REVIEW

Deep learning techniques to classify agricultural crops through UAV imagery: a review

Abdelmalek Bouguettaya[1] · Hafed Zarzour[2] (ID) · Ahmed Kechida[1] · Amine Mohammed Taberkit[1]

Abstract

During the last few years, Unmanned Aerial Vehicles (UAVs) technologies are widely used to improve agriculture productivity while reducing drudgery, inspection time, and crop management cost. Moreover, they are able to cover large areas in a matter of a few minutes. Due to the impressive technological advancement, UAV-based remote sensing technologies are increasingly used to collect valuable data that could be used to achieve many precision agriculture applications, including crop/plant classification. In order to process these data accurately, we need powerful tools and algorithms such as Deep Learning approaches. Recently, Convolutional Neural Network (CNN) has emerged as a powerful tool for image processing tasks achieving remarkable results making it the state-of-the-art technique for vision applications. In the present study, we reviewed the recent CNN-based methods applied to the UAV-based remote sensing image analysis for crop/plant classification to help researchers and farmers to decide what algorithms they should use accordingly to their studied crops and the used hardware. Fusing different UAV-based data and deep learning approaches have emerged as a powerful tool to classify different crop types accurately. The readers of the present review could acquire the most challenging issues facing researchers to classify different crop types from UAV imagery and their potential solutions to improve the performance of deep learning-based algorithms.

1 Introduction

Agriculture is the hand that feeds more than seven billion people around the globe, which is first appeared 13,000 years before present in an area between the Tigris and Euphrates rivers, north of what is now called Iraq [32]. It was limited to the only harvesting of spontaneous wild wheat and other plant species using rudimentary means. At that time, there were only a few people to be fed. However, according to the United Nations (UN), the population number will be around 10 billion by 2050 [111], which is a big challenge that faces the farmers to feed all of these mouths. On the other hand, the farmlands are decreasing due to many issues, including urban expansion and desertification. Moreover, the COVID-19 pandemic has provoked a serious threat to economic growth and food security [50]. In order to overcome such issues, we need to develop more reliable solutions to ensure the necessary amount of food with very little human resources.

Over time, many researchers have proposed various innovative techniques to improve agriculture productivity such as greenhouses [64], vertical agriculture [10, 31], and even exploiting new technologies such as satellites and airplanes [16, 69]. The continuous evolution of different sciences such as chemistry, biology, agronomy, and

Abdelmalek Bouguettaya and Hafed Zarzour have contributed equally to this work.

✉ Hafed Zarzour
   hafed.zarzour@univ-soukahras.dz

   Abdelmalek Bouguettaya
   a.bouguettaya@crti.dz

   Ahmed Kechida
   a.kechida@crti.dz

   Amine Mohammed Taberkit
   a.taberkit@crti.dz

[1] Research Centre in Industrial Technologies (CRTI), P.O. Box 64, 16014 Cheraga, Algiers, Algeria

technology, provides to the farmers the ability to know precisely the plants' nutrient and water needs, their diseases. The future is changing and if we look 10–20 years from now the agriculture sector would be very different from the way we perceive it today. Nowadays, farmers are turning to modern technologies that provide accurate and timely information about crop states ensuring productivity increasement. In addition, in the last years, agricultural technologies are attracting investment like never before, even from big companies such as Google and Amazon. These companies do not have the same classical investments vision anymore. They are making big investments in many other fields, including agriculture to increase global food production.

Accurate and timely information about the crop state is an important component to improve crop productivity, which could be done using efficient remote sensing technologies. Moreover, powerful algorithms are crucial to exploit remote sensing data efficiently. Crop classification is one of the fundamental keys in modern agriculture, which aims classify crop and plant types into different categories while defining their spatial distribution. It can help government authorities and farmers to have efficient information about their crops that could be used to improve their abilities of decision-making. Abundant research has been carried out on precise crop classification from satellite-based remote sensing imagery using different machine learning and deep learning algorithms achieving remarkable results [2, 56, 70, 117, 120]. However, they have many drawbacks such as low spatial/temporal resolutions that should have a harmful impact on data quality, and different weather conditions that should make data collection very hard. These limitations could decrease the algorithm performance leading to wrong crop classification. In addition, traditional approaches to classify different crop/plant types from aerial imagery have relied on classical machine learning algorithms, including Support Vector Machine (SVM) [83] and Random Forrest (RF) [106]. These techniques are based on extracting features manually using different features extraction methods, including Local Binary Pattern (LBP), Histogram of Oriented Gradients (HOG), and Scale Invariant Feature Transform (SIFT), to name a few. These properties make traditional methods time-consuming and inefficient in the case of complex data. However, deep learning algorithms have emerged as an interesting solution to overcome these issues.

In recent years, UAV-based remote sensing and deep learning have emerged as new technologies that can play a crucial role in future agriculture and global food productivity. They are new efficient ways that should help farmers to automate many tasks, including plant/crop identification. UAVs provide many advantages over other available remote sensing platforms (Table 1), such as high flexibility, low cost, small size, real-time data acquisition, the best tradeoff between spatial, temporal, and spectral resolution. Moreover, UAVs are non-destructive technologies while analyzing different crops, unlike Unmanned Ground Vehicles (UGVs) that could destruct some plants and impact the field through soil compaction, which has a direct effect on crop productivity. These characteristics make UAVs typical for crop monitoring and classification.

Combining UAVs and deep learning models should provide instantaneous information about crop status, soil type, and disease/pest attack, which were impossible during the past millennia. Precise and automatic crop classification using UAV-based remote sensing imagery and deep learning techniques represent a fundamental task for many smart farming applications, including crop yield estimation [5, 119], crop surveying/monitoring [107, 113], water stress monitoring [28], precise pesticide and liquid fertilizers spraying [35]. These tasks could lead to crop production increasement, cost reduction, and save a lot of precious time providing precious information that should help the farmers to make instant decisions. In the literature, there are a few survey studies that investigated the applications of either UAV or various deep learning algorithms in several fields, including agriculture. Mittal et al. [71] investigated the performance of different deep learning-based object detection algorithms to detect objects from low altitudes, which could be very helpful to understand the concept of recent object detection algorithms. The authors in [51, 98] provide review studies on the applications of deep learning techniques to solve different challenges related to agriculture and food production. Similarly, Gao et al. [28] reviewed the most recent deep learning-based computer vision algorithms to detect crop stress. However, none of these reviews have focused on the use of UAV-based images to perform agricultural tasks. Other studies focused on the use of UAV technologies to achieve different agricultural tasks [36, 115] but they did not investigate the use of deep learning algorithms in detail. To the best of our knowledge, this is the first review that investigates crop classification through the combination of deep learning and UAV.

In this review paper, we are interested in investigating deep learning techniques applied to UAV-based remote sensing images, which should achieve interesting results on the crop classification task. To the best of our knowledge, this is the first review paper that highlights crop classification from UAV imagery using CNN-based algorithms. The main contributions of this paper include:

- Presenting the importance of UAV-based remote sensing technologies and Deep Learning techniques used to improve the agricultural field as well as solve many related issues, especially crop/plant classification.

Table 1 Advantages and drawbacks of the existing remote-sensing platforms for agriculture

Remote sensing technologies	Advantages	Drawbacks
Satellite		Very low spatial resolution
	Very large area coverage	Cloud sensitivity
	Very high spectral resolution	High cost
		Data not available all time
Manned aircraft		Cloud sensitivity
	Large area coverage	High cost
	High spectral resolution	Low spatial resolution
		Affected by weather conditions
Unmanned aerial vehicle	Low cost	Medium area coverage
	High spatial/spectral resolution	Low endurance
	Data available all time whenever we want it	Affected by weather conditions
	Not sensitive to clouds	Require more time than satellite to cover very large areas
	Accurate position	Taught flying laws and regulations
	Difficult areas accessibility	
Unmanned/manned ground vehicle + other on-ground technologies		Very low area coverage
	Very high spatial resolution	Direct impact on the field
		Require very long time to cover even small areas

- Presenting state-of-the-art object-based and pixel-based deep learning techniques to achieve high crop/plant classification performance.
- Helping researchers and farmers to decide what algorithms they should apply accordingly to their studied crops and the used hardware.

The remainder of this paper is organized as follows. In Sect. 2, we present different CNN architectures helping to realize the crop classification task. Advantages of UAVs, as a reliable remote sensing platform, over other available technologies are presented in Sect. 3. Section 4 highlights the object-based and pixel-based crop/plant classification algorithms, while presenting the pros and cons of each one. In Sect. 5, we present different UAV-based datasets and evaluation metrics. Discussions and conclusions are presented in Sects. 6 and 7, respectively.

2 Convolutional neural networks overview

Computer vision is in continuous progress thanks to the tremendous advances in deep learning techniques and hardware processing technologies. Nowadays, deep learning-based computer vision is one of the most powerful tools that can help farmers to solve and facilitate many agriculture-related applications. In this section, we are going to present some deep learning-based techniques that should help with plant/crop classification. Deep Learning is

a machine learning subfield, which is based on Deep Artificial Neural Networks. It has emerged as an efficient solution for many remote sensing image processing issues, including UAV-based images. Among the various deep learning techniques, CNN represents the state-of-the-art deep learning methods for many image processing and computer vision-based applications. Nowadays, due to its powerful capabilities for automatic feature extraction from input images, CNN architecture is widely used in a wide range of image processing applications achieving state-of-the-art results, including object detection [114], face recognition [40], and action recognition [110]. Moreover, these capabilities make CNN an effective architecture to solve many vision-based agricultural issues, including weed detection [7], plant disease identification [95], yield estimation [119], precise crop classification [128].

CNN is a neural network type with a special structure, which gained popularity in the field of image analysis. It could be used to make robots (including UAVs) able of analyzing, processing, and understanding the content of digital images to make decisions according to the extracted information. A typical CNN architecture consists of three main types of layers. A series of convolution and pooling layers repetitively stacked one after the other followed by one or more dense layers (also called fully connected layers) (Fig. 1). The main role of convolution and pooling layers is to find patterns in the input data, whereas the dense layers are responsible for classifying these features into their corresponding category.

A wide range of deep CNN models have been proposed over the last few years to solve different computer vision problems. The majority of deep CNNs follow a simple architecture by concatenating convolutional and pooling layers followed by one or more fully connected layers. Among these models we can find LeNet [59], AlexNet [55], ZFNet [121], and VGGnet [99]. Other CNN architectures are deeper, more complex, and provide better results, including GoogLeNet [103], ResNet [37], Dense-Net [44], Inception V3 [104], Inception V4 [105], Inception-ResNet [105], among others. These aforementioned CNN architectures represent the most used ones in the literature. However, there are other efficient and lightweight architectures proposed just recently. These models have very few trainable parameters making them more suitable for small devices with low computational power achieving better processing speed. Among the lightweight versions we can find the different MobileNet versions [41, 42, 96] and ShuffleNet [65, 124] to name a few. Figure 2, shows the performance evolution of the CNN architectures tested on the famous ImageNet dataset.

Many studies have shown that 2D-CNN-based models can achieve remarkable results in plant/crop classifications from high spatial and spectral resolutions [4, 8, 97, 102]. Some of these models are based on object detection techniques such as YOLOv3 [87] and Faster R-CNN [89]. Others are based on segmentation techniques, including Mask R-CNN [38], SegNet [6], U-Net [90]. The main difference between object detection-based techniques and segmentation-based approaches is that the first one only predicts bounding boxes around each plant of interest in the input image. However, segmentation-based techniques predict an exact mask of each crop or plant. Many other architectures are increasingly being applied for multitemporal crop classification from satellite imagery. Among these methods, one dimensional CNN (Conv-1D), and

Recurrent Neural Network (RNN, including Long Short-Term Memory (LSTM)) was used to capture temporal characteristics [127]. However, only few studies have addressed multitemporal crop classification from UAV imagery, such as the works in [24, 77]. In fact, multitemporal crop classification was beyond the scope of this study. Therefore, the present study has only investigated crop classification, applying 2D-CNN-based techniques, from spatial and spectral viewpoints using RGB and spectral UAV imagery.

3 UAVs as remote sensing platforms

Remote sensing technology, for crop classification, is about measuring crop characteristics by observing them from different distances that could range from few meters to hundreds of kilometers from the targeted object. It has been proved to be one of the most promising technologies for crops and farmlands monitoring. Satellites, manned airplanes, UGVs, and UAVs are among the most reliable remote sensing technologies that provide better performance than conventional methods for crop scouting, where we need most of the time a large number of workers.

Even with the continuous improvements of satellites and airplanes-based remote sensing technologies, there are still limitations concerning spatial/temporal resolution, cost, and smallholders crop classification performance [13]. The same thing for UGVs that provide high spatial resolution images, but they take very long time to cover large farmlands. In the case of small scale and mixed planting agricultural field, very high spatial, temporal, and spectral image resolutions are required to identify crop/plant types all along with large area monitoring in reasonable time ability.
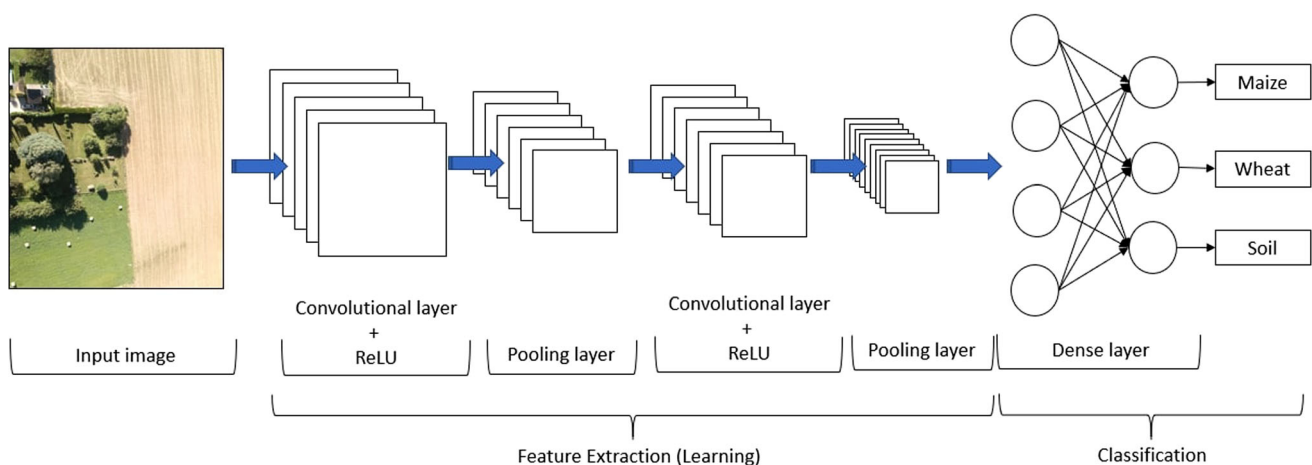

Fig. 1 CNN architecture

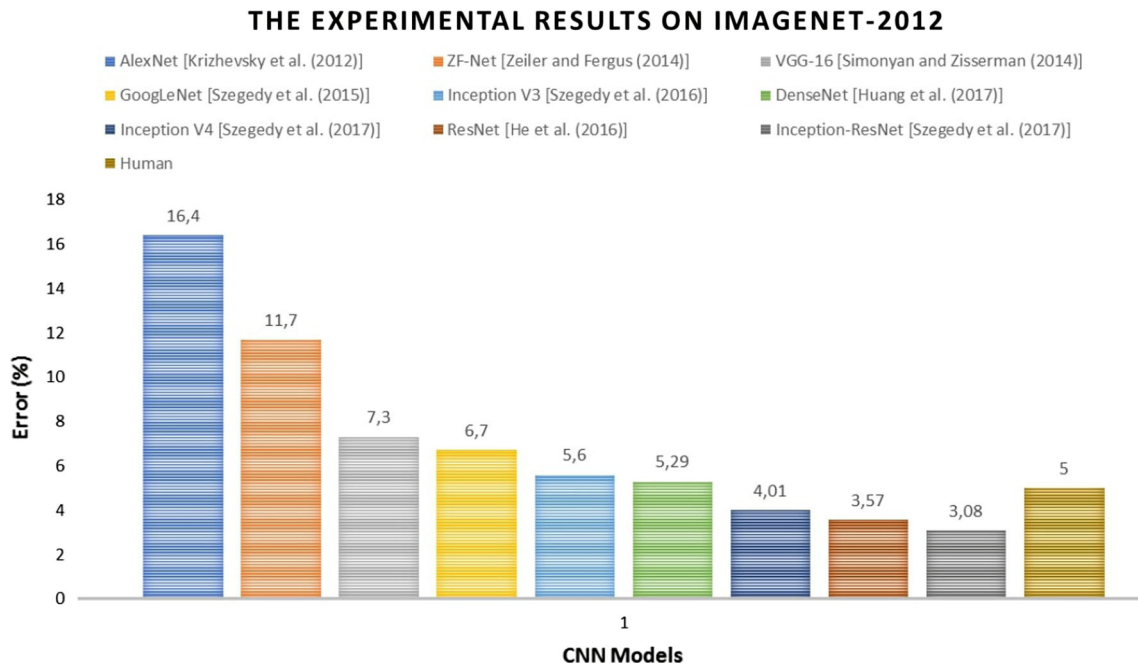## THE EXPERIMENTAL RESULTS ON IMAGENET-2012

Fig. 2 Performance evolution of different Deep CNN architectures for Image net classification challenge

In the last years, UAVs equipped with multi-sensor cameras have received much attention and becoming more and more important as a data acquisition platform for precision agriculture. Like any other robotic system, UAVs perform the three main concepts, which are perception, decision making, and actions performing. In the agricultural field, UAVs can perceive the external world through different sensors, like cameras to understand the world visually, GPS to know its geographical position, barometer to detect altitude. The decision-making depends on applying different algorithms to achieve certain objective. The last concept is to take actions like invasive plant fighting, pesticide spraying, communicate with farmers or other machines.

Todays' technological advancements make UAVs capable of acquiring high-quality images at different band ranges using a variety of image acquisition tools, including RGB, Multispectral, and Hyperspectral camera sensors. Therefore, there are several camera sensor characteristics that should be considered according to the targeted applications such as cost and spatial/spectral resolution. These sensors have shown great potential for various precision agriculture and smart farming applications. Each type of these sensors can sense a portion of the electromagnetic radiation reflectance from agricultural fields, thus providing valuable vegetation characteristics, including the color, the texture, and the reflected radiation in certain wavebands [13, 24].

RGB cameras are the most used sensors in agricultural UAVs due to their properties, including Low price,

lightweight, flexibility, high-spatial resolution, easy data analysis. They provide images in three wavebands, which are Red, Green, and Blue. However, due to various factors, RGB cameras have lower performance than other available cameras to achieve different agricultural tasks, including disease detection and crop classification. They are extremely susceptible to environmental factors such as sunlight angle and shadows, resulting in performance reduction of the developed classifier. Multispectral and hyperspectral imaging systems could be an effective solution to overcome some of the RGB cameras issues. They are less sensitive to environmental conditions making them more suitable and efficient for crop classification. Multispectral and hyperspectral cameras can capture more valuable information that varies from a few wavebands (more than three bands) up to thousands in both visible and NIR spectral regions. However, in addition to their very high prices and hard operation, hyperspectral cameras provide complex data making their processing and analysis complicated. Thermal infrared cameras are other highly used types of sensors for agricultural field monitoring. This sort of camera may be utilized at any time of day or night due to its qualities. Thermal cameras, unlike RGB cameras that detect visible light, are sensitive to infrared spectra, which provide additional information about the plant's health that is not available in the visible spectrum making them the best choice for different crop stress monitoring, including water stress and plant diseases. On the other hand, the Light Detection and Ranging (LiDAR) imaging systems provide 3D data that could be very useful for biomass and

crop height estimation [68, 73]. Similar to hyperspectral cameras, LiDAR systems are very expensive. Figure 3 shows some examples of each type of the aforementioned camera sensors that could be used for crop classification and many other agricultural applications, Zenmuse x3 (Fig. 3a), Mecasense RedEdgeMX (Fig. 3b), Nano-Hyperspec (Fig. 3c), FLIR Vue Pro R (Fig. 3d), and Phoenix LiDAR (Fig. 3e).

4 UAV and deep learning for accurate crop classification

The problem of automatic crop classification and recognition is based especially on classical machine learning techniques, like SVM [19], Decision Trees [26], and Random Forest [15]. However, these techniques facing several limitations in terms of performance. Deep learning algorithms have emerged as an effective solution for many agricultural applications allowing farmers to take important decisions at the right time. They provide several advantages over the classical machine learning algorithms, such as their high ability to extract highly relevant features automatically. Also, recent advances in hardware and software technologies all along with the high availability of data make it possible to train and deploy such powerful techniques. Thus, the use of deep learning algorithms in crop classification tasks has greatly increased in the last

few years, where we find several serious efforts to improve this task using recent deep learning algorithms.

In order to achieve smart and precision agriculture, several studies adopted deep learning algorithms to process satellite images [11, 125]. However, most of the satellite images have low spatial and temporal resolutions making them not suitable to analyze small-scale plants. The low spatial resolution satellite images are mostly caused by the sensor types and the very high altitude that could be hundreds of kilometers, whereas the low temporal resolution is due to the long revisit time and weather conditions that can cause the missing of an entire growing season data [34]. The missed spatial and temporal information can affect the performance of the deep learning model. The continuous advances in remote sensing technologies make satellites and airbornes achieve higher spatial and temporal resolutions. For example, the WorldView-3 satellite provides a relatively high spatial resolution of around 0.3 meters and a revisit frequency of less than 1 day [33, 94]. However, even with these impressive advances, they are still facing several challenges, including the high cost and weather sensitivity which could affect the deep learning model efficiency. Thus, UAV platforms flying at low altitudes can capture more detailed information about crop types and status that were impossible using satellite data. This detailed information giving us the possibility to build more accurate deep learning algorithms that can even detect small pests and diseases at their early stages before their large spreading. Fusing satellite and UAV technologies is

Fig. 3 Examples of different camera sensors **a** RGB, **b** Multispectral, **c** Hyperspectral, **d** FLIR Vue Pro R, and **e** Phoenix LiDAR

(a) (b) (c)

(d) (e)
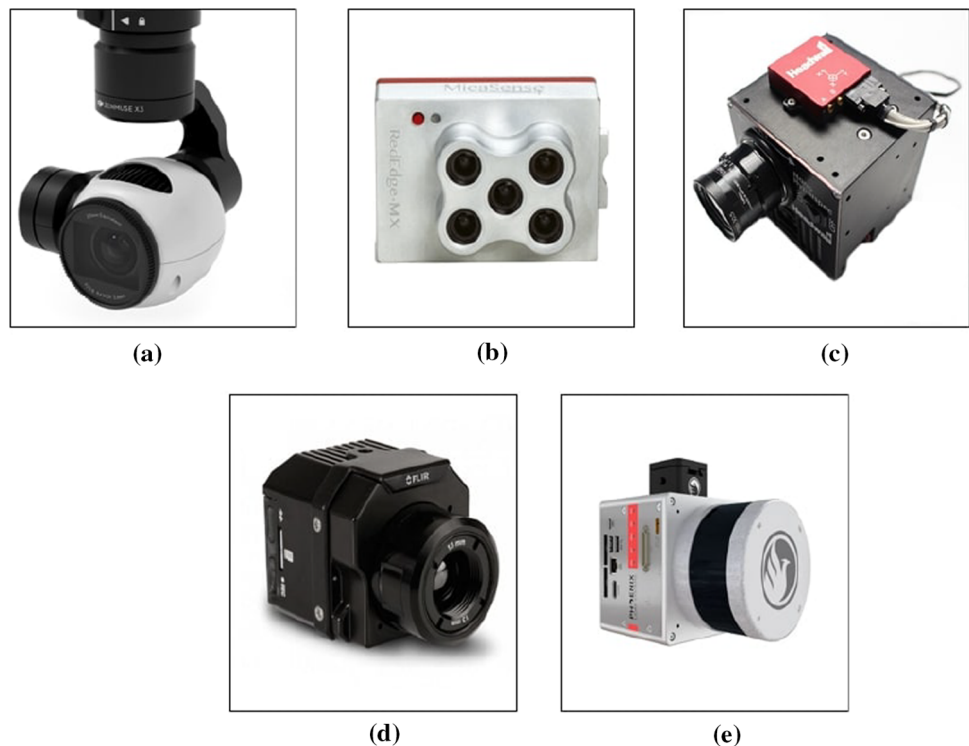
another recent solution that may improve the deep learning model performance further as shown in the studies of [126, 129].

Recently, Deep learning and UAV-based remote sensing technologies are changing the way we are looking to agriculture. They have emerged as new powerful innovative techniques for automatic and accurate crop classification task. UAVs provide to the farmers the ability to inspect small and medium farmlands and crops using different camera sensors. On the other hand, deep learning allows automatic crop classification and decision-making. Deep learning-based automatic crop classification from UAV-based images have been widely used in many agriculture applications, like precise agrochemicals spraying [35], water stress monitoring [28], crop yield estimation [119], among other needs. Recently, many researchers have focused on CNN architectures for large crop and single plant segmentation and detection from UAV imagery.

Deep learning can help UAVs to fly autonomously above agricultural fields and provide instantly information about crop state. Deep learning techniques are becoming more sophisticated, enabling UAVs to understand the real world through images and to classify different crop types precisely. Crop classification is a fundamental task in agricultural field, which can be used to perform many other tasks. Fusing UAV and Deep Learning should help the farmers to monitor their crops, fight invasive plants, estimate the crop productivity. In order to increase crop productivity, farmers are increasingly using different agrochemical products. However, the wrong use of these products is dangerous and have many side effects on the environment and the human health. Deep learning-based precise crop classification methods through UAV-based remotely sensed data has received much attention over the last decade. These techniques can help farmers to reduce the amount of the used chemical products, such as fertilizers, herbicides and pesticides, by applying them at the right place and the right time. This reduction has a direct effect on the overall farming cost.

Recently, there has been considerable interest in crop/plant types classification and identification through UAV imagery using deep learning techniques. Most of the recent studies on crop classification using CNN-based deep learning algorithms and UAVs targeted cash crops, such as legumes and fruits. However, to address food security issues, most countries focus their efforts to increase the major grain crops productivity, which are representing the most consumed food all around the world, especially in the African and Asian countries [80]. To the best of our knowledge, other studies, in a smaller number, targeted the major grain crops such as maize, wheat, and rice. He et al. [39] adopted YOLOv4 to detect wheatears from images acquired by the UAV platform, which could be used for yield estimation. The studies in [118] and [21] targeted the classification of rice crops, which is one of the most important grain crops in Asian countries. Also, several studies [8, 18, 21, 53, 81, 109, 129] targeted the classification of maize/corn crops using different deep learning-based image segmentation. Thus, using deep learning and remote sensing technologies to classify and monitor the major grain crops is a fundamental topic that we should focus on in the future to achieve food security due to their importance to a large number of communities. In this section, we are going to present the state-of-the-art deep learning-based methods for crop classification through UAV imagery.

4.1 Object-based crop classification

Object detection through remote sensing technologies is one of the fundamental tasks in computer vision applications, which aims to identify objects of interest within digital images/videos and placing a bounding box around each one of them (Fig. 4). Object detection algorithms have been widely studied in the last years even from UAV imagery, where CNN-based algorithms achieved the state-of-the-art results. In the last years, CNN-based object detection algorithms represent the sate-of-the-art techniques in image processing and computer vision applications achieving remarkable results even in the agriculture field. Various object detection approaches have been proposed over the few past years, which can be divided into two major classes: - two-stage detectors and - one-stage detectors. The two-stage algorithms mainly include R-CNN [30] and its variants, including Fast R-CNN [29], and Faster R-CNN [89]. Two-stage algorithms split the detection framework into two parts. The first stage is responsible for candidate regions generation, and the second one is about making classification predictions for each of these regions. These algorithms achieved remarkable results in terms of accuracy. However, the inference speed still very limited making them not suitable for real-time applications, especially for small devices with low computation power, such as UAVs.

In order to improve detection speed and memory efficiency, one-stage methods were proposed. The idea behind one-stage detectors is that the algorithm only needs to look to the input image once making the detection faster. Redmond et al. [88] introduced the first real-time object detection algorithm called You Only Look Once (YOLO), which is improved in YOLOv2 [86], and YOLOv3 [87]. Other popular one-stage detectors were proposed over the time, including SSD [62] and RetinaNet [61]. Bochkovskiy et al. [12] proposed YOLOv4 achieving the state-of-the-art accuracy of 43.5% on the challengings MS COCO dataset while keeping a high interference speed of 65 frame per
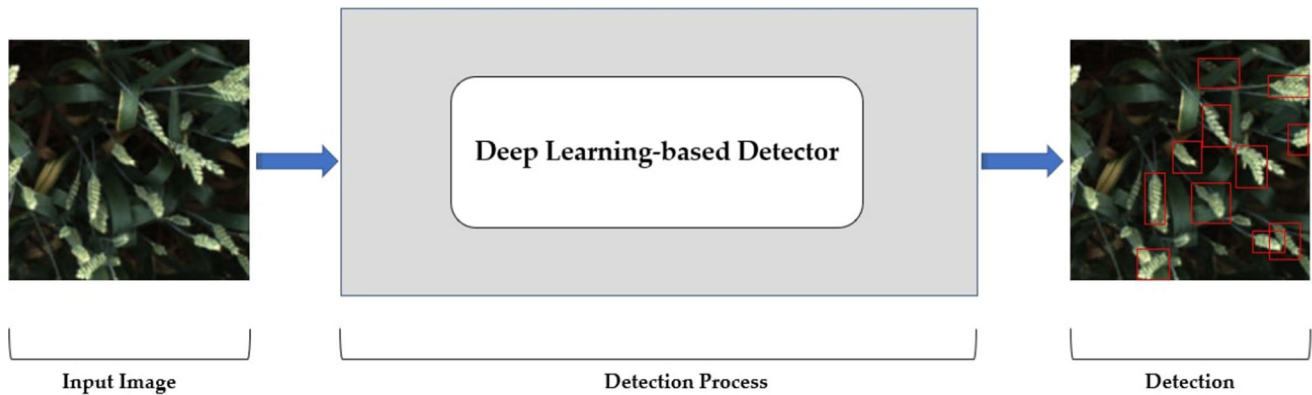
Fig. 4 Example of wheat detection framework from UAV imagery.

second (fps) on Tesla V100 GPU making it the optimal detection and speed solution. Just after the release of YOLOv4, Glenn Jocher released an open source implementation of YOLOv5 on his GitHub [48]. YOLOv5 achieved higher inference speed than YOLOv4 of 140 FPS on Tesla P100 GPU [75]. CNN-based object detection algorithms have realized huge enhancement in terms of accuracy and speed overcoming traditional techniques [9]. These algorithms are mostly used for per-plant crop classification, such as tree and individual plant/fruit detection [9, 54, 116]. Table 2 summarizes object-based crop/plant classification methods that are available in the literatures.

In [9], the authors investigated the performance of two state-of-the-art object detection algorithms (YOLOv3 and RetinaNet) to detect and count ornamental plants from UAV-based RGB images. As shown in Table 2, YOLOv3 outperformed RetinaNet detector achieving a mean average precision (mAP) of around 80% and 73%, respectively.

Other studies have investigated the performance of object-based classification algorithms on different trees detection, where three state-of-the-art CNN-based detectors were evaluated in [97] for individual tree detection from UAV-based high-resolution RGB images. RetinaNet provided the best average precision of 92.64% against 85.88% and 82.48% for YOLOv3 and Faster R-CNN, respectively. Similarly, Csillik et al. [20] utilized CNN architecture followed by a Simple Linear Iterative Clustering (SLIC) algorithm [1] for classification refinement to identify citrus crops from UAV multispectral imagery. They were able to achieve an F1-Score, Precision, and Recall rates of 96.24%, 94.59%, and 97.94%, respectively. Also, Ampatzidis and Partel [4] used two CNN-based algorithms to detect and count citrus trees automatically in large multispectral UAV imagery. They achieved remarkable precision and recall rates of 99.9% and 99.7%, respectively, using YOLOv3 as the main detector. Another study targeted apple tree detection using Faster R-CNN

was adopted in the work of Wu et al. [112], where they achieved, also, good results.

Several researches have adopted object detection algorithms to classify different fruit types. The performance of YOLOv2 was investigated in [116] for mango fruits detection from UAV-based color images in real-time. Remarkable results were achieved, even in various lightening conditions, achieving a precision and recall rates of 96.1% and 89%, respectively, on a dataset that was collected from a distance of 1.5–2 meters. Another study that targeted mango fruits detection in trees, but not from UAV imagery, were presented by Koirala et al. [54]. They proposed "MangoYOLO" architecture that combines YOLOv3 and YOLOv2-tiny properties to detect mango fruits (at night) in trees from a distance of 2 m for fruit load estimation achieving an overall accuracy of 98.3%. However, the used detectors suffer to detect small mango fruits from farther distances due to the YOLO architecture that has limitation to detect small objects [87]. Also, the MangoYOLO detector provides real-time detection on a large on-ground vehicle that can carry sophisticated hardware for the processing task. Unfortunately, it is not suitable to be implemented on UAV platforms that have a limited computational power. In order to detect and count banana plant from high-resolution UAV-based RGB images, Neupane et al. [76] adopted Faster R-CNN based on Inception-V2 as the main features extractor. The results in Table 2 showed that their proposed approach provides an acceptable detection performance achieving F1-scores of 97.82%, 91.05%, and 85.67% from different altitudes of 40, 50, and 60 meters, respectively. Also, they achieved an even better F1-Score of 98.64% by fusing data collected from different altitudes.

Improvements on YOLOv4 algorithm was made by He et al. [39] to detect small wheatear in dense fields from UAV-based aerial images. They achieved remarkable results even in different environment conditions, such as different colors at different growing stages, light changes,

Table 2 Different object-based crop/plant classification methods

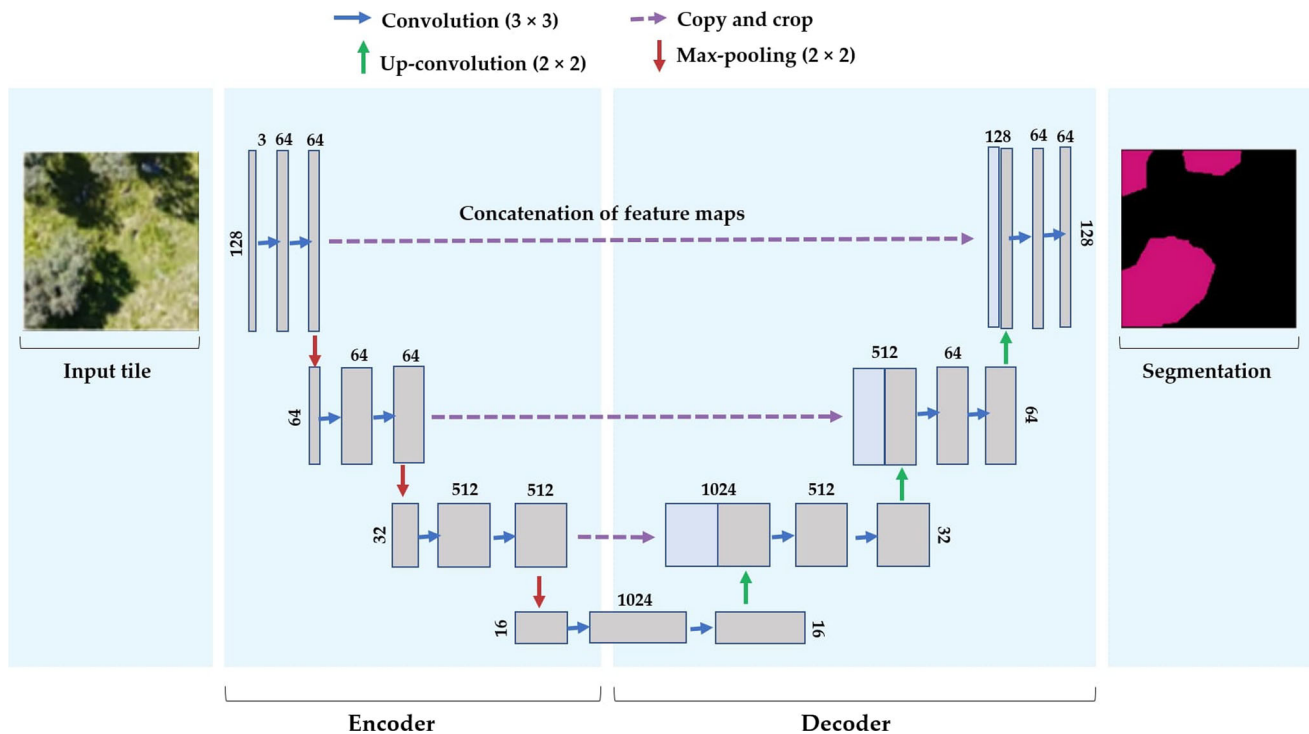| Ref.	Method	Backbone	Plant types	Sensor	GSD (cm)	Altitude (m)	Average Precision (%)	Precision (%)	Recall (%)	F1-Score (%)	Inference time (ms)	FPS
[76]	Faster R-CNN	Inception-v2	Banana plant	RGB	1.78	40		99.3	96.4	97.82		
					2.03	50		97.9	85.1	91.05		
					2.54	60		98.5	75.8	85.67		
						40 + 50		98.3	99	98.64		
						40 + 50 + 60		97.9	98.6	98.24		
[9]	RetinaNet	ResNet	Ornamental plants	RGB			73.41					
	YOLOv3	DarkNet-53					79.85					
[116]	YOLOv2		Mango fruits	RGB		1.5 − 2	86.4	96.1	89	92.41	80	40
[54]	MangoYOLO	/	Mango fruits	RGB		2	98.3			96.8	15	14
	YOLOv2	DarkNet-19					95.9			93.3	20	
	YOLOv2-tiny	/					95.3			91.7	10	
	YOLOv3	DarkNet-53					96.7			95.1	25	
	SSD	VGG-16					98.3			95.9	70	
	Faster R-CNN	VGG					95.3			94.5	67	
	Faster R-CNN	ZF					95			93.9	37	
[97]	Faster R-CNN	ResNet-50	Trees	RGB	0.82	20 − 40	82.48				163	
	YOLOv3	DarkNet-53					85.88				26	
	RetinaNet	FPN					92.64				67	
[39]	YOLOv4		Wheatears	RGB	0.01 − 0.06	1.2 − 3	62.75			88.23		57
	Modified YOLOv4						77.81			96.71		72
[4]	YOLOv3	DarkNet-53	Citrus trees	Multi-spectral		75		99.9	99.7	99.79		
[20]	CNN + Classifier (with refinement)		Citrus trees	Multi-spectral		104		94.59	97.94	96.24		
	CNN + Classifier (without refinement)							65	98	78		
[78]	RetinaNet		Citrus trees	Multi-spectral	12.9	120		62	92	74		
	Faster R-CNN							86	39	54		
	Proposed approach	VGG-16						95	96	95	40 (8 stages)	25
[112]	Faster R-CNN		Apple trees	RGB			91.1		94.1	92.5		

Fig. 5 Example of semantic segmentation from UAV imagery using U-Net architecture [52].

overlap. The improved version of YOLOv4 overcame the original YOLOv4 architecture providing an F1-Score and average precision of 96.71% and 77.81% against 88.23% and 62.75%, respectively.

These detectors could be very effective for plant detection and counting when they have some distance between them. However, they should have limitation in high-density planting crops due to the complex features. In order to overcome some bounding box limitations in counting number of trees in crowdy images, the authors in [78] presented another CNN-based approach to identify and estimate the number of citrus trees in dense fields from multispectral UAV imagery. They compared their proposed method with two state-of-the-art bounding box-based object detection algorithms, which are Faster R-CNN and RetinaNet. The proposed approach applied CNN architecture to estimate a confidence map that gives the probability of occurrence of a plant at each pixel instead of using rectangular bounding box, which could affect the performance of detection in crowdy images. The proposed approach achieved an F1-Score of 95% against 74% and 54% of RetinaNet and Faster R-CNN, respectively.

Several review papers have investigated the performance of object detection algorithms [14, 47, 108] showing that these algorithms can achieve good results in different domains. Unfortunately, they are limited to estimate the objects locations drawing bounding boxes around each of them, which could affect the detection performance in

crowdy images [14, 47, 108]. Moreover, they did not perform a precise location of the object, which is essential for some applications where we need more precision like vegetation harvesting or mechanical weed fighting. In this regard, a more precise algorithms will be presented in the next section, which are based on image segmentation techniques.

4.2 Pixel-based crop classification

After the big success of CNN architectures in object detection task, many researchers adopted them for more challenging and efficient tasks, such as semantic and instance segmentation. Semantic segmentation of UAV imagery has emerged as an effective pixel-based method to monitor agriculture fields and classify different crop types within these fields. Unlike object detection algorithms, where we need to draw a rectangular bounding box around each object of interest, semantic segmentation algorithms aim to classify each pixel in an input image to a specific class resulting a mask or a heatmap that identifies different crop classes (Fig. 5). Thus, semantic segmentation could be more powerful tool than object detection, which aims, in our case, to improve crop/plant classification performance. However, they are much more complex than bounding box-based object detection algorithms and consume a lot of time in the phase of data labeling, where we need to

Table 3 Different pixel-based crop/plant classification methods

Ref.	Data types	GSD (cm)	Altitude (m)	Method	Backbone	Target crops	Overall Accuracy (%)	Precision (%)	Recall (%)	Kappa	F1-Score (%)	Time (s)
[118]	RGB	5.3	230	SegNet	VGG-16	Rice paddy, Rice lodging		83.10	69.06		75.43	101
	RGB+ExG							71.03	89.64		79.26	108
	RGB+ExGR							73.96	83.36		78.38	109
	RGB+ExG+ExGR							87.66	57.38		69.36	106
	RGB			FCN	AlexNet			84.73	82.43		83.56	59
	RGB+ExG							84.92	80.85		82.84	65
	RGB+ExGR							77.02	88.80		82.49	66
	RGB+ExG+ExGR							82.02	84.44		83.21	72
	RGB	5.7	200	SegNet	VGG-16	Rice paddy, Rice lodging		87.55	38.65		53.63	99
	RGB+ExG							57.06	67.50		61.84	109
	RGB+ExGR							81.35	58.59		68.12	107
	RGB+ExG+ExGR							82.47	29.07		42.99	113
	RGB			FCN	AlexNet			99.12	39.59		56.58	57
	RGB+ExG							95.03	59.18		72.94	67
	RGB+ExGR							95.32	66.39		78.27	68
	RGB+ExG+ExGR							93.19	67.03		77.97	71
[102]	RGB	6.4	150	FCN		Sunflower	78					
				SegNet			79					
				Improved SegNet	VGG-16		81.95					
	Multispectral	8.1		FCN			77.45					
				SegNet			78.25					
				Improved SegNet	VGG-16		80.5					
	Fusion (RGB + FNIR)	9.5		FCN			81.55					
				SegNet			82.65					
				Improved SegNet	VGG-16		86.55					
[57]	RGB	9.5	20	FCRN-MTL		Full-grown citrus trees	98.8	99.2	98.4		98.8	
						Citrus tree seedlings	56.6	47.8	65.4		55.23	
[27]	RGB		20	Deep convolutional Encoder-decoder network		Fig Plant	93.84	92	95.05		93.50	
				SegNet-basic			93.82	93.49	93.22		93.35	
[53]	RGB		20	U-Net		Corn	99.4					
[8]	RGB		10	FCN		Beet		49.69	26.25		32.15	
				SegNet				80.24	67.4		71.79	

Table 3 (continued)

Ref.	Data types	GSD (cm)	Altitude (m)	Method	Backbone	Target crops	Overall Accuracy (%)	Precision (%)	Recall (%)	Kappa	F1-Score (%)	Time (s)
[21]	RGB			CR-Hough-SLIC				86.58	85.67		85.40	
				CRowNet				90.37	90.56		90.39	
				CR-Hough-SLIC		Maize		85.14	97.74		82.13	
				CRowNet				84.57	80.93		82.5	
[23]	Multispectral	2.21		SegNet	VGG-16	Rice, Corn	89.44					
				FCN	AlexNet		88.48					
				U-Net	VGG-16	Sugar beet	99					
				SegNet			98					
[112]	RGB			U-Net		Apple tree crown	97.1	84.2	84.5		84.2	
[123]	RGB	0.37, 0.56, 0.74	20	U-Net		Purple rapeseed leaves		∼94	∼90		91.56	
[52]	RGB	3	100	U-Net		Pinus radiata, Ulex europaeus	85.5					
[66]	RGB	1.7–2	40–60	Mask R-CNN		Potato		99.72	82.50		90.3	
						Lettuce		100	95.43		97.7	
[81]	RGB	1		Mask Scoring R-CNN		Maize V5 growth stage	95.8	82.8				
	Multispectral	2.5				Maize V4 growth stage	82.7	79.9				
[25]	RGB	4	120	DeepLabv3+	ResNet-18	Amazonian palms		86	88		87	
[74]	RGB			DeepLabv3+		Mauritia flexuosa	98.036	96.688	95.616		96.14	
				U-Net1			95.973	91.381	92.632		92	
				U-Net2			97.682	94.858	95.953		95.4	
				U-Net3			96.843	92.534	94.886		93.7	
				U-Net4			97.512	95.166	95.028		95.1	
[128]	Hyperspectral	46.3	500	CNN-CRF		WHU-Hi-LongKou: 6 crop types	98.91			0.9857		
		10.9	250			WHU-Hi-HanChuan: 7 crop types	93.95			0.9290		
		4.3	100			WHU-Hi-HongHu: 17 crop types	93.74			0.9217		
[60]	RGB	0.47, 0.90, 1.43, and 1.76		FDN-92		12 plants	87					
				FDN-29			84					
				FDN-17			86					
				Inception-V1			80					

Table 3 (continued)

Ref.	Data types	GSD (cm)	Altitude (m)	Method	Backbone	Target crops	Overall Accuracy (%)	Precision (%)	Recall (%)	Kappa	F1-Score (%)	Time (s)
				Inception-V2			79					
				Inception-V3			77					
				ResNet-17			73					
				ResNet-50			74					
				ResNet-101			76					
				DenseNet-21			82					
				DenseNet-36			80					
				DenseNet-121			82					
[43]	Hyperspectral			CNN		19 crop types	88.62			0.8557		
				CNN-CRF			91.79			0.8957		
[82]	RGB	25		2D-CNN		Highland Kimchi cabbage, Cabbage, Potato	86.56					
[18]	RGB	3		DNN	VGG-16	Maize, Bananas, Legumes	86	86	86		86	
[109]	RGB + Multi-spectral		100	LeNet	LeNet	Corn	67.7		46.4	0.82		
			180				86.8		81.8			
			100–180				72.6		54.5			
			100				47.7		12.8			
[85]	RGB	5		Hybrid CNN-HistNN		22 crops						
[129]	UAV			ANN		Peanut, Maize, Honeysuckle, Tree	78.53			0.73	90	
	Fused						85.72			0.81		

annotate each single pixel into the corresponding category [8].

The main architecture of semantic segmentation algorithms consists of two parts, an encoder network (compression path) followed by a decoder network (expansion path). The encoder part is responsible for feature extractions from the input image using the aforementioned CNN architectures, like VGGnet and ResNet, without the fully connected layers. On the other hand, the decoder 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 to define each crop type. The state-of-the-art semantic segmentation techniques for crop types classification from UAV imagery are highlighted in this section.

During the last years, several efficient image segmentation techniques were proposed to classify crop types, and still evolving more and more with the advance of CNN architectures. The authors from the reviewed papers adopted different CNN-based image segmentation algorithms. J. Long et al. [63] proposed the first CNN-based image segmentation technique called Fully Convolutional Network (FCN), that uses convolution and deconvolution layers to produce semantic mask of each class in the input image. However, FCN architecture has some limitations due to the max-pooling layers that reduce spatial resolution, which could lead to lose some valuable information about different crops or plants [123]. Thus, more sophisticated algorithms were proposed over the past few years, including U-Net [90], SegNet [6], DeepLab [17], which are providing good results for crop classification. Table 3 summarizes the available CNN-based algorithms in various studies that target pixel-based crop classification from UAV imagery.

Several studies have already presented the potential of various segmentation architectures for various crop classification. The authors in [21, 118] adopted two semantic segmentation algorithms (SegNet and FCN) for rice crop identification from RGB images acquired through UAV platform. As shown in Table 3, the performance of SegNet was slightly outperformed FCN achieving good results. However, the adopted procedure in [21] for data collection and transmission to the on-ground server for processing task could cause some problems, including loss of data during wireless transmission in the case of bad 4G connectivity, which is the case in many rural areas. Moreover, in [102], the authors compared FCN and SegNet algorithms with their improved SegNet version for sunflower crop classification from UAV imagery achieving an overall accuracy, between two studied fields, of 78%, 79%, and 81.95%, respectively, on RGB images, while achieving 77.45%, 78.25%, and 80.5% on RGB+NIR imagery. Also, they fused RGB and FNIR bands to improve the classification performance achieving better results of 81.55%, 82.65%, and 86.55% using FCN, SegNet, improved SegNet, respectively, while keeping real-time crop classification of 83 frame per second for the proposed approach. The authors used a combination of three approaches, including skip connection, Convolutional Conditional Random Field (ConvCRF), and deep-wise separable convolutions, to improve SegNet performance. To solve the vanishing problem, the skip connection was implemented. The ConvCRF is used to reduce processing complexity and minimize the noise in sunflower lodging. They also used deep-wise separable convolutions instead of generic convolution procedures to boost model speed while reducing computational power.

Bah et al. [8], proposed CRowNet architecture that combines the characteristics of SegNet and Hough transform (HoughCNet) to detect and highlight crop rows from UAV imagery. The proposed technique shows a clear advantage over the other applied techniques achieving an F1-Score of 90.39% and 82.5% for sugar beet classification, while FCN and SegNet achieve only 32.15% and 72%, respectively (Table 3). In [27], the authors focused on fig crop segmentation from UAV-based RGB images, where they developed a Deep Convolutional Encoder-Decoder Network inspired from SegNet architecture. The encoder part employed a shallow CNN architecture with seven learnable layers, four of which are convolution layers and the other three are pooling layers. The decoder, on the other hand, has three upsampling and three pooling layers. Because there are only two classes, the softmax function was replaced with a sigmoid function to estimate the probability of each pixel's class. Because of its simplicity, the proposed SegNet version may reach faster processing rates. Moreover, they achieved slightly better classification performance than SegNet-Basic achieving an overall accuracy of 93.84% and 93.82%, respectively.

Other studies adopted U-Net as the main architecture to classify different crops from UAV imagery [23, 52, 53, 112]. As shown in Table 3, Wu et al. [112] achieved good results using Faster R-CNN algorithm to detect apple trees from UAV imagery. Moreover, they adopted U-Net architecture, as a semantic segmentation algorithm, to classify each pixel of the detected trees inside the generated bounding boxes achieving an overall accuracy of 97.1%. However, the adopted process is very costly in terms of time and computation power. Another example of using U-Net architecture is the one proposed by Kattenborn et al. [52], where they adopted it to predict plant species accurately from high-resolution UAV-based RGB images. They fed a (128×128) image to a U-Net architecture to identify different types of trees. The input image is passed through a series of (3×3) convolutional and (2×2) max-pooling layers to extract relevant features. The

Table 4 UAV-based datasets

Ref.	Dataset	Data type	Number of images	Image size	Altitude/ GSD (m)	Crop type	Acquisition time
	WHU-Hi-HongHu			550×400	500 / 0.463	Corn, cotton, sesame, broad-leaf soybean, narrow-leaf soybean, and rice	July 17, 2018, from 13:49 to 14:37
[128]	WHU-Hi-HanChuan	Hyperspectral	–	1217×303	250 / 0.109	Strawberry, cowpea, soybean, sorghum, water spinach, watermelon, and greens	June 17, 2016, from 17:57 to 18:46
	WHU-Hi-HongHu			940×475	100 / 0.043	17 crop types, including cotton, rape, and cabbage	November 20, 2017, from 16:23 to 17:37
[93]	Weedmap (000)	Multispectral	10,196	5995×5854	10 / 1.04	Sugar beet & weed	18 September 2017, 9:18–40 a.m.
	Weedmap (001)			4867×5574	10 / 0.94		
	Weedmap (002)			6403×6405	10 / 0.96		
	Weedmap (003)			5470×5995	10 / 0.99		
	Weedmap (004)			4319×4506	10 / 1.07		
	Weedmap (005)			7221×5909	10 / 0.85		5–18 May 2017, around 12:00 p.m.
	Weedmap (006)			5601×5027	10 / 1.18		
	Weedmap (007)			6074×6889	10 / 0.83		
[60]	VOAI	RGB	2293 (No data augmentation)	224×224	20 / 0.47	12 tree species	–
			21,863 (with data augmentation)		30 / 0.90		
					40 / 1.43		
					50 / 1.76		
[92]	WeedNet	Multispectral	465	–	2 / –	Sugar beet & weed	–

resulting feature maps are then fed to the encoder part that consists of (2×2) up-convolution (also called transposed convolution) and (3×3) convolutional layers along with the concatenation operation. Unlike SegNet, the whole feature maps are transferred from the encoder part to the decoder part in the U-Net architecture making use of the concatenation concept. However, this increases the size of the model and necessitates more memory. The concatenation concept represents the main benefit of U-Net architecture by giving it the ability to localize the targeted object, which is in our case the crop type. Its principal purpose is to conserve information that might otherwise be lost as the convolution blocks shrink the input image. As shown in Figure 5, the output of U-Net architecture produces pixel-

based maps that classify each pixel that represents the targeted type of trees.

DeepLabv3+ was adopted in [74] and [25] to identify Mauritia flexuosa and different Amazonian palms, respectively, from UAV RGB images. Fuentes-Pacheco et al. [27] proposed an end-to-end Deep Convolutional Encoder-Decoder Network, which can accurately perform semantic segmentation of fig plant from UAV imagery. Also, they compared their model with another state-of-the-art semantic segmentation network (SegNet). The experimental results showed that their proposed approach provides slightly better performance than SegNet architecture achieving an overall accuracy of 93.84% against 93.82% of SegNet (Table 3). In [57], the authors developed Fully Convolutional Regression Networks and Multi-Task

Learning (FCRN-MTL) to detect citrus trees from UAV imagery, where they achieved an overall accuracy of 98.8% for full-grown trees classification. However, they achieved only an overall accuracy of 56.6% for tree seedlings classification, which could be due to its small size from high altitude. Some studies adopted instance segmentation algorithm called Mask R-CNN [66, 81] which is an extended version of Faster R-CNN algorithm combined with FCN algorithm producing a pixel mask for each bounding box locating each pixel that belong to the object.

Other approaches based on CNN were also proposed in the literatures, such as the work of Rebetez et al. [85], where they proposed a hybrid architecture that combines deep CNN with an RGB histogram (HistNN) to classify different crops distributed in small parcels. They were able to classify 22 crop types efficiently from RGB images acquired through UAV platform achieving F1-scores of around 84% and 87%, respectively; using CNN and HistNN separately. However, by combining CNN and HistNN, they were able to achieve better results with an F1-score of 90%. Also, the authors in [109] applied LeNet architecture to identify corn crops from UAV imagery. In [18], the authors applied CNN-based algorithm and transfer learning to identify some strategic crop types in Rwanda. They used a pretrained VGG-16 network, without the dense layers, for feature extraction followed by one-layer neural network, dropout, and Softmax layers to classify each of the crop types. They achieved remarkable results of 96% and 90% on the classification of bananas and maize crops, respectively. However, their approach seems to have some limitations to classify intercropping such as legumes achieving only an F1-Score of 49%, which affect the overall accuracy decreasing it to 86%. The classification accuracy reduction could be affected by the diversity of legumes types in the images.

5 Available UAV-based datasets and evaluation metrics

Since most datasets are not created to fit specific scenarios, to provide UAVs better computer vision capabilities to achieve agricultural tasks, large aerial image datasets are crucial to meet the deep learning models' training requirements. In the agriculture field, the lack of large datasets that consist of aerial images collected from UAVs to train deep learning models is a serious problem facing farmers, researchers, and developers. Several methods were adopted to overcome such a problem, including the different techniques of data augmentation. However, even with such techniques, it still not sufficient in many cases. In the literature, some available public datasets can be used as a benchmark to evaluate the developed models. Most of them do not contain aerial images such as the famous PlantVillage dataset [72]. Therefore, in this section, we aim to provide readers with the available UAV-based datasets of different crops all along with the methods used to evaluate the performance of the trained models.

5.1 UAV-based datasets

Wuhan UAV-borne hyperspectral image (WHU-Hi) dataset: The Wuhan UAV-borne hyperspectral image (WHU-Hi) dataset was built and published in 2020 as a benchmark dataset by Zhong et al. [128] to train and evaluate deep learning models for the crop classification task. Hyperspectral images acquired through a Headwall Nano-Hyperspec sensor mounted on two UAV types (DJI Matrice 600 Pro and Leica Aibot X6 UAV V1) were collected at different locations and altitudes, where the images at different locations were kept separate to build different UAV-based datasets for each location. These datasets were named after their corresponding location as follows: WHU-Hi-LongKou, WHU-Hi-HanChuan, and WHU-Hi-HongHu. As shown in Table 4, the WHU-Hi-LongKou dataset

Table 5 Evaluation metrics

Metric	Formula	Description
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$	Indicates the percentage of true predictions among all predictions.
Precision	$\frac{TP}{TP+FP}$	Determines how well a model is in predicting positive labels.
Recall (sensitivity)	$\frac{TP}{TP+FN}$	Measures the percentage of true positives successfully detected by a model.
F1-score (F-measure)	$\frac{2*Precision*Recall}{Precision+Recall}$	Represents the harmonic mean of Precision and Recall rates
Kappa coefficient	$\frac{P_o-P_e}{1-P_e}$	Measures inter-annotator agreement.
FPS	$\frac{Number\ of\ Frames}{Current\ time-Start\ time}$	Determines the detection speed.

consists of aerial images of six types of crops that were collected at an altitude of 500 m above ground with a spatial resolution of 46.3 cm on July 17, 2018, under clear and cloudless weather conditions. Similarly, the WHU-Hi-HanChuan dataset consists of seven crop types collected using the Leica Aibot X6 UAV V1 that flew at an altitude of 250 m above ground with a spatial resolution of 10.9 cm on June 17, 2016, under clear and cloudless weather conditions. However, the collected images have a lot of shadow-covered parts because they were obtained in the afternoon. The last dataset, WHU-Hi-HongHu, consists of seventeen different cultivars of three main crop types, which are cotton, rape, and cabbage. The collected images were acquired on November 20, 2017, through a DJI Matrice 600 Pro that flew at an altitude of 100 m above the soil with a Ground Sampling Distance (GSD) of 4.3 cm under cloudy weather conditions. Table 4, shows more information about the WHU-Hi dataset. The dataset is available at http://rsidea.whu.edu.cn/e-resource_WHUHi_sharing.htm.

Weedmap dataset: The Weedmap dataset [93] is one of the most known and largest multispectral datasets that can be used to classify sugar beet crops and weeds. It consists of more than 10,000 multispectral aerial images with different resolutions collected through two camera types at different locations and times. The two adopted cameras are RedEdge-M and Sequoia that are capable of capturing 5 and 4 raw image channels, respectively. This dataset can be used to train deep learning models to detect and classify sugar beet and weeds from aerial images. The dataset is available at https://projects.asl.ethz.ch/datasets/doku.php?id=weedmap:remotesensing2018weedmap.

VOAI dataset: The vegetational optical aerial image (VOAI) dataset [60] consists of aerial images of 12 tree species captured using a UAV platform that flew at different low altitudes of 20, 30, 40, and 50 meters above the ground. As presented in Table 4, the VOAI dataset contains around 2,300 cropped and normalized images without data augmentation and more than 21,800 after applying data augmentation techniques.

WeedNet dataset: The WeedNet dataset [92] consists of 465 multispectral images collected using a UAV platform equipped with a four-channel Sequoia camera flying at 2 meters above a sugar beet crop. The collected images are divided into a training set that contains 132 images for sugar beet crop and 243 images for weed, and a test set of 90 images for the mixture of both crop and weed. The dataset is available at https://github.com/inkyusa/weedNet.

Even with these datasets, there is still a lack of datasets targeting crop classification. Therefore, most of the reviewed articles in this survey used their own collected datasets through different UAV platforms and camera

sensors technologies [39, 53, 78, 85]. Unfortunately, these datasets are not available online.

5.2 Evaluation metrics

After the development of the deep learning-based crop classification models, we should evaluate their performance using several metrics. Therefore, various evaluation metrics were proposed to determine the effectiveness of the developed crop classification model. Several evaluation metrics could be measured to evaluate the deep learning models, depending on the study's purpose. Table 5 lists the most used evaluation metrics all along with their formulas and descriptions.

Note that TP, TN, FN, and FP refer to True Positive, True Negative, False Negative, and False Positive, respectively. Also, P_o refers to the observed proportional agreement, while P_e is the expected agreement by chance [58].

6 Discussion

In this study, various deep learning-based crop classification techniques were presented. Either object-based or pixel-based crop/plant classification achieved a remarkable performance. However, there still several factors that could affect crop/plant classification performance from UAV imagery, including image resolution, flight altitude, sensors characteristics, same spectral information from various materials, crop classes with similar feature properties, dataset size and quality, and choice of the network architecture and the baseline [109, 128].

6.1 The effect of spatial resolution on the crop classification performance

Several studies investigated the impact of flight altitude, image resolution, used sensors types, and GSD, which are strongly related to each other, on the performance of crop classification and plant detection. Low altitudes flights should provide lower GSD thus increase spatial resolution, which can improve classification and detection accuracy. As shown in [20, 46], a lower GSD provides more information about the targeted crop, which should lead to better classification performance. However, achieving lower GSD requires flying at lower altitudes, which leads to a lower area coverage [79, 100]. In order to demonstrate how the spatial resolution could affect the detection accuracy, the authors in [81] adopted different spatial resolution simulating different flight altitude. According to the obtained results, they conclude that the lower spatial resolution leads to the poor detection accuracy. Similarly, the poor results
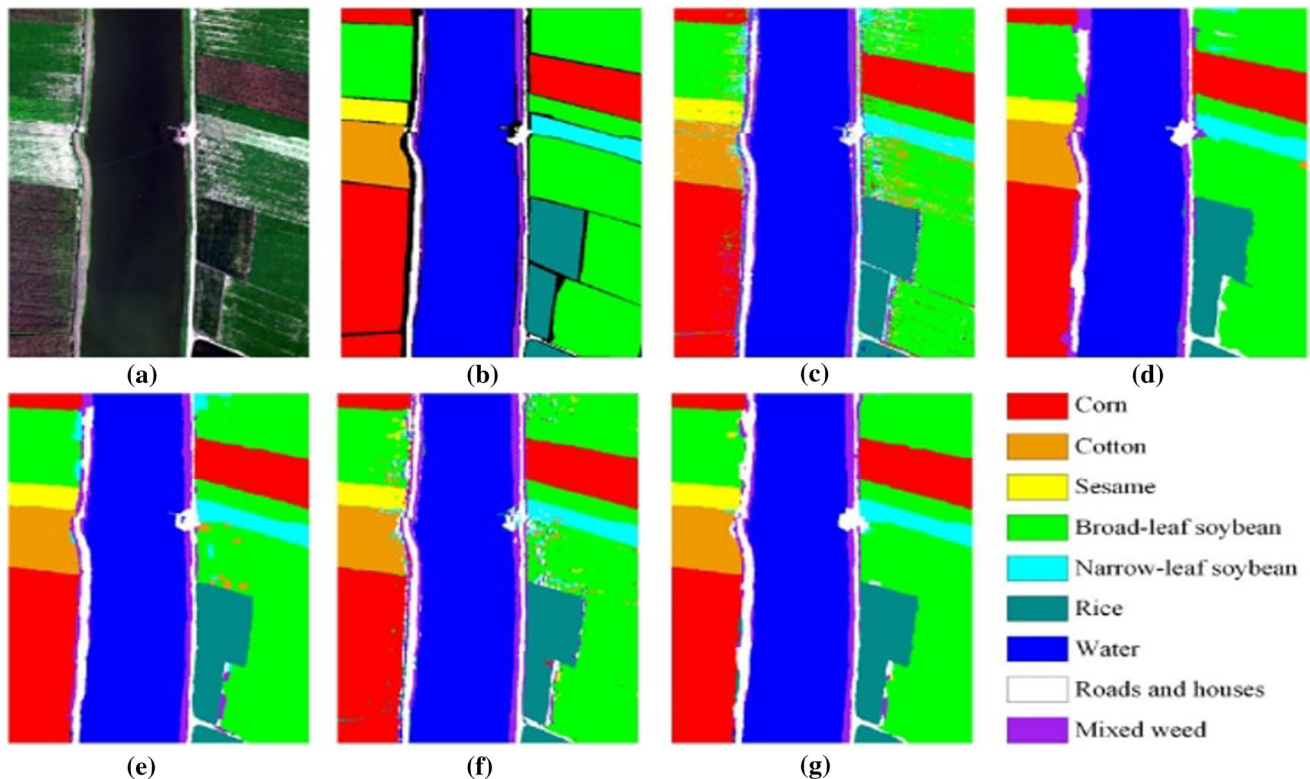
Fig. 6 Comparison of different algorithms to classify various crop types **a** True-color image, **b** Ground-truth image, **c** SVM, **d** FNEA-OO, **e** SVRFMC, **f** Benchmark CNN, and **g** CNNCRF [128].

provided by the adopted FCRN-MTL architecture [57] to identify citrus tree seedlings could be due to the limited spatial resolution of these seedlings and their small sizes. To improve the classification performance, the input patch size should be carefully selected. Combining different altitude could also improve banana plant identification achieving the best F1-Score of 98.64% combining 40m and 50m altitudes. However, the flight altitude is not the only factor that has an impact on the classification performance, but also the camera specification [54, 93]. Therefore, we need to take all of these parameters into account for achieving better results.

6.2 The effect of the selected model on the crop classification performance

Another major challenge in deep learning field is to find the best CNN architecture to solve the targeted problem. The classification performance could be affected by the network architecture and also the depth of these networks. The choice of the right architecture depends on many factors, including targeted crop/plant types, interference speed, among others. In [60], the authors investigated the impact of different CNN architectures with various number of layers for plant classification from UAV-based optical images. They proposed a CNN-based architecture that

outperformed the state-of-the-art CNN architectures achieving an overall accuracy of 86%, 84%, and 87% for FDN-17, FDN-29, and FDN-92, respectively, against 80%, 76%, and 82% for Inception-V1, ResNet-101, and Dense-Net-121 [44], respectively. Another work presented in [74] that investigate the impact of the chosen U-Net-based architectures on the overall classification performance.

Similarly, the performance of YOLOv3 outperformed RetinaNet detector in [9], which are trained on the same dataset achieving an average precision of 79.85% against 73.41% of the RetinaNet. This should be due to the different backbone architectures used by the two detectors, which are DarkNet-53 for YOLOv3 and ResNet for RetinaNet. From the results in Tables 2 and 3, we conclude that, most of the times, the deeper the network, the higher the classification performance. However, deeper networks require larger memory for data storing, more computation capabilities, and more processing time. Some studies have adopted shallow architectures that should improve the classification speed [21], while reducing the accuracy. Also, the authors in [27] developed a SegNet-based architecture with fewer layers resulting less trainable parameters, which make it faster and simpler. Moreover, a sigmoid function was adopted instead of Softmax, because they have only two classes "fig" and "not fig". Doing so, they achieved slightly better results than original SegNet

architecture achieving an overall accuracy of around 93.84% and 93.82%, respectively, but it must be faster due to the reduced number of learnable parameters. However, such shallow architecture could affect the classification performance, especially in the case of multi-crop classification, where it has a lack in relevant features extraction. These architectures are more suitable for machines with low computation power and for applications where we do not need very high accuracies.

According to Table 2, it is clear that YOLOv3 provides better results than RetinaNet and Faster R-CNN achieving an F1-score of around 99.8%. Also, according to Table 2, RGB images are preferred for tree detection and fruit detection. For example, faster R-CNN achieved a higher F1-score of 92% for apple tree detection while it achieved only 54% using multispectral images. Also, Santos et al. [97] showed that Faster R-CNN provides the worst inference time of 163 ms/image, due to the two detection stages, making it not suitable for real-time operations nether for tiny devices. However, one-stage algorithms (YOLOv3 and RetinaNet), as expected, achieved the lowest computational cost allowing real-time detection. According to Table 3, CrowNet [8] that is based on SegNet architecture provide slightly better results on maize crop row segmentation than Mask Scoring R-CNN architecture [81] achieving a precision rate of around 85% and 83%, respectively. However, the lower precision provided by Mask Scoring R-CNN could be resulted due to the higher flying altitude that is varying between 40 and 60 meters above the ground. Also, the developed model was trained on only two growing stages (V4 and V5), which could affect its performance to detect maize crop rows at the other growth stages. Also, the study of [118] showed that the FCN algorithm provides better results on rice crop

classification achieving an F1-score of 83.21% with an inference time of 72 seconds/image, while Segnet achieved only 69.36% with an inference time of 106 seconds/image. Moreover, SegNet has a very high false detection rate with a recall rate that could be less than 30% using RGB+Ex-G+ExGR data and a GSD equal to 5.7 cm/pixel. However, Song et al. [102] showed that SegNet architecture provides slightly better overall accuracy than FCN in the case of sunflower crop classification (Table 3). Similarly, Bah et al. [8] found that FCN architecture performs very badly to detect beet rows. Figure 6, shows a comparison between various algorithms used to classify different crop types.

6.3 The effect of data size and type on the crop classification performance

In order to train deep learning models accurately, we usually require a huge amount of data, which is a common point between different deep learning algorithms. Dataset type, size, and quality are other parameters that could improve the performance of crop classification. The impact of different dataset was investigated in [128], where they adopted three different datasets. The HanChuan and HongHu datasets have more complicated structure than the LongKou dataset, which explains the obtained results. As shown in Table 3, they achieved an overall accuracy of around 94% on the two first datasets against 98.91% on the last one. Also, the authors in [109] tested 4 different datasets to show their impact on corn classification from UAV imagery (see Tables 1 and 2). However, collecting and labelling large datasets is a time-consuming task and not always feasible. In [18, 22], the authors adopted transfer learning technique as a solution to overcome the lack in dataset. Transfer learning gives the possibility to
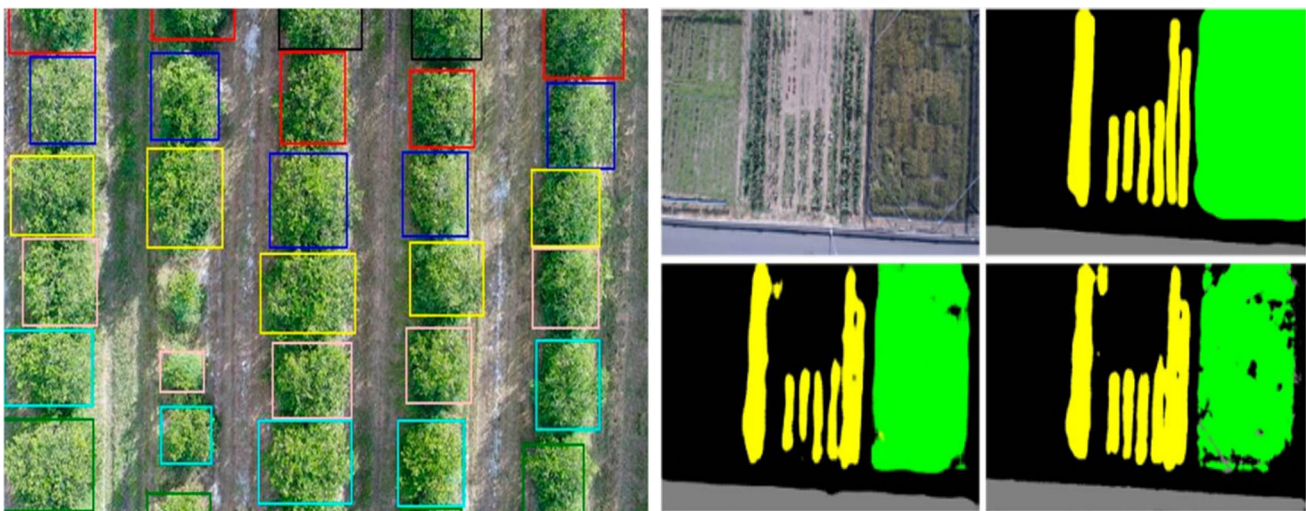

Fig. 7 Examples of object-based (left) [4] and pixel-based (right) [21] crop classification.

train deep neural network models with small dataset. Data augmentation is another effective solution that aims to increase the data size, which is well studied in many research papers [39, 109, 128].

The training patch size is another fundamental parameter that could affect the classification performance. In [82], the authors investigated the impact of class purity and patch size on the classification performance of homogenous crops using 2D-CNN. Patches with different sizes and class purity values are used to evaluate the CNN-based classifier performance on two different datasets collected from two platforms, which are Landsat-8 OLI satellite and senseFly eBee UAV. Similarly, the authors in [27] showed that the patch size has a slight impact on the classification performance achieving an accuracy of around 91% for patch of 256 × 256 pixels against 89.55% for 32 × 32 pixels. Moreover, the patch size has a significant importance on the training time. As the patch size increase the training time will also increase.

Efficient techniques were proposed over the few past years to improve the performance of crop classification. Several studies have used images acquired at different spectral bands, such as the works of [13, 82]. Sa et al. [93] investigated the impact of the chosen camera and the used channels on the classification performance. Also, in [34], the authors showed the impact of combining NIR and RGB channels to improve the classification accuracy of maize. Other researchers have proposed to apply various vegetation indices to enhance crop identification [93, 118]. In order to distinguish between small and large trees, the authors in [20] thresholding NDVI value taking only the values superior than 0.7 achieving an F1-Score of around 96%. In addition, FCN and SegNet algorithms performance were also investigated in the work of Yang et al. [118] for rice lodging identification using different RGB-based vegetation indices combinations that are produced from UAV imagery.

Also, some studies have proposed to use data fusion from satellite and UAV. The spectral information provided by UAV-based sensors provide less details than satellites-based ones [67]. Many studies investigated satellites and UAVs data fusion to increase the spectral information, which should help to improve crop classification [126, 129]. As example, in [129], the authors used three different approaches (SVM, ML, ANN) to investigate the impact of satellite and UAV data fusion on crop classification performance. By fusing the two platforms-based remote sensing information, they achieved better overall accuracy of around 86% on the fused data using ANN against 78.53% obtained from the original UAV data. Moreover, the results obtained applying ANN technique outperform the other two algorithms.

6.4 The impact of crop and UAV characteristics on the classification performance

The deep learning model performance is highly dependent on the characteristics of different crop types that vary according to the different growing stages and health, including texture, color, size, and spatial distribution. According to Table 1, object detection algorithms are preferred for tree crop monitoring. Most of the available object detection-based studies targeted fruits and tree crops classification due to their characteristics that are very suitable for object detection algorithms, especially the low density and the relatively large distance between trees (Fig, 7 (left)). However, the best results achieved for the major crop classification, like maize, wheat, legumes, are based on image segmentation techniques like U-Net, SegNet, FCN, and Mask R-CNN (Tab. 3) due to their characteristics in terms of texture, color, and canopy density (Fig. 7 (right)). According to Table 3, the authors in [91] showed that the geometric features of citrus trees have an impact on the deep learning model performance, where they achieved an overall accuracy of around 99% for full-grown citrus trees while they achieved only 56.6% to classify citrus tree seedlings. Also, the method developed in [66], which is based on Mask R-CNN architecture, performs better on lettuce crops than on potato crops, which could be due to the texture and density level of the two types of crops. In addition, as shown in Tables 2 and 3, vegetation indices, either obtained from RGB or multispectral data, can play a crucial role to improve the deep learning model accuracy due to the specific information about crop properties provided by each type of vegetation indices.

Moreover, UAV properties are also very important that could affect the deep learning model performance. Endurance, payload weight, flight speed, and altitude play crucial roles to improve classification accuracy and UAV efficiency. Fixed-wings UAVs are preferred for large crop monitoring due to their long endurance, speed, and high altitude [45, 84]. However, their high speed can lead to missing some valuable information due to the motion blur, and it is very difficult to monitor small and medium fields and need space to land and take-off [122]. Recently, hybrid VTOL UAVs were adopted as an effective solution to bypass the need for landing and taking-off space problems [101]. Also, multirotor UAVs are considered the best choice to cover small and medium fields, especially those located in difficult environments. To cover larger fields through multirotor UAVs, several studies proposed the use of multiple UAVs in swarm [3, 49]. Using UAV swarms allow farmers to examine very large farms in a matter of minutes. Also, their ability to hover and fly at very low altitudes above the targeted crop makes them the best

choice for small plants and fruits classification [43]. Sensor type is another fundamental parameter that we should take care of because each crop has different characteristics.

6.5 Choice of deep learning model

The choice of the right deep learning model, among a range of different models, depends on many factors like the targeted crop type, the deserved output that we want, the dataset size and type, the targeted hardware, to name a few. In this section, we go through the most crucial factors to think about while choosing the deep learning algorithm.

- *Crop type*: The targeted crop type and its characteristics play a fundamental role to select the right algorithm. According to the reviewed papers, it seems that object detection algorithms are preferred in the case of tree crop classification or identifying fruit types from images acquired at low altitudes. However, these algorithms still have some limitations to detect small objects in dense crops. For example He et al. [39] show that object detection algorithms achieved an average precision of only 78% to detect wheatears even at very low altitude. Also, it is not practical to cover large fields at very low altitude, which takes a very long time to cover the whole area. Therefore, as shown in Table 3, pixel-based techniques are preferred to monitor and classify large fields at very high altitudes.
- *Data size; nature; and quality*: The nature, size, and quality of data play a fundamental role in deciding which algorithm that we need to choose. To achieve the high performance of the algorithm, it is usually recommended to collect a large amount of data. However, it is a very challenging task to have such a large amount of data. Therefore, we can select pretrained models and adapt them with the targeted problem using transfer learning and fine tuning techniques. Also, data augmentation could be an effective solution to improve the algorithm performance.
- *Computational time*: The available computation time is another crucial factor that we should consider. Usually, we need a very long time and powerful GPUs to train accurate deep learning algorithms. The selection of the right algorithm is highly dependent on the hardware that we have. For tasks that require high accuracy, we need to select deep and complex algorithms that need high computation performance. The current state-of-the-art CNN architectures such as GoogLeNet, VGGnet, and ResNet are based on several trainable layers (convolutional and fully connected layers) making them very heavy and need powerful GPUs and large memory. These algorithms are not suitable to be implemented on the drone itself due to the low

processing ability of tiny devices. Therefore, very deep models are not the best choice to perform crop classification in real-time. On the other hand, light-weight models that are based on shallow feature extractors, like YOLOv2 and YOLOv2-tiny adopted in [54], are preferred to achieve real-time classification. Also, the authors in [118] showed that FCN is faster than SegNet architecture due to the adopted backbone which are AlexNet that consists only of 8 trainable layers and VGG-16 that contains 16 trainable layers, respectively.

7 Conclusion and future outlooks

Automatic crop classification using new technologies and techniques is recognized as one of the most important keys in today's smart farming improvement. It plays a crucial role in reducing the overall financial cost in reducing the number of workers, the amount of applied agrochemical products, among others. Most of previous studies are based on satellites and UGVs as remote sensing platforms and classical machine learning algorithms as processing techniques. However, UAVs have emerged as new effective remote sensing platforms that provide valuable data through the carried sensors. UAV-based Aerial image processing using different CNN-based algorithms have been applied successfully in crop/plant classification even though there still many challenges. So, in this paper, the advantages of UAVs over traditional technologies were highlighted. Moreover, different object-based and pixel-based algorithms were investigated in this review to help researchers and farmers to choose the right classifier according to the targeted crop, the used camera sensors, among other parameters. The choice of the right model is a fundamental key that should improve the accuracy, speed, and reliability to classify different crop types. Furthermore, possible challenges and their solutions for crop identification from UAV imagery were presented in this paper.

The agriculture industry is moving very fast toward Agriculture 4.0, which is based mainly on recent artificial intelligence algorithms and remote sensing technologies, especially deep learning and UAVs. The high advancement in new technologies and processing techniques, including camera sensors, 5G connection, cloud computing, GPUs, quantum computing, and different deep learning algorithms, should provide new opportunities and future outlooks in the field of precision agriculture. UAV platforms still have many issues that we should overcome to achieve the full capabilities required for smart farming. Many research directions, that have a relevant impact on the deep learning algorithms and UAV performances in the field of

agriculture, are being highly investigated. Time to cover large fields is one of the most disturbing problems that face the adoption of UAV platforms to handle large area covering. One solution to overcome such a problem is to build a whole intelligent Internet of Drone (IoD) system that could monitor farmlands to improve crop productivity with less human intervention. Moreover, we should focus on the development of powerful lightweight versions of deep learning algorithms that can be implemented on small devices with low computational power while reducing the energy consumption to increase the flying time of UAVs, which is a crucial problem.

Declarations

Conflict of interest The authors declare that they have no conflict of interest concerning the publication of this manuscript.

References

1. Achanta R, Shaji A, Smith K et al (2012) Slic superpixels compared to state-of-the-art superpixel methods. IEEE Trans Pattern Anal Machine Intell 34(11):2274–2282. https://doi.org/10.1109/TPAMI.2012.120

2. Adrian J, Sagan V, Maimaitijiang M (2021) Sentinel sar-optical fusion for crop type mapping using deep learning and google earth engine. ISPRS J Photogramm Remote Sens 175:215–235. https://doi.org/10.1016/j.isprsjprs.2021.02.018

3. Albani D, Nardi D, Trianni V (2017) Field coverage and weed mapping by uav swarms. In: 2017 IEEE/RSJ international conference on intelligent robots and systems (IROS), pp 4319–4325, https://doi.org/10.1109/IROS.2017.8206296

4. Ampatzidis Y, Partel V (2019) Uav-based high throughput phenotyping in citrus utilizing multispectral imaging and artificial intelligence. Remote Sens 11(4):410. https://doi.org/10.3390/rs11040410

5. Ashapure A, Jung J, Chang A et al (2020) Developing a machine learning based cotton yield estimation framework using multi-temporal uas data. ISPRS J Photogramm Remote Sens 169:180–194. https://doi.org/10.1016/j.isprsjprs.2020.09.015

6. Badrinarayanan V, Kendall A, Cipolla R (2017) Segnet: a deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans Pattern Anal Machine Intell 39(12):2481–2495. https://doi.org/10.1109/TPAMI.2016.2644615

7. Bah MD, Hafiane A, Canals R (2018) Deep learning with unsupervised data labeling for weed detection in line crops in uav images. Remote Sens 10(11):1690. https://doi.org/10.3390/rs10111690

8. Bah MD, Hafiane A, Canals R (2019) Crownet: deep network for crop row detection in uav images. IEEE Access 8:5189–5200. https://doi.org/10.1109/ACCESS.2019.2960873

9. Bayraktar E, Basarkan ME, Celebi N (2020) A low-cost uav framework towards ornamental plant detection and counting in the wild. ISPRS J Photogramm Remote Sens 167:1–11. https://doi.org/10.1016/j.isprsjprs.2020.06.012

10. Beacham AM, Vickers LH, Monaghan JM (2019) Vertical farming: a summary of approaches to growing skywards.

J Hortic Sci Biotechnol 94(3):277–283. https://doi.org/10.1080/14620316.2019.1574214

11. Bhosle K, Musande V (2020) Evaluation of cnn model by comparing with convolutional autoencoder and deep neural network for crop classification on hyperspectral imagery. Geocarto International 1–15. https://doi.org/10.1080/10106049.2020.1740950

12. Bochkovskiy A, Wang CY, Liao HYM (2020) Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:200410934

13. Böhler JE, Schaepman ME, Kneubühler M (2018) Crop classification in a heterogeneous arable landscape using uncalibrated uav data. Remote Sens 10(8):1282. https://doi.org/10.3390/rs10081282

14. Bouguettaya A, Zarzour H, Kechida A, et al (2021) Vehicle detection from uav imagery with deep learning: A review. IEEE Trans Neural Netw Learn Syst pp 1–21. https://doi.org/10.1109/TNNLS.2021.3080276

15. Breiman L (2001) Random forests. Machine Learn 45(1):5–32. https://doi.org/10.1023/A:1010933404324

16. Chamorro Martinez JA, Cué La Rosa LE, Feitosa RQ et al (2021) Fully convolutional recurrent networks for multidate crop recognition from multitemporal image sequences. ISPRS J Photogramm Remote Sens 171:188–201. https://doi.org/10.1016/j.isprsjprs.2020.11.007

17. Chen LC, Papandreou G, Kokkinos I et al (2017) Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE Trans Pattern Anal Machine Intell 40(4):834–848. https://doi.org/10.1109/TPAMI.2017.2699184

18. Chew R, Rineer J, Beach R, et al (2020) Deep neural networks and transfer learning for food crop identification in uav images. Drones 4(1). https://doi.org/10.3390/drones4010007

19. Cortes C, Vapnik V (1995) Support-vector networks. Machine Learn 20(3):273–297. https://doi.org/10.1007/BF00994018

20. Csillik O, Cherbini J, Johnson R et al (2018) Identification of citrus trees from unmanned aerial vehicle imagery using convolutional neural networks. Drones 2(4):39. https://doi.org/10.3390/drones2040039

21. Der Yang M, Tseng HH, Hsu YC, et al (2020) Real-time crop classification using edge computing and deep learning. In: 2020 IEEE 17th annual consumer communications & networking conference (CCNC), IEEE, pp 1–4, https://doi.org/10.1109/CCNC46108.2020.9045498

22. Duong-Trung N, Quach LD, Nguyen MH, et al (2019) A combination of transfer learning and deep learning for medicinal plant classification. In: Proceedings of the 2019 4th international conference on intelligent information technology. Association for computing machinery, New York, NY, USA, ICIIT '19, p 83-90, https://doi.org/10.1145/3321454.3321464, https://doi.org/10.1145/3321454.3321464

23. Fawakherji M, Potena C, Bloisi DD, et al (2019) Uav image based crop and weed distribution estimation on embedded gpu boards. In: International conference on computer analysis of images and patterns, Springer, pp 100–108, https://doi.org/10.1007/978-3-030-29930-9_10

24. Feng Q, Yang J, Liu Y et al (2020) Multi-temporal unmanned aerial vehicle remote sensing for vegetable mapping using an attention-based recurrent convolutional neural network. Remote Sens 12(10):1668. https://doi.org/10.3390/rs12101668

25. Ferreira MP, de Almeida DRA, de Almeida Papa D et al (2020) Individual tree detection and species classification of amazonian palms using uav images and deep learning. Forest Ecol Manag 475(118):397. https://doi.org/10.1016/j.foreco.2020.118397

26. Friedl MA, Brodley CE (1997) Decision tree classification of land cover from remotely sensed data. Remote Sens Environ 61(3):399–409. https://doi.org/10.1016/S0034-4257(97)00049-7

27. Fuentes-Pacheco J, Torres-Olivares J, Roman-Rangel E et al (2019) Fig plant segmentation from aerial images using a deep convolutional encoder-decoder network. Remote Sens 11(10):1157. https://doi.org/10.3390/rs11101157

28. Gao Z, Luo Z, Zhang W et al (2020) Deep learning application in plant stress imaging: a review. AgriEng 2(3):430–446. https://doi.org/10.3390/agriengineering2030029

29. Girshick R (2015) Fast r-cnn. In: Proceedings of the IEEE international conference on computer vision, pp 1440–1448, https://doi.org/10.1109/ICCV.2015.169

30. Girshick R, Donahue J, Darrell T, et al (2014) 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, https://doi.org/10.1109/CVPR.2014.81

31. Goodman W, Minner J (2019) Will the urban agricultural revolution be vertical and soilless? a case study of controlled environment agriculture in new york city. Land Use Policy 83:160–173. https://doi.org/10.1016/j.landusepol.2018.12.038

32. Gray H, Nuri KR (2020) Differing visions of agriculture: Industrial-chemical vs. small farm and urban organic production. Am J Econ Sociol 79(3):813–832. https://doi.org/10.1111/ajes.12344

33. Guo X, Li P (2020) Mapping plastic materials in an urban area: Development of the normalized difference plastic index using worldview-3 superspectral data. ISPRS J Photogramm Remote Sens 169:214–226. https://doi.org/10.1016/j.isprsjprs.2020.09.009, https://www.sciencedirect.com/science/article/pii/S09242716203002562

34. Hall O, Dahlin S, Marstorp H, et al (2018) Classification of maize in complex smallholder farming systems using uav imagery. Drones 2(3). https://doi.org/10.3390/drones2030022

35. Hasan M, Tanawala B, Patel KJ (2019) Deep learning precision farming: Tomato leaf disease detection by transfer learning. In: Proceedings of 2nd international conference on advanced computing and software engineering (ICACSE), https://doi.org/10.2139/ssrn.3349597

36. Hassler SC, Baysal-Gurel F (2019) Unmanned aircraft system (uas) technology and applications in agriculture. Agronomy 9(10). https://doi.org/10.3390/agronomy9100618, https://www.mdpi.com/2073-4395/9/10/618

37. He K, Zhang X, Ren S, et al (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778, https://doi.org/10.1109/CVPR.2016.90

38. He K, Gkioxari G, Dollár P, et al (2017) Mask r-cnn. In: Proceedings of the IEEE international conference on computer vision, pp 2961–2969

39. He MX, Hao P, Xin YZ (2020) A robust method for wheatear detection using uav in natural scenes. IEEE Access 8:189,043–189,053. https://doi.org/10.1109/ACCESS.2020.3031896

40. Herrmann C, Willersinn D, Beyerer J (2016) Low-resolution convolutional neural networks for video face recognition. In: 2016 13th IEEE international conference on advanced video and signal based surveillance (AVSS), IEEE, pp 221–227, https://doi.org/10.1109/AVSS.2016.7738017

41. Howard A, Sandler M, Chen B, et al (2019) Searching for mobilenetv3. In: 2019 IEEE/CVF international conference on computer vision (ICCV), pp 1314–1324, https://doi.org/10.1109/ICCV.2019.00140

42. Howard AG, Zhu M, Chen B, et al (2017) Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:170404861

43. Hu X, Zhong Y, Luo C, et al (2018) Fine classification of typical farms in southern china based on airborne hyperspectral remote sensing images. In: 2018 7th international conference on agro-geoinformatics (Agro-geoinformatics), pp 1–4, https://doi.org/10.1109/Agro-Geoinformatics.2018.8475977

44. Huang G, Liu Z, Van Der Maaten L, et al (2017) Densely connected convolutional networks. In: 2017 IEEE conference on computer vision and pattern recognition (CVPR), pp 2261–2269, https://doi.org/10.1109/CVPR.2017.243

45. Hunt ER, Stern AJ (2019) Evaluation of incident light sensors on unmanned aircraft for calculation of spectral reflectance. Remote Sens 11(22). https://doi.org/10.3390/rs11222622, https://www.mdpi.com/2072-4292/11/22/2622

46. Jiang R, Wang P, Xu Y, et al (2020) Assessing the operation parameters of a low-altitude uav for the collection of ndvi values over a paddy rice field. Remote Sens 12(11). https://doi.org/10.3390/rs12111850, https://www.mdpi.com/2072-4292/12/11/1850

47. Jiao L, Zhang F, Liu F, et al (2019) A survey of deep learning-based object detection. IEEE Access 7:128,837–128,868. https://doi.org/10.1109/ACCESS.2019.2939201

48. Jocher G (2021) yolov5. https://github.com/ultralytics/yolov5, (Accessed: 10-06-2021)

49. Ju C, Son HI (2018) Multiple uav systems for agricultural applications: Control, implementation, and evaluation. Electronics 7(9). https://doi.org/10.3390/electronics7090162, https://www.mdpi.com/2079-9292/7/9/162

50. Jámbor A, Czine P, Balogh P (2020) The impact of the coronavirus on agriculture: first evidence based on global newspapers. Sustainability 12(11). https://doi.org/10.3390/su12114535

51. Kamilaris A, Prenafeta-Boldú FX (2018) Deep learning in agriculture: a survey. Comput Electron Agric 147:70–90. https://doi.org/10.1016/j.compag.2018.02.016, https://www.sciencedirect.com/science/article/pii/S0168169917308803

52. Kattenborn T, Eichel J, Fassnacht FE (2019) Convolutional neural networks enable efficient, accurate and fine-grained segmentation of plant species and communities from high-resolution uav imagery. Sci Rep 9(1):1–9

53. Kitano BT, Mendes CC, Geus AR et al (2019) Corn plant counting using deep learning and uav images. IEEE Geosci Remote Sens Lett. https://doi.org/10.1109/LGRS.2019.2930549

54. Koirala A, Walsh K, Wang Z et al (2019) Deep learning for real-time fruit detection and orchard fruit load estimation: benchmarking of 'mangoyolo'. Precis Agric 20(6):1107–1135. https://doi.org/10.1007/s11119-019-09642-0

55. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. Adv Neural Info Process Syst 25:1097–1105. https://doi.org/10.1145/3065386

56. Kussul N, Lavreniuk M, Skakun S et al (2017) Deep learning classification of land cover and crop types using remote sensing data. IEEE Geosci Remote Sens Lett 14(5):778–782. https://doi.org/10.1109/LGRS.2017.2681128

57. La Rosa LEC, Zortea M, Gemignani BH, et al (2020) Fcrn-based multi-task learning for automatic citrus tree detection from uav images. In: 2020 IEEE Latin American GRSS ISPRS remote sensing conference (LAGIRS), pp 403–408, https://doi.org/10.1109/LAGIRS48042.2020.9165654

58. Lan Y, Huang Z, Deng X et al (2020) Comparison of machine learning methods for citrus greening detection on uav multi-spectral images. Comput Electron Agric 171(105):234. https://doi.org/10.1016/j.compag.2020.105234, https://www.sciencedirect.com/science/article/pii/S0168169919313857

59. LeCun Y, Bottou L, Bengio Y et al (1998) Gradient-based learning applied to document recognition. Proc IEEE 86(11):2278–2324. https://doi.org/10.1109/5.726791

60. Lin C, Ding Q, Tu W, et al (2019) Fourier dense network to conduct plant classification using uav-based optical images.

IEEE Access 7:17,736–17,749. https://doi.org/10.1109/ACCESS.2019.2895243

61. Lin TY, Goyal P, Girshick R, et al (2017) Focal loss for dense object detection. In: Proceedings of the IEEE international conference on computer vision, pp 2980–2988, https://doi.org/10.1109/TPAMI.2018.2858826

62. Liu W, Anguelov D, Erhan D, et al (2016) Ssd: single shot multibox detector. In: European conference on computer vision, Springer, pp 21–37, https://doi.org/10.1007/978-3-319-46448-0_2

63. Long J, Shelhamer E, Darrell T (2015) Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3431–3440

64. van Loon MP, Hijbeek R, ten Berge HFM, et al (2019) Impacts of intensifying or expanding cereal cropping in sub-saharan africa on greenhouse gas emissions and food security. Global Change Biol 25(11):3720–3730. https://doi.org/10.1111/gcb.14783

65. Ma N, Zhang X, Zheng HT et al (2018) Shufflenet v2: Practical guidelines for efficient cnn architecture design. In: Ferrari V, Hebert M, Sminchisescu C et al (eds) Comput Vision - ECCV 2018. Springer International Publishing, Cham, pp 122–138

66. Machefer M, Lemarchand F, Bonnefond V, et al (2020) Mask r-cnn refitting strategy for plant counting and sizing in uav imagery. Remote Sens 12(18). https://doi.org/10.3390/rs12183015

67. Maimaitijiang M, Sagan V, Sidike P, et al (2020a) Crop monitoring using satellite/uav data fusion and machine learning. Remote Sens 12(9). https://doi.org/10.3390/rs12091357

68. Maimaitijiang M, Sagan V, Sidike P et al (2020) Soybean yield prediction from uav using multimodal data fusion and deep learning. Remote Sens Environ 237(111):599. https://doi.org/10.1016/j.rse.2019.111599, https://www.sciencedirect.com/science/article/pii/S0034425719306194

69. Mazzia V, Comba L, Khaliq A, et al (2020a) Uav and machine learning based refinement of a satellite-driven vegetation index for precision agriculture. Sensors 20(9). https://doi.org/10.3390/s20092530

70. Mazzia V, Khaliq A, Chiaberge M (2020) Improvement in land cover and crop classification based on temporal features learning from sentinel-2 data using recurrent-convolutional neural network (r-cnn). Appl Sci 10(1):238. https://doi.org/10.3390/app10010238

71. Mittal P, Singh R, Sharma A (2020) Deep learning-based object detection in low-altitude uav datasets: a survey. Image Vision Comput 104(104):046. https://doi.org/10.1016/j.imavis.2020.104046, https://www.sciencedirect.com/science/article/pii/S0262885620301785

72. Mohanty SP, Hughes DP, Salathé M (2016) Using deep learning for image-based plant disease detection. Front Plant Sci 7:1419. https://doi.org/10.3389/fpls.2016.01419, https://www.frontiersin.org/article/10.3389/fpls.2016.01419

73. Montagnoli A, Fusco S, Terzaghi M et al (2015) Estimating forest aboveground biomass by low density lidar data in mixed broad-leaved forests in the italian pre-alps. Forest Ecosyst 2(1):1–9. https://doi.org/10.1186/s40663-015-0035-6

74. Morales G, Kemper G, Sevillano G, et al (2018) Automatic segmentation of mauritia flexuosa in unmanned aerial vehicle (uav) imagery using deep learning. Forests 9(12). https://doi.org/10.3390/f9120736

75. Nelson J, Solawetz J (2020) Yolov5 is here: State-of-the-art object detection at 140 fps. Roboflow com https://blog roboflow com/yolov5-is-here Accessed 17

76. Neupane B, Horanont T, Hung ND (2019) 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):e0223,906. https://doi.org/10.1371/journal.pone.0223906

77. Nogueira K, dos Santos JA, Menini N et al (2019) Spatio-temporal vegetation pixel classification by using convolutional networks. IEEE Geosci Remote Sens Lett 16(10):1665–1669. https://doi.org/10.1109/LGRS.2019.2903194

78. Osco LP, de Arruda MdS, Junior JM et al (2020) A convolutional neural network approach for counting and geolocating citrus-trees in uav multispectral imagery. ISPRS J Photogramm Remote Sens 160:97–106. https://doi.org/10.1016/j.isprsjprs.2019.12.010

79. Pádua L, Vanko J, Hruška J et al (2017) Uas, sensors, and data processing in agroforestry: a review towards practical applications. Int J Remote Sens 38(8–10):2349–2391. https://doi.org/10.1080/01431161.2017.1297548

80. Panday US, Pratihast AK, Aryal J, et al (2020) A review on drone-based data solutions for cereal crops. Drones 4(3). https://doi.org/10.3390/drones4030041, https://www.mdpi.com/2504-446X/4/3/41

81. Pang Y, Shi Y, Gao S et al (2020) Improved crop row detection with deep neural network for early-season maize stand count in uav imagery. Comput Electron Agric 178(105):766. https://doi.org/10.1016/j.compag.2020.105766

82. Park S, Park NW (2020) Effects of class purity of training patch on classification performance of crop classification with convolutional neural network. Appl Sci 10(11). https://doi.org/10.3390/app10113773

83. Piiroinen R, Heiskanen J, Mõttus M et al (2015) Classification of crops across heterogeneous agricultural landscape in kenya using aisaeagle imaging spectroscopy data. Int J Appl Earth Observ Geoinfo 39:1–8.https://doi.org/10.1016/j.jag.2015.02.005, https://www.sciencedirect.com/science/article/pii/S0303243415000379

84. Raeva PL, Šedina J, Dlesk A (2019) Monitoring of crop fields using multispectral and thermal imagery from uav. Euro J Remote Sens 52(sup1):192–201. https://doi.org/10.1080/22797254.2018.1527661, https://doi.org/10.1080/22797254.2018.1527661

85. Rebetez J, Satizábal HF, Mota M, et al (2016) Augmenting a convolutional neural network with local histograms-a case study in crop classification from high-resolution uav imagery. In: ESANN

86. Redmon J, Farhadi A (2017) Yolo9000: better, faster, stronger. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 7263–7271, https://doi.org/10.1109/CVPR.2017.690

87. Redmon J, Farhadi A (2018) Yolov3: an incremental improvement. arXiv preprint arXiv:180402767

88. Redmon J, Divvala S, Girshick R, et al (2016) You only look once: Unified, real-time object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 779–788

89. Ren S, He K, Girshick R, et al (2015) Faster r-cnn: towards real-time object detection with region proposal networks. arXiv preprint arXiv:150601497

90. Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional networks for biomedical image segmentation. In: International conference on medical image computing and computer-assisted intervention, Springer, pp 234–241, https://doi.org/10.1007/978-3-319-24574-4_28

91. Rosa LECL, Oliveira DAB, Zortea M, et al (2020) Learning geometric features for improving the automatic detection of citrus plantation rows in uav images. IEEE Geosci Remote Sens Lett pp 1–5. https://doi.org/10.1109/LGRS.2020.3024641

92. Sa I, Chen Z, Popović M et al (2018) weednet: dense semantic weed classification using multispectral images and mav for smart farming. IEEE Robot Autom Lett 3(1):588–595. https://doi.org/10.1109/LRA.2017.2774979

93. Sa I, Popović M, Khanna R, et al (2018b) Weedmap: A large-scale semantic weed mapping framework using aerial multi-spectral imaging and deep neural network for precision farming. Remote Sens 10(9). https://doi.org/10.3390/rs10091423

94. Sagan V, Maimaitijiang M, Bhadra S et al (2021) Field-scale crop yield prediction using multi-temporal worldview-3 and planetscope satellite data and deep learning. ISPRS J Photogramm Remote Sens 174:265–281. https://doi.org/10.1016/j.isprsjprs.2021.02.008, https://www.sciencedirect.com/science/article/pii/S0924271621000411

95. Saleem MH, Potgieter J, Arif KM (2019) Plant disease detection and classification by deep learning. Plants 8(11):468. https://doi.org/10.3390/plants8110468

96. Sandler M, Howard A, Zhu M, et al (2018) Mobilenetv2: inverted residuals and linear bottlenecks. In: 2018 IEEE/CVF Conference on computer vision and pattern recognition, pp 4510–4520, https://doi.org/10.1109/CVPR.2018.00474

97. Santos AAd, Marcato Junior J, Araújo MS et al (2019) Assessment of cnn-based methods for individual tree detection on images captured by rgb cameras attached to uavs. Sensors 19(16):3595. https://doi.org/10.3390/s19163595

98. Santos L, Santos FN, Oliveira PM et al (2020) Deep learning applications in agriculture: a short review. Robot 2019: Fourth Iberian robotics conference. Springer International Publishing, Cham, pp 139–151

99. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:14091556

100. Sona G, Passoni D, Pinto L, et al (2016) Uav multispectral survey to map soil and crop for precision farming applications. In: Remote sensing and spatial information sciences congress: international archives of the photogrammetry remote sensing and spatial information sciences congress: 19 July, International Society for Photogrammetry and Remote Sensing (ISPRS), pp 1023–1029

101. Song B, Park K (2020) Detection of aquatic plants using mul-tispectral uav imagery and vegetation index. Remote Sens 12(3). https://doi.org/10.3390/rs12030387, https://www.mdpi.com/2072-4292/12/3/387

102. Song Z, Zhang Z, Yang S et al (2020) Identifying sunflower lodging based on image fusion and deep semantic segmentation with uav remote sensing imaging. Comput Electron Agric 179(105):812. https://doi.org/10.1016/j.compag.2020.105812

103. Szegedy C, Wei Liu, Yangqing Jia, et al (2015) Going deeper with convolutions. In: 2015 IEEE Conference on computer vision and pattern recognition (CVPR), pp 1–9, https://doi.org/10.1109/CVPR.2015.7298594

104. Szegedy C, Vanhoucke V, Ioffe S, et al (2016) Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2818–2826, https://doi.org/10.1109/CVPR.2016.308

105. Szegedy C, Ioffe S, Vanhoucke V, et al (2017) Inception-v4, inception-resnet and the impact of residual connections on learning. In: Proceedings of the AAAI conference on artificial intelligence

106. Tatsumi K, Yamashiki Y, Canales Torres MA et al (2015) Crop classification of upland fields using random forest of time-series landsat 7 etm+ data. Comput Electron Agric 115:171–179. https://doi.org/10.1016/j.compag.2015.05.001, https://www.sciencedirect.com/science/article/pii/S0168169915001234

107. Théau J, Gavelle E, Ménard P (2020) Crop scouting using uav imagery: a case study for potatoes. J Unmanned Vehicle Syst 8(2):99–118. https://doi.org/10.1139/juvs-2019-0009

108. Tong K, Wu Y, Zhou F (2020) Recent advances in small object detection based on deep learning: a review. Image Vision Comput 97(103):910. https://doi.org/10.1016/j.imavis.2020.103910

109. Trujillano F, Flores A, Saito C, et al (2018) Corn classification using deep learning with uav imagery. an operational proof of concept. In: 2018 IEEE 1st Colombian conference on applications in computational intelligence (ColCACI), pp 1–4, https://doi.org/10.1109/ColCACI.2018.8484845

110. Ullah A, Muhammad K, Haq IU et al (2019) Action recognition using optimized deep autoencoder and cnn for surveillance data streams of non-stationary environments. Future Gener Comput Syst 96:386–397. https://doi.org/10.1016/j.future.2019.01.029

111. United-Nation (2020) Growing at a slower pace, world population is expected to reach 9.7 billion in 2050 and could peak at nearly 11 billion around 2100. https://www.un.org/development/desa/en/news/population/world-population-prospects-2019.html

112. Wu J, Yang G, Yang H et al (2020) Extracting apple tree crown information from remote imagery using deep learning. Comput Electron Agricul 174(105):504. https://doi.org/10.1016/j.compag.2020.105504

113. Wu S, Wang J, Yan Z et al (2021) Monitoring tree-crown scale autumn leaf phenology in a temperate forest with an integration of planetscope and drone remote sensing observations. ISPRS J Photogramm Remote Sens 171:36–48. https://doi.org/10.1016/j.isprsjprs.2020.10.017

114. Wu X, Sahoo D, Hoi SC (2020) Recent advances in deep learning for object detection. Neurocomputing 396:39–64. https://doi.org/10.1016/j.neucom.2020.01.085

115. Xie C, Yang C (2020) A review on plant high-throughput phenotyping traits using uav-based sensors. Comput Electron Agricul 178(105):731. https://doi.org/10.1016/j.compag.2020.105731, https://www.sciencedirect.com/science/article/pii/S0168169919320046

116. Xiong J, Liu Z, Chen S et al (2020) Visual detection of green mangoes by an unmanned aerial vehicle in orchards based on a deep learning method. Biosyst Eng 194:261–272. https://doi.org/10.1016/j.biosystemseng.2020.04.006

117. Yan Y, Ryu Y (2021) Exploring google street view with deep learning for crop type mapping. ISPRS J Photogramm Remote Sens 171:278–296. https://doi.org/10.1016/j.isprsjprs.2020.11.022

118. Yang MD, Tseng HH, Hsu YC, et al (2020a) Semantic segmentation using deep learning with vegetation indices for rice lodging identification in multi-date uav visible images. Remote Sens 12(4). https://doi.org/10.3390/rs12040633

119. Yang Q, Shi L, Han J et al (2019) Deep convolutional neural networks for rice grain yield estimation at the ripening stage using uav-based remotely sensed images. Field Crops Res 235:142–153. https://doi.org/10.1016/j.fcr.2019.02.022

120. Yang S, Gu L, Li X, et al (2020b) Crop classification method based on optimal feature selection and hybrid cnn-rf networks for multi-temporal remote sensing imagery. Remote Sens 12(19). https://doi.org/10.3390/rs12193119

121. Zeiler MD, Fergus R (2014) Visualizing and understanding convolutional networks. In: European conference on computer vision, Springer, pp 818–833, https://doi.org/10.1007/978-3-319-10590-1_53

122. Zhang H, Zhang B, Wei Z, et al (2020a) Lightweight integrated solution for a uav-borne hyperspectral imaging system. Remote Sens 12(4). https://doi.org/10.3390/rs12040657, https://www.mdpi.com/2072-4292/12/4/657

123. Zhang J, Xie T, Yang C et al (2020) Segmenting purple rapeseed leaves in the field from uav rgb imagery using deep learning as an auxiliary means for nitrogen stress detection. Remote Sens 12(9):1403. https://doi.org/10.3390/rs12091403

124. Zhang X, Zhou X, Lin M, et al (2018) Shufflenet: An extremely efficient convolutional neural network for mobile devices. In: 2018 IEEE/CVF conference on computer vision and pattern recognition, pp 6848–6856, https://doi.org/10.1109/CVPR.2018.00716

125. Zhao H, Chen Z, Jiang H, et al (2019a) Evaluation of three deep learning models for early crop classification using sentinel-1a imagery time series-a case study in zhanjiang, china. Remote Sens 11(22). https://doi.org/10.3390/rs11222673, https://www.mdpi.com/2072-4292/11/22/2673

126. Zhao L, Shi Y, Liu B, et al (2019b) Finer classification of crops by fusing uav images and sentinel-2a data. Remote Sens 11(24). https://doi.org/10.3390/rs11243012

127. Zhong L, Hu L, Zhou H (2019) Deep learning based multi-temporal crop classification. Remote Sens Environ 221:430–443. https://doi.org/10.1016/j.rse.2018.11.032

128. Zhong Y, Hu X, Luo C et al (2020) Whu-hi: Uav-borne hyperspdectral with high spatial resolution (h2) benchmark datasets and classifier for precise crop identification based on deep convolutional neural network with crf. Remote Sens Environ 250(112):012. https://doi.org/10.1016/j.rse.2020.112012

129. Zou Y, Li G, Wang S (2018) The fusion of satellite and unmanned aerial vehicle (uav) imagery for improving classification performance. In: 2018 IEEE international conference on information and automation (ICIA), pp 836–841, https://doi.org/10.1109/ICInfA.2018.8812312