.g., satellite RS, proximal RS by ground robotics systems and IoT devices), are able to provide a low-cost and flexible platform for aerial data acquisition. The collected data in PA framework mainly include images of various types for AI perception and numerical values for decision making (e.g., regarding weather, soil and crop variables). Upon data collection, data analysis and interpretation can be done by various AI algorithms so that the spatial, temporal variability of crop and soil can be achieved. Then decisions can be made through crop (simulation) modelling [5] or data-driven approaches along with knowledge base of crop and soil sciences so that targeted interventions (e.g., the “four right”: treatment, amount, place and time) can be implemented by robotics systems (e.g., ground tractors/robotics, UAVs), generating economic, social and environmental impacts. The main shift of PA over conventional agriculture is mainly threefold: the calendar based intervention is changed to information based one, the uniform (or blanket) approach is changed to precision (or site-specific) one and the reactive approach is changed to proactive one. Typical PA elements that can be achieved by AI empowered UAV perception systems are detailed as below.

Crop planting area estimation: A timely and accurate estimation of crop planting area is important in crop management, crop

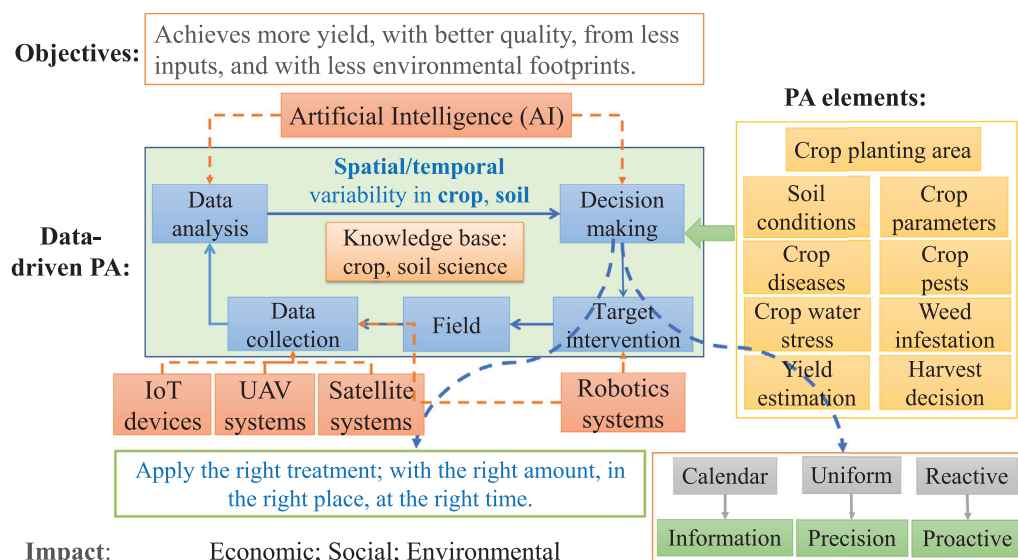


Fig. 1. PA framework including objectives, data-driven PA management framework, PA elements and various impacts.

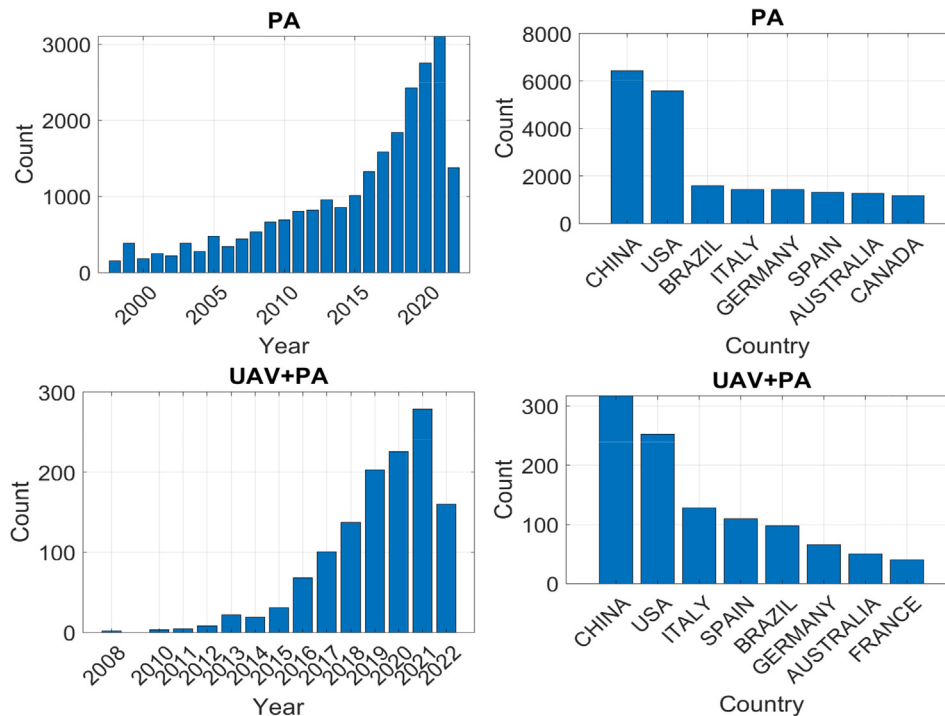


Fig. 2. Number of publications by year (left) and country (right), by keyword “Precision Agriculture”(upper plots) and keywords “UAV” AND “Precision Agriculture”(lower plots): results returned from WoS database on 25/Jul/2022.

production estimation and decision making for various stakeholders. Medium and low spatial-resolution RS images from satellites are suitable for large-scale applications [31], but may not be suitable for small plots due to their poor spatial resolution and cloud problem, particularly for mountainous areas. UAV RS is therefore becoming popular in small-scale crop planting area estimation due to their flexibility in spatial resolutions and sensor selections. The problem of crop planting area estimation is usually formulated as a supervised classification problem, where both Conventional Machine Learning (CML) and recent DL approaches can be applied. For example, UAV hyperspectral images have been applied in [32] for vegetation classification, where the so-called object-based image analysis (OBIA) is applied by integrating unsupervised image segmentation (i.e., segment generation) and CML based classification with various extracted features (e.g., spectral features, shape feature, texture feature) of the segments. UAV RGB and multispectral imaging are applied in [33] for soybean planting area estimation by comparing deep semantic segmentation networks (e.g., DeepLabv3 + and U-Net) against CML (e.g., Random Forest (RF) and SVM based classifier), where U-Net semantic segmentation network obtains the best performance. The advantages of UAV RS with a high spatial resolution and satellite RS with a wider spectral range are combined in [34] by using Gram-Schmidt transformation (i.e., image information fusion) for the purpose of crop classification via pixel-wise RF classifier. It is noted that crop planting area estimation based on one-off aerial images is relatively simple in terms of logistics, however, the performance may be limited as crops may share very similar characteristics in some stages (e.g., early stages in particular). Therefore, time-series approaches integrating measurements at different stages, capturing various crop phenology characteristics, are also being used, particularly in satellite RS applications due to the easy availability of time-series RS data (e.g., Sentinel-2 series) [31].

Soil conditions: Soil conditions, e.g., soil moisture content (SMC), salinity, erosion, organic matter content (e.g., nitrogen), clay content, pH, bulk density are vital for crop development. Therefore,

soil conditions ought to be monitored regularly, which can be achieved by UAV RS if these soil conditions can be reflected by soil surface reflection. Soil condition assessment is usually formulated as a regression problem, where UAV spectral measurements (e.g., multispectral, hyperspectral, thermal) are to derive suitable features and ground samplings provide groundtruth soil parameter values. For example, different regression approaches are adopted in [35] (e.g., BP neural network) and [36] (e.g., XGBoost) to derive the SMC from UAV-based multispectral and hyperspectral images, respectively. Multiple linear regression is also used in [37] to estimate soil salinity from UAV-based hyperspectral, multispectral and thermal images, and also LiDAR data. It should be noted that the main challenges for soil condition assessment by UAV RS lie in whether the soil surface reflectance can reflect soil properties under soil surface or not, and the problem of occlusion and shadows brought by crop structures. Therefore, soil experts should be involved in determining the suitability of UAV RS for soil condition assessment, and the suitable crop development stage for soil assessment (due to occlusion and shadow issue) should also be defined.

Crop parameters: Crop parameters throughout the vegetation period are important indices in monitoring crop growth, yield and quality. Depending on the crops of interest, these crop parameters may include crop height [38], biomass [39,40], nitrogen content [41], canopy cover [42], leaf area index (LAI) [43] among others. Estimation of these crop parameters is either formulated as a regression problem [39] (e.g., biomass, nitrogen content, LAI) or classification problem (e.g., canopy cover) [42]. For example, in [39], a regression model is adopted to estimate biomass of barley from plant height, where plant height is estimated as the difference between digital (crop) surface model and digital elevation model. Both crop surface and ground models are generated from overlapped UAV color images via image mosaicking, point cloud generation and digital surface model. In [42], Machine Learning (ML) based classification (soil/vegetation classification) and UAV multispectral imaging are used to derive canopy cover for winter

wheat, which is defined by the vegetation percentage in given areas after classification.

Crop diseases: An accurate and timely monitoring of crop diseases enables a timely targeted fungicide application to stop their spread to surrounding regions so that their adverse effects can be effectively minimized with a reduced use of pesticides. This can be achieved by capturing the physical and chemical changes (e.g., Chlorophyll content reduction, water loss, visual symptoms) on crop canopy by spectral sensors onboard UAV systems. The task of crop diseases detection/quantification can generally be formulated as either a regression or classification problem as below.

- In term of regression, the target/dependent variable is usually chosen as the disease severity value (e.g., disease index/score via expert inspection), and the independent variables are different image-derived features (e.g., vegetation indices at pixel-level, patch-level [44] or plot-level [45], deep features by CNN [45]). Then various regression algorithms (e.g., linear regression [45], multilayer perception, support vector regression, RF regression [44]) can be applied to build the regression relationship;
- In term of classification, the target variable is defined as the different categories of crop diseases (e.g., healthy vs diseased or different levels of disease severity). Then both CML algorithms with handcrafted features (e.g., vegetation indices) [30,46] and DL algorithms with automatically learnt deep features [11] can be applied to classify the healthy or diseased regions of different disease severity levels. It should be noted that in comparison to crop parameters estimation, crop diseases detection has a higher requirement on image resolution (e.g., spatial resolution, spectral range and spectral resolution) to capture the subtle visible or non-visible changes caused by crop diseases, particularly for early crop disease detection.

Crop pests: Early diagnosis of crop pests (e.g., splices, population) is essential because the damage may spread quickly in favourable conditions. The crop pest monitoring approaches by using UAVs can be broadly categorized into two approaches including direct and indirect crop pest monitoring. Moreover, the problem is usually formulated as object detection [47], regression [48] or classification [49] problem, which are detailed as below.

- In direct crop pest monitoring, UAVs play the role of either a moving platform for insect sampling or an aerial imaging device for image acquisition in support of follow-up image analysis. The UAV aerial sampling approach may provide some advantages over conventional manual sampling in term of accessibility and efficiency. For example, [50] uses UAV with remote-controlled insect net for insect population monitoring in rice field. UAV and sticky traps are used in [47] to estimate the population of soft-skinned fruit pests, where UAV as a moving platform is to take aerial images of the sticky traps for object detection and counting, while [49] applies UAV RGB camera to capture soybean pest images for pest classification;
- In indirect crop pest monitoring, the differences among healthy crop fields and different levels of pest-infected crop fields are usually assessed. [51] considers the detection of rice sheath blight using UAV with color and multispectral images, where regression analysis is conducted between vegetation indices and disease severity [48]. There is also a survey paper on the use of RS for forest pest damage assessment [52]. However, it should be noted that in the indirect crop pest monitoring extra agronomy knowledge and expertise (i.e., pest outbreak conditions) should be incorporated, since it would be a challenge to

tell the reflectance differences of crop damaged by pest, diseases, nutrient or drought without incorporating crop, soil and pest characteristics.

Crop water stress: Water stress has adverse effects on crops such as closing stomata, reducing transpiration rate and up-taking less carbon dioxide, inevitably diminishing crop growth, productivity and resulting in less biomass and yield. The existing UAV based studies on crop water stress can be broadly divided into two categories including index based approaches [53] and regression approaches. A review paper on UAV RS for water stress assessment is also available in [54], which highlights different UAVs systems and indices for such a purpose.

- In index approaches, various vegetation indices are first derived from UAV image bands or wavelengths so that the most sensitive one can be chosen as the crop water stress index (CWSI) [55]. For example, [56] maps the spatial variability of vineyard water deficits by using CWSI derived from UAV thermal imaging. [53] proposes the adaptive CWSI by using UAV thermal image for nectarine and peach orchards;
- In regression approaches [57], the functional relationship between UAV image-derived features (e.g., vegetation indices) and crop water content measurements via field sampling is derived. For example, maize water content is estimated in [57] by using various regression approaches (SVR, RF regression, DT regression, ANN and PLSR) and UAV multispectral imaging.

Weed infestation: Weeds, aggressively competing with crops for water, nutrients, sunlight and growing space, are responsible for an approximate 35% reduction in potential global crop yields. Accurate weed mapping is a prerequisite for site-specific weed management, reducing the use of herbicides thus contributing to an increased agriculture sustainability. Weed infestation monitoring can be usually formulated as a regression, classification or object detection problem. In particular, a survey paper on DL for weed detection is also available in [58], highlighting data acquisition, dataset preparation, DL techniques for detection, localisation and classification, and also evaluation metrics.

- In regression approaches [59], weed density is chosen as the response variable and various UAV image-derived information is chosen as feature variables so that the regression relationship can be learnt for mapping purpose;
- While weed classification can be at either image-level (i.e., image-wise classification) or pixel-level (e.g., semantic segmentation). Both CML (e.g., RF classifier) with handcrafted features (e.g., vegetation indices) [60] and deep semantic segmentation networks (e.g., SegNet [61]) have been successfully applied to different crop weeds;
- Considering the weed patch characteristics, in object detection approach for weed detection [62], weed patches can be labelled by using bounding box so that object detection networks can be applied such as Faster Region-based Convolutional Neural Networks (RCNN) and Single Shot Detector (SSD) among others [58].

Yield prediction: An accurate and early yield estimation is important in plant growing situation monitoring, productivity maximisation, crop insurance, planning of harvest and storage requirement, etc. There are mainly two yield prediction approaches according to the plant/crop of interest including regression approach [63] and plant detection and counting approach, where the latter is mainly for fruit plants [64,65].

- Yield regression is to establish a mapping from various input features (e.g., various vegetation indices or their combination, plant height, etc.) to yield variable based on data of various crop plots, where each plot denotes a sampling in the regression model. For example, UAV-based multimodal data (e.g., RGB, multispectral, thermal) is applied to estimate soybean grain yield by using both CML based regression (e.g., PLSR, RF regression, SVR) and also Deep ANN [63];
- Detection and counting is to automatically detect the number of plants per unit area so that decision making can be made at various stages (e.g., seedling stage [66] or yield stage). This task is usually challenging because of appearance variations due to illumination changes, occlusions from background (e.g., foliage), and overlapping issues. Therefore, the existing approaches are mainly aimed at maximally separating the crop (or fruits) from background by using various measurement information (e.g., color, shape, temperature) and designing advanced detection algorithm for an accurate and robust detection and counting. Various algorithms have been applied for such a task, such as two-stage object detector Faster R-CNN, single-shot detector (SSD) and YOLO [64,67] to different fruits such as citrus [64], oil palm fruit [67]. A survey paper is also available on this topic [65].

Harvest decision: At the end of crop growing season, identifying the maturity is important for advancement decision (e.g., harvest date estimation, breeding variety selection). Considering the limitations of resources in harvest decision (e.g., time window, labor, money), it is a trend to deploy advanced imaging devices with data analysis algorithms for such a labor-intensive and time-consuming task. This task can be formulated into two different ML tasks according to different crops of interest including regression and object detection.

- In term of regression, a mapping between image-derived features (e.g., vegetation indices, deep features learnt by CNN) and maturity level should be established, which can be solved by various regression algorithms such as multiple linear regression, partial least square regression [68], Artificial Neural Networks (ANNs) [69]. For example, CNN-LSTM is adopted in [70] for soybean relative maturity estimation, where CNN acts as a feature extractor (against handcrafted features by a domain expert in CML) and LSTM makes prediction of the relative maturity (days) based on time-series CNN features;
- In term of object detection, the crop of interest is generally categorised into different levels of maturity and annotated by using bounding box, so that object detection algorithms can be constructed. For example, five object detectors are applied in [71] for maturity level classification of broccoli crops using UAV RGB images.

The different PA elements are summarized and compared in Table 2, which includes representative examples, UAV sensors and the corresponding AI problems. To sum up, this section introduces the PA framework including its flowchart highlighting the role of UAV sensing for data collection and AI algorithms for data analysis in support of agricultural decision making, and on this basis, presents different PA elements along with the corresponding AI problems. It should also be noted that there exist a number of common challenges in these PA elements faced by AI empowered UAV perception systems, which are summarized as below.

- In term of agriculture operational environments, there are various field changes (e.g., seasons, geolocation, crop varieties, crop and stress development stages), illumination changes (e.g., sun-

exposed, partly or completely shadowed due to time of day, weather conditions), wind and gust (e.g., adverse effects on UAV flight and imaging), communication access issue, etc.

- In term of agriculture requirements, the fields to be monitored are large-scale and the operations should be usually done in a short time window in support of timely intervention (e.g., for disease control, harvest decision).

These significant challenges have unseen requirements on UAV sensing systems (i.e., UAV systems and sensing units) and also AI algorithms, which will be addressed in the following sections separately.

3. UAV sensing systems

UAV sensing systems play a vital role in data collection module of the PA framework in Fig. 1. As depicted in Fig. 3, they are mainly comprised of UAV platforms (e.g., airframe, navigation system, power system), sensing units (e.g., internal ones for their own operation and external ones for environmental perception), and ground control systems (e.g., UAV information monitoring, flight display, task management, wireless communication system). The in-depth analysis of the platform, control system and communication system is out of the scope of this study, and the interested readers may refer to the existing survey papers [9,83] among others for further information. But different UAV sensing platforms and external sensing units for PA applications are detailed in subSections 3.1 and 3.2.

3.1. UAV sensing platforms

The mostly used UAVs for non-military applications are mini- and micro-UAV with a payload less than 30 kg [83], and there is no exception for UAVs in PA applications. There are various UAV sensing platforms for PA applications, which can be classified into three categories including rotor-wing, fixed-wing and hybrid UAVs. A survey paper on this topic is also available at [84], which highlights the platforms, sensors, and regulations in RS and scientific research.

- **Rotor-wing UAVs** rely on spinning rotor(s) to provide aerodynamic lift. These rotors control UAV direction and elevation by varying each rotor speed to alter the amount of thrust and torque produced. This gives a unique set of handling characteristics, such as vertical flight capability to facilitate take-offs and landings, and the hovering capability to fly in narrow spaces and around obstacles. According to the number of main rotors available, rotor-wing UAVs can be further divided into single-rotor UAVs and multi-rotor UAVs as below.

- **Single-rotor UAVs** have just one rotor plus a tail rotor to control direction and stability. Their mechanical design is generally complex but much more efficient than multi-rotor UAVs and therefore makes a much better trade-off between payload and endurance. Their disadvantages, on the other hand, include a high cost, vibration and instability, the requirement of maintenance due to their mechanical complexity, and a generally long and dangerous heavy spinning blade. In agriculture, they are generally used for aerial LIDAR scanning and drone spraying [85], where heavy payloads are required;

- **Multi-rotor UAVs** (e.g., tricopters with 3 rotors, quadcopters with 4 rotors, hexacopters with 6 rotors and octocopters with 8 rotors) rely on multiple rotating propellers to obtain lift, where their structure is generally simple, flexible and convenient. They offer greater control over position

Table 2
Summary of different PA elements, representative examples, UAV sensors and AI problems.

PA elements	Representatives	Sensors	AI problems
Area estimation	- soybean [33]; - several vegetation types [32,34].	- RGB [33]; - multispectral [33]; - hyperspectral [32].	classification [33]
Soil conditions	- water content [35]; - salinity [37].	- multispectral [35]; - hyperspectral [36]; - thermal [37].	regression [35]
Crop parameters	- crop height [38]; - canopy cover [42]; - biomass [39]; - leaf area index [43]; - leaf nitrogen content [41]; - survey papers [20,21].	- RGB [39]; - multispectral [42]; - hyperspectral [43]; - LiDAR [38].	- regression [38]; - classification [42];
Crop diseases	- yellow rust [30]; - citrus canker [72]; - bacterial wilt [73]; - foliage disease [45]; - late blight [44]; - survey paper [74].	- RGB [45]; - multispectral [30]; - hyperspectral [75].	- regression [45]; - classification [30].
Crop pests	- fruit pests [49]; - sheath blight [51]; - soybean pests [49].	- RGB [49]; - multispectral [51]; - hyperspectral [48].	- regression [51]; - classification [49]; - object detection [47].
Crop water stress	- barley [76]; - maize [57]; - vineyard [55]; - wheat [77]; - survey paper [54].	- RGB [78]; - multispectral [57]; - thermal [56].	- index [53]; - regression [57].
Weeds	- blackgrass in wheat [59,60]; - weed in sugar beet [61]; - weed in soybean [62]; - survey paper [58].	- RGB [59]; - multispectral [60]; - hyperspectral [79].	- regression [59]; - classification [60]; - object detection [62].
Yield prediction	- soybean [63]; - cotton [80]; - citrus [64]; - oil palm fruit [67]; - survey on fruit [65].	- RGB [63]; - multispectral [63]; - thermal [63]; - stereo camera [81].	- regression [63]; - object detection [64].
Harvest decision	- soybean [68,70]; - peanut [69]; - broccoli [71]; - canola [82].	- RGB [70]; - multispectral [68]; - hyperspectral [82].	- regression [68]; - object detection [71].

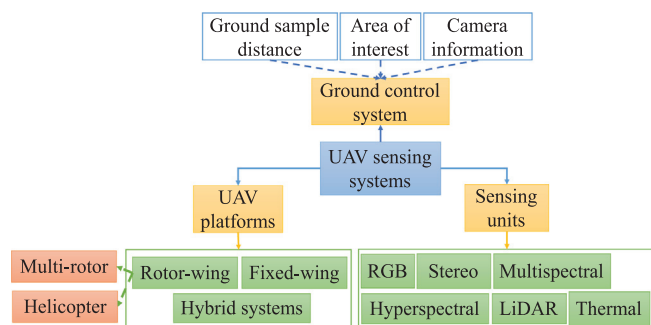


Fig. 3. UAV sensing systems including various UAV platforms, ground control system and different external sensing units.

during the flight, such as hovering, move up and down on the same vertical line, back to front, side to side and rotate in its own axis. The high manoeuvrability makes them ideal for aerial photography and close inspection. Their disadvantages, on the other hand, include a limited endurance (e.g., 30 min) [17] and speed (unsuitable for large-scale agricultural sensing), energy-inefficiency due to aerodynamics (i.e., severe payload-endurance trade-off). Rotor-wing UAVs have been widely used in low-altitude remote sensing including visual inspections and 3-D modelling, where high-resolution images (e.g., centimetre-level) are required [86].

- **Fixed-wing UAVs** look and work like an aeroplane with rigid wings providing the lift, rather than using vertical lift rotors in rotor-wing UAVs. As a result, they are more energy-efficient since energy is mostly needed to move forward instead of holding itself in the air (as rotor-wing UAVs). As a result, they are of high speed while with a much longer flight time and so are able to cover much larger areas efficiently [87]. For example, typical battery powered fixed-wings can fly about 1 h while gas engine powered (fuel) can stay aloft for several hours. The disadvantages of fixed-wing UAVs include a relatively high price, operation difficulty (e.g., the requirement of open and flat ground for take-off and landing, and no hovering), relatively high flying attitude (e.g., 40 m + and therefore with a reduced image resolution). Fixed-wing UAVs have been widely used in aerial mapping, drone surveying, agriculture, inspection, where medium-resolution images (ten(s) centimetre-level) are required.
- **Hybrid UAVs** take the advantages of fixed-wing and rotor-based designs concurrently, where rotors are attached to the fixed wings allowing it to hover and take off and land vertically, and meanwhile are able to fly as a fixed-wing UAV by switching the flight mode. As a result, they have a long flight duration while with a low requirement on landing site, and so greatly expand the application scope of both rotor-wing and fixed-wing UAVs. However, this technology is still in the nascent stage, where further development is urgently required to guarantee their stability, safety and reliability such as aerodynamic interference of tilting propeller, tilting structure design, and

power system matching in various flight states. A brief review paper on VTOL UAVs and their application prospects in PA is available in [88].

A comparison of representative UAV sensing platforms against the ground- (e.g., tractor or robots), manned aircraft- and satellite-based sensing platforms is made in Table 3. In particular, the comparisons are made in terms of payload, flight endurance, revisit-time (or temporal resolution), flight efficiency, coverage, spatial resolution, environmental constraint, beyond surface assessment or not, soil compaction, cost, operational complexity, and labor-intensity, which should be taken into account in choosing the most suitable one for the target applications of interest. It follows from Table 3 that UAVs make an excellent trade-off between ground-based and satellite-based systems, and therefore are suitable for field-scale applications with high spatial-temporal image resolutions while with a reasonable cost, operational complexity and labor-intensity.

3.2. UAV sensing units for PA

This section aims to introduce different UAV external sensing units for PA applications. It is noted that the “external sensors” here are relative to “internal sensors”, which refer to sensors observing agriculture fields rather than for its own operations (e.g., GPS for position, inertial measurement unit (IMU) for attitude, barometer/altimeter for altitude). One of the most commonly exploited sensing technology in agriculture sector is optical sensing, where different optical sensing units with different spectral wavelengths (mainly in visible, NIR and thermal infrared ranges) investigated in this study are shown in Fig. 4. The metrics/factors that should be considered in choosing the most suitable sensor for applications of interest mainly include: weight, price, spectral range, spectral resolution, operational and data processing difficulty. In particular, the optical sensing units mainly include true color camera (or RGB camera), stereo camera system, color infrared camera, multispectral camera, hyperspectral camera, LiDAR and thermal camera, where their details in PA applications are presented as below.

- **True color camera:** also termed visible light sensor or RGB camera, refers to standard consumer camera with Red, Green and Blue bands. It is able to capture an image in such a way that it reproduces the same effect as seen with a human eye. Mean-

while, it is relatively affordable, of light weight, with a high spatial resolution, simple data processing and video output. Therefore, it has been widely applied in PA applications, where image resolution and visible symptoms are the main priorities, such as crop parameter estimation (e.g., canopy cover, LAI, height, biomass [39]), crop damage assessment (e.g., lodging [90]), crop maturity [70,91], fruit segmentation, detection and counting [64,92], pest detection [47,93], etc. The main drawbacks of true color camera include the broad and only visible spectral bands, which limits its performance in early crop stress detection (i.e., when visible color symptoms have not appeared in early stages).

- **Stereo camera system:** is comprised of two color cameras with a fixed layout, which can simulate human binocular vision to capture three-dimension images. Particularly, colored point cloud (e.g., color with depth) can be generated via the principle of triangulation, and in comparison to single camera, the occlusion problem can also be alleviated. Therefore, stereo camera system is usually applied for the scenarios where both true color information and depth information are important. For example, a stereo camera comprised of two SONYA6000 cameras has been applied in [94] for canopy height and above-ground biomass estimation. A commercial stereo camera has also been used in [95] for fast detection and location of longan fruits. Other applications of stereo camera system include navigation and localization, which is out of the scope of this study and interested readers may refer to the survey paper [96] for further information. The main limitations of stereo camera system are its limited sensing range (generally within 20 meters) and high sensitivity to illumination changes.
- **Color infrared camera:** refers to a standard camera with near-infrared (NIR) blocking filter removed. Representatives include Green-Red-NIR three-band camera ADC Micro, Blue-Green-NIR three-band camera XNiteCanon SX230 NDVI [97]. It can be seen as a transition from true color camera to multispectral camera, and so its price and usability in PA also fall into true color and multispectral cameras. Therefore, it is usually applied to the scenarios where non-visible bands such as RedEdge or NIR are important but with a limited budget on multispectral camera. For example, different color infrared cameras have been used in VI calculation [98], crop parameter estimation [97] and nitrogen nutrition estimation [99]. Its main limitations lie in the calibration required for spectral information, and worse performance than commercial multispectral camera.

Table 3

Representative UAV sensing platforms against ground-, manned aircraft- and satellite-based sensing systems: pros and cons.

Platform	UAVs			Ground	Aircraft	Satellite
Sub-type	Rotor-wing	Fixed-wing	VTOL			
Representative	• DJI Phantom 4 Pro; • DJI Matrice 300.	• Sensefly eBee; • Sentera PHX.	WingtraOne	CLAAS crop sensor	ER-2	• Sentinel-2; • Landsat 8.
Payload	Low (0.5 kg)	Low (0.9 kg)	Low (0.8Kg)	High	High	High
Endurance	Low (30 min)	Low (45 min)	Low (59 min)	Long (hours)	Long (hours)	Long (years)
Revisit-time	User-defined	User-defined	User-defined	User-defined	User-defined	Long (5 days)
Efficiency	R low	Medium	Medium	Low	High	High
Coverage	Plot-level	Field	Field	Plot-level	Large Field	Region
Resolution	High (cm)	High (5 cm+)	High (5 cm+)	High (mm)	Low (10 cm+)	Low (10 m+)
Constraint	Rain/Wind	Rain/Wind	Rain/Wind	Terrain	Strong wind	Cloud
Beyond surface	No	No	No	Yes	No	No
Soil compaction	NA	NA	NA	Yes	NA	NA
Cost	Low	R Low	R Low	R High	High	V High
Operational complexity	Simple	R complex	R simple	R simple	Complex	Complex
Labor-intensity	R low	R low	R low	High	High	Low

(1) “Beyond surface” means measuring the information under crop/soil surface, as RS can only sense the surface reflectance, but proximal sensing (e.g., tractor, robotics) is able to detect soil/crop attributes under their surface via specific sensors [89]; (2) The values in “()” denote the example values for the first representative platforms; (3) “R” means “Relative”, and “V” means “Very”.

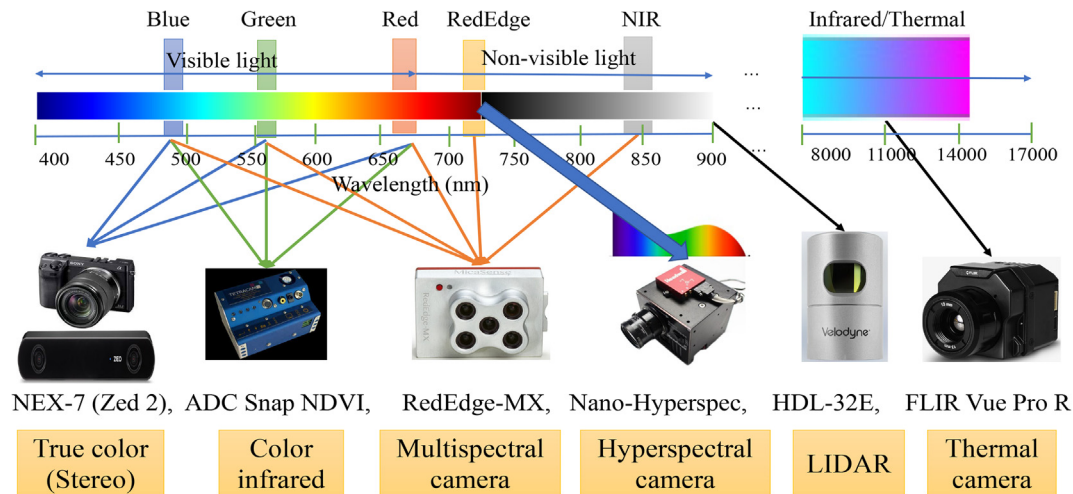


Fig. 4. UAV sensing units for PA applications with different spectral wavelengths/bands investigated in this study. Note: the top spectral image is from MicaSense and the example sensors' images are directly from their official websites.

- Multispectral camera:** refers to a camera with more than three RGB bands, where additional band(s) may include RedEdge or NIR(s) or their combinations. Different multispectral cameras are available in the market, e.g., four-band Sequoia, five-band RedEdge [30] and DJI P4 multispectral, six-band Sentera Quad [100], Buzzard camera, MiniMCA6, where a comparison of Sequoia and Mini-MCA6 is available in [101] in terms of accuracy and applicability. UAV based multispectral imaging has been used for various PA applications, where extra RedEdge or NIR introduces additional spectral information, such as VI calculation, crop parameters estimation, weed mapping [60], weed species identification [100], crop maturity and damage, etc. The main limitations of multispectral camera is its high price over true color or color infrared cameras, medium spatial resolution compared against true color camera, and low spectral resolution compared against hyperspectral camera.
- Hyperspectral camera:** combining the advantages of spectroscopy (i.e., point-based spectral measurement) and digital imaging, is able to obtain a large number (e.g., hundreds) of narrow and continuous wavelengths (images) from visible to NIR ranges. Several commercial hyperspectral cameras are available on the market, such as Nano-Hyperspec, Rikola, Pika NIR-640, Hyperspec@SWIR (please refer to [28] for more hyperspectral sensors). These wavelengths are generally much narrower than the bands of color image, color-infrared or multispectral image and so the increased spectral resolution would be more sensitive to crop characteristics changes. Therefore, hyperspectral cameras have been widely used in accurate VI calculation, crop parameter estimation and early crop stress assessment, among others in Table 4. The main limitations of hyperspectral camera are its high price, relative complex data processing pipeline due to calibration and large data volume. A review paper on UAV based hyperspectral imaging for agriculture and forestry is available in [28], which highlights hyperspectral sensors, inherent data processing and application both on agriculture and forestry.
- Thermal camera:** is a device that creates an image using infrared radiation for temperature signature, where temperature plays a critical role in plant physiological processes such as transpiration, leaf water potential and photosynthesis [102]. At the foliar level, stomata closure reduces transpiration and evaporative cooling, leading to an increased leaf temperature. Therefore, in practice, temperature difference between crop canopy (T_c) and air (T_a) (its normalized version is also termed

Crop Water Stress Index) is usually used to assess crop water status and consequently act as a tool for irrigation management. Typical applications of UAV thermal imaging include crop stress (e.g., water, disease) monitoring, soil characteristics (e.g., moisture, texture), crop maturity and yield estimation. Recent survey papers are also available in [3,27], which highlight the correct use, applications, issues and concerns of UAV thermal imaging for PA applications. As highlighted in [103], two key considerations for (UAV) thermal imaging include the optimal spatial resolution (as typical thermal image resolutions are low ranging from 320×240 to 640×512 pixels, leading to the issue of mixed pixels, i.e., pixel containing signals from multiple sources) for the identification of crop canopy from background and the appropriate time of day for image acquisition so that the object of interest can easily separated from the background and environment.

- LiDAR** stands for Light Detection And Range, which sends out laser light pulses and measures the time it takes the light to hit an object and bounce back to the sensor. As a result, it is able to measure the distance to the surface and build a 3-D point cloud for various applications. UAV LiDAR measurements are gaining popularity due to their increased spatial resolution, better penetration (than image-based approaches), higher throughput and independence of air temperature, illumination condition and a decreasing weight and price [104,105]. As a result, they are finding various applications in PA such as plant height estimation [38], crop lodging mapping, etc., where 3-D point cloud information is important. In comparison against image-based point cloud generation, LiDAR based approach generally has better density but at the price of a higher equipment cost [106]. It is noted that a trade-off is usually made between laser point density (resolution) and LiDAR price [105], in addition, the large data processing and water-absorbing problem should also be considered in practical applications. A comparison of these UAV sensing units (externally observing agriculture fields) for PA applications are summarized in Table 4 in terms of representative examples, typical applications, advantages and practical considerations. It follows from Table 4 that the selection of different sensors for an application of interest would depend on a number of factors, which are summarized as below.
- Scientific factors to be considered in sensor selection include spectral range/resolution and spatial resolution.

Table 4

Comparisons of UAV external sensing units for PA applications: representatives, applications, advantages and considerations.

Sensor	Representative	Applications	Advantages	Considerations
True colour	• Sony NEX-7 [78], • Nikon D850.	• Crop parameters (e.g., canopy cover, LAI, height, biomass [39]); • crop damage (lodging) [90]; • Crop maturity [70]; • Pre-harvest yield [70]; • 3-D point cloud [107].	• High availability; • Easy operation and integration in various platforms; • High spatial resolution; • Easy visual interpretation; • Video output.	• Only visible and broad spectrum; • Lack of proper calibration; • Very sensitive to sunlight reflection.
Stereo camera	• Zed 2i, • Two SONYA6000 [94]	• Crop parameters (e.g., height, biomass [94]); • Count and yield [95].	• Obtain color, texture, geometric information in 3D; • Distance information; • Less occlusion problem than single RGB camera.	• Only visible and broad spectrum; • Limited operation distance (e.g., 15 m); • Very sensitive to sunlight.
Colour infrared	• ADC-SNAP [108] • XNiteCanon SX230 [97]	• VI calculation [98]; • Crop parameters (e.g., canopy cover, crop height) [97]; • Nitrogen [99].	• Cheap infrared camera.	• Calibration required for spectral information; • Worse performance than multispectral camera.
Multispectral	• Parrot Sequoia • Micasense RedEdge, • DJI P4 Multispectral	• VI calculation [30]; • Crop parameters (e.g., crop height, biomass [41], canopy cover [42]); • Crop stress (e.g., weed, disease, pest, nitrogen [41]); • Crop maturity [68]; • Crop damage (lodging) [90]; • pre-harvest yield [63].	• Wider spectrum range and narrower bandwidth; • Well-established preprocessing steps and software; • Sub-decimeter multispectral mapping.	• A high price (e.g., €10K); • Radiometric calibration is required for spectral reflectance; • Lower spatial resolution than RGB camera.
Hyperspectral	• Nano-Hyperspec • Rikola, • Pika NIR-640, • Hyperspec@SWIR	• VI via narrow wavelengths [40]; • Crop parameters (e.g., LAI [43], biomass [40]); • Early crop stress monitoring (e.g., weed, disease, nutrient [43], chlorophyll content, crop water content); • Soil conditions [36].	• Many (e.g., hundreds of) narrow, continuous and wavelengths in wide-range spectrum; • Better stress monitoring performance;	• A much higher price (e.g., €50 K) and weight; • Relative complex data preprocessing (e.g., calibration, stitching); • Large volume of data to process.
Thermal	• FLIR Vue Pro R, • WIRIS Pro • DJI Zenmuse XT2	• Plant stress (e.g., water, disease) [3]; • Soil parameters (moisture, texture) [3]; • Crop maturity and yield [80];	• Extra temperature information; • Work in both day and night.	• Sensitive to temperature related environmental variations (e.g., time of day, cloud); • Complex data collection (e.g., calibration, ground control point), processing (e.g., low resolution and contrast, strong optical distortion).
LIDAR	• Velodyne HDL-32E, • Livox Mid-40 • LeddarTech VU8	• Crop parameters (height [105], biomass [38]); • Crop stress (e.g. nitrogen [41]); • 3-D point cloud [109].	• High-performance geometric measurements; • Robust against environmental changes (e.g., illumination, cloud).	• A high price; • Resolution and price trade-off; • Large volume of data to process.

- For some applications (e.g., fruit detection, crop lodging, direct pest detection, 3-D point cloud) where color, texture and image resolution are the priority factors, true color or stereo camera would be the choice due to their high spatial resolution and true color information;

- For some applications (e.g., early crop stress detection, crop maturity estimation, accurate vegetation indices) where spectral range and spectral resolution are important, multispectral (i.e., relatively wider spectral bandwidths with a lower price) or hyperspectral camera (i.e., narrow spectral bands/wavelengths with a higher price) would be the candidates, particularly if RedEdge and (or) NIR are the important bands, multispectral camera (with a higher price) and colour infrared (with a lower price) would be the choice;

- If temperature information is important (e.g., soil/crop drought assessment), camera with thermal information would be the choice such as thermal camera or multispectral camera with thermal bands (e.g., MicaSense Altum-PT, DJI Zenmuse H20);

- In applications where high-quality distance information (e.g., plant height estimation, 3-d point cloud generation) is important, the candidates would be stereo camera (with a low price and poorer performance) or LiDAR (with a high price but better performance);

- In terms of spatial resolution, thermal camera has the poorest spatial resolution followed by multispectral, colour

infrared and hyperspectral cameras, while true colour, stereo and LiDAR have very high resolutions.

• Operational factors to be considered in sensor selection mainly include weight, price, operation environments, operation difficulty, data processing expertise, etc.

- True color, stereo camera, color infrared, multispectral and thermal cameras usually have a low weight and therefore are suitable for small UAVs, however, hyperspectral camera and LiDAR with a relatively high weight should be put on UAVs with a large payload (e.g., DJI M300, S1000);

- The price of true colour, stereo and color infrared cameras are relatively low, however, multispectral, hyperspectral, thermal cameras and LiDAR are relatively expensive;

- Stereo camera with a limited operation range (e.g., 15 m) is usually applied for low-altitude applications, however, other sensors do not have a particular requirement on operational height and mainly depend on the spatial resolution required for the specific applications;

- Thermal camera and LiDAR are able to work at night, however, other cameras should be operated at day time with sufficient illumination;

- Hyperspectral camera and LiDAR generally involve relatively complex data processing pipeline with a large amount of data, while other camera images can be pre-processed

easily with off-the-shelf software (e.g., Pix4D, DJI TERRA, Agisoft Metashape, etc.).

4. AI algorithms for UAV based PA

Upon obtaining a large amount of data (e.g., images) from UAVs with imaging sensors, it is important to interpret these data into semantically meaningful information for follow-up applications, which is time-consuming and labor-intensive by conventional manual annotation or expert interpretation. There is a trend to adopt AI algorithms to automate this process. So this section aims to categorize the literature according to AI algorithms/models involved in the direct use of UAVs for PA applications. The “AI” algorithms in this context mainly refer to ML methods that are used for regression, classification and object detection although unsupervised learning algorithms, relatively sparse compared against supervised learning algorithms, are also briefly covered. According to the learning methods, ML algorithms can be broadly categorised into three types including supervised, unsupervised and reinforcement learning, which are further detailed as below.

- In supervised learning, a prediction model is built upon some labelled data (e.g., each sample in the labelled dataset is with a label or target) to mapping features to labels so that it is able to predict labels of new and unseen data automatically. This approach dominates AI algorithms for UAV based PA applications, which include regression, classification and object detection problems;
- Unsupervised learning is to extract meaningful representations and explain key features of the dataset. No labels are necessary in order to learn from the data, although labels are sometimes useful for algorithm evaluation. The objective is to divide the unlabelled dataset into separate groups or clusters, which share similar proprieties within clusters but differ from other clusters. The applications of this approach in UAV based PA mainly include feature dimension reduction, simple classification and data labelling;
- While in reinforcement learning, intelligent agents learn to take actions in a real or simulated environment in order to maximize the notion of cumulative reward. By trial-and-error, the agent explores environment and exploits feedback to improve its performance. In reinforcement learning, contrary to supervised learning paradigm, no feature-label pairs are required. The application of reinforcement learning in UAV based PA mainly focuses on planning layer [110] and so not covered in this review since this review is mainly focused on perception layer.

4.1. Supervised learning

Supervised learning aims to establish a mapping (i.e., functional relationship) from feature (or independent, predictor, explanatory) variables to response (or dependent, target) variable. Assuming that Y is response variable of interest, and $X = (x_1, \dots, x_k)$ is feature variable with k being the sample number, the model to describe the relationship between Y and X in supervised learning can be written as Eq. (1).

$$Y = f(X, \beta) + \epsilon, \quad (1)$$

where β is a parameter vector, ϵ is an error term whose distribution may or may not be normal, $f(\cdot)$ is some function that describes the relationship between Y and X . Depending on data-type of Y , supervised learning algorithms can be broadly categorized into regression problem (i.e., with continuous data-type) and classification problem (i.e., with categorical data-type).

In addition, depending on characteristics of the function f in Eq. (1) selected for the mapping, supervised learning algorithms can

also be broadly divided into two types including parametric and non-parametric approaches, which are detailed as below.

- *Parametric model* simplifies the learning task by presuming that the form of mapping function is known as a prior. The learnt model, therefore, has a set of parameters β with a fixed size (independent of the number of training samples), and the main task is to learn these parameters from training data. This approach is simple, more interpretable, requires less data and does well if the functional assumption is correct. However, the learnt model is constrained to the selected functional form and may perform badly, e.g., a large bias, if the functional assumptions are wrong;
- *Non-parametric model*, on the other hand, is able to relax the assumption of functional form and enables one to explore data more flexibly. Instead, model form is learnt from dataset automatically, and the number of parameters are also flexible and not fixed as a prior. This approach is suitable for the case with a lot of data but little or no prior knowledge, and may result in better performance than parametric models. It is also noted that this approach generally requires more data, learning time and more subject to the problem of overfitting if handled inappropriately.

4.1.1. Regression

Regression approaches are widely used in PA applications when response variables are continuous. In regression, the response variable Y in Eq. (1) is continuous, X denotes feature variables, β are unknown parameters to be estimated by using the feature-response pairs (X_i, Y_i) with $i = 1, \dots, k$. The typical evaluation metrics for regression algorithms, as given in Eq. (2), include R-Squared (or Coefficient of Determination) R^2 showing how well the model fits the dependent variable, Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) illustrating model fitness, Mean Absolute Error (MAE) measuring model prediction error, or their combinations such as Huber loss function [70].

$$R^2 = 1 - \frac{SS_{\text{residual}}}{SS_{\text{total}}}, SS_{\text{residual}} = \sum_{i=1}^k (y_i - \hat{y}_i)^2, \\ SS_{\text{total}} = \sum_{i=1}^k (y_i - \bar{y})^2, \text{ with } \bar{y} = \frac{1}{k} \sum_{i=1}^k y_i, \\ MAE = \frac{1}{k} \sum_{i=1}^k |y_i - \hat{y}_i|, MSE = \frac{1}{k} \sum_{i=1}^k (y_i - \hat{y}_i)^2, RMSE = \sqrt{\frac{\sum_{i=1}^k (y_i - \hat{y}_i)^2}{k}} \quad (2)$$

However, the above metrics are sensible to the inclusion of additional variables even if they make little contribution in explaining the outcome. To this end, more robust metrics for model evaluation and selection are proposed such as Adjusted R-squared $\bar{R}^2$ (adjusting the R^2 for having too many variables in the model), Akaike Information Criterion (AIC) and Bayesian Information Criteria (BIC) adding a penalty that increases the error when including additional variables, defined as below.

$$\bar{R}^2 = 1 - \left(1 - R^2\right) \frac{k-1}{k-v}, \\ AIC = 2p - 2\ln(\hat{L}), BIC = p\ln(k) - 2\ln(\hat{L}) \quad (3)$$

where v denotes the total number of explanatory variables, p is the total number of model parameters used, $\hat{L}$ is the log-likelihood estimate. Regression algorithms can also be broadly divided into two categories including parametric and non-parametric approaches, detailed as below.

- In parametric approaches, the unknown parameters β are a vector in finite p-dimensional space and the $f(\cdot)$ is with known functional forms [111]. Typical parametric regression methods include linear regression (LR) (linear in the parameters β) such as multiple LR, polynomial regression, Poisson regression, non-linear regression with given functional forms, ridge regression, Least Absolute Shrinkage and Selection Operator (LASSO), Partial Least Squares Regression (PLSR), Principal component regression (PCR), etc. In particular, ridge regression, LASSO, PLSR and PCR are applicable to the cases with high levels of multicollinearity or the automated variable selection is desirable, which is a common problem in hyperspectral imaging due to the large number of wavelengths being highly correlated in adjacent spectral ranges [28,40].
- While in non-parametric approaches, parameter set β are a subset of infinite dimensional vector spaces, where their dimension and estimation should be determined by the labelled data on-the-fly. Typical non-parametric regression models include Support Vector Machine Regression (SVR), K-Nearest Neighbor (KNN) regression, decision tree regression, RF Regression (RFR), Gaussian Process Regression (GPR), Artificial Neural Network (ANN), Deep Neural Network (e.g., Long Short-Term Memory (LSTM) [70]). These nonparametric approaches generally require a relatively large number of training samples to build a reliable and generalizable model.

There are various packages available to implement these different regression algorithms such as scikit-learn in Python language, Caret Package in RStudio software, regression learner in MATLAB. As shown in Table 2, typical applications of regression approaches in PA application include soil moisture estimation [35], various plant parameters estimation (such as LAI [43], crop height [38], biomass [39]), crop disease severity [45], crop pest severity [51], leaf water content [57], weed density [59], yield estimation [63] and maturity estimation [68]. The features in regression models mainly include spectral bands, spectral wavelengths, various vegetation indices or their combinations; however, some studies also include other variables (or information) such as height, temperature, texture or deep features learnt by CNN or ANN approaches for better regression performance [70,63]. As an illustrating example, please refer to Fig. 5 for the workflow diagram of soybean yield prediction from [63] based on UAV multimodal data fusion and various regression algorithms such as PLSR, SVR, RFR and Deep ANN models. However, it should be noted that the significantly increased feature dimensionality also requires a larger number of training samples in order to better learn the model parameters and therefore the problem of feature dimension reduction should also be carefully considered [112,113].

4.1.2. Image classification

The other main approach in UAV based PA applications is image classification. Different from regression problems, the response variable in image classification problems (i.e., Y in Eq. 1) is categorical in a finite set. The main classification methods can be divided into two categories including conventional ML approaches (i.e., via human-engineered features) and recent DL approaches (i.e., via learnt deep features), where their comparisons in term of workflow are displayed in Fig. 6 and presented as below. A large number of evaluation metrics are available for image classification algorithms, where the commonly used ones include Precision, Recall, Accuracy, F score, ROC curve and Area Under the ROC Curve (ROC AUC score), defined as below.

$$\text{Precision} = \frac{TP}{TP+FP}, \text{Recall} = \frac{TP}{TP+FN}, \text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN},$$

$$F_\beta = (1 + \beta^2) \times \frac{\text{Precision} \times \text{Recall}}{\beta^2 \times \text{Precision} + \text{Recall}}, \text{ROCcurve, ROCAUCscore} \quad (4)$$

where true positive (TP) denotes the correctly predicted positive values, true positive (TN) denotes the correctly predicted negative values, false positive (FP) represents the falsely predicted positive values, false negative (FN) is falsely predicted negative values. Confusion matrix is also used to present Precision, Recall and Accuracy in matrix format. When choosing β in F_{β} score the more you care about recall over precision the higher β you should choose. ROC curve is a chart visualizing the tradeoff between TP rate and FP under every threshold. ROC AUC score is a value evaluating how good the ROC curve is, where the more top-left ROC curve is the higher the area and hence higher ROC AUC score. In addition to the aforementioned model performance metrics, model size and inference time should also be considered for embedded system (or real-time) applications.

- It follows from Fig. 6 that CML and DL based approaches share the same steps of data acquisition and data preparation as below.

- Regarding data acquisition with UAV for PA applications, flight plan is usually generated before missions with dedicated planning software (e.g., open-source software MissionPlanner, and commercial ones including DJI GS Pro, PIX4D Capture, DroneDeploy, UGCS, Altizure, Waypoint Master) by simultaneously considering the area of interest (i.e., boundary shape, terrain), the required ground sample distance and the characteristics (e.g., image resolution, field of view) of UAV cameras [86].

- Upon data collection, the next step would be data preparation for ML modelling. Different data preparation steps can be done depending on different applications of interest, such as image pre-processing, data labelling, data augmentation or synthesis among others, which are briefly discussed as below.

- * Typical data pre-processing steps may include image geometric correction, spectral radiation processing, orthomosaic generation, etc. [20]. Many off-the-shelf commercial software are available to automate these processes, such as PIX4Dmapper, Agisoft Metashape, DJI Terra, DroneDeploy, and open-access software OpenDroneMap, etc.;

- * Different data labelling methods should be adopted according to the problems of interest, such as image level labelling for image-wise classification, pixel-wise labelling for pixel-wise semantic segmentation, bounding box annotation for object detection and localization, etc. Off-the-shelf software are also available for this task, such as labelme, labellmg, CVAT, hasty.ai, etc;

- * Data augmentation and data synthesis are to boost ML model performance by expanding training datasets in terms of scale and variations [114]. A survey paper on image data augmentation for DL is available at [115], and a systematic review of Generative Adversarial Networks (GANs) for image augmentation in agriculture is also available at [116].

- In particular, conventional data augmentation mainly refers to various image operations on the input space such as geometric transformations (e.g., flipping, cropping, rotation, translation), color space transformation, kernel filters, mixing images and random erasing;

- While data synthesis mainly refers to augmenting training datasets by generating samples via DL approaches such as feature space augmentation, adversarial training, GANs, neural style transfer and meta-learning.

- The extra module of CML over DL is feature engineering, which is to extract hand-engineered features (e.g., color, shape, texture) from raw images for the purpose of ML model construc-

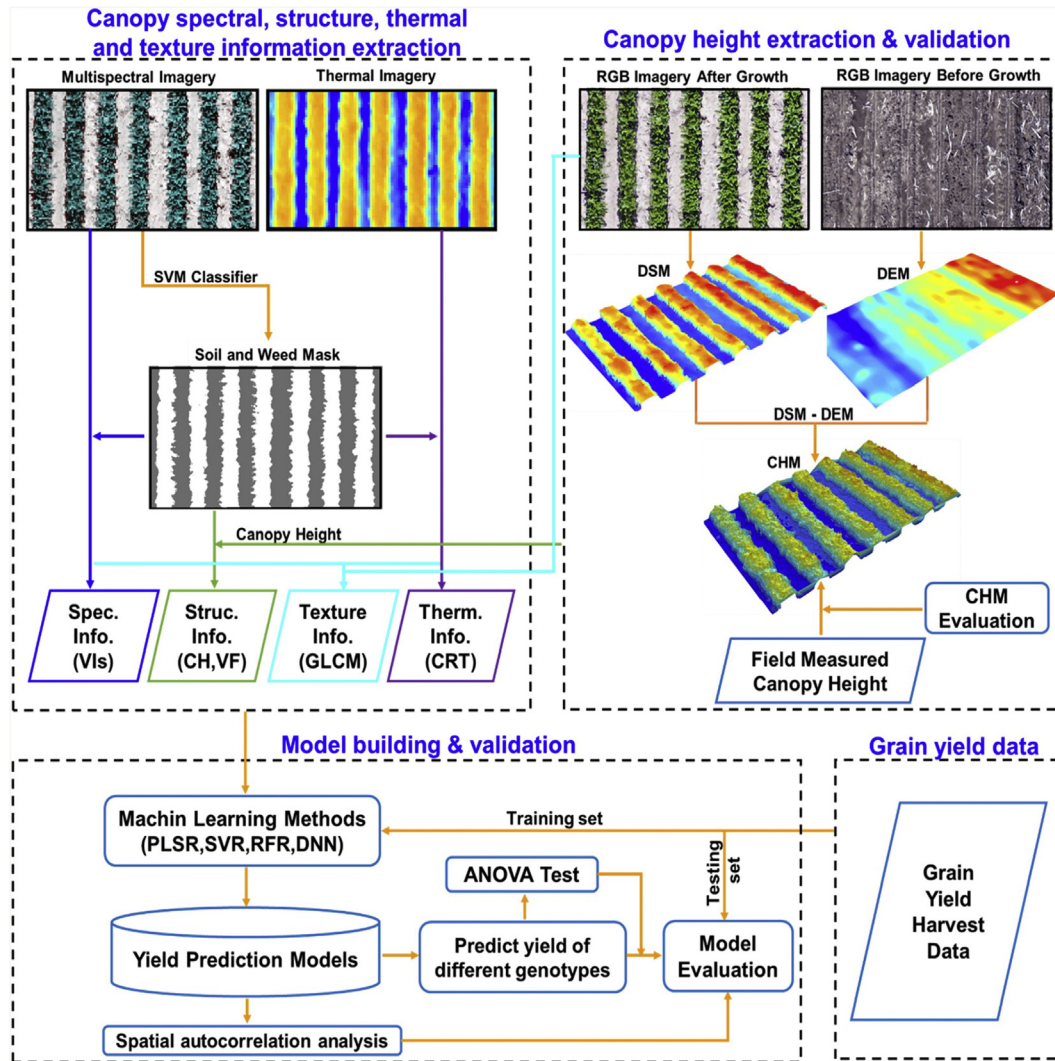


Fig. 5. Workflow diagram of data processing, feature extraction and regression modelling (via PLSR, SVR, RFR and Deep ANN) from [63] for soybean yield prediction based on UAV multimodal data fusion.

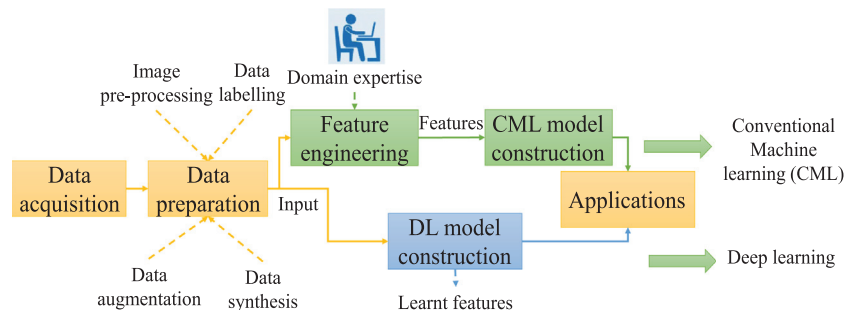


Fig. 6. A comparison between CML and DL approaches in image based classification problems.

tion. This is generally a cumbersome, trial-and-error process to choose the “right” features without guaranteeing optimal model performance. However, it should also be noted that feature engineering in the area of Computer Vision (CV) is still important and may be very useful for efficient implementation of DL approaches [117]. While, DL approaches combine feature learning (i.e., the most descriptive and salient features for the task of interest) and model training in one DL model construction module, which is also termed end-to-end learning. Unlike

the handcrafted features (e.g., HOG, SIFT [118], SURF [119]) in CML, the learnt deep features, however, are generally constrained to the training datasets. Therefore, “representative” training datasets are vital in DL approaches, and the deep features learnt in one application (dataset) are very likely to be not suitable to another application (dataset).

- The shared CML/DL model construction module can be further detailed in Fig. 7. In particular, the labelled dataset can be divided into three subsets including training set, validation set

and test set. Training set is to learn model parameters by defining some suitable loss functions, and validation set is to support model evaluation to avoid the problem of overfitting, and testing data, mimicking real-life applications, is to confirm the usability of the learnt model. The most commonly used train/validation/test partition is randomly splitting the labelled data by fixed proportions. However, a more reasonable approach for PA applications would be temporal or spatial split [11] (instead of random split) of the labelled images into different time periods or regions so that model generality (i.e., time-wise, location-wise) can be practically tested. In model training, the error between predicted outputs and ground truth (GT) labels is calculated in the loss function, so that model parameters can be learnt/updated recursively. For example, in DL parameter learning is achieved via back-propagation using chain rule and gradient descent.

The approaches for image classification can also be divided into two categories including image-wise classification and pixel-wise classification, where a diagram is given in Fig. 8 to display typical methods for image-wise classification and pixel-wise classification. In particular, CML approaches are highlighted in yellow, CNN based approaches are highlighted in green and CNN derived deep semantic segmentation approaches are highlighted in red.

Image-wise classification is to assign a label to the whole image [120], where the main approaches include CML and the SOTA deep CNN approaches, which are presented as below.

- CML: before the popularity of DL (CNN in particular), different image-wise classification approaches have been developed in the community of computer vision and pattern recognition. As displayed in Fig. 8, the typical parametric CML methods include discriminant analysis (DA), Naive Bayes (NB), Logistic Regression (LR), Perception, Simple Neural Networks (SNN); while the nonparametric CML methods include K-Nearest Neighbor (KNN), Decision Tree (DT), Support Vector Machine (SVM), Random Forest (RF), AdaBoost among others. There are also many packages to implement these CML models such as scikit-learn in Python, Classification Learner in Matlab. As shown in Fig. 6, the main challenges of CML based image-wise classification is to define the representative features for model learning. For example, [121] applied a textron-based linear classifier to image paths from high-resolution UAV images for invasive weed classification, where a bank of image filters were used for feature extraction. [49] compares SVM, RF, NB, KNN, AdaBoost, and local descriptors (e.g., SIFT and SURF) in soybean pest classification via UAV derived RGB image segments (by SLIC superpixel segmentation). This approach is useful when the image classification task is relatively simple and a limited number of labelled samples are available. However, the expressiveness (complexity) of these shallow models is usually limited and therefore the performance is easily saturated even with a large dataset for model training.

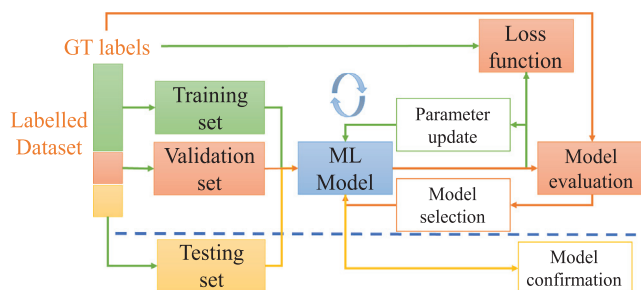


Fig. 7. Flowchart of ML model construction: training, validation and test.

- However, CNN has been the dominant method for image-level classification since 2012 with the success of AlexNet [122]. CNN is a special type of deep neural network (DNN), which excels at image-based tasks due to their advantages in learning spatial structure of the pixels automatically. It also requires a fewer number of learning parameters over full connected DNN models, achieving superior computational performance and reducing the risk of overfitting due to the use of convolution filters. The general structure of a CNN is composed of multiple convolutional layers interlaced with pooling layers, followed by some fully-connected layers in the end. In particular, the convolutional layer automatically extracts the hierarchical and high-level feature representation, while pooling layer reduces the image resolution to increase receptive field and help achieve spatial invariance, and the fully-connected layer reduces the image dimension and outputs a 1-D distribution over classes.

The main development of CNN is briefed for the sake of completeness. Back in 1989, [123] constructed a CNN for handwritten zip code recognition and first used the term “convolution”, which is the original version of LeNet. Then various shallow neural networks were successfully proposed. In particular, in 1998, a pioneering 7-level convolutional network named LeNet-5 [124] was proposed to classify hand-written digits in small grey-scale images. However, with the increasing number of layers of neural networks, traditional backpropagation networks encountered several problems, such as local optimum, overfitting, gradient vanishing, and gradient exploding. In 2012, AlexNet [122] made a major breakthrough in ImageNet competition by using “layer-wise pre-training”, which attracted enormous research interests. Since then, several CNN architectures (e.g., deeper, wider architectures) have been proposed such as Visual Geometry Group (VGG) [125] (via small convolution filters) and Inception (or GoogLeNet) [126] in 2014 (deeper network structures than AlexNet by suitably factorized convolutions and aggressive regularization), ResNet [127] in 2015 (i.e., addressing the degradation problems so that the training of deeper network becomes possible via learning residual functions with reference to the layer inputs instead of learning unreferenced functions), DenseNet [128] in 2016 (thinner and compact networks), ResNeXt [129] in 2017 (i.e., exposing a new dimension termed cardinality in addition to the dimensions of depth and width), EfficientNet [130] in 2019 (i.e., uniformly scaling all dimensions of depth/width/resolution using a compound coefficient), Vision Transformer (ViT) in 2021 [131,132] (i.e., employing a Transformer-like architecture [133] over patches of the image), achieving better performance even outperforming human being in the well-known large-scale ImageNet dataset. In particular, transformer based network architectures have recently surpassed CNN network architectures with its massive network architecture and model parameters to capture the global information (i.e., with a higher computational and memory cost). It should be noted that generally these deeper and wider architectures achieve better performance at the expense of requirement of a larger labelled dataset, a longer training time, a larger model size and a longer inference time.

There are also efficient and lightweight CNN architectures, which have fewer trainable parameters and so are more applicable on small devices with a low computation power. In this context, MobileNet of three versions (V1 [134], V2 [135], V3 [136]) and ShuffleNet of two versions V1 [137], V2 [138] have been proposed. In these approaches, the computation cost of conventional convolution operations is improved by using separable convolution, channel shuffle, etc. There are also recent survey papers on CNNs [139,140], which covers CNN history,

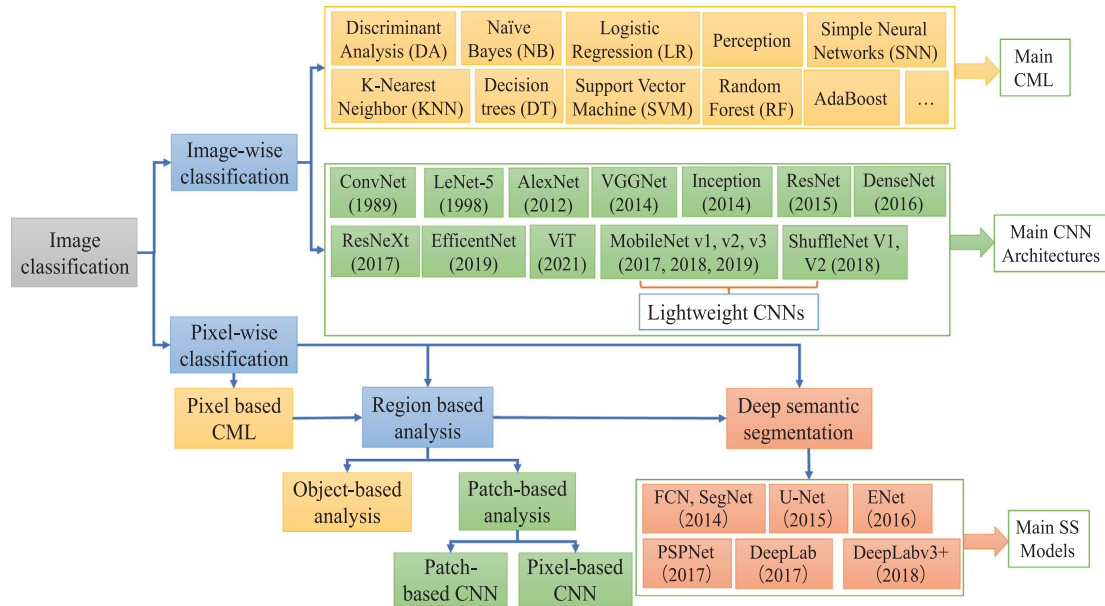


Fig. 8. Diagram displaying the main approaches for image-level and pixel-wise classification approaches, where CML approaches are highlighted in yellow, CNN based DL approaches at image level are highlighted in green and CNN derived deep SS approaches are highlighted in red. CML: conventional Machine Learning, CNN: Convolutional Neural Network, SS: Semantic segmentation.

classic and advanced models and their various applications. There are also many applications of various CNN architectures in UAV based PA applications. For example, [49] compares different CNN architectures (e.g., Inception-v3, ResNet-50, VGG-16, VGG-19 and Xception) for soybean pest classification, and also detection based on the sliding window technique (please refer to Fig. 9 for the workflow of image-level classification). More importantly, these CNNs become the backbone network in object detection or semantic segmentation networks, which are more widely used in PA applications.

Pixel-wise classification is to achieve fine-grained inference by assigning a label to each pixel of the image. This problem is also

termed semantic segmentation in computer vision community and has found a wide range of applications in various applications such as self-driving vehicles, medical image processing, etc. (please refer to a recent survey paper [141] for further information). To achieve semantic segmentation for UAV based PA applications, both CML and DL approaches have been applied, which are detailed as below.

- Conventionally, with a low-resolution remote sensing image, pixel information (e.g., pixel color and spectral information, surface orientation, height, HOG, SURF) is directly used in traditional pixel-based classification method. In this approach, features for each pixel are first generated via feature engineer-

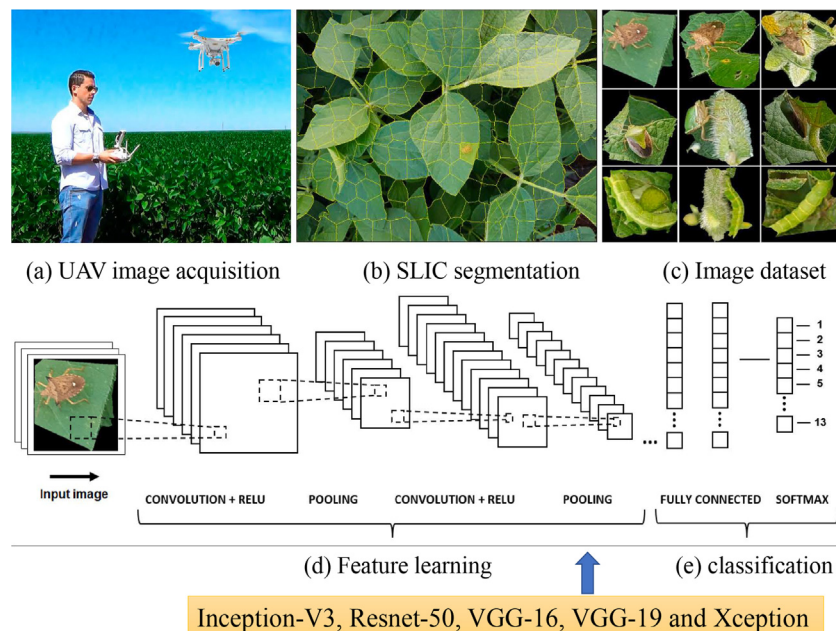


Fig. 9. Workflow of soybean pest classification at image-level based on UAV images and various CNN architectures [49].

ing, which are fed into different ML based classifiers (e.g., shallow neural network, SVM, RF) for classification. For example, RF has been applied to UAV multispectral images in [30] for wheat yellow rust disease mapping (please refer to Fig. 10 for its framework including data collection and preprocessing). Spectral, height (i.e., via digital surface model) and texture information are fused at pixel-level and applied for rice lodging classification by using a decision tree in [142]. However, with the significantly increased image resolution, the performance of this approach is limited (e.g., large noise, poor semantic meaning) since the important spatial and semantic information is overlooked in traditional pixel-based (or spectral-based) approaches.

- To incorporate spatial information, different approaches have been proposed, where the main methods in UAV based PA applications include object-based image analysis and patch-based image analysis, summarized as below.

- Object-based image analysis [143] follows the “segmentation-classification” mode, and has been available in several commercial software such as eCognition and ENVI. This approach starts with unsupervised image segmentation (e.g., via Multi-Resolution, Full-Lambda Schedule, Mean-shift, Quadtree-Seg [144]), aggregating information through coherent and spatially connected pixels into diverse image segments. It then conducts classification analysis (via CML such as RF, SVM and neural network) on these segments instead of individual pixels. In particular, various features can be defined for the segments, such as statistics (e.g., mean, standard derivation, maximum, and minimum values) of the spectral bands, texture features (e.g., Gray-Level Co-Occurrence Matrix, Gray Level Difference Vector), geometric features (e.g., object boundary, area, shape index) and contextual features [144,145]. There are also tunable parameters in segmentation algorithms for segment generation, which are able to determine the granularities of the segments but meanwhile are hard to choose optimally. Efforts are also made to address the scale differences among different objects including feature level and image level approaches [25]. However, (i) the quality of image object segmentation directly affects the classification performance,

which is generally based on trial and error; (ii) the empirical selection/extraction of the representative features is also the bottleneck to improving the segmentation performance.

- In patch-based analysis [146–148], patches/windows (i.e., rectangular regions of the image) are treated as the samples for training the classifier. There are two representative approaches in this regard including pixel-based CNN and patch-based CNN, which are further detailed as below.

- * In patch-based CNN [146], the large image is separated into small patches, then the common CNN model is applied on each patch to generate feature vector, which is rearranged into matrix format and then up-sampled to the original patch size. However, in this approach, each image patch is input to the model independently and the correlations between patches are ignored, leading to obvious gaps between patches, especially for objects with strong continuity;

- * In pixel-based CNN [148], the patch label is defined by the label of the center pixel for training. In the inference stage, the class label at each pixel is derived by classifying the window centered at the pixel via CNN. This approach, however, is extremely time-consuming due to the significant number of patches to be classified. There are also techniques for speeding up the approach such as applying a stride (i.e., reducing the number of patched to be classified) and interpolating the results via regression [149].

To summarize, the limitations of patch-based analysis are mainly as below: (1) there is no systematic approach to optimally choose the patch size, and it does not support arbitrary sizes, either; (2) the receptive field of the patch-based analysis is limited by the patch size (e.g., “intra-patch” context information) without considering overall context information, although multiple patch sizes can be adopted [147]. There is also a study integrating object-based and patch-based image analysis together [145], where an adaptive window exactly enclosing the objects (or segments) from image segmentation step is extracted and used in patch-based analysis after resizing to a fixed image size.

- Recently, there is also a trend to apply DL based semantic segmentation networks to pixel-wise classification problem in UAV based PA applications. This is because: (1) the spatial resolution of UAV images is becoming higher and consequently more detailed and “semantic” information is available for better classification performance; (2) DL based semantic segmentation networks, exploiting context by various techniques such as encoder-decoder architecture, spatial pyramid pooling and dilated convolution, possess strong capability in automatically learning the deep and global features for better classification performance over traditional shallow learning methods. The main architecture of semantic segmentation network is comprised of two parts, i.e., an encoder network (compression path) followed by a decoder network (expansion path). The encoder network is to extract features from input image via different CNN architectures, e.g., VGG and ResNet by removing the fully connected layers. While, the decoder network does the reverse of the encoder using convolution and up-sampling operations to generate an image of the same size as the input image with masks for pixel-wise classification [16].

For example, the pioneering Fully Convolutional Network (FCN) [150], stacking a bunch of convolutional layers in an encoder-decoder fashion (i.e., encoder downsampling via strided convolution, and decoder upsampling via transpose convolution), has been applied in [151] for crop weed mapping from UAV RGB imagery. The SegNet (i.e., encoder via convolutions and max pooling, decoder via upsampling by the max pooling indices

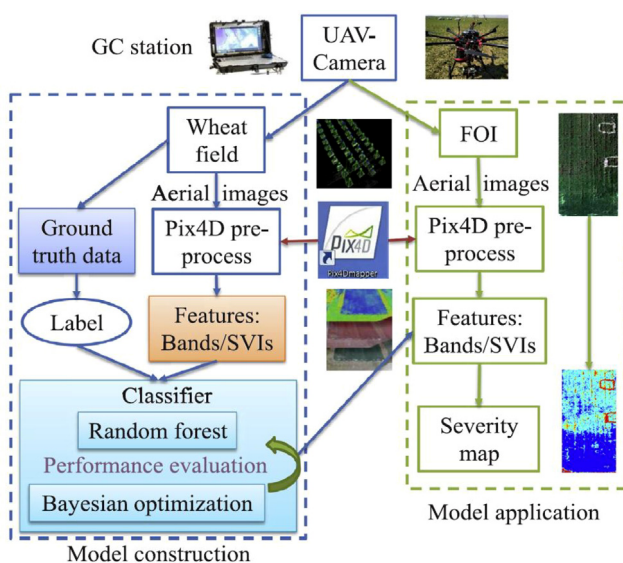


Fig. 10. Framework of UAV based yellow rust monitoring system for use at field scales in [30]: CML based pixel-wise classification model construction (left part) and model application for ROI (right part).

and convolutions) [152] is also applied in [61] for weed mapping by using multispectral image (please refer to Fig. 11 for the framework of weedNet). However, the encoder downsampling layers (reducing spatial resolution) tend to produce round corners and smoothed edges in increasing the receptive fields to capture contextual information. Therefore, U-Net [153], which can take into account the features from multiple layers, was then proposed. It can not only achieve good localization but also utilize contextual information concurrently. It has been applied in [11,154] for wheat yellow rust disease mapping by using UAV multispectral images. Then many other semantic segmentation networks have also been proposed such as ENet [155] in 2016 (i.e., a lightweight architecture, excluding downsampling operations in the last stage of the network, to obtain a faster inference time), PSPnet [156] in 2017 (i.e., exploiting global context information, see, [157] for its application in yellow rust disease detection) and DeepLab [158] in 2017. Recently, DeepLabv3 + network [159] addresses the limitations of FCN (a serious loss of spatial information due to conventional convolution and downsampling operations) by using dilated/Atrous convolution and Atrous Spatial Pyramid Pooling (ASPP), which is able to consider objects at different scales. It has been shown in [90] that DeepLabv3 + outperforms the U-Net in wheat lodging area segmentation at multiple growth stages.

The main challenge of pixel-wise classification over image-level classification lies in a significantly increased workload in collecting the representative datasets with pixel-level annotation instead of image-level annotation. The time for image-level annotation is generally less than a second, however, the time for pixel-wise annotation is in the range of minutes to even hours depending on image size, scene complexity and class number. Therefore, (1) data-level approaches, e.g., semi-automatic annotation approaches, various data augmentation and synthesis techniques, are paramount to pixel-wise classification to reduce the time and workload in manual image annotation; (2) algorithmic approaches (e.g., transfer learning, semi-, self- and unsupervised learning [160]) discussed in SubSection 5.2.3 are also important in reducing algorithm reliance on a large volume of labelled dataset.

Apart from pixel-level labelled data scarcity, another challenge is the problem of class imbalance (i.e., some classes have a significantly higher number of examples in the training set than other classes), which widely exists in various application domains such as computer vision, medical diagnosis, fraud detection. In PA applications, in general, the healthy crop pixels significantly outnumber the stressed crop pixels, early crop stress detection in particular. Imbalanced dataset degrades algorithm performance as the overall accuracy is biased to the majority class, and therefore leading to

poor predictive performance for the minority class [161]. Various techniques are available in the literature to tackle this issue, such as sampling approaches at data level and various approaches at classifier level. The data-level approaches include under-sampling and over-sampling to balance the dataset. While classifier-level approaches include thresholding, modified cost function by assigning a higher weight to the minority class misclassification, different cost functions in CNN learning (e.g., pixel-wise cross entropy loss, dice loss), one-class classification and ensemble approaches. The interested readers may refer to the recent survey paper [162] for more details.

4.1.3. Object detection

Object detection, including object localization and classification, aims to mark bounding boxes around objects of interest in an image and classify each one with a predefined class label [163]. There are already several review papers on this topic. For example, [164] reviews object detection approaches in optical images with a focus on traditional approaches including ML based ones. [140] reviews generic CNNs frameworks including CNN based object detection networks and their applications. [165] reviews recent advances in DL for object detection including detection components, learning strategies, applications and benchmarks. Recently, [166] investigates the performance of different CNNs based object detection networks for low-altitude UAV datasets. [16] reviews DL techniques to classify agricultural crops through UAV images including object detection based approaches. As displayed in Fig. 12, the existing object detection approaches can be mainly divided into conventional object detection and CNN based object detection, while the latter can be further divided into two categories including two-stage detectors (or region proposal-based methods) and one-stage detectors (or regression based methods). In order to understand performance evaluation metrics for object detection, we first define confidence score and Intersection of Union (IoU). Confidence score is the probability that an anchor box contains an object, which is usually returned by a classifier, while IoU is defined as the area of the intersection divided by the area of the union of a predicted bounding box B_p and a ground-truth box B_{gt} , given by Eq. 1

$$IoU = \frac{area(B_p \cap B_{gt})}{area(B_p \cup B_{gt})} \quad (5)$$

A detection is considered to be a true positive (TP) only if it satisfies three conditions: confidence score is greater than a predefined threshold; the predicted class matches the class of a ground truth; the predicted bounding box has an IoU greater than a predefined threshold (e.g., 0.5) with the ground-truth. And if IoU is less than

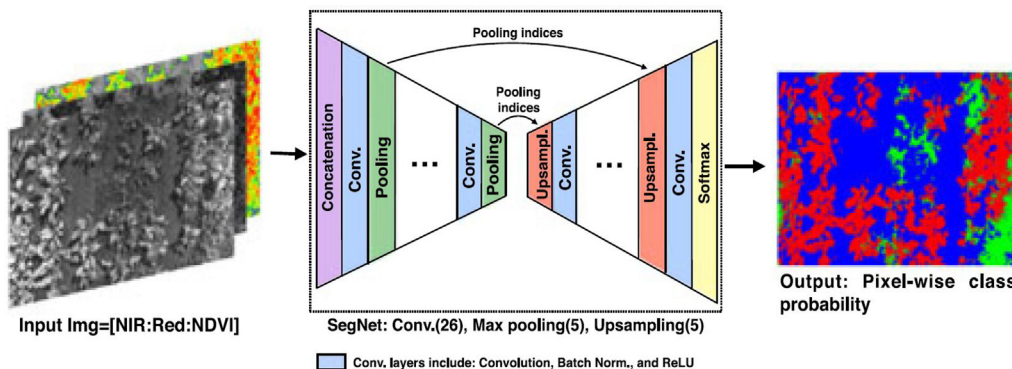


Fig. 11. Framework of weedNet [61] based on an encoder-decoder cascaded dense semantic segmentation framework SegNet [152]: UAV based multispectral imaging for weed segmentation.

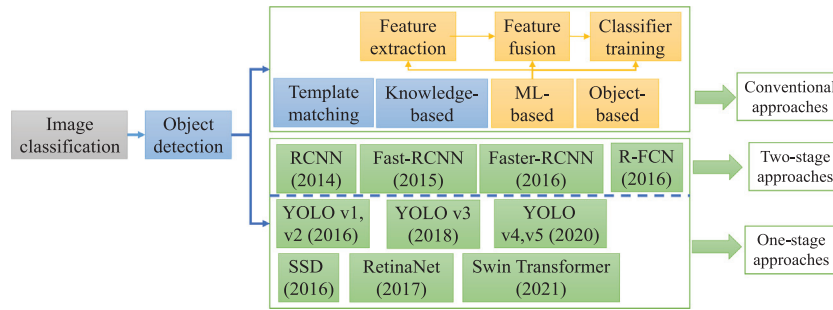


Fig. 12. Diagram displaying main approaches for object detection, where CML ones are highlighted in yellow and CNN based DL approaches are in green. It is noted that the diagram of conventional object detection approaches is directly from the survey paper [164].

a threshold, then it is a wrong detection and is classified as False Positive (FP), while when a ground truth is present in the image and the model fails to detect the object, then it is classified as False Negative (FN). As a result, some metrics such as Precision and Recall in Eq. (4) can be defined accordingly after defining TP and FN. Then some commonly used performance metrics for object detection can be further defined such as average precision (AP), mean average precision (mAP), average recall (AR) and mean average recall (mAR). In particular, AP is the precision averaged across all unique recall levels, and AR is the recall averaged over all $IoU \in [0.5, 1.0]$. While mAP and mAR are defined as the mean of AP and AR across all k classes, given by Eq. (4)

$$mAP = \frac{1}{k} \sum_{i=1}^k AP_i; mAR = \frac{1}{k} \sum_{i=1}^k AR_i. \quad (6)$$

It is also noted that different threshold values can be chosen for the IoU such as 0.5, 0.75 in metrics calculation.

- Following the review paper [164] on object detection in optical RS images, the conventional object detection approaches are mainly divided into template matching-, knowledge-, object- and ML- based approaches. These approaches, however, highly depend on human involvement in feature design to guarantee satisfying performance.
- Recently, CNNs have been applied to object detection problems, which have achieved the SOTA performance. The two main categories of CNNs based object detection methods are briefed as below.
 - Two-stage object detection is to generate region proposals (i.e., candidate rectangular bounding boxes) on the feature map, then use CNNs to classify each proposal into an object class label and refines the proposals with a bounding box regression. Typical two-stage approaches include RCNN in 2014 [167] (i.e., using region proposal to replace conventional sliding window approach), Fast-RCNN in 2015 [168] (i.e., with lessons from SPP-Net [169] to introduce ROI pooling layer, making the network faster), Faster-RCNN in 2016 [170] (i.e., introducing Region Proposal Network (RPN) [156] at the end of convolutional layer) and RFCN in 2016 [171] (i.e., adopting position-sensitive score maps to address a dilemma between translation-invariance in image classification and translation-variance in object detection);
 - One-stage approach, on the other hand, directly makes a classification and detects object location, which is able to achieve a high detection speed but usually with a reduced accuracy. In this approach, the images are divided into a grid system, where each cell in the grid is responsible for detecting objects within itself. Typical one-stage approaches include You Only Look Once (YOLO) in 2016 [172] and its family (YOLO V2 in 2017 [173] (with darknet-19), V3 in 2018 [174] (Feature Pyramid Network to achieve multi-

scale feature fusion [175]), V4 in 2020 [176], and V5), and Single Shot multibox Detector (SSD) in 2016 [177], RetinaNet in 2017 [178] (i.e., introducing focal loss function to address class imbalance during training), Swin Transformer [179] in 2021.

As reviewed in [16] and also the Table 2, applications of object detection in UAV based PA mainly focus on vegetation or fruit plants where boundary or shape is visually available for the individual. Some application examples include banana plant detection by Faster RCNN [180], tree detection by YOLOv3 [181], citrus fruit by Faster RCNN with RGB [64] and citrus tree by Faster RCNN with multispectral imaging [182], apple tree crown by Faster RCNN [183] (please refer to Fig. 13 for its flowchart), broccoli by Faster R-CNN and CenterNet [71]. There are also some studies on crop detection such as wheat ear detection by RetinaNet and YOLOV4 [184], weed detection by Faster RCNN and SSD [62]. Most of the studies directly draw different SOTA object detection algorithms in CV and DL community with minor modifications and then identify the most suitable one for the applications of interest.

It is noted that most of the existing object detection applications are performed on RGB images for the purpose of detection and counting, this is because detection and counting more relies on spatial, texture, pattern, and shape characteristics of the object. More research should be conducted on object objection via non-visible images, particularly for tasks of stressed crop/fruit detection and classification, where non-visible spectral and multimodal information (e.g., multispectral, thermal, and hyperspectral imaging) plays a more important role in discriminating stressed plants from healthy ones. A challenge to be addressed in object detection by non-visible images is the lack of well-trained back-borne models in non-visible image domains, since most existing back-borne models are trained via RGB images in CV community.

4.2. Unsupervised learning

Application of unsupervised learning in UAV based PA is relatively sparse compared against supervised learning. As displayed in Fig. 14, the role of unsupervised learning in UAV based PA is mostly to ease the major issue of supervised learning, i.e., the need of large amounts of labelled dataset for model construction, which is time-consuming, labor-intensive and costly. In this context, the existing unsupervised learning algorithms for UAV based PA applications can be divided into several categories including unsupervised learning for feature dimension reduction, direct unsupervised learning for simple classification tasks and unsupervised learning for data labelling, which are further detailed as below.

- A large number of features usually require a large volume of labelled datasets to learn the large amount of parameters in supervised learning. For example, there are usually hundred of

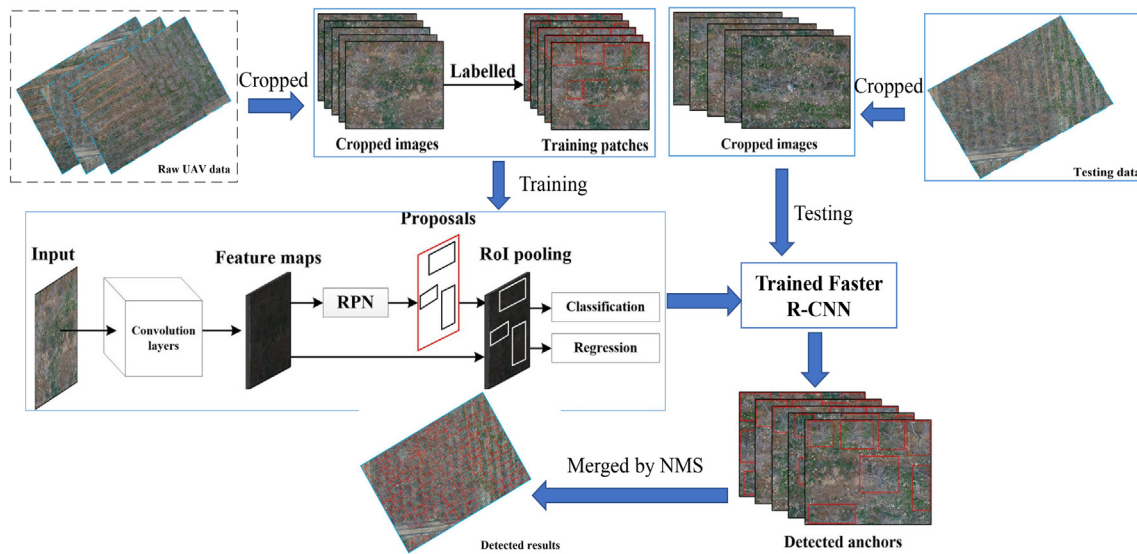


Fig. 13. Flowchart of Faster R-CNN for apple tree detection by UAV images in support of tree crown information extraction [183].

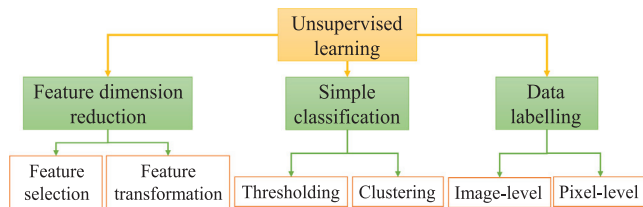


Fig. 14. Unsupervised learning in UAV based PA applications: feature dimension reduction, simple classification tasks, and data labelling.

narrow spectral bands in hyperspectral imaging, which are highly correlated and will also increase the computation cost in model training and inference [28,112]. Therefore, feature dimension reduction approaches are usually adopted including supervised and unsupervised approaches. In term of unsupervised approach for dimension reduction, different algorithms are available in the literature including feature selection and feature transformation approaches. Typical feature transformation approaches include Principal Component Analysis (PCA) maximizing the variance with a limited number of features, and independent component analysis. While feature selection approaches include optimum index factor and standard false color composite. For example, [185] compares different PCAs for hyperspectral imaging. [186] evaluates different unsupervised dimension reduction methods for individual tree crown delineation by using UAV multispectral imaging.

- In direct unsupervised learning, unsupervised learning algorithms are directly used for different PA tasks, relatively simple classification tasks in particular. For example, [108] compares two unsupervised algorithms for vineyard canopy segmentation including thresholding via Otsu's algorithm and K-means clustering. Different colour features and texture features are compared to identify the best one for lodged and non-lodged wheat segmentation via thresholding approach in [187]. [188] assesses different bands and vegetation indices derived from UAV multispectral imaging, and applies thresholding approach for wheat yellow rust disease mapping. It is noted that this approach is generally applicable to some simple tasks with moderate classification performance, since index thresholding approach only uses one index with limited information and the clusters derived from clustering approaches are not necessarily linked to classification task of interest.

- In term of data labelling, the main methods can be further categorized into image-level and pixel-level approaches.

- At image-level, [189] applied two unsupervised DL algorithms (i.e., deep clustering) on two image-level weed classification datasets, and then designed a semi-automatic data labelling method, where human labelling is only done on the clusters generated instead of all images. It should be noted that the same object in real world may belong to several classes depending on the context of interest. Therefore, a larger number of clusters than the number of classes (in supervised learning task) but significantly less than the number of images should be defined, where a trade-off is usually made between clustering quality and manual labelling cost reduction;
- At pixel-level, by considering the "row" characteristics of line crops, [190] automatically labels inter-row weeds by detecting the crop rows along with superpixel operations (i.e., simple linear iterative cluster (SLIC)), so that crop/weed dataset are automatically labelled at pixel-level for follow-up supervised learning (i.e., semantic segmentation).

It should be noted that more advanced unsupervised learning algorithms (e.g., data augmentation and synthesis, domain adaptation models as discussed in Section 5.2) in image processing and machine learning communities should be exploited for UAV based PA applications so that the workload of data labelling in supervised learning and better algorithm performance can be resulted.

5. Challenges with possible solutions

This section discusses challenges along with possible solutions of AI-enabled UAV perception systems for PA applications, which are organized in terms of UAVs sensing systems, AI algorithms, and user acceptance. The structure of this section is displayed in Fig. 15 for the ease of reading.

5.1. UAVs sensing systems

There are still many challenges to be addressed in order to fully realize the potentials of UAV sensing systems, which, in term of UAV sensing platforms, include flight stability and reliability, limited payload and flight endurance, image resolution, surface reflectance and illumination issues.

5.1.1. Flight stability and reliability

Flight stability and reliability is important to obtain high-quality orthorectified or tilted images, particularly for small UAVs (i.e., with a small size and low weight) under the wind effects (turbulence/gust in particular) in field conditions. Flight stability includes accurate path following control and also stable flight attitude under uncertainties and disturbances. This is because many post-processing algorithms rely on sufficient image overlap for image stitching, and a significant derivation of flight path or flight attitude, leading to missed regions or images of low quality (e.g., image blur), will make image stitching and post-processing a challenge. Therefore, better airframes and flight control algorithms (e.g., disturbance rejection algorithms [191]) should be further developed to improve flight stability under disturbances and uncertainties. In addition, the field terrains are not necessarily flat, rugged or mountainous regions in particular, therefore, terrain following and obstacle avoidance functions should also be further developed and enhanced to improve flight reliability and safety so that the operators' workload can be significantly reduced.

5.1.2. Payload and endurance

Small UAVs usually have a limited payload due to their specific structure characteristics, which may limit their applications in multimodal data collection (i.e., with a number of different payloads (e.g., multispectral camera, LIDAR) of different sizes and weights). Moreover, flight endurance is also a limiting factor of small UAVs for large-scale agricultural applications. This is mainly due to a limited battery capacity and power consumption of payload and computing units. To this end, the following solutions can be considered.

- An apparent solution would be developing high energy-density (i.e., efficient) battery for UAVs, where factors such as weight, price and longevity should be considered simultaneously in choosing suitable batteries; alternative power sources can also be considered such as hydrogen fuel cells, combustion engines (pesticide spraying drones in particular) and solar power (expensive) [192];
- A reduced payload (e.g., with miniaturization and weight reduction of sensing units) and computing units with low-power consumption can also extend flight endurance;
- A large UAV platform (e.g., fixed wing over rotor-wing) can be adopted, however, price, safety, operation complexity (e.g., landing requirements) and regulatory issues should be considered in using large-scale UAV platforms;
- Better coverage mission/path planning algorithms should also be developed by considering field (e.g., geometric and topological) characteristics and operation conditions (e.g., wind direction) [193] so that energy can be saved maximally;

- Multiple UAVs (or swarming technology [194]) can be deployed to cover fields of interest more efficiently than a single UAV with battery replacement or recharge [195,196].

5.1.3. Image resolution

Image resolution is vital in crop management such as early crop stress detection and crop counting. For example, a better spectral resolution (e.g., hyperspectral image over multispectral image) is able to provide more sensitive spectral bands to detect physical, chemical and biological changes of crops under stresses. A higher spatial resolution can provide more detailed information in crop classification, detection and counting. A shorter revisit time can provide a better temporal resolution, which is important for early crop stress monitoring and crop modelling via time-series data. However, a better image resolution is usually at a price of a higher operational cost. For instance, flying lower obtains images with a higher spatial resolution but inevitably reduces flight efficiency (e.g., coverage per unit time); increasing spectral resolution or ranges requires more expensive sensors; and a shorter revisit time generally requires more resources and human efforts. Therefore, a trade-off should be made between image resolution and cost/efficiency. In practice, it is paramount to investigate “suitable” image resolutions for particular applications of interest while with acceptable algorithm performance. The following solutions can be exploited in resolution determination.

- Spectral resolution: hyperspectral imaging can be first used to determine the most sensitive bands/wavelengths so that tailored multispectral cameras can be designed and applied with a significantly lower price and weight, reduced operational difficulty and data volume;
- Spatial resolution: one can fly UAVs at different altitudes or downsampling high-resolution images to determine the most suitable spatial resolution for applications of interest, so that flight efficiency can be improved;
- Temporal resolution: instead of conducting revisit survey with a fixed period, one can design a suitable revisit protocol by considering the key crop or stress development stages determined by agronomists.

5.1.4. Surface reflectance

It should also be highlighted that although UAV (remote) sensing is very useful in evaluating soil/crop conditions, it actually provides a poor representation of the root/under-canopy zone environment, this is because (remote) sensing data represent reflectance of surface materials, which might be bare topsoil, plant material, or a mixture of both. This may limit their applications, where crop stress appears in root zone or under-canopy. For example, some pests/diseases may develop from bottom canopy before

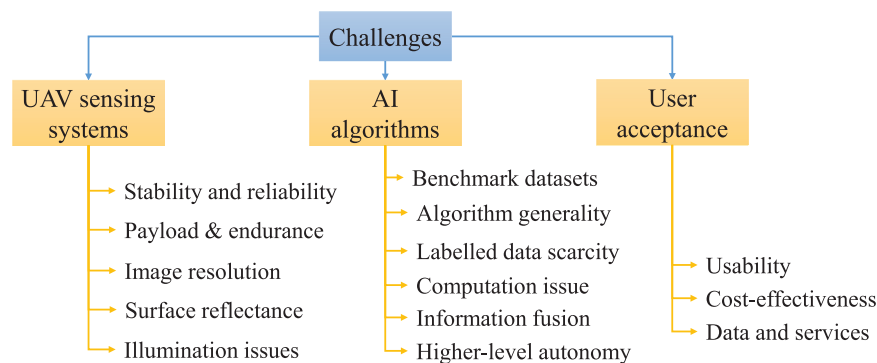


Fig. 15. Challenges faced by AI-empowered UAV perception systems in terms of UAV sensing systems, AI algorithms and user acceptance.

spreading to the top, and so remote sensing may fail to detect them in early stages. Therefore,

- UAV sensing system is actually a good complement to (instead of a replacement of) ground-based sensing systems, and therefore should be combined with ground-based approaches in crop management;
- Agronomists should be actively involved in determining the suitable PA application scenarios where UAV remote sensing systems can be applied;
- There are also ongoing research efforts to develop UAV based sampling systems for air, water [197] and soil based on dedicated sensors beyond optical sensing, which can further expand the application domains of UAV sensing systems in PA and beyond.

5.1.5. Illumination issues

Although UAV RS is less affected by the cloud issue in comparison against satellite RS, non-consistent illumination issues caused by moving cumulus clouds should be carefully considered. This is because shaded areas by moving clouds on a bright sunny day can vary significantly between images obtained during the same mission. As a result, some automated image matching algorithms used in both triangulation and digital elevation model generation may fail for some images [198,199], leading to poor image mosaics and digital elevation models. Different strategies can be adopted to alleviate this issue for better image quality.

- Avoiding UAV surveying under moving cumulus clouds would be the best choice, but instead performing the task in illumination consistent conditions;
- It is suggested in [198] that one can generate band ratio images from the original images (i.e., less sensitive to shadow) and use them (instead of the original images) for triangulation instead;
- Different algorithmic approaches can be adopted to alleviate the adverse effects of shadow, for example, shadow removal algorithms can be considered to pre-process UAV images with cloud shadows [200].

5.2. AI algorithms

In addition to UAV sensing systems, there is an ever more room for further improvement in AI algorithms in the context of AI-empowered UAV perception systems for PA applications, which is organized in terms of benchmark PA dataset, algorithm generality, labelled data scarcity, computation issue, multiple-sources information fusion and higher-level autonomy.

5.2.1. Benchmark PA datasets

AI algorithms, deep CNNs in particular, generally require a large volume of dataset (labelled dataset in particular for supervised learning) to train their models. There is no exception to UAV sensing for PA application with AI algorithms as well. In the DL or computer vision community, there are several large-scale benchmark datasets for algorithm benchmark and development such as ImageNet [201] for image classification, MS COCO [202] for object detection and segmentation. There is also an attempt in agriculture community such as the PlantVillage dataset [203], which supports leaf-level crop species and diseases identification. In addition, some UAV-derived PA benchmark datasets are also available such as the Wuhan UAV-borne hyperspectral image dataset for crop classification [204], Weedmap multispectral dataset in [205] to classify sugar beet and weeds from aerial images, VOAI dataset for 12 tree species classification with UAV RGB imaging [206], WeedNet dataset [61] for semantic segmentation of crop and weed with 4-band UAV multispectral images and UAV based rice seed-

ling detection dataset [207]. A compilation of publicly available UAV dataset with labels for PA applications are also listed in [208].

However, there is still a lack of large-scale UAV-derived benchmark datasets for PA applications [16], particularly for crop stress detection and mapping. The possible reasons may be the significant efforts required in term of logistics (e.g., crop planting, management and crop stress control experiments) and ground-truthing (pixel-level annotation in particular). Therefore, future attempts should be made to prepare large-scale UAV benchmark datasets by considering environmental variations (e.g., background clutter, illumination changes, occlusion, etc.), crop development stages, crop varieties, stress development stages (please refer to [209] for the ongoing effort towards this end).

5.2.2. Algorithm generality

Algorithm generality is paramount for agricultural applications so that their potentials can be fully exploited. This is however challenging because the developed algorithms are facing very complex agricultural operating environments. Therefore, future efforts should be made to guarantee the algorithm performance under different crop fields, crop varieties, crop growth stages, stress development stages, cameras with different resolutions and lighting/environmental conditions. To improve model generality, the following approaches can be considered.

- Dataset collected for model training should be representative for the target applications of interest to improve algorithm accuracy and robustness;
- Regularization methods (e.g., reducing model complexity, using validation set to early training stopping, weight decay, dropout, batch normalization) can be adopted to improve algorithm generality;
- Transfer learning [90] of a pre-trained model with sufficient datasets in related domains can be adopted, particularly for the case with a limited number of labelled samples in the target domains;
- Domain adaptation can be further developed [210], which is to transfer classification system to target (but related) domains where there exist distribution discrepancies in source and target domains. In particular, unsupervised domain adaptation aims to adapt models across domains without the need for an additional labelling effort in target domains [211].

5.2.3. Labelled data scarcity

Obtaining large-scale datasets with accurate labels is critical for supervised ML but is generally challenging (sometimes prohibitively expensive) in agriculture applications. This is because: (1) different from many other application domains, the period of obtaining agricultural datasets (field datasets in particular) is at least in the unit of season or even year; (2) similar to many other application domains, agricultural data labelling highly involves expert participation or multiple experts for obtaining consensus labels and therefore is resource-intensive (e.g., financially, logistically). To address this issue, the following approaches are identified according to different applications:

- Communication (i.e., crowd-sourced non-experts) system to build large training datasets [212]; however, the problem of label noise should be carefully considered;
- Deploying semi-automated data labelling so that more labelled data can be generated [113], which is particularly important for semantic segmentation problems since pixel-level labels are required, being extremely labor-intensive and time-consuming;
- Data augmentation (e.g. geometrical transformation such as rotation, scaling, translation, shearing, flipping; color manipulation such as gamma correction, colour space, colour space

transformation, noise injection, kernel filtering [58,213]) and data generation/synthesis [214,215] to enhance representativeness of the collected datasets;

- Algorithmic approaches for small or even no labelled datasets [160], e.g., transfer learning (i.e., transferring the model learnt from a large labelled dataset to related domains) [90], semi-supervised learning (i.e., learning from both labelled and unlabelled images for better performance) [216,217] or weak supervised learning (i.e., semantic segmentation by different levels of labels including image-level annotation and coarse annotations such as scribble, bounding box and point/spot) [218], self-supervised learning (i.e., generating labelled data on its own), unsupervised learning (e.g., unsupervised domain adaptation [211]).

5.2.4. Computation issue

ML algorithms in image analysis, deep CNNs in particular, generally involve a high training time (not a big concern as this is normally done offline) and inference time (this is the main issue, e.g., a few frames per second, fps), leading to a large system latency for real-time applications. As a result, there are still a limited number of studies [61] in using ML algorithms onboard a UAV for real-time PA applications (e.g., UAV based weed detection and spraying). A survey paper is available on the use of edge computing in UAV RS for PA applications [208], where several PA application examples are discussed. To achieve real-time PA applications, the following attempts should be further made:

- More efficient ML models/architectures, e.g., light-weight models by employing model compression techniques (e.g., parameter pruning and quantization) [219] or deploying light-weight CNNs, should be designed to shorten the inference time, where a speed/memory/accuracy trade-off can be made (see [220] for convolutional object detectors);
- The hybrid/fusion of CML and DL may accelerate the inference time without degrading the performance [117]. For example, CML with domain knowledge may significantly reduce the search region of DL; the use of CML for feature selection may significantly reduce computation load (i.e., training and inference) of DL models [113];
- Better or dedicated computation resources (e.g., CPU, DSP, GPU, ASIC, FPGA, MCUs, [221,222], Vision Processing Unit (VPU)) should be deployed onboard the UAV (with an optimized workload assignment) to accelerate the computing, meanwhile, energy consumption issue should also be considered due to a limited UAV battery capacity;
- Image spatial resolution can be tuned to make a trade-off between computing time and algorithm performance. It is also noted that a higher image resolution does not necessarily lead to better algorithm performance, this is because some information may be unrelated to the task of interest or plays the role of noises.

5.2.5. Multiple-sources information fusion

In real-life PA applications, UAV-derived information can seldom be used alone. They should be combined with other sources of information for better agricultural decision making. For example,

- UAV multimodal (complementary) data (e.g., RGB images, multispectral and hyperspectral information, height, thermal, LIDAR) should be fused for a richer representation and therefore much improved performance. As reported in [223], multimodal learning has widely been used in human activity recognition, medical applications, autonomous driving [224] and satellite remote sensing [225]; however, it is less exploited in UAV based

PA applications [63]. The fusion can be achieved by conventional multimodal learning (via modality-specific features) or deep multimodal learning, where the later can not only provides data-level fusion (or early fusion) and decision-level fusion (or late fusion) but also provides much flexibility at feature-level fusion since DL architectures are able to learn a hierarchical representation of the underlying data across their hidden layers;

- Data from multiple (heterogeneous) sources and sensors such as satellites, drones, robotics and IoT devices should also be integrated together to provide holistic insights. For example, fusion of UAV images (high spatial resolution) and Sentinel-2A multispectral images (wider spectral range and spectral resolution) outperforms individual image source [34] for RS classification. In this regard, the challenges of integrating multiple-sources datasets should be addressed, such as heterogeneity in spatial/temporal resolutions, data synchronization issue;
- Crop models can be applied for crop monitoring, prediction, management, and intervention assessment, by coupling agronomic (e.g., climate, crop information) knowledge and expertise and RS information [42,226]. Crop model calibration problem for regions of interest should be addressed by suitable optimization [227] or Bayesian approaches so that the developed model can be applied in different regions and over different years [228].

5.2.6. Higher-level autonomy

It is also noted that current use of AI and UAV in PA is mostly “reactive”, where UAV mainly plays the role of data collection and AI mainly plays the role of data analysis (mostly offline) for crop stress management. This approach is less efficient in routine crop monitoring as in each time the whole regions of interest should be covered by UAVs and huge volumes of data are collected, transferred, analyzed by AI algorithms for decision making. Future efforts should be made to develop “proactive” farming by developing higher-level autonomy. For example,

- Multiple-sources information fusion in Section 5.2.5 (e.g., historical crop/soil/climate information, crop model information, current measurements from satellite/UAVs/ground IoT devices) should be able to identify the targeting areas for assessment instead of whole areas for each time;
- Real-time analysis capability onboard UAV should also be further developed so that real-time decision making and target intervention can be done on-the-fly [229], which is critical to optimising crop stress control (e.g., timely stress detection and target intervention). Moreover, the targeting areas should also be updated dynamically upon each routine monitoring.

5.3. User acceptance

User trust and acceptance of a particular technology is paramount to realize its full potentials. There is no exception to AI empowered UAV perception systems for PA. At present, mainstream acceptance of this technology across the agricultural sector is not yet being realised, with concerns on usability, economic and agronomic justifications of the technology, and data & services available [230]. As highlighted in [230], the main reasons are threefold:

- In term of data collection, due to the intensity of human operation involved in large-scale applications along with regulation issues, current cost of UAV data collection is still very high;
- In term of “data to actionable information”, the acquired data do not have tailored precision, integrity or quality to derive adequately actionable information for agricultural practice;

- In term of “information to action”, the automated integration of information into the existing precision farming systems is still missing, as translating the information into appropriate actions still requires manual intervention by a specialised analyst.

Therefore, an appropriate engagement between the UAV sector and the agriculture sector is urgent to allow UAV-derived services to mature into useful PA tools and techniques. To this end,

- UAV sensing technologies (e.g., platforms, communication, positioning, monitoring and observation systems) need to be further developed to ensure safe, reliable, resilient and cost-effective operations;
- Further development in sensors, data storage & transfer and AI algorithms is required to transform raw sensing data into actionable agricultural information more easily;
- UAV derived information should also be integrated with information of other sources (e.g., historical information of soil or yield, climatic information) so that better decisions (e.g., application map for fertilisers) can be made via crop models or data-driven models, and the decisions can be understood and implemented by existing intervention systems (e.g., farm management software packages of different farming machines);
- It is also paramount to understand and meet the needs and requirements of different users from across the agricultural sector so that a clear agronomic and economic benefit (cost-effectiveness) for UAV technologies can be presented.

The future research directions highlighted in Section 5.1 and 5.2 will help to support the developments and advances along these listed areas.

6. Conclusions

This survey paper performs a thorough review on recent developments in the integration of UAV sensing systems and AI algorithms in PA applications, as well as the challenges and prospects for their future development in order to realize their full potentials in boosting crop production, reducing costs and environmental footprint, and improving workplace safety and welfare. In particular, an integrated PA framework is first presented, highlighting the roles of UAV sensing and AI algorithms, and also presenting different PA elements sharing some common challenges in crop development lifecycle. Then UAV sensing systems are discussed, which include the pros and cons of different UAV platforms and also various UAV sensing units for PA applications. Selection of UAV platforms and sensing units is discussed according to requirements of different applications. Then a wide range of AI algorithms in UAV based PA applications are presented, which are categorized according to the learning methods, mainly comprising of supervised learning and a few unsupervised learning algorithms. The supervised algorithms, mainly including regression, image classification (e.g., image-wise, pixel-wise and object-wise) and image object detection, are the dominant approaches in UAV based PA applications; while unsupervised learning is mainly to reduce the need of a large amount of labelled datasets in supervised learning. Finally, the underlying challenges and possible solutions are presented in terms of UAV sensing systems, AI algorithms, and user acceptance. By demystifying a variety of UAV sensing and AI algorithms in PA applications, it is envisioned that this review is able to promote rapid research, deployment and successful exploitation of AI empowered UAV sensing systems for PA applications, and therefore contributing to addressing the future agricultural and human nutrition challenges in the long run.

CRediT authorship contribution statement

Jinya Su: Conceptualization, Investigation, Writing - original draft. **Xiaoyong Zhu:** Writing - review & editing. **Shihua Li:** Writing - review & editing. **Wen-Hua Chen:** Supervision, Writing - review & editing, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by UK Science and Technology Facilities Council (STFC) under Newton fund with Grant No. ST/N006852/1 and ST/V00137X/1.

References

- [1] J.A. Foley, N. Ramankutty, K.A. Brauman, E.S. Cassidy, J.S. Gerber, M. Johnston, N.D. Mueller, C. O'Connell, D.K. Ray, P.C. West, et al., Solutions for a cultivated planet, *Nature* 478 (7369) (2011) 337–342.
- [2] A. Sridhar, A. Balakrishnan, M.M. Jacob, M. Sillanpää, N. Dayanandan, Global impact of covid-19 on agriculture; role of sustainable agriculture and digital farming, *Environmental Science and Pollution Research* (2022) 1–17.
- [3] S. Khanal, J. Fulton, S. Shearer, An overview of current and potential applications of thermal remote sensing in precision agriculture, *Computers and Electronics in Agriculture* 139 (2017) 22–32.
- [4] R. Gebbers, V.I. Adamchuk, Precision agriculture and food security, *Science* 327 (5967) (2010) 828–831.
- [5] J. Jung, M. Maeda, A. Chang, M. Bhandari, A. Ashapure, J. Landivar-Bowles, The potential of remote sensing and artificial intelligence as tools to improve the resilience of agriculture production systems, *Current Opinion in Biotechnology* 70 (2021) 15–22.
- [6] T. Duckett, S. Pearson, S. Blackmore, B. Grieve, W.-H. Chen, G. Cielniak, J. Cleaversmith, J. Dai, S. Davis, C. Fox, et al., Agricultural robotics: the future of robotic agriculture, arXiv preprint arXiv:1806.06762, 2018.
- [7] A. Carrio, C. Sampedro, A. Rodriguez-Ramos, P. Campoy, A review of deep learning methods and applications for unmanned aerial vehicles, *Journal of Sensors* 2017 (2017).
- [8] D.R. Green, *Unmanned aerial remote sensing: UAS for environmental applications*, CRC Press, 2020.
- [9] J. Kim, S. Kim, C. Ju, H.I. Son, Unmanned aerial vehicles in agriculture: A review of perspective of platform, control, and applications, *Ieee Access* 7 (2019) 105100–105115.
- [10] Z. Mi, X. Zhang, J. Su, D. Han, B. Su, Wheat stripe rust grading by deep learning with attention mechanism and images from mobile devices, *Frontiers in Plant Science* 11 (2020).
- [11] J. Su, D. Yi, B. Su, Z. Mi, C. Liu, X. Hu, X. Xu, L. Guo, W.-H. Chen, Aerial visual perception in smart farming: Field study of wheat yellow rust monitoring, *IEEE transactions on industrial informatics* 17 (3) (2020) 2242–2249.
- [12] L. Ma, Y. Liu, X. Zhang, Y. Ye, G. Yin, B.A. Johnson, Deep learning in remote sensing applications: A meta-analysis and review, *ISPRS journal of photogrammetry and remote sensing* 152 (2019) 166–177.
- [13] A. Kamilaris, F.X. Prenafeta-Boldú, Deep learning in agriculture: A survey, *Computers and electronics in agriculture* 147 (2018) 70–90.
- [14] D. Karimi, H. Dou, S.K. Warfield, A. Gholipour, Deep learning with noisy labels: Exploring techniques and remedies in medical image analysis, *Medical Image Analysis* 65 (2020).
- [15] A. Bashar et al., Survey on evolving deep learning neural network architectures, *Journal of Artificial Intelligence* 1 (02) (2019) 73–82.
- [16] A. Bouguettaya, H. Zarzour, A. Kechida, A.M. Taberkit, Deep learning techniques to classify agricultural crops through uav imagery: a review, *Neural Computing and Applications* (2022) 1–26.
- [17] P.K.R. Maddikunta, S. Hakak, M. Alazab, S. Bhattacharya, T.R. Gadekallu, W.Z. Khan, Q.-V. Pham, Unmanned aerial vehicles in smart agriculture: Applications, requirements, and challenges, *IEEE Sensors Journal* 21 (16) (2021) 17608–17619.
- [18] D.C. Tsouros, S. Bibi, P.G. Sarigiannidis, A review on uav-based applications for precision agriculture, *Information* 10 (11) (2019) 349.
- [19] W.H. Maes, K. Steppe, Perspectives for remote sensing with unmanned aerial vehicles in precision agriculture, *Trends in plant science* 24 (2) (2019) 152–164.
- [20] G. Yang, J. Liu, C. Zhao, Z. Li, Y. Huang, H. Yu, B. Xu, X. Yang, D. Zhu, X. Zhang, et al., Unmanned aerial vehicle remote sensing for field-based crop phenotyping: current status and perspectives, *Frontiers in plant science* 8 (2017) 1111.

- [21] X. Jin, P.J. Zarco-Tejada, U. Schmidhalter, M.P. Reynolds, M.J. Hawkesford, R.K. Varshney, T. Yang, C. Nie, Z. Li, B. Ming, et al., High-throughput estimation of crop traits: A review of ground and aerial phenotyping platforms, *IEEE Geoscience and Remote Sensing Magazine* 9 (1) (2020) 200–231.
- [22] A.D. Boursianis, M.S. Papadopoulou, P. Diamantoulakis, A. Liopa-Tsakalidi, P. Barouchas, G. Salahas, G. Karagiannidis, S. Wan, S.K. Goudos, Internet of things (IoT) and agricultural unmanned aerial vehicles (uavs) in smart farming: A comprehensive review, *Internet of Things* 18 (2022).
- [23] A. Mukherjee, S. Misra, N.S. Raghuvanshi, A survey of unmanned aerial sensing solutions in precision agriculture, *Journal of Network and Computer Applications* 148 (2019).
- [24] N. Delavarpour, C. Koparan, J. Nowatzki, S. Bajwa, X. Sun, A technical study on uav characteristics for precision agriculture applications and associated practical challenges, *Remote Sensing* 13 (6) (2021) 1204.
- [25] H. Yao, R. Qin, X. Chen, Unmanned aerial vehicle for remote sensing applications—a review, *Remote Sensing* 11 (12) (2019) 1443.
- [26] L.P. Osco, J.M. Junior, A.P.M. Ramos, L.A. de Castro Jorge, S.N. Fatholahi, J. de Andrade Silva, E.T. Matsubara, H. Pistori, W.N. Gonçalves, J. Li, A review on deep learning in uav remote sensing, *International Journal of Applied Earth Observation and Geoinformation* 102 (2021).
- [27] G. Messina, G. Modica, Applications of uav thermal imagery in precision agriculture: State of the art and future research outlook, *Remote Sensing* 12 (9) (2020) 1491.
- [28] T. Adão, J. Hruška, L. Pádua, J. Bessa, E. Peres, R. Morais, J.J. Sousa, Hyperspectral imaging: A review on uav-based sensors, data processing and applications for agriculture and forestry, *Remote sensing* 9 (11) (2017) 1110.
- [29] Y. Meng, W. Zhong, C. Liu, J. Su, J. Su, Y. Lan, Z. Wang, M. Wang, Uav spraying on citrus crop: impact of tank-mix adjuvant on the contact angle and droplet distribution, *PeerJ* 10 (2022).
- [30] J. Su, C. Liu, M. Coombes, X. Hu, C. Wang, X. Xu, Q. Li, L. Guo, W.-H. Chen, Wheat yellow rust monitoring by learning from multispectral uav aerial imagery, *Computers and electronics in agriculture* 155 (2018) 157–166.
- [31] Y. Pan, L. Li, J. Zhang, S. Liang, X. Zhu, D. Sulla-Menashe, Winter wheat area estimation from modis-evi time series data using the crop proportion phenology index, *Remote Sensing of Environment* 119 (2012) 232–242.
- [32] M. Liu, T. Yu, X. Gu, Z. Sun, J. Yang, Z. Zhang, X. Mi, W. Cao, J. Li, The impact of spatial resolution on the classification of vegetation types in highly fragmented planting areas based on unmanned aerial vehicle hyperspectral images, *Remote Sensing* 12 (1) (2020) 146.
- [33] Q. Yang, B. She, L. Huang, Y. Yang, G. Zhang, M. Zhang, Q. Hong, and D. Zhang, Extraction of soybean planting area based on feature fusion technology of multi-source low altitude unmanned aerial vehicle images, *Ecological Informatics*, p. 101715, 2022.
- [34] L. Zhao, Y. Shi, B. Liu, C. Hovis, Y. Duan, Z. Shi, Finer classification of crops by fusing uav images and sentinel-2a data, *Remote Sensing* 11 (24) (2019) 3012.
- [35] X. Bai, Y. Chen, J. Chen, W. Cui, X. Tai, Z. Zhang, J. Cui, J. Ning, Optimal window size selection for spectral information extraction of sampling points from uav multispectral images for soil moisture content inversion, *Computers and Electronics in Agriculture* 190 (2021).
- [36] X. Ge, J. Ding, X. Jin, J. Wang, X. Chen, X. Li, J. Liu, B. Xie, Estimating agricultural soil moisture content through uav-based hyperspectral images in the arid region, *Remote Sensing* 13 (8) (2021) 1562.
- [37] K. Ivushkin, H. Bartholomeus, A.K. Bregt, A. Pulatov, M.H. Franceschini, H. Kramer, E.N. van Loo, V.J. Roman, R. Finkers, Uav based soil salinity assessment of cropland, *Geoderma* 338 (2019) 502–512.
- [38] J. ten Harkel, H. Bartholomeus, L. Kooistra, Biomass and crop height estimation of different crops using uav-based lidar, *Remote Sensing* 12 (1) (2019) 17.
- [39] J. Bendig, A. Bolten, S. Bennertz, J. Broscheit, S. Eichfuss, G. Bareth, Estimating biomass of barley using crop surface models (csms) derived from uav-based rgb imaging, *Remote sensing* 6 (11) (2014) 10395–10412.
- [40] J. Yue, G. Yang, C. Li, Z. Li, Y. Wang, H. Feng, B. Xu, Estimation of winter wheat above-ground biomass using unmanned aerial vehicle-based snapshot hyperspectral sensor and crop height improved models, *Remote Sensing* 9 (7) (2017) 708.
- [41] Y. Shendryk, J. Sofonia, R. Garrard, Y. Rist, D. Skocaj, P. Thorburn, Fine-scale prediction of biomass and leaf nitrogen content in sugarcane using uav lidar and multispectral imaging, *International Journal of Applied Earth Observation and Geoinformation* 92 (2020).
- [42] T. Zhang, J. Su, C. Liu, W.-H. Chen, Bayesian calibration of aquacrop model for winter wheat by assimilating uav multi-spectral images, *Computers and Electronics in Agriculture* 167 (2019).
- [43] H.T. Nguyen, B.-W. Lee, Assessment of rice leaf growth and nitrogen status by hyperspectral canopy reflectance and partial least square regression, *European Journal of Agronomy* 24 (4) (2006) 349–356.
- [44] J.M. Duarte-Carvajalino, D.F. Alzate, A.A. Ramirez, J.D. Santa-Sepulveda, A.E. Fajardo-Rojas, M. Soto-Suárez, Evaluating late blight severity in potato crops using unmanned aerial vehicles and machine learning algorithms, *Remote Sensing* 10 (10) (2018) 1513.
- [45] M. Bhandari, A.M. Ibrahim, Q. Xue, J. Jung, A. Chang, J.C. Rudd, M. Maeda, N. Rajan, H. Neely, J. Landivar, Assessing winter wheat foliage disease severity using aerial imagery acquired from small unmanned aerial vehicle (uav), *Computers and Electronics in Agriculture* 176 (2020).
- [46] J. Su, C. Liu, X. Hu, X. Xu, L. Guo, W.-H. Chen, Spatio-temporal monitoring of wheat yellow rust using uav multispectral imagery, *Computers and electronics in agriculture* 167 (2019).
- [47] P.P. Roosjen, B. Kellenberger, L. Kooistra, D.R. Green, J. Fahrenttrapp, Deep learning for automated detection of drosophila suzukii: potential for uav-based monitoring, *Pest Management Science* 76 (9) (2020) 2994–3002.
- [48] F. Vanegas, D. Bratanov, K. Powell, J. Weiss, F. Gonzalez, A novel methodology for improving plant pest surveillance in vineyards and crops using uav-based hyperspectral and spatial data, *Sensors* 18 (1) (2018) 260.
- [49] E.C. Tetila, B.B. Machado, G. Astolfi, N.A. de Souza Belete, W.P. Amorim, A.R. Roel, H. Pistori, Detection and classification of soybean pests using deep learning with uav images, *Computers and Electronics in Agriculture* 179 (2020).
- [50] H.G. Kim, J.-S. Park, D.-H. Lee, Potential of unmanned aerial sampling for monitoring insect populations in rice fields, *Florida Entomologist* 101 (2) (2018) 330–334.
- [51] D. Zhang, X. Zhou, J. Zhang, Y. Lan, C. Xu, D. Liang, Detection of rice sheath blight using an unmanned aerial system with high-resolution color and multispectral imaging, *PLoS one* 13 (5) (2018).
- [52] R. Hall, G. Castilla, J. White, B. Cooke, R. Skakun, Remote sensing of forest pest damage: A review and lessons learned from a canadian perspective, *The Canadian Entomologist* 148 (S1) (2016) S296–S356.
- [53] S. Park, D. Ryu, S. Fuentes, H. Chung, E. Hernández-Montes, M. O'Connell, Adaptive estimation of crop water stress in nectarine and peach orchards using high-resolution imagery from an unmanned aerial vehicle (uav), *Remote Sensing* 9 (8) (2017) 828.
- [54] J. Gago, C. Douthe, R.E. Coopman, P.P. Gallego, M. Ribas-Carbo, J. Flexas, J. Escalona, H. Medrano, Uavs challenge to assess water stress for sustainable agriculture, *Agricultural water management* 153 (2015) 9–19.
- [55] J. Baluja, M.P. Diago, P. Balda, R. Zorer, F. Meggio, F. Morales, J. Tardaguila, Assessment of vineyard water status variability by thermal and multispectral imagery using an unmanned aerial vehicle (uav), *Irrigation Science* 30 (6) (2012) 511–522.
- [56] J. Bellvert, P.J. Zarco-Tejada, J. Girona, E. Fereres, Mapping crop water stress index in a 'pinot-noir'vineyard: comparing ground measurements with thermal remote sensing imagery from an unmanned aerial vehicle, *Precision agriculture* 15 (4) (2014) 361–376.
- [57] H.S. Ndlovu, J. Odindi, M. Sibanda, O. Mutanga, A. Clulow, V.G. Chimonyo, T. Mabhaudhi, A comparative estimation of maize leaf water content using machine learning techniques and unmanned aerial vehicle (uav)-based proximal and remotely sensed data, *Remote Sensing* 13 (20) (2021) 4091.
- [58] A.M. Hasan, F. Soheli, D. Diepeveen, H. Laga, M.G. Jones, A survey of deep learning techniques for weed detection from images, *Computers and Electronics in Agriculture* 184 (2021).
- [59] J. Lambert, H. Hicks, D. Childs, R. Freckleton, Evaluating the potential of unmanned aerial systems for mapping weeds at field scales: a case study with alopecurus myosuroides, *Weed research* 58 (1) (2018) 35–45.
- [60] J. Su, D. Yi, M. Coombes, C. Liu, X. Zhai, K. McDonald-Maier, W.-H. Chen, Spectral analysis and mapping of blackgrass weed by leveraging machine learning and uav multispectral imagery, *Computers and Electronics in Agriculture* 192 (2022).
- [61] I. Sa, Z. Chen, M. Popović, R. Khanna, F. Liebisch, J. Nieto, R. Siegwart, weednet: Dense semantic weed classification using multispectral images and mav for smart farming, *IEEE robotics and automation letters* 3 (1) (2017) 588–595.
- [62] A.N. Veeranampalayam Sivakumar, J. Li, S. Scott, E. Psota, A.J. Jhala, J.D. Luck, Y. Shi, Comparison of object detection and patch-based classification deep learning models on mid-to late-season weed detection in uav imagery, *Remote Sensing* 12 (13) (2020) 2136.
- [63] M. Maimaitijiang, V. Sagan, P. Sidike, S. Hartling, F. Esposito, F.B. Fritsch, Soybean yield prediction from uav using multimodal data fusion and deep learning, *Remote sensing of environment* 237 (2020).
- [64] O. Apolo-Apolo, J. Martínez-Guanter, G. Egea, P. Raja, M. Pérez-Ruiz, Deep learning techniques for estimation of the yield and size of citrus fruits using a uav, *European Journal of Agronomy* 115 (2020).
- [65] L. He, W. Fang, G. Zhao, Z. Wu, L. Fu, R. Li, Y. Majeed, J. Dhupia, Fruit yield prediction and estimation in orchards: A state-of-the-art comprehensive review for both direct and indirect methods, *Computers and Electronics in Agriculture* 195 (2022).
- [66] Y. Bai, C. Nie, H. Wang, M. Cheng, S. Liu, X. Yu, M. Shao, Z. Wang, S. Wang, N. Tuohuti, et al., A fast and robust method for plant count in sunflower and maize at different seedling stages using high-resolution uav rgb imagery, *Precision Agriculture*, pp. 1–23, 2022.
- [67] M.H. Junos, A.S. Mohd Khairuddin, S. Thannirmalai, M. Dahari, Automatic detection of oil palm fruits from uav images using an improved yolo model, *The Visual Computer* (2021) 1–15.
- [68] J. Zhou, D. Yungbluth, C.N. Vong, A. Scaboo, J. Zhou, Estimation of the maturity date of soybean breeding lines using uav-based multispectral imagery, *Remote Sensing* 11 (18) (2019) 2075.
- [69] A.F. Santos, L.N. Lacerda, C. Rossi, L. d. A. Moreno, M.F. Oliveira, C. Pilon, R.P. Silva, and G. Vellidis, Using uav and multispectral images to estimate peanut maturity variability on irrigated and rainfed fields applying linear models and artificial neural networks, *Remote Sensing* 14 (1) (2021) 93.
- [70] S. Moenizade, H. Pham, Y. Han, A. Dobbels, G. Hu, An applied deep learning approach for estimating soybean relative maturity from uav imagery to aid plant breeding decisions, *Machine Learning with Applications* 7 (2022).

- [71] V. Psiroukis, B. Espejo-Garcia, A. Chitos, A. Dedousis, K. Karantzalos, S. Fountas, Assessment of different object detectors for the maturity level classification of broccoli crops using uav imagery, *Remote Sensing* 14 (3) (2022) 731.
- [72] I. Pölönen, H. Saari, J. Kaivosoja, E. Honkavaara, and L. Pesonen, Hyperspectral imaging based biomass and nitrogen content estimations from light-weight uav, in *Remote Sensing for Agriculture, Ecosystems, and Hydrology XV*, vol. 8887, pp. 141–149, SPIE, 2013.
- [73] T. Chen, W. Yang, H. Zhang, B. Zhu, R. Zeng, X. Wang, S. Wang, L. Wang, H. Qi, Y. Lan, et al., Early detection of bacterial wilt in peanut plants through leaf-level hyperspectral and unmanned aerial vehicle data, *Computers and Electronics in Agriculture* 177 (2020).
- [74] K. Neupane, F. Baysal-Gurel, Automatic identification and monitoring of plant diseases using unmanned aerial vehicles: A review, *Remote Sensing* 13 (19) (2021) 3841.
- [75] J. Abdulridha, O. Batuman, Y. Ampatzidis, Uav-based remote sensing technique to detect citrus canker disease utilizing hyperspectral imaging and machine learning, *Remote Sensing* 11 (11) (2019) 1373.
- [76] H. Hoffmann, R. Jensen, A. Thomsen, H. Nieto, J. Rasmussen, T. Friberg, Crop water stress maps for an entire growing season from visible and thermal uav imagery, *Biogeosciences* 13 (24) (2016) 6545–6563.
- [77] Y. Zhou, C. Lao, Y. Yang, Z. Zhang, H. Chen, Y. Chen, J. Chen, J. Ning, N. Yang, Diagnosis of winter-wheat water stress based on uav-borne multispectral image texture and vegetation indices, *Agricultural Water Management* 256 (2021).
- [78] J. Su, M. Coombes, C. Liu, Y. Zhu, X. Song, S. Fang, L. Guo, W.-H. Chen, Machine learning-based crop drought mapping system by uav remote sensing rgb imagery, *Unmanned systems* 8 (01) (2020) 71–83.
- [79] F. López-Granados, M. Jurado-Expósito, J.M. Peña-Barragán, L. García-Torres, Using remote sensing for identification of late-season grass weed patches in wheat, *Weed Science* 54 (2) (2006) 346–353.
- [80] A. Feng, J. Zhou, E.D. Vories, K.A. Sudduth, M. Zhang, Yield estimation in cotton using uav-based multi-sensor imagery, *Biosystems Engineering* 193 (2020) 101–114.
- [81] Q. Wang, S. Nuske, M. Bergerman, S. Singh, Automated crop yield estimation for apple orchards, in: *Experimental robotics*, Springer, 2013, pp. 745–758.
- [82] K.D. Singh, H.S. Duddu, S. Vail, I. Parkin, S.J. Shirliffe, Uav-based hyperspectral imaging technique to estimate canola (brassica napus l.) seedpods maturity, *Canadian Journal of Remote Sensing* 47 (1) (2021) 33–47.
- [83] D. Giordan, M.S. Adams, I. Aicardi, M. Alicandro, P. Allasia, M. Baldo, P. De Berardinis, D. Dominici, D. Godone, P. Hobbs, et al., The use of unmanned aerial vehicles (uavs) for engineering geology applications, *Bulletin of Engineering Geology and the Environment* 79 (7) (2020) 3437–3481.
- [84] A.C. Watts, V.G. Ambrosia, E.A. Hinkley, Unmanned aircraft systems in remote sensing and scientific research: Classification and considerations of use, *Remote Sensing* 4 (6) (2012) 1671–1692.
- [85] Y. Meng, J. Su, J. Song, W.-H. Chen, Y. Lan, Experimental evaluation of uav spraying for peach trees of different shapes: Effects of operational parameters on droplet distribution, *Computers and Electronics in Agriculture* 170 (2020).
- [86] F. Nex, F. Remondino, Uav for 3d mapping applications: a review, *Applied geomatics* 6 (1) (2014) 1–15.
- [87] J. Smith, J. Su, C. Liu, W.-H. Chen, Disturbance observer based control with anti-windup applied to a small fixed wing uav for disturbance rejection, *Journal of Intelligent & Robotic Systems* 88 (2) (2017) 329–346.
- [88] M. Zhou, Z. Zhou, L. Liu, J. Huang, and Z. Lyu, Review of vertical take-off and landing fixed-wing uav and its application prospect in precision agriculture, *International Journal of Precision Agricultural Aviation*, vol. 3, no. 4, 2020.
- [89] R.V. Rossel, V. Adamchuk, K. Sudduth, N. McKenzie, C. Lobsey, Proximal soil sensing: An effective approach for soil measurements in space and time, *Advances in agronomy* 113 (2011) 243–291.
- [90] D. Zhang, Y. Ding, P. Chen, X. Zhang, Z. Pan, D. Liang, Automatic extraction of wheat lodging area based on transfer learning method and deeplabv3+ network, *Computers and Electronics in Agriculture* 179 (2020).
- [91] L. Shen, S. Chen, Z. Mi, J. Su, R. Huang, Y. Song, Y. Fang, B. Su, Identifying veraison process of colored wine grapes in field conditions combining deep learning and image analysis, *Computers and Electronics in Agriculture* 200 (2022).
- [92] L. Shen, J. Su, R. Huang, W. Quan, Y. Song, Y. Fang, and B. Su, Fusing attention mechanism with mask r-cnn for instance segmentation of grape cluster in the field, *Frontiers in plant science*, p. 2528, 2022.
- [93] D. Yi, J. Su, W.-H. Chen, Probabilistic faster r-cnn with stochastic region proposing: Towards object detection and recognition in remote sensing imagery, *Neurocomputing* 459 (2021) 290–301.
- [94] W. Li, Z. Niu, H. Chen, D. Li, M. Wu, W. Zhao, Remote estimation of canopy height and aboveground biomass of maize using high-resolution stereo images from a low-cost unmanned aerial vehicle system, *Ecological indicators* 67 (2016) 637–648.
- [95] D. Li, X. Sun, H. Elkhouchlaa, Y. Jia, Z. Yao, P. Lin, J. Li, H. Lu, Fast detection and location of longan fruits using uav images, *Computers and Electronics in Agriculture* 190 (2021).
- [96] Y. Lu, Z. Xue, G.-S. Xia, L. Zhang, A survey on vision-based uav navigation, *Geo-spatial information science* 21 (1) (2018) 21–32.
- [97] S. Sankaran, L.R. Khot, A.H. Carter, Field-based crop phenotyping: Multispectral aerial imaging for evaluation of winter wheat emergence and spring stand, *Computers and electronics in agriculture* 118 (2015) 372–379.
- [98] H. Zheng, T. Cheng, D. Li, X. Zhou, X. Yao, Y. Tian, W. Cao, Y. Zhu, Evaluation of rgb, color-infrared and multispectral images acquired from unmanned aerial systems for the estimation of nitrogen accumulation in rice, *Remote Sensing* 10 (6) (2018) 824.
- [99] N. Lu, Y. Wu, H. Zheng, X. Yao, Y. Zhu, W. Cao, and T. Cheng, An assessment of multi-view spectral information from uav-based color-infrared images for improved estimation of nitrogen nutrition status in winter wheat, *Precision Agriculture*, pp. 1–22, 2022.
- [100] A. Shirzadifar, S. Bajwa, J. Nowatzki, A. Bazrafkan, Field identification of weed species and glyphosate-resistant weeds using high resolution imagery in early growing season, *Biosystems Engineering* 200 (2020) 200–214.
- [101] L. Deng, Z. Mao, X. Li, Z. Hu, F. Duan, Y. Yan, Uav-based multispectral remote sensing for precision agriculture: A comparison between different cameras, *ISPRS Journal of Photogrammetry and Remote Sensing* 146 (2018) 124–136.
- [102] V. Sagan, M. Maimaitijiang, P. Sidike, K. Eblimit, K.T. Peterson, S. Hartling, F. Esposito, K. Khanal, M. Newcomb, D. Pauli, 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* 11 (3) (2019) 330.
- [103] V. Gonzalez-Dugo, P. Zarco-Tejada, E. Nicolás, P.A. Nortés, J. Alarcón, D.S. Intrigliolo, E. Fereres, Using high resolution uav thermal imagery to assess the variability in the water status of five fruit tree species within a commercial orchard, *Precision Agriculture* 14 (6) (2013) 660–678.
- [104] S. Mader, F. Baret, B. De Solan, S. Thomas, D. Dutartre, S. Jezequel, M. Hemmerlé, G. Colombeau, A. Comar, High-throughput phenotyping of plant height: comparing unmanned aerial vehicles and ground lidar estimates, *Frontiers in plant science* 8 (2017) 2002.
- [105] S. Luo, W. Liu, Y. Zhang, C. Wang, X. Xi, S. Nie, D. Ma, Y. Lin, G. Zhou, Maize and soybean heights estimation from unmanned aerial vehicle (uav) lidar data, *Computers and Electronics in Agriculture* 182 (2021).
- [106] L. Wallace, A. Lucieir, C. Watson, D. Turner, Development of a uav-lidar system with application to forest inventory, *Remote sensing* 4 (6) (2012) 1519–1543.
- [107] J. Torres-Sanchez, A.I. de Castro, J.M. Pena, F.M. Jimenez-Brenes, O. Arquero, M. Lovera, F. Lopez-Granados, Mapping the 3d structure of almond trees using uav acquired photogrammetric point clouds and object-based image analysis, *Biosystems engineering* 176 (2018) 172–184.
- [108] P. Cinat, S.F. Di Gennaro, A. Berton, A. Matese, Comparison of unsupervised algorithms for vineyard canopy segmentation from uav multispectral images, *Remote Sensing* 11 (9) (2019) 1023.
- [109] M.P. Christiansen, M.S. Laursen, R.N. Jørgensen, S. Skovsen, R. Gislum, Designing and testing a uav mapping system for agricultural field surveying, *Sensors* 17 (12) (2017) 2703.
- [110] S. Pourroostaei Ardakani, A. Cheshmehzangi, Reinforcement learning-enabled uav itinerary planning for remote sensing applications in smart farming, *Telecom*, vol. 2, MDPI, 2021, pp. 255–270.
- [111] H.F. Mahmoud, Parametric versus semi and nonparametric regression models, *arXiv preprint arXiv:1906.10221*, 2019.
- [112] J. Su, D. Yi, C. Liu, L. Guo, W.-H. Chen, Dimension reduction aided hyperspectral image classification with a small-sized training dataset: experimental comparisons, *Sensors* 17 (12) (2017) 2726.
- [113] Y. Wang, J. Su, X. Zhai, F. Meng, C. Liu, Snow coverage mapping by learning from sentinel-2 satellite multispectral images via machine learning algorithms, *Remote Sensing* 14 (3) (2022) 782.
- [114] L. Perez and J. Wang, The effectiveness of data augmentation in image classification using deep learning, *arXiv preprint arXiv:1712.04621*, 2017.
- [115] C. Shorten, T.M. Khoshgoftaar, A survey on image data augmentation for deep learning, *Journal of big data* 6 (1) (2019) 1–48.
- [116] E. Olaniyi, D. Chen, Y. Lu, and Y. Huang, Generative adversarial networks for image augmentation in agriculture: a systematic review, *arXiv preprint arXiv:2204.04707*, 2022.
- [117] N. O'Mahony, S. Campbell, A. Carvalho, S. Harapanahalli, G.V. Hernandez, L. Krpalkova, D. Riordan, J. Walsh, Deep learning vs. in: traditional computer vision, in *Science and information conference*, Springer, 2019, pp. 128–144.
- [118] D.G. Lowe, Object recognition from local scale-invariant features, in *Proceedings of the seventh IEEE international conference on computer vision*, vol. 2, pp. 1150–1157, IEEE, 1999.
- [119] H. Bay, A. Ess, T. Tuytelaars, L. Van Gool, Speeded-up robust features (surf), *Computer vision and image understanding* 110 (3) (2008) 346–359.
- [120] H. Li, N. Zeng, P. Wu, K. Clawson, Cov-Net: A computer-aided diagnosis method for recognizing COVID-19 from chest X-ray images via machine vision, *Expert Systems with Applications* 207 (2022).
- [121] C. Hung, Z. Xu, S. Sukkarieh, Feature learning based approach for weed classification using high resolution aerial images from a digital camera mounted on a uav, *Remote Sensing* 6 (12) (2014) 12037–12054.
- [122] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, *Advances in neural information processing systems* 25 (2012).
- [123] Y. LeCun, B. Boser, J.S. Denker, D. Henderson, R.E. Howard, W. Hubbard, L.D. Jackel, Backpropagation applied to handwritten zip code recognition, *Neural computation* 1 (4) (1989) 541–551.
- [124] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, Gradient-based learning applied to document recognition, *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [125] K. Simonyan and A. Zisserman, Very deep convolutional networks for large-scale image recognition, *arXiv preprint arXiv:1409.1556*, 2014.

- [126] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, Going deeper with convolutions, in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1–9, 2015.
- [127] K. He, X. Zhang, S. Ren, and J. Sun, Deep residual learning for image recognition, in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778, 2016.
- [128] G. Huang, Z. Liu, L. Van Der Maaten, and K.Q. Weinberger, Densely connected convolutional networks, in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4700–4708, 2017.
- [129] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, Aggregated residual transformations for deep neural networks, in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1492–1500, 2017.
- [130] M. Tan, Q. Le, Efficientnet: Rethinking model scaling for convolutional neural networks, in: International conference on machine learning, PMLR, 2019, pp. 6105–6114.
- [131] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, et al., An image is worth 16x16 words: Transformers for image recognition at scale, arXiv preprint arXiv:2010.11929, 2020.
- [132] S. Khan, M. Naseer, M. Hayat, S.W. Zamir, F.S. Khan, and M. Shah, Transformers in vision: A survey, ACM Computing Surveys (CSUR), 2021.
- [133] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, Advances in neural information processing systems 30 (2017).
- [134] A.G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, Mobilenets: Efficient convolutional neural networks for mobile vision applications, arXiv preprint arXiv:1704.04861, 2017.
- [135] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, Mobilenetv2: Inverted residuals and linear bottlenecks, in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4510–4520, 2018.
- [136] A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan, et al., Searching for mobilenetv3, in Proceedings of the IEEE/CVF international conference on computer vision, pp. 1314–1324, 2019.
- [137] X. Zhang, X. Zhou, M. Lin, and J. Sun, Shufflenet: An extremely efficient convolutional neural network for mobile devices, in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 6848–6856, 2018.
- [138] N. Ma, X. Zhang, H.-T. Zheng, and J. Sun, Shufflenet v2: Practical guidelines for efficient cnn architecture design, in Proceedings of the European conference on computer vision (ECCV), pp. 116–131, 2018.
- [139] A. Khan, A. Sohail, U. Zahoor, A.S. Qureshi, A survey of the recent architectures of deep convolutional neural networks, Artificial intelligence review 53 (8) (2020) 5455–5516.
- [140] Z. Li, F. Liu, W. Yang, S. Peng, J. Zhou, A survey of convolutional neural networks: analysis, applications, and prospects, IEEE transactions on neural networks and learning systems (2021).
- [141] S. Hao, Y. Zhou, Y. Guo, A brief survey on semantic segmentation with deep learning, Neurocomputing 406 (2020) 302–321.
- [142] M.-D. Yang, K.-S. Huang, Y.-H. Kuo, H.P. Tsai, L.-M. Lin, Spatial and spectral hybrid image classification for rice lodging assessment through uav imagery, Remote Sensing 9 (6) (2017) 583.
- [143] T. Blaschke, Object based image analysis for remote sensing, ISPRS journal of photogrammetry and remote sensing 65 (1) (2010) 2–16.
- [144] G. Fu, C. Liu, R. Zhou, T. Sun, Q. Zhang, Classification for high resolution remote sensing imagery using a fully convolutional network, Remote Sensing 9 (5) (2017) 498.
- [145] T. Liu, A. Abd-Elrahman, J. Morton, V.L. Wilhelm, Comparing fully convolutional networks, random forest, support vector machine, and patch-based deep convolutional neural networks for object-based wetland mapping using images from small unmanned aircraft system, GIScience & remote sensing 55 (2) (2018) 243–264.
- [146] V. Mnih, Machine learning for aerial image labeling, University of Toronto (Canada), 2013.
- [147] M. Långkvist, A. Kiselev, M. Alirezaie, A. Loutfi, Classification and segmentation of satellite orthoimagery using convolutional neural networks, Remote Sensing 8 (4) (2016) 329.
- [148] A. Sharma, X. Liu, X. Yang, D. Shi, A patch-based convolutional neural network for remote sensing image classification, Neural Networks 95 (2017) 19–28.
- [149] S. Bittel, V. Kaiser, M. Teichmann, and M. Thoma, Pixel-wise segmentation of street with neural networks, arXiv preprint arXiv:1511.00513, 2015.
- [150] J. Long, E. Shelhamer, and T. Darrell, Fully convolutional networks for semantic segmentation, in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3431–3440, 2015.
- [151] H. Huang, J. Deng, Y. Lan, A. Yang, X. Deng, L. Zhang, A fully convolutional network for weed mapping of unmanned aerial vehicle (uav) imagery, PloS one 13 (4) (2018).
- [152] V. Badrinarayanan, A. Kendall, R. Cipolla, Segnet: A deep convolutional encoder-decoder architecture for image segmentation, IEEE transactions on pattern analysis and machine intelligence 39 (12) (2017) 2481–2495.
- [153] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical image computing and computer-assisted intervention, Springer, 2015, pp. 234–241.
- [154] T. Zhang, Z. Xu, J. Su, Z. Yang, C. Liu, W.-H. Chen, J. Li, Ir-unet: Irregular segmentation u-shape network for wheat yellow rust detection by uav multispectral imagery, Remote Sensing 13 (19) (2021) 3892.
- [155] A. Paszke, A. Chaurasia, S. Kim, and E. Culurciello, Enet: A deep neural network architecture for real-time semantic segmentation, arXiv preprint arXiv:1606.02147, 2016.
- [156] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, Pyramid scene parsing network, in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2881–2890, 2017.
- [157] Q. Pan, M. Gao, P. Wu, J. Yan, S. Li, A deep-learning-based approach for wheat yellow rust disease recognition from unmanned aerial vehicle images, Sensors 21 (19) (2021) 6540.
- [158] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A.L. Yuille, Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs, IEEE transactions on pattern analysis and machine intelligence 40 (4) (2017) 834–848.
- [159] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, Encoder-decoder with atrous separable convolution for semantic image segmentation, in Proceedings of the European conference on computer vision (ECCV), pp. 801–818, 2018.
- [160] L. Schmarje, M. Santarossa, S.-M. Schröder, R. Koch, A survey on semi-, self- and unsupervised learning for image classification, IEEE Access 9 (2021) 82146–82168.
- [161] F. Panella, A. Lipani, J. Boehm: Semantic segmentation of cracks: Data challenges and architecture, Automation in Construction 135 (2022).
- [162] M. Buda, A. Maki, M.A. Mazurowski, A systematic study of the class imbalance problem in convolutional neural networks, Neural networks 106 (2018) 249–259.
- [163] N. Zeng, P. Wu, Z. Wang, H. Li, W. Liu, X. Liu, A small-sized object detection oriented multi-scale feature fusion approach with application to defect detection, IEEE Transactions on Instrumentation and Measurement 71 (2022) 1–14.
- [164] G. Cheng, J. Han, A survey on object detection in optical remote sensing images, ISPRS Journal of Photogrammetry and Remote sensing 117 (2016) 11–28.
- [165] X. Wu, D. Sahoo, S.C. Hoi, Recent advances in deep learning for object detection, Neurocomputing 396 (2020) 39–64.
- [166] P. Mittal, R. Singh, A. Sharma, Deep learning-based object detection in low-altitude uav datasets: A survey, Image and Vision Computing 104 (2020).
- [167] R. Girshick, J. Donahue, T. Darrell, and J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 580–587, 2014.
- [168] R. Girshick, Fast r-cnn, in Proceedings of the IEEE international conference on computer vision, pp. 1440–1448, 2015.
- [169] K. He, X. Zhang, S. Ren, J. Sun, Spatial pyramid pooling in deep convolutional networks for visual recognition, IEEE transactions on pattern analysis and machine intelligence 37 (9) (2015) 1904–1916.
- [170] S. Ren, K. He, R. Girshick, J. Sun, Faster r-cnn: Towards real-time object detection with region proposal networks, Advances in neural information processing systems 28 (2015).
- [171] J. Dai, Y. Li, K. He, J. Sun, R-fcn: Object detection via region-based fully convolutional networks, Advances in neural information processing systems 29 (2016).
- [172] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, You only look once: Unified, real-time object detection, in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 779–788, 2016.
- [173] J. Redmon and A. Farhadi, Yolo9000: better, faster, stronger, in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 7263–7271, 2017.
- [174] J. Redmon and A. Farhadi, Yolo3: An incremental improvement, arXiv preprint arXiv:1804.02767, 2018.
- [175] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, Feature pyramid networks for object detection, in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2117–2125, 2017.
- [176] A. Bochkovskiy, C.-Y. Wang, and H.-Y.M. Liao, Yolo3: Optimal speed and accuracy of object detection, arXiv preprint arXiv:2004.10934, 2020.
- [177] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, A.C. Berg, Ssd: Single shot multibox detector, in: European conference on computer vision, Springer, 2016, pp. 21–37.
- [178] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, Focal loss for dense object detection, in Proceedings of the IEEE international conference on computer vision, pp. 2980–2988, 2017.
- [179] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, Swin transformer: Hierarchical vision transformer using shifted windows, in Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 10012–10022, 2021.
- [180] B. Neupane, T. Horanont, N.D. Hung, Deep learning based banana plant detection and counting using high-resolution red-green-blue (rgb) images collected from unmanned aerial vehicle (uav), PloS one 14 (10) (2019).
- [181] A.A. d. Santos, J. Marcato Junior, M.S. Araújo, D.R. Di Martini, E.C. Tetila, H.L. Siqueira, C. Aoki, A. Eltner, E.T. Matsubara, H. Pistori, et al., Assessment of cnn-based methods for individual tree detection on images captured by rgb cameras attached to uavs, Sensors, vol. 19, no. 16, p. 3595, 2019.
- [182] L.P. Osco, M. d. S. De Arruda, J.M. Junior, N.B. Da Silva, A.P.M. Ramos, É. A.S. Moryia, N.N. Imai, D.R. Pereira, J.E. Creste, E.T. Matsubara, et al., A

- convolutional neural network approach for counting and geolocating citrus-trees in uav multispectral imagery, *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 160, pp. 97–106, 2020.
- [183] J. Wu, G. Yang, H. Yang, Y. Zhu, Z. Li, L. Lei, C. Zhao, Extracting apple tree crown information from remote imagery using deep learning, *Computers and Electronics in Agriculture* 174 (2020).
- [184] M.-X. He, P. Hao, Y.-Z. Xin, A robust method for wheat ear detection using uav in natural scenes, *IEEE Access* 8 (2020) 189043–189053.
- [185] G. Licciardi, P.R. Marpu, J. Chanussot, J.A. Benediktsson, Linear versus nonlinear pca for the classification of hyperspectral data based on the extended morphological profiles, *IEEE Geoscience and Remote Sensing Letters* 9 (3) (2011) 447–451.
- [186] X. Xi, K. Xia, Y. Yang, X. Du, H. Feng, Evaluation of dimensionality reduction methods for individual tree crown delineation using instance segmentation network and uav multispectral imagery in urban forest, *Computers and Electronics in Agriculture* 191 (2021).
- [187] J.-J. Wang, H. Ge, Q. Dai, I. Ahmad, Q. Dai, G. Zhou, M. Qin, C. Gu, Unsupervised discrimination between lodged and non-lodged winter wheat: a case study using a low-cost unmanned aerial vehicle, *International Journal of Remote Sensing* 39 (8) (2018) 2079–2088.
- [188] J. Su, C. Liu, and W.-H. Chen, Uav multispectral remote sensing for yellow rust mapping: Opportunities and challenges, *Unmanned Aerial Systems in Precision Agriculture*, pp. 107–122, 2022.
- [189] A. dos Santos Ferreira, D.M. Freitas, G.G. da Silva, H. Pistori, M.T. Folhes, Unsupervised deep learning and semi-automatic data labeling in weed discrimination, *Computers and Electronics in Agriculture* 165 (2019).
- [190] M.D. Bah, A. Hafiane, R. Canals, Deep learning with unsupervised data labeling for weed detection in line crops in uav images, *Remote sensing* 10 (11) (2018) 1690.
- [191] W.-H. Chen, J. Yang, L. Guo, S. Li, Disturbance-observer-based control and related methods—an overview, *IEEE Transactions on industrial electronics* 63 (2) (2015) 1083–1095.
- [192] A. Townsend, I.N. Jiya, C. Martinson, D. Bessarabov, R. Gouws, A comprehensive review of energy sources for unmanned aerial vehicles, their shortfalls and opportunities for improvements, *Heliyon* 6 (11) (2020).
- [193] M. Coombes, T. Fletcher, W.-H. Chen, C. Liu, Decomposition-based mission planning for fixed-wing uavs surveying in wind, *Journal of Field Robotics* 37 (3) (2020) 440–465.
- [194] D. Albani, D. Nardi, V. Trianni, Field coverage and weed mapping by uav swarms, in: in 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2017, pp. 4319–4325.
- [195] C. Ju, H.I. Son, Multiple uav systems for agricultural applications: Control, implementation, and evaluation, *Electronics* 7 (9) (2018) 162.
- [196] C. Qu, J. Boubin, D. Gafurov, J. Zhou, N. Aloysius, H. Nguyen, and P. Calyam, Uav swarms in smart agriculture: Experiences and opportunities.
- [197] C. Koparan, A.B. Koc, C.V. Privette, C.B. Sawyer, J.L. Sharp, Evaluation of a uav-assisted autonomous water sampling, *Water* 10 (5) (2018) 655.
- [198] K. Whitehead, C.H. Hugenholtz, Remote sensing of the environment with small unmanned aircraft systems (uass), part 1: A review of progress and challenges, *Journal of Unmanned Vehicle Systems* 2 (3) (2014) 69–85.
- [199] H. Shakhathreh, A.H. Sawalmeh, A. Al-Fuqaha, Z. Dou, E. Almaita, I. Khalil, N.S. Othman, A. Khreishah, M. Guizani, Unmanned aerial vehicles (uavs): A survey on civil applications and key research challenges, *IEEE Access* 7 (2019) 48572–48634.
- [200] H. Le and D. Samaras, Shadow removal via shadow image decomposition, in: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 8578–8587, 2019.
- [201] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, Imagenet: A large-scale hierarchical image database, in: in 2009 IEEE conference on computer vision and pattern recognition, IEEE, 2009, pp. 248–255.
- [202] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C.L. Zitnick, Microsoft coco: Common objects in context, in: *European conference on computer vision*, Springer, 2014, pp. 740–755.
- [203] S.P. Mohanty, D.P. Hughes, M. Salathé, Using deep learning for image-based plant disease detection, *Frontiers in plant science* 7 (2016) 1419.
- [204] Y. Zhong, X. Hu, C. Luo, X. Wang, J. Zhao, L. Zhang, Whu-hi: Uav-borne hyperspectral with high spatial resolution (h2) benchmark datasets and classifier for precise crop identification based on deep convolutional neural network with crf, *Remote Sensing of Environment* 250 (2020).
- [205] I. Sa, M. Popović, R. Khanna, Z. Chen, P. Lottes, F. Liebisch, J. Nieto, C. Stachniss, A. Walter, R. Siegwart, Weedmap: A large-scale semantic weed mapping framework using aerial multispectral imaging and deep neural network for precision farming, *Remote Sensing* 10 (9) (2018) 1423.
- [206] C.-W. Lin, Q. Ding, W.-H. Tu, J.-H. Huang, J.-F. Liu, Fourier dense network to conduct plant classification using uav-based optical images, *IEEE Access* 7 (2019) 17736–17749.
- [207] M.-D. Yang, H.-H. Tseng, Y.-C. Hsu, C.-Y. Yang, M.-H. Lai, D.-H. Wu, A uav open dataset of rice paddies for deep learning practice, *Remote Sensing* 13 (7) (2021) 1358.
- [208] J. Liu, J. Xiang, Y. Jin, R. Liu, J. Yan, L. Wang, Boost precision agriculture with unmanned aerial vehicle remote sensing and edge intelligence: A survey, *Remote Sensing* 13 (21) (2021) 4387.
- [209] N. Mylonas, I. Malounas, S. Mouseti, E. Vali, B. Espejo-Garcia, S. Fountas, Eden library: A long-term database for storing agricultural multi-sensor datasets from uav and proximal platforms, *Smart Agricultural Technology* 2 (2022).
- [210] M. Wang, W. Deng, Deep visual domain adaptation: A survey, *Neurocomputing* 312 (2018) 135–153.
- [211] D. Gogoll, P. Lottes, J. Weyler, N. Petrinic, C. Stachniss, Unsupervised domain adaptation for transferring plant classification systems to new field environments, crops, and robots, in: in 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2020, pp. 2636–2642.
- [212] B. Kellenberger, D. Marcos, D. Tuia, Detecting mammals in uav images: Best practices to address a substantially imbalanced dataset with deep learning, *Remote sensing of environment* 216 (2018) 139–153.
- [213] E.D. Cubuk, B. Zoph, D. Mane, V. Vasudevan, and Q.V. Le, Autoaugment: Learning augmentation policies from data, *arXiv preprint arXiv:1805.09501*, 2018.
- [214] M. Di Cicco, C. Potena, G. Grisetti, A. Pretto, Automatic model based dataset generation for fast and accurate crop and weeds detection, in: in 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2017, pp. 5188–5195.
- [215] N. Lv, H. Ma, C. Chen, Q. Pei, Y. Zhou, F. Xiao, J. Li, Remote sensing data augmentation through adversarial training, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14 (2021) 9318–9333.
- [216] H. Jiang, C. Zhang, Y. Qiao, Z. Zhang, W. Zhang, C. Song, Cnn feature based graph convolutional network for weed and crop recognition in smart farming, *Computers and Electronics in Agriculture* 174 (2020).
- [217] W. Wang, C. Su, Semi-supervised semantic segmentation network for surface crack detection, *Automation in Construction* 128 (2021).
- [218] Z.-H. Zhou, A brief introduction to weakly supervised learning, *National science review* 5 (1) (2018) 44–53.
- [219] S. Han, H. Mao, and W.J. Dally, Deep compression: Compressing deep neural networks with pruning, trained quantization and huffman coding, *arXiv preprint arXiv:1510.00149*, 2015.
- [220] J. Huang, V. Rathod, C. Sun, M. Zhu, A. Korattikara, A. Fathi, I. Fischer, Z. Wojna, Y. Song, S. Guadarrama, et al., Speed/accuracy trade-offs for modern convolutional object detectors, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7310–7311, 2017.
- [221] K. Abdelouahab, M. Pelcat, J. Serot, and F. Berry, Accelerating cnn inference on fpgas: A survey, *arXiv preprint arXiv:1806.01683*, 2018.
- [222] M. Fawakherji, C. Potena, D.D. Bloisi, M. Imperoli, A. Pretto, D. Nardi, Uav image based crop and weed distribution estimation on embedded gpu boards, in: *International Conference on Computer Analysis of Images and Patterns*, Springer, 2019, pp. 100–108.
- [223] D. Ramachandram, G.W. Taylor, Deep multimodal learning: A survey on recent advances and trends, *IEEE signal processing magazine* 34 (6) (2017) 96–108.
- [224] A. Jain, A. Singh, H.S. Koppula, S. Soh, A. Saxena, Recurrent neural networks for driver activity anticipation via sensory-fusion architecture, in: *2016 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2016, pp. 3118–3125.
- [225] N. Audebert, B. Le Saux, S. Lefèvre, Beyond rgb: Very high resolution urban remote sensing with multimodal deep networks, *ISPRS journal of photogrammetry and remote sensing* 140 (2018) 20–32.
- [226] T. Zhang, J. Su, C. Liu, W.-H. Chen, State and parameter estimation of the aquacrop model for winter wheat using sensitivity informed particle filter, *Computers and Electronics in Agriculture* 180 (2021).
- [227] N. Zeng, Z. Wang, W. Liu, H. Zhang, K. Hone, X. Liu, A dynamic neighborhood-based switching particle swarm optimization algorithm, *IEEE Transactions on Cybernetics* (2020).
- [228] X. Jin, L. Kumar, Z. Li, H. Feng, X. Xu, G. Yang, J. Wang, A review of data assimilation of remote sensing and crop models, *European Journal of Agronomy* 92 (2018) 141–152.
- [229] B.H.Y. Alsalam, K. Morton, D. Campbell, F. Gonzalez, Autonomous uav with vision based on-board decision making for remote sensing and precision agriculture, in: in 2017 IEEE Aerospace Conference, IEEE, 2017, pp. 1–12.
- [230] M. Jarman, J. Vesey, and P. Febvre, Unmanned aerial vehicles (uavs) for uk agriculture: Creating an invisible precision farming technology, *White Paper*, July 2016.



Jinya Su holds a Professorship at the School of Automation, Southeast University, China. His research interests mainly include control, filtering, vision and applied artificial intelligence and their real-world applications in robotics systems (e.g., perception and control under uncertain and dynamic environments), where in these areas he has published over 60 journal, conference and book chapter papers. He is now an associate editor for two international journals, editorial board member for three internal journals, and program committee member for five international conferences. He has also organized several special issues on international peer reviewed journals in the areas of Control, AI and Smart Agriculture.



more than 70 referred technical papers, and holds 12 patents.

Xiaoyong Zhu received the B.Sc. and M.Sc. degrees in electrical engineering from Jiangsu University, Zhenjiang, China, in 1997 and 2002, respectively, and the Ph. D degree from the School of Electrical Engineering, Southeast University, Nanjing, China, in 2008. He has been with Jiangsu University since 1999, where he is currently a Professor with the School of Electrical Information Engineering. His current research interests include design and drive control of electric machines with wide-speed range, less rare-earth permanent magnet motor, and multi-port permanent magnet motor. In these areas, he has authored and co-authored



Institution of Engineering and Technology (IET) and Institute of Measurement and Control (IMC). Prof. Li is a Fellow of IEEE and IET.

Shihua Li received the B.E., M.E., and Ph.D. degrees in automatic control from Southeast University, Nanjing, China, in 1995, 1998, and 2001, respectively. Since 2001, he has been with the School of Automation, Southeast University, China, where he is currently a Professor and the Director of Mechatronic Systems Control Laboratory (MSCL). His main research interests include modelling, analysis, and nonlinear control theory with applications to mechatronic systems, including manipulator, robot, alternating current motor, power electronic systems, and others. In these areas, he has published over 200 journal papers including best journal papers from



Sciences Research Council (EPSRC) Established Career Fellowship in developing new control theory for robotics and autonomous systems.

Wen-Hua Chen holds Chair in Autonomous Vehicles with the Department of Aeronautical and Automotive Engineering, Loughborough University, U.K. He is the founder and the Head of the Loughborough University Centre of Autonomous Systems. He is interested in control, signal processing and artificial intelligence and their applications in robots, aerospace, and automotive systems. Dr. Chen is a Chartered Engineer, and a Fellow of IEEE, the Institution of Mechanical Engineers and the Institution of Engineering and Technology, U.K. He has authored or coauthored near 300 papers and 2 books. Currently he holds the UK Engineering and Physical