Smart Agricultural Technology 9 (2024) 100625

Contents lists available at ScienceDirect

Smart Agricultural Technology

journal homepage: www.journals.elsevier.com/smart-agricultural-technology

A survey on deep learning in UAV imagery for precision agriculture and wild flora monitoring: Datasets, models and challenges

Lorenzo Epifani, Antonio Caruso *

Department of Mathematics and Physics 'Ennio De Giorgi', Palazzo Fiorini, Campus Ecotekne, Lecce, 73100, Italy

ARTICLE INFO

Keywords:
Machine learning
Deep neural networks
Image analysis
Unmanned aerial vehicles
Agritech

ABSTRACT

Machine learning is the state of the art for many recurring tasks in several heterogeneous domains. In the last decade, it has been also widely used in Precision Agriculture (PA) and Wild Flora Monitoring (WFM) to address a set of problems with a big impact on economy, society and academia, heralding a paradigm shift across the industry and academia. Many applications in those fields involve image processing and computer vision stages. Remote sensing devices are very popular choice for image acquisition in this context, and in particular, Unmanned Aerial Vehicles (UAVs) offer a good tradeoff between cost and area coverage. For these reasons, research literature is rich of works that face problems in Precision Agriculture and Wild Flora Monitoring domains with machine learning/computer vision methods applied to UAV imagery. In this work, we review this literature, with a special focus on algorithms, model sizing, dataset characteristics and innovative technical solutions presented in many domain-specific models, providing the reader with an overview of the research trend in recent years.

1. Introduction

In recent decades, interest of industry and academia for agroforestry has gradually grown, for several economic, social and environmental reasons [1]. Agricultural production has always been economically and livelihood relevant: the primary sector is critical to sustaining the world's population, which in Food and Agriculture Organization projections is expected to increase to nearly 10 billion by 2050 [2], which will inevitably cause an increase in demand for food [3]. Urbanization and the gradual depopulation of rural areas is a well-known phenomenon, and now it is affecting developing countries, limiting the areas devoted to agriculture and the people willing to care for them [4]. The pandemic of COVID-19 and recent geopolitical crisis (Ukrainian War, Suez Canal problems) has shown the weaknesses of the global supply chain and the dependence of agriculture on human labor [5]. It is clear how modern agriculture requires automated, sustainable, and reliable systems especially when labor shortages occur. Automated and environmentally friendly crop management solutions[1] are therefore increasingly in demand, as it is required to scale food production with world population growth without generating adverse side effects. With the fight against climate change, the desire to reduce human environmental footprint has attracted further interest in precision agriculture.

As well as interest in the primary sector, there has been a growing interest in the development of smart, reliable and sustainable monitoring techniques to safeguard the biodiversity and conservation status of wild plant species, as they also have a direct impact on the animal world and the preservation of the delicate environmental balance of our planet [6,7]. As global temperatures rise, the ecosystem is changing rapidly to the detriment of local fauna and flora in many areas of the world [8]. Thus, automated and noninvasive observation of plant species is a task with cross-domain relevance. There are several open challenges for which there is a constant research (both in industry and academia) for the optimal solution (listed in the column "Goal" in Table 8). Information required to face these challenges can be obtained from images, therefore, academia and industry have questioned which imagery systems are most suitable for these use cases. Unmanned Ground Vehicles (UGVs) allow images to be acquired at a very close range, maximizing the information obtained for each plant specimen, however, they directly affect the terrain, and their usage is expensive because of slow acquisition and low coverage [9]. For this reason, the use cases where this technology is the optimal choice are limited.

Remote sensing technologies, i.e. Unmanned Aerial Vehicles (UAVs) and satellites, have proven to have many characteristics that make them an attractive and versatile choices. Satellite imagery is a popular choice,

* Corresponding author.
E-mail addresses: lorenzo.epifani@unisalento.it (L. Epifani), antonio.caruso@unisalento.it (A. Caruso).

[1] Croptracker, AgOS Crop Planning, Agrinavia, Agvance Grain, Dairyone Crop Management, FarmRexx, agCOMMANDER, etc.

https://doi.org/10.1016/j.atech.2024.100625
Received 20 August 2024; Received in revised form 18 October 2024; Accepted 24 October 2024
Available online 29 October 2024
2772-3755/© 2024 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

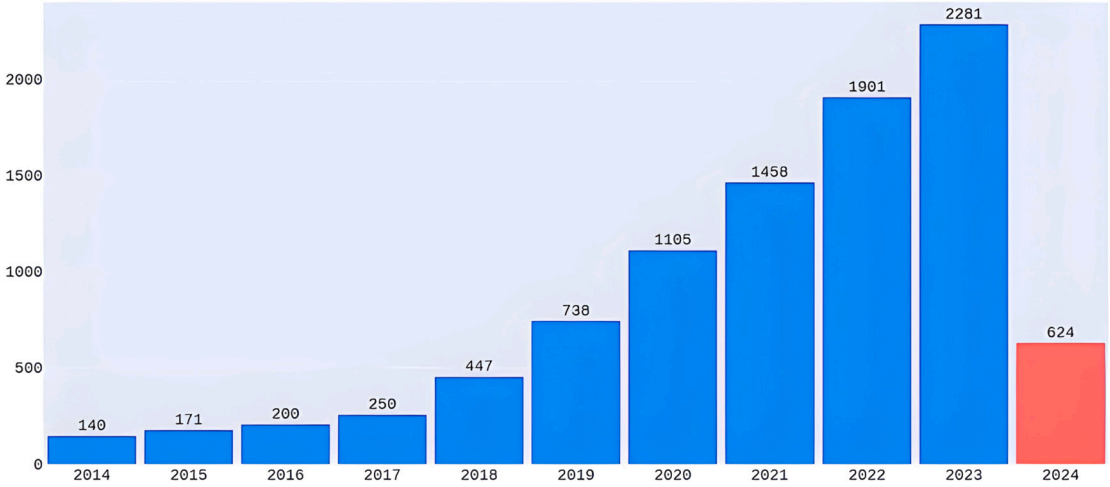


Fig. 1. Articles per year that matches the query. Source: Scopus. The last bar is updated to 1 Apr 2024. Query is the one written in Section 2.

Table 1
Table summarizing the surveys that we compared to our one. We highlighted the main topics.

Reference	Year	Core subject(s)
Zhang and Kovacs [18]	2012	UAVs in precision agriculture
Seng et al. [20]	2018	Computer vision and Machine Learning in Viticulture
Kamilaris and Prenafeta-Boldú [22]	2018	Deep Learning in Agriculture
Lu and Young [19]	2020	Public datasets in PA
Osco et al. [24]	2021	Deep learning techniques in UAV remote sensing
Bouguettaya et al. [21]	2022	Deep learning for crop classification in UAV imagery
Su et al. [23]	2022	AI empowered UAV perceptron systems for precision agriculture

[10–12] as it minimizes acquisition costs, allowing to obtain data anywhere on the planet with very large time coverage. However, due to the coarse resolution, while they are appropriate with large-scale observable phenomena [13,14], specific patterns of various plant species cannot be observed [15]; this aspect is critical for several challenges approached in literature and industry: species classification, yield prediction, specimen coverage ratio of target areas, biovolume estimation, instance counting, specific species monitoring are all examples of challenges that can only be accomplished by observing at Areas Of Interest (AOI) with higher resolution. The usage of UAVs, with the appropriate sensors/equipment, has proven to meet the necessary scale requirements while increasing acquisition costs sustainably [16]. Machine learning (specifically, deep learning) has been the state of the art in most computer vision tasks since the publication of Alex Krizhevsky's work in 2012 [17]. For all the reasons outlined so far, scientific literature regarding the use of remote sensing-acquired forestry and crop field images processed with deep-learning-based computer vision methods is very large, and it is continuously growing, attracting interest in all parts of the world as shown in Figs. 1, 3.

In this survey, we focus on UAV scale phenomenon, providing an overview of the state of the art of the methods used, the typical challenges that arise, the characteristics of the datasets and the cautions one must have when dealing with these use-case. The rest of this paper is structured as follows: In the following subsection we review recent surveys on topics close to the one proposed in this work, and highlight the novelty and merit of our proposal, Section 2 provide a general literature analysis for deep learning in PA & WFM, our article selection criteria and a *recurrent research pattern* observed in most articles. Section 3 describes all issues and cautions related to PA&WFM datasets and common data pre/post-processing methods. Section 4 contains an in-depth analysis of the various specific solutions (data-driven and non data-driven) proposed, while also providing a meta-analysis for benchmarking-oriented papers regarding machine learning in this domain. Finally, we end with Section 5 that provides an overview of the domain's potential, industrial outlets, and directions that will fuel this research area. At the end, to help the reader, we report in Tables 5, 6 and 7 a list of acronyms and abbreviations commonly used.

1.1. Previous surveys

Several surveys on those topics are already available in literature, but they differ in content and/or method from the work we propose, we recollect them in Table 1. One of the earliest works found in the literature is that of Zhang and Kovacs [18] which provide an exhaustive overview on application of UAVs in PA, however, at the time (2012) most of the research strands that led to modern deep-learning techniques, that now result state of the art, had not yet sufficiently matured. Lu and Young [19] proposed a survey that provides an overview of publicly available datasets for computer vision tasks in PA, with respect to this, our work also provides a detailed description of domain-specific models. Seng et al. [20] proposed a survey about computer vision and machine learning for viticulture; although it provides an accurate and comprehensive analysis of the state of the art, datasets and tasks, the latter are limited to ones that can be addressed with UGV (e.g. presence, evolution of diseases). Instead, we focus on UAVs and additionally provide a description of models. Bouguettaya et al. [21] proposed a survey on deep learning for classification of crops from UAV imagery; although this paper has some points in common with our article, our work covers a larger amount of tasks, challenges and models. Kamilaris and Prenafeta-Boldú [22] analyze 40 papers on deep learning for agriculture, covering different acquisition devices and machine learning techniques; with respect to this, we also discuss the WFM domain, and, since this work is more satellite oriented, it also deals with different categories

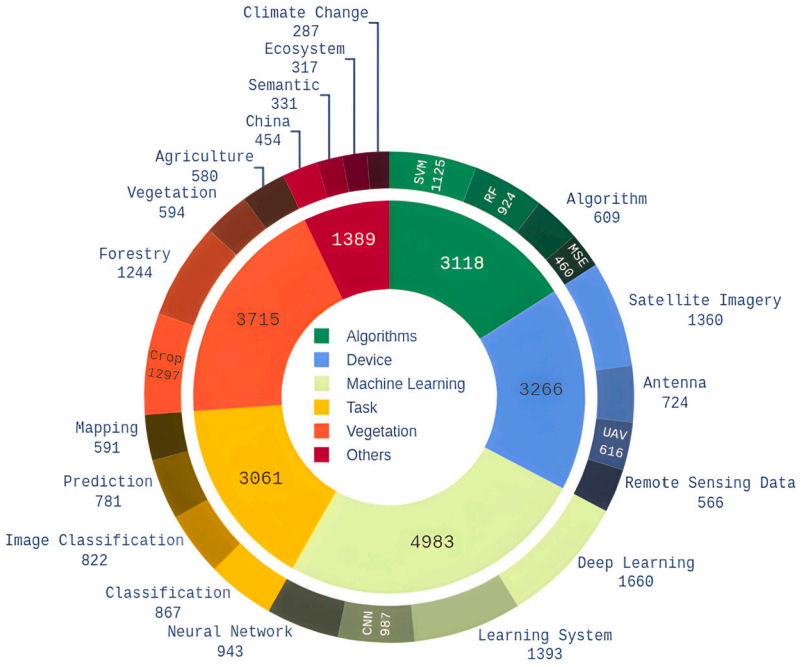


Fig. 2. Most recurrent indexing keywords associated to metadata of the query shown in Section 2. Source: Scopus. For each category, only the 4 most popular words are shown. Total results available in Table 2. Query is the one written in Section 2.

of models (e.g., recurrent networks) that are better suited to analyze time-sequenced satellite images. In Su et al. [23] the authors proposed a survey on AI empowered UAV perceptron systems for PA: although this work has many points in common with our paper, we also discuss WFM domain, and even if it provides a good overview of deep learning models for computer vision, our paper also provides an analysis of specific domain-tailored machine learning models and an in-depth analysis of dataset characteristics to consider during the data acquisition stage. Osco et al. [24] proposed a survey on deep learning techniques in UAV remote sensing with a strong focus on tasks, domains and models taxonomy. However, even if it provides a comprehensive and exhaustive overview, our work is PA & WFM centered and thus covers domain-specific models and concepts related to data acquisition and datasets. In conclusion, arguments that make our work useful despite the presence of previous valuable works are the following:

- The identification and schematization of a shared research framework/pattern emerging from the state-of-the-art of works dealing with UAV imagery in PA & WFM,
- The analysis and consideration on the trend of research and development of new data-driven methods, with some considerations and connections also with theoretical research and models proposed in other domains,
- The effort put into building a taxonomy of dataset characteristics, which we think can be very useful to those planning an image acquisition campaign, so that it can be managed in a way that fits the target goal,
- The level of detail in analyzing and comparing models specifically proposed for this domain.

2. Deep learning in PA & WFM: literature analysis

Given the large (and increasing) number of articles available in the literature (as shown in Fig. 1), we propose a prior screening of the Scopus meta-data. After performing this query:

```
TITLE-ABS-KEY ( ( ("deep learning" OR
                   "Artificial intelligence" OR
                   "neural network" OR
                   "machine learning" )
                  AND
                  ( "agriculture" OR "crop" OR
                  "farm" OR "trees" OR
                  "Forestry" OR "Vegetation" )
                   AND
                  ( "ugv" OR "remote sensing" ) ) )
```

we exported and processed the results as follows:

- Indexing keywords were counted and the 100 most used were chosen,
- We chose six semantic clusters and used them as a grouping criterion for keywords,
- We merged keywords pertaining to the same concept (and their count added up),
- We discarded very generic keywords with a very high count (e.g. “machine learning”).

Fig. 2 shows a summary of the most common words for each category, while Table 2 contains all the data. The semantic clusters are “Algorithms”, “Device”, “Machine Learning”, “Task”, “Vegetation” and “Others”. “Algorithms related keywords refer to specific algorithms (data driven and non-data driven) and technical details related to the algorithms itself. “SVM” turns out to be very popular keyword as it is one of the first widely successful machine learning algorithms used in many different domains [25]. Device related keywords concern all hardware devices that are used in the chosen domain. As mentioned earlier, “UAV” and “Satellite” are very popular in the literature but they are tailored to deal with phenomena on different scales. Machine learning related keyword pertain to all those machine learning concepts that are not specific

Table 2
Complete list of keywords from Scopus metadata. Query is the one written in Section 2. Semantic clusters ("Category" column) were created to group keywords. Acronyms are explained in Table 5.

(a)

Category	Keyword	Count
Algorithms	Decision Tree	1914
	SVM	1125
	RF	924
	Algorithm	609
	MSE	460
	Adaptive Boosting	191
Task	Classification	867
	Image Classification	822
	Prediction	781
	Mapping	591
	Semantic Segmentation	515
	Land Use	478
	Image Enhancement	418
	Image Processing	394
	Regression Analysis	371
	Spectroscopy	306
	Feature Extraction	286
	Accuracy Assessment	244
	Classification Accuracy	234
	Data Mining	233
	Time Series	232
	Reflection	208
	Nearest Neighbor Search	207
	Decision Making	201
	Object Detection	188
	Land Cover	171
	Aircraft Detection	164
Vegetation	Crop	1297
	Forestry	1244
	Vegetation	594
	Agriculture	580
	Soil	267
	Soil Moisture	250
	Vegetation Mapping	216
	VIs	194
	Biomass	193
	Cultivation	190
	Farm	188
	Crop Yield	188
	NDVI	180

(b)

Category	Keyword	Count
Device	Satellite Imagery	1360
	Antenna	724
	UAV	616
	Remote Sensing Data	566
	Landsat	455
	Sentinel	447
	Agricultural Robot	375
	Synthetic Aperture Radar	316
	Satellite	276
	Optical Radar	240
	Aerial Vehicle	196
	Infrared Device	194
	Radiometer	193
	Optical Remote Sensing	189
	Unmanned Vehicle	178
	Radar Imaging	175
	Modis	173
	Space Optic	164
Machine Learning	Deep Learning	1660
	Learning System	1393
	CNN	987
	Neural Network	943
	Learning Algorithm	798
	ANN	567
	Convolution	553
	AI	451
	Machine Learning Method	342
	DNN	339
Others	China	454
	Semantic	331
	Ecosystem	317
	Climate Change	287
	Pixel	287
	Extraction	283
	Food Supply	257
	Texture	235
	Image Resolution	205
	Dataset	204
	Hyperspectral	194
	USA	173
	Geology	172
	Article	166

or do not refer to unique models/concepts (otherwise they would fall under the Algorithm cluster). Since the quantity of articles has a strongly increasing trend each year (as shown in Fig. 1), most of the articles have been published in the last 5 years, and are those that use keywords related to more modern machine learning methods ("deep learning", "neural network"). Task related keywords refer to specific problems faced in the literature selected. Tasks that can be performed at different scales are most common, like "classification" or "semantic segmentation", while ones applicable to specific scales of resolution ("land use", "land coverage" for large scale, "object detection" on small scale) are also available at gradually decreasing counts. Vegetation related keywords have to do with concepts pertaining to the agricultural/vegetable domain. Lastly, the "Others" cluster was created by putting words with a non-negligible count and that do not fit into the other clusters. It is interesting how words pertaining to issues of common public interest, such as "climate change", "food supply", "ecosystem", and "geology", emerge in this category in addition to the two main competitors ("U.S.A." and "China").

2.1. Article selection criteria

We chose several articles among the results of the following query, made on Scopus:

```
TITLE-ABS-KEY ( ( ("deep learning" OR
                   "Artificial intelligence" OR
                   "neural network" OR
                   "machine learning" )
                  AND
                  ( "agriculture" OR "crop" OR
                  "farm" OR
                  "trees" OR "Forestry" OR
                  "Vegetation" )
                   AND
                  ( "UAV" ) ) )
```

We highlight that this query differs from that used in Section 2 only for the last line: this one selects only those items that rely on UAV as an acquisition device, while the previous one, as it aims to build the literature overview, selects both those that use UGV and remote sensing in general. From the results of this query, we chose 3 papers at a time according to the following criteria until we felt that the amount of material reviewed was enough:

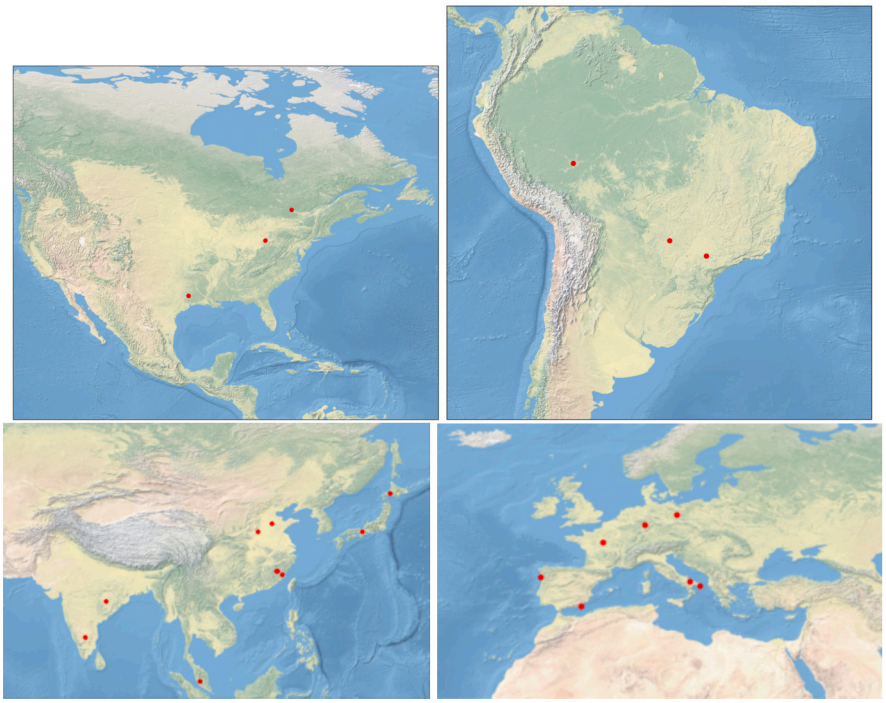


Fig. 3. Location of areas of interest observed in the works examined.

1. Sort the articles by citation and take the most cited (ever),
2. Filter articles from the last 5 years, sort them by citation and take the most cited,
3. Filter the articles from the last 2 years and make a content-based selection.

After careful reading and analysis of the articles iteratively selected using these criteria, we chose to mention those that best highlight state-of-the-art emerging topics discussed in Sections 2, 3 and 4. The inclusion and exclusion criteria (in addition to query, year of publication, and number of citations) are based primarily on the cutting edge of the implemented machine learning techniques and the thoroughness of both the benchmarking and the data acquisition campaign. In the core of the paper (Sections 2, 3, 4) we examine the chosen articles from different perspectives, and we summarize in Table 8 the results of our analysis.

2.2. Common research pipeline

In almost every article analyzed, we observe a recurring research framework: Each author focus his work in a specific geographic area (AOI) and particular specimens, depending on the use case (i.e. endangered plant species or industrial crops to be monitored) and therefore everyone provides for themselves in the construction of the dataset. In general, the research framework follows these steps:

1. The research team plans/performs flights over AOI in order to acquire orthomosaics,
2. Orthomosaics are split into tiles/patches and preprocessed,
3. Agronomists/domain experts assign the ground truth to each tile/patch.
4. A supervised machine learning model is trained with these tiles/patches.
5. Model is validated and its performances evaluated.

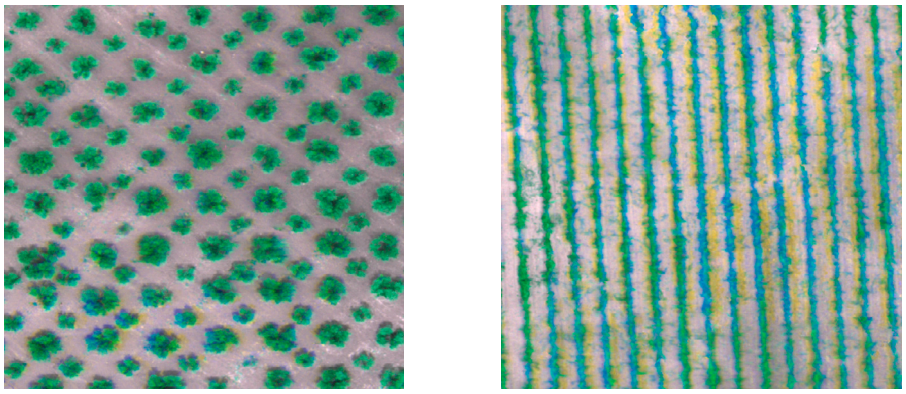
All articles roughly follow these steps, along with, more or less complex intermediate data processing stages. We discuss the first 3 steps in the following subsections.

2.2.1. Data acquisition stage

Before the initial image acquisition stage, authors schedule flights based on the need to capture certain characteristics of the AOI. We discuss these characteristics in Section 3.1. Behera et al. [26] are the only authors that use a public dataset, therefore, they did not need to perform the image acquisition autonomously. Some authors perform several flights of the same AOI to increase the representativeness of certain aspects (phenological state of the plant, different seasons, different light conditions). Egli and Höpke [15] adopt the Leave-Location-And-Time-Out procedure (acquire test images in different zones and periods than those used for training) as described by Meyer et al. [27], this method increases the reliability and representativeness of the data, enhancing the overall quality of the dataset, as it allows to minimize space-time correlation of the samples. Gurumurthy et al. [13] focus on capturing images of trees of variable ages, also using images with plants different from the target plant (mango), nevertheless, they choose to remove ambiguous samples. In [28,15] authors acquire samples with variable conditions such weather, light and phenological stages. Ye et al. [29] point particular attention to vegetation pattern distribution: authors identify 4 types of pattern distributions of olive trees and keep the proportion of these patterns balanced in both training and testing partitions. In general, in the context of industrial cultivation, crops of the same species have fairly regular distribution patterns, as shown in Fig. 4, for this reason, there are cases in which authors only acquire patterns strictly necessary in their research work [30,31]. In these works, target plants are regularly arranged, and since the goal is to count their number, acquiring additional images with specimens arranged in a different pattern may be redundant [31–33]. In some works [31,32,34], in addition to UAV flights, the research team performs local physical inspections on the plants in order to acquire ground truth of specific parameters (such as height and crown size).

2.2.2. Image tiling

Tiling is a mandatory step, regardless of the task, as it allows the global analysis of an orthomosaic to be broken down into several less complex tasks that focus on local features of relatively small areas. Moreover, from a technical point of view, it allows to limit the resources needed for training and inference. The use of pre-trained net-



(a) Olive (b) Vineyard

Fig. 4. Typical patterns of industrial crop. Source: TEBAKA project [35].

works/transfer learning introduce an additional constraint on the size of the tiles, as they must be coherent with that of the pre-trained model. Tiling is just one of the data pre-processing steps, which are rather numerous and heterogeneous in the literature. Several pre-processing methods are covered in Section 3.

2.2.3. Data labeling and ground truth

Data labeling is another one of those mandatory steps when collecting training data for supervised models, the shape of ground truth changes depending on the task, and is almost always collected by relying on manual labeling by domain experts. In this subsection, we describe works that approaches labeling in a nonstandard way, respect to the *vanilla* way one would expect with datasets of this type. Gurumurthy et al. [13] proposed two labeled versions of the same dataset to solve an instance segmentation problem: the individual crown detection in the target crop, using a double step fully convolutional semantic segmentation network. In the first version, they label samples with the "canopy/background" classes, while in the second they use three classes: "canopy/overlapping/background"; an U-Net inspired model is trained with the first version of the dataset and produces a score describing the probability that the given pixel belongs to a canopy. In the second step, they retrain the model by replacing the last layer so as to produce confidence scores according to the second version of the dataset: for the canopy, the background, and the overlapping surface between the canopies.

In [35] Epifani and Caruso propose an automatic labeling method implemented with several stacked traditional computer vision techniques (erosion, dilation, fast non local means denoising and thresholding) applied to the NDVI channel. In this way, they significantly reduce the time it takes to label the dataset, limiting human intervention to the manual tuning of the pipeline and to a quick visual inspection of the labeled samples. Ground truth is not always represented by truth masks associated with an image, and in some cases, it is related to physical/chemical properties of vegetation acquired with local inspections. Hao et al. [33] propose a method to detect crowns for Chinese firs and estimate the height of each specimen, therefore they measure height for 265 trees, selected in order to form a reliable distribution set respect to height. Safonova et al. [31] use the results of instance segmentation to estimate the height of olive trees and their biovolume, therefore they had to measure those values for 6 different trees. In this case, few samples are sufficient, as the final estimation of biovolume and height is not performed with data-driven methods but with traditional algorithms. Yu et al. [36] measure several physical values of the plants in their AOI. After that, they develop (using images and these physical values) several regression methods for estimation of the Above Ground Biomass (AGB) of plants. In our literature analysis, the only completely unsupervised system is the one proposed by [37], even though they still use ground truth to compare their unsupervised system and several, state of the art, supervised models for semantic segmentation.

3. UAV datasets for agroforestry: features, methods and issues

In this section, we describe some characteristics of the datasets that must be known to properly schedule and perform the images acquisition campaign. We classify these characteristics into Reference Distribution (RD) and technical characteristics of samples (discussed respectively in Section 3.1, 3.2), and both represented and schematized in Fig. 5. By *dataset* we mean a list of tiles - i.e. - a "sample" means a single tile and a "batch" a list of tiles extracted from the dataset for a training cycle. Generally, in the use case analyzed, the role of the model is to infer some kind of *knowledge* about each sample given as input. It is reasonable to assume that these samples come from an unknown probability RD, on which the dataset represents a statistical sampling. In PA & WFM domain, RD have recurring characteristics for which it is required to ensure their representativeness in the dataset (and therefore, also during flights planning and the samples selection). The RD characteristics are the set of all these characteristics, and we describe them in the following subsection. During flights planning, the effort required in ensuring the representativeness of RD characteristics varies according to the use case, and it is often necessary to find a tradeoff between the model's expected ability to generalize, and the effort spent in the data acquisition. Technical characteristics of the samples also influence the quality of the dataset, and they usually are homogeneous within the same dataset.

3.1. Reference distribution characteristics in agroforestry

The RD directly influence the effort required for data acquisition and the design, training and inference phases of the model. In our meta-analysis, we identify three main categories of RD characteristics:

- Positioning: everything related with the positioning of instances (plant specimen) within tiles. The variability of the positioning patterns and whether or not the canopies overlap are the two most important information. Canopies overlap makes the biggest difference in the difficulty to detect different instances belonging to the same semantic class [13,29].
- Organic: All features influencing the biophysical status and appearance of specimens. Vegetation appearance (even in the same species) can change more or less drastically according to several factors (i.e. irrigation treatment supplied to plants [34], fertilizers [36], phenological status [15] etc).
- Environmental: All characteristics that influence the context around the instances (weather, background, lighting conditions and so on).

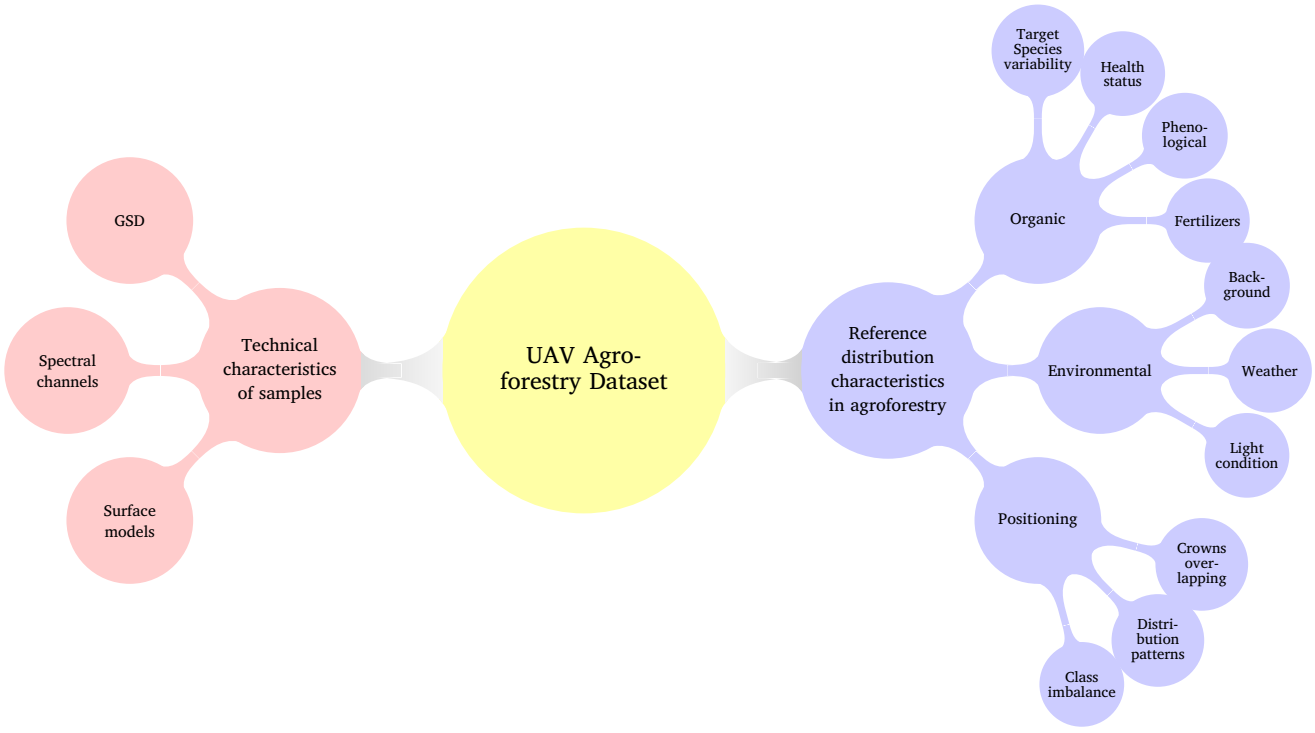


Fig. 5. Diagram representing the taxonomy of the datasets characteristics we proposed in this article.

Characteristics in the last category are more generic and well known even outside the PA & WFM literature. The representativeness of all these 3 groups of features should be managed in the scheduling phase of image acquisition, and many authors take this into account during the flight planning, as discussed in Section 2.2.1 [15,29,28,16,13]. We can find some exceptions, for example Ye et al. [29] propose a data augmentation procedure to simulate plant color variation (occurring at different phenological stages) and light conditions. They transform the color space of the original dataset from the RGB model to the Hue Saturation Value (HSV) mode, measuring the variation range of these parameters inside the dataset. Therefore, during data augmentation, they add noise to the HSV values, always keeping the parameters of the augmented tiles within the previously estimated range. Yu et al. [36] exploit the HSV space in similar way. Background also plays an important role, as it contributes directly to the disorder/noise available in the tiles, as evidenced by the fact that it is more difficult to classify tree species in an urban context than in an agricultural one [16].

3.2. Technical characteristics of samples and augmentation

Technical aspects of the acquired samples play a key role in the behavior of the overall systems (not only data drive ones), as they directly affect the quantity and the quality of information conveyed by each tile/patch. The most important are the GSD (Ground Sampling Distance), type/number of spectral channels available and the presence/absence of surface models.

3.2.1. Ground sample distance

As expected, the GSD plays a crucial role. Across the UAV imagery literature for PA&WFM, the GSD of the orthomosaics varies at most in the range $[10^{-1}, 10^{1}]$ cm^2/pixel. In the "GSD" column of Table 8 we reported this value for each paper analyzed. Many authors focus on this parameter trying to improve it or studying how it affects the overall performances. Ye et al. [29] implement an Enhanced Super-Resolution Generative Adversarial Network (ESRGAN [38]), that is a super resolution imaging technique, capable of generating realistic textures during single image super-resolution. In this way they increase the resolution of their samples. This technique has already proven interesting results in remote sensing [39]. In [15], Egli et al. perform 10 downsampling runs by means of bilinear interpolation on the test tiles, starting from 0.27 cm/pixel and reaching 54.78 cm/pixel, a resolution in which each tile corresponds to a single pixel (in RGB, this correspond to 3 scalar values, 1 per channel). Given that their classification task is a 4-class one and that they achieve a 56% accuracy with the lowest tile resolution, authors conclude that the mean spectral values of the tiles already provide an explanatory content of 31%, assuming a base rate accuracy of 25%. In [40], Zhang et al. lower the resolution of the dataset using blurring kernels of several sizes (3,5 and 7). They assess the impact of the resolution by training their DefoNet model on these new datasets observing a considerable drop in performance. While precision has a better tolerance over resolution loss, recall decreased from 80.6% to 73.8% (with a 7×7 kernel). Guirado et al. [14] use a dataset with tiles coming from images acquired by different devices (over the same AOI), namely satellite (Google earth) an helicopter and UAV. The resolutions are 50,10 and 3 cm^2/pixel respectively.

3.2.2. Spectral bands and indexes

Acquisition costs heavily depend on the sensors with which UAVs are equipped: hyper-spectral/multi-spectral acquisition systems have significantly higher costs but are able to provide more information about vegetation than RGB cameras [36,32] especially in the Near Infra Red (NIR) and red edge channels. In Table 8, in the column "Sensors" we summarize bands used in the literature reviewed. However, the information contained in the human observable patterns/spectrum at a resolution on the order of centimeters has proven to be sufficient to solve the computer vision tasks required for the high level goals mentioned so far, and therefore RGB cameras, are a good cost/result trade-off in many cases [41]. Extra-RGB bands are very useful in the context of vegetation identification as they carry a lot of information related to plants [15,36,32]. In fact, various spectral indices are known in the literature to be obtained as a linear combination of other channels [42], for which correlation with the presence and vital parameters of vegetation has been widely demonstrated. In [34,36] authors make intensive use of these indexes 12 and 29 respectively. Epifani and Caruso in [35] exploit the proper-

ties of NDVI for automatic tile labeling. However, since these indexes do not take spatial correlation into account, being limited to the information independently conveyed by each pixel, they had to include intermediate stages of image processing based on classic computer vision methods. Some authors that use other spectral bands in addition to the RGB, such as NIR and Infrared [35,31,30,33] test the variation in model performance by using different subsets of available bands, for example Safonova et al. [31] perform several cross validation training runs with different combination of data fusion performed between RGB and spectral indices maps like NDVI and GNDVI. We provide some of these spectral indexes in Table 7.

3.2.3. Surface models

Surface models convey a different type of information respect to spectral bands, and they can be exploited in different stages of the research pipeline. In column "Surface Models" of Table 8, we specify which papers use this type of data and which do not. Egli and Höpke [15] exploit the DSM (Digital Rurface Model) for tree marking in order to schedule flights in an optimized way. In [41,43], authors exploit the DSM to carry out the object proposal stage for the following image classification CNN, resulting in an instance segmentation system.

Hao et al. [33] have at their disposal 5 spectral channels (RGB, red edge and NIR) and 2 surface models: the DSM and the Canopy Height Model. They perform several cross-validated training runs of a Mask R-CNN using different combination of spectral channels and surface models. Therefore they use the surface models as a raw additional channel for the input tiles.

Before being fed as input to the models, UAV images/tiles are processed in several ways, depending on how authors interpret the problem and how the model input is expected to be. There are many preprocessing techniques that are recurrent in several works exploiting the use of surface models. Simple Linear Iterative Clustering is an unsupervised K-means clustering superpixel generation method used in [37,44]. Superpixels are small cluster of pixels that share similar properties, allowing to simplify images with a great number of pixels reducing the effort required in subsequent stages. In [43,36,37] authors use Marker Controlled Watershed Algorithm. This method is commonly used together with local maxima filtering. Both algorithms perform well in canopy identification when applied to surface models, indeed they can be used as object proposal upstream stage. Gaussian filters, erosion and dilations are a popular choice to remove artifacts and spurious elements in images or spectral maps [35,37,45].

4. Deep learning models and benchmarking

In this section, we analyze the solution proposed in the PA & WFM domain, also describing the connections with the rest of the reference deep learning and computer vision literature. All the use-case faced in the agroforestry literature reviewed require to elaborate captured raw images, in order to extract valuable additional information. When the goal is the prediction of a specific dependent variable, data coming from UAV sensors can be processed with more or less complex regression methods. In these cases, independent variables can be extracted by feature engineering (classical method) or with a black box deep learning based approach. In addition to these regression methods, four computer vision tasks were identified, and all together they cover all the use case addressed in the literature reviewed. These 5 categories of tasks represent the core part where most of the valuable information is extracted/processed. The four task of computer vision mentioned above are: image classification, semantic segmentation, instance segmentation and object detection. In the literature, different definitions of the same tasks can be found. In this paper we are using the taxonomy proposed by X. Shen in [46] as baseline, represented in Fig. 6.

In addition, we denote by *object proposal* the upstream task that deals with the identification of bounding boxes of potential objects of interest, without the class label. This task is described in several works dealing with object detection and instance segmentation problems [55]. In our analysis It is useful and consistent with the PA & WFM literature, the evolving state of the art, and from a practical point of view to merge object detection and instance segmentation tasks. A well known model for offline instance segmentation (Mask R-CNN [51]) was developed from a set of models dedicated to object detection (Region Convolutional Neural Networks, R-CNN [56–58]) on which there was a rather fruitful line of research. Moreover, for several tasks approached in the literature reviewed, there is no difference in using an object detection or an instance segmentation network (for example, for counting specific specimen). Therefore, we use the acronym IS/OD to denote the tasks of instance segmentation e object detection. With the advent machine learning, state of the art in these four tasks is constantly evolving as they are milestone problems for computer vision. Increasingly high-performance architectures are being proposed as time goes on, often refining solutions proposed by other researchers. State of the art models are designed as general purpose solutions [59,53,51,47], so they are a baseline for those who want to solve those tasks on specific domains. Exploiting domain specificities by translating them into adjustments/variations on the model or methodology enables better results and less parameters compared to the *vanilla* versions, at the cost of developing domain-specific solutions, losing cross-domain versatility. In Table 3, we summarize the models for which the number of parameters was available in the reference paper (or we calculated them ourselves from the structure); we also added the number of parameters of the general purpose models for the reference task. In analyzing the models proposed in the UAV imagery for PA&WFM literature, we identify 3 main categories of works:

- those proposing variants of pre-existing or traditional models used in a domain-adapted manner (e.g. making ensembles, or changing some processing stages),
- those proposing novel (though sometimes inspired by well-known models) deep networks to solve specific tasks,
- those focusing on benchmarking and comparing multiple known models (sometimes even the one proposed in the article itself) on a reference dataset,

These categories should not be taken cleanly, as sometimes, works may fall into more than one category, but they have been chosen to facilitate the exposition of the results of our analysis.

Table 3
Number of parameters of different models, categorized by task. For some architectures, parameter numbers are variable. For others, we computed from the schematics provided by the authors. For CD-CNN this was not possible.

Task	Model	# Parameters (10^6)
Semantic Segmentation	[26] LW Aerial Segnet	2.48
	[13] Mango Tree Net	0.663
	[47] U-Net	10 to 30
	[48] SegNet	10 to 20
IS/OD	[30] Osco et al.	0.009+0.0031*t
	[49] Improved Yolo	14.5
	[50] SEYOLOX	5.04
	[51] Mask R-CNN	63.7
Image Classification	[15] Egli et al.	2.98
	[40] DefoNet	4.43
	[52] CD-CNN	Unknown
	[53] ResNet-50	23
	[54] VGG-16	138

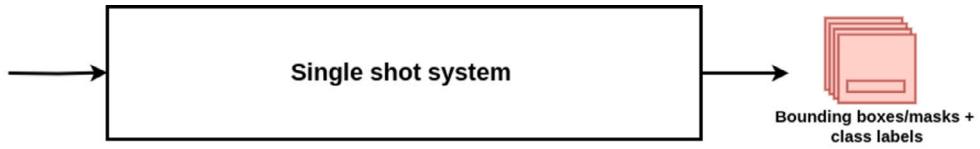
4.1. Variants for object proposal and traditional methods

In this section, we describe domain specific methods found in PA & WFM, proposed to carry out the instance IS/OD task. Fig. 7a) and Fig. 7b) show the approach of the two most popular general purpose

(a) Image classification (b) Object detection (c) Semantic segmentation (d) Instance segmentation

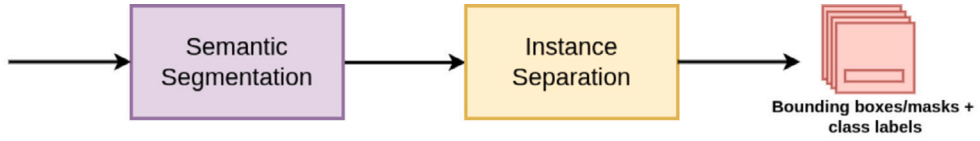
Fig. 6. Computer vision tasks solved by architectures presented in this work. Image taken from [46].

(a) Standard two stage architectures, e.g. R-CNN [56, 58, 57, 51] family.



(b) Standard single shot architectures, e.g. YOLO [60, 61, 62, 63] family.

(c) Instance segmentation carried out performing an upstream object proposal (non-data driven) followed by an element-wise image classification task.



(d) Instance segmentation carried out by separating instances from the output of a semantic segmentation system.

Fig. 7. A summary of the various approaches found in literature (for agroforestry domain) used to carry out IS/OD.

IS/OD architectures, namely the Region Convolutional Neural Networks [56,58,57,51] and the YOLO [60–63] family. In many works analyzed, data driven components don't operate in an end to end fashion: the usage of machine learning is limited only to certain stages of the pipeline, while the others are implemented with non data-driven methods. Many authors perform object proposal and image classification separately in order to carry out IS/OD, where only the second step is carried out with data-driven methods (as shown in Fig. 7c). Onishi and Ise in [41] propose a machine vision system for a 7 class CNN based classificator for tree crowns. In this article, they use a DSM to obtain a slope model. After that, this slope model, the DSM itself and the RGB spectral bands are fed together to the eCognition Developer software, that returns the patches containing the crowns. In this use case, the eCognition software behaves as object proposal upstream stage. After manually labeling the patches in the 7 different classes, they train and compare 4 different standard image classification networks: VGG-16, ResNet-18, ResNet-152 and AlexNet.

Xiao et al. [64] besides a fully data driven object detection pipeline based on YOLO v5, propose a method based on the Otsu algorithm [65] applied on DSM and vegetation indices (NDVI and NDRE) in order to separate instances of corn trees. Natesan et al. [43] propose an object proposal pipeline based on classical computer vision techniques. They apply Gaussian Smoothing on the DSM and then marker controlled watershed segmentation algorithm, obtaining patches of tree crowns (as in [66]). Forestry experts provide ground truth labels for each patch extracted with this method, and authors train with these data (extracted patches and ground truth) a ResNet50 for patches classification. Ferreira et al. [45] perform a downstream instances separation, after carrying out semantic segmentation with a Deeplab v3+. They propose a post-processing pipeline based on morphological operations (such as dilation and erosion) to force separation between nearby group of pixels belonging to the same semantic class. In this way they count the instances of each class. This is an example of the pattern shown in the Fig. 7d. Epifani and Caruso in [35] also use erosion and dilation (together with thresholding methods and fast non-local means denoising algorithm) but they implement them in a refinement process of the NDVI channel in order to use it as ground truth. In this way, they significantly reduce the time it takes to label the dataset, limiting human intervention to a quick visual inspection of the labeled samples. Fei et al. [34] use 180 wheat plots of fixed size as a sample, and for each plot they measure the yield, using it as ground truth. They also divide these 180 plots into 3 groups, each one irrigated in a different way (low, moderate, high). Finally, the conduct UAV flyovers with different types of sensors (thermal, multi-spectral). From these data, they extract 29 different types of well known spectral

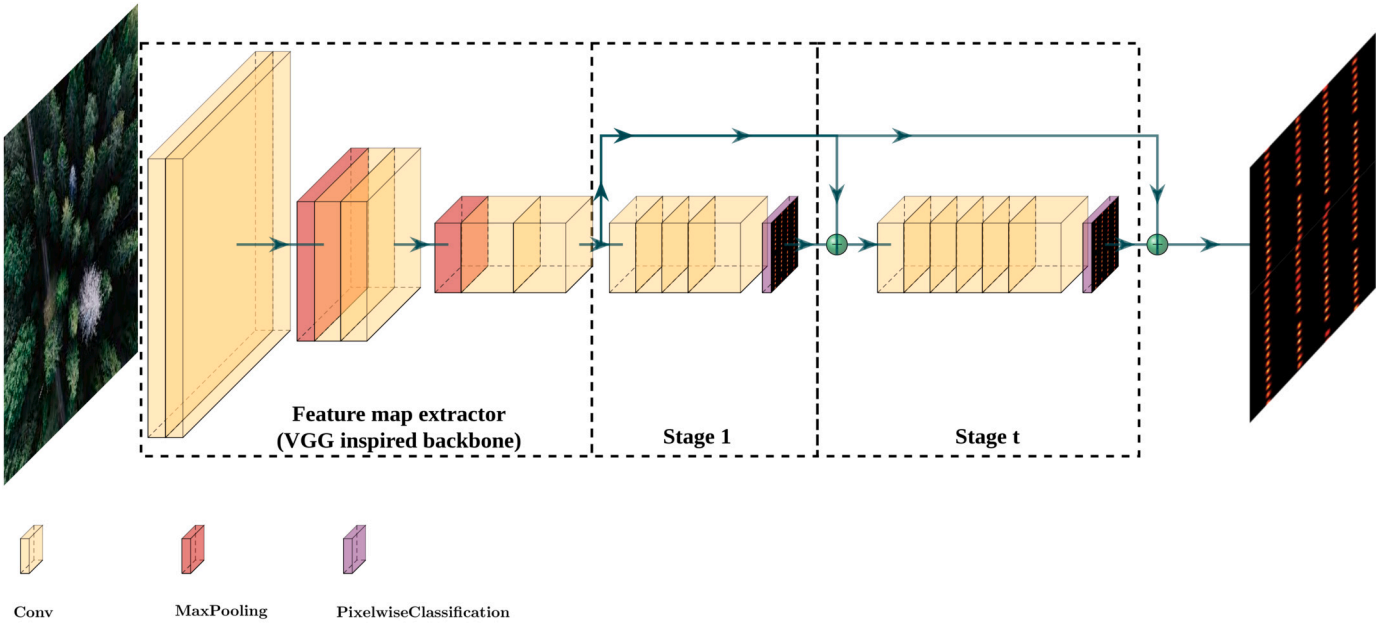


Fig. 8. Network proposed by Osco et al. [30]. Reprocessing of the image in the original article. The number of t stages is variable; the authors train networks with different numbers in t stages in the head of the network. As t increases, the cost of training/inference increases linearly, while performance gradually saturates.

indexes (some of them described in Section 3.2.2). The model is an ensemble of 5 different classical machine learning algorithms. They feed each sub-models is with the features, and it returns a yield estimate. In this work authors do not address any computer vision tasks of those previously exposed, but they extract correlation patterns between the handcrafted features (spectral indexes) and the variable to be inferred (yield).

4.2. Domain-tailored neural networks

Several authors propose their own architecture that has usually been built starting from a well known model, or by developing a new one from scratch. Osco et al. [30] propose a new object detection architecture: their final goal was to estimate the number of Valencia-orange trees inside a region of interest. They generate ground truth mask associated with each UAV image by placing a Gaussian kernel centered on each tree. Their architecture, represented in Fig. 8, has 3 parts: the encoder, the segmentation stage 1, and a variable length tail. The tail is made by several stages similar to the stage 1. Authors refer to the output of the encoder as F. The first segmentation stage (S_1) performs several convolutions (without reducing the output size), with the last two of size 1×1. The output of this part is compared with the ground truth map (rescaled, in order to achieve the same resolution) using a loss function. A generic segmentation stage t (called S_t with $t \geq 2$) behaves in the same way as S_1, (it applies more convolutions with different settings) and returns an output that still has the same height and width of the ground truth map. The input of the S_t is the concatenation between the output of stage S_{t-1} and the output F of the encoder. With this solution, authors test the architecture with an increasing complexity by only tuning a single hyperparameter (i.e. the number of tail stages t_{max}). They also perform loss computation and back propagation on every segmentation map produced in each stage. This is coherent to the idea behind the residual block, that manages to solve the gradient vanish side effect.

Egli and Höpke [15] propose a multi-class 4-layer lightweight CNN classifier. Compared to state-of-the-art classifiers, this solution is simpler and the one with fewer parameters. It consists of four consecutive convolution/pooling blocks (conv1 to conv4), a fully connected layer (fc) and a final output layer that maps the four tree species considered in their study. The original schema of this solution is shown in Fig. 9. This network is fed with tiles obtained from the original orthomosaic.

They extract tiles following a uniform grid. In this way, by assigning a class to each tile, the model produces a semantic map of the orthomosaic. Although the map is coarse compared to other works, this work is coherent with the idea of performing a fully automated online classification over an area, relying on IoT devices with limited computing power. In [26], Behera et al. propose a lightweight segmentation network called Lightweight Aerial SegNet. The final goal behind this work is to propose a fast and light solution that can be deployed on the internet of things edge devices to perform real time segmentation. This solution is an improvement of another network previously proposed by the same authors in another paper [67]. LW Aerial SegNet architecture implements depthwise separable convolution, which reduces the number of parameters compared to classical convolution. The authors compare several state-of-the-art architectures. Although their solution has fewer parameters, the performance is similar. Their solution has only 2.48 M parameters, while all the other networks compared have between 9.42 M and 16 M parameters. The structure of the network is shown in Fig. 10, it is similar to other well known architectures [47,48]: it has the encoding/decoding paths whose stages communicate specularly with the skip connections.

Gurumurthy et al. [13] in their model, claim to have been inspired to the hour-glass shape of the U-Net. They used larger kernels (up to 7×7) in order to increase the receptive field of the layers. A diagram of this network is shown in Fig. 11.

In [49] Junos et al. propose a system for the automatic detection of palm oil fruits from UAV images. They propose a variation of YOLO v3 tiny, introducing Densenet as feature extractor, a feature Pyramid Network as a multi-scale target detection network and a learnable Swish activation function. The diagram of this network is shown in Fig. 12. All these changes increase the complexity of the model compared to Yolo v3 Tiny, but it still retains a quarter as many parameters as standard Yolo v3. In their tests, they compare the proposed model with several object detection architectures. Zhang et al. [40] explore the usage of several machine learning techniques to describe plant defoliation from UAV images. They test both classical techniques (Naive Bayes, KNN, RF, SVM, Gaussian Process) and those based on deep learning for image classification (VGG-16 and ResNet-50). Finally, the authors propose DefoNet

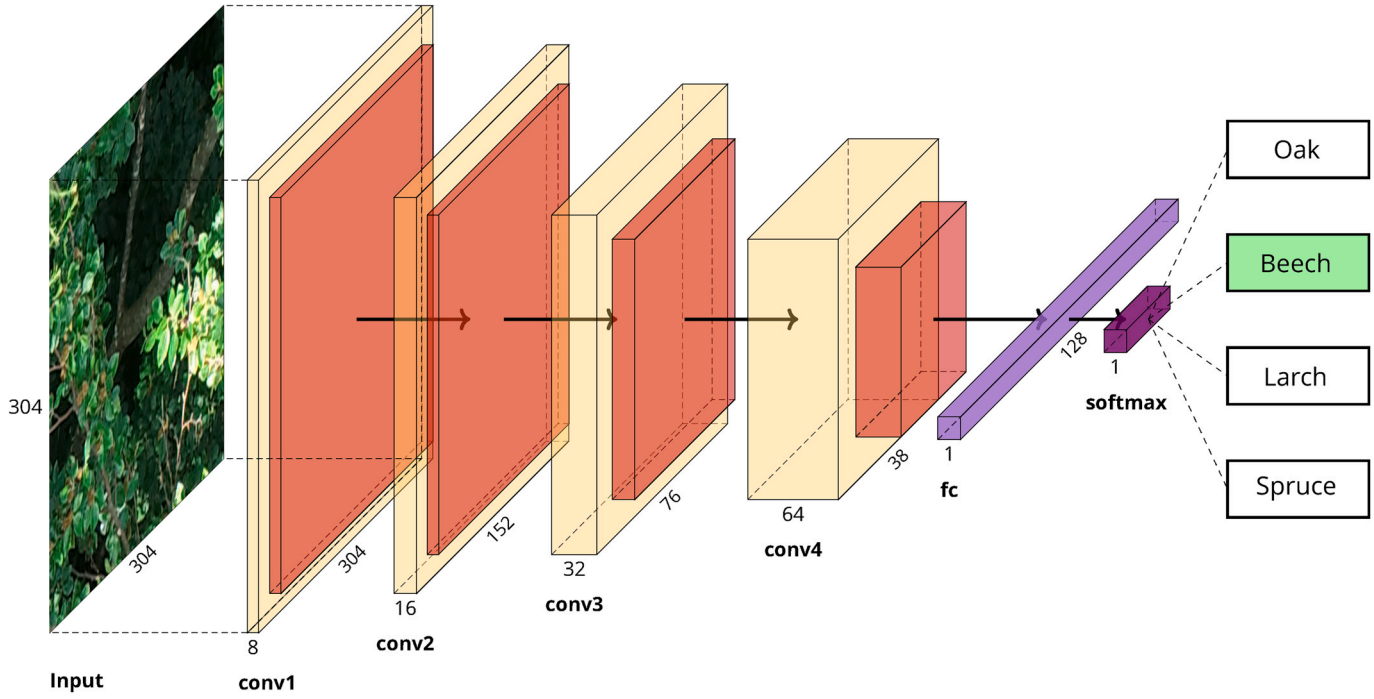


Fig. 9. Network proposed by Egli et Al. Image from [15].

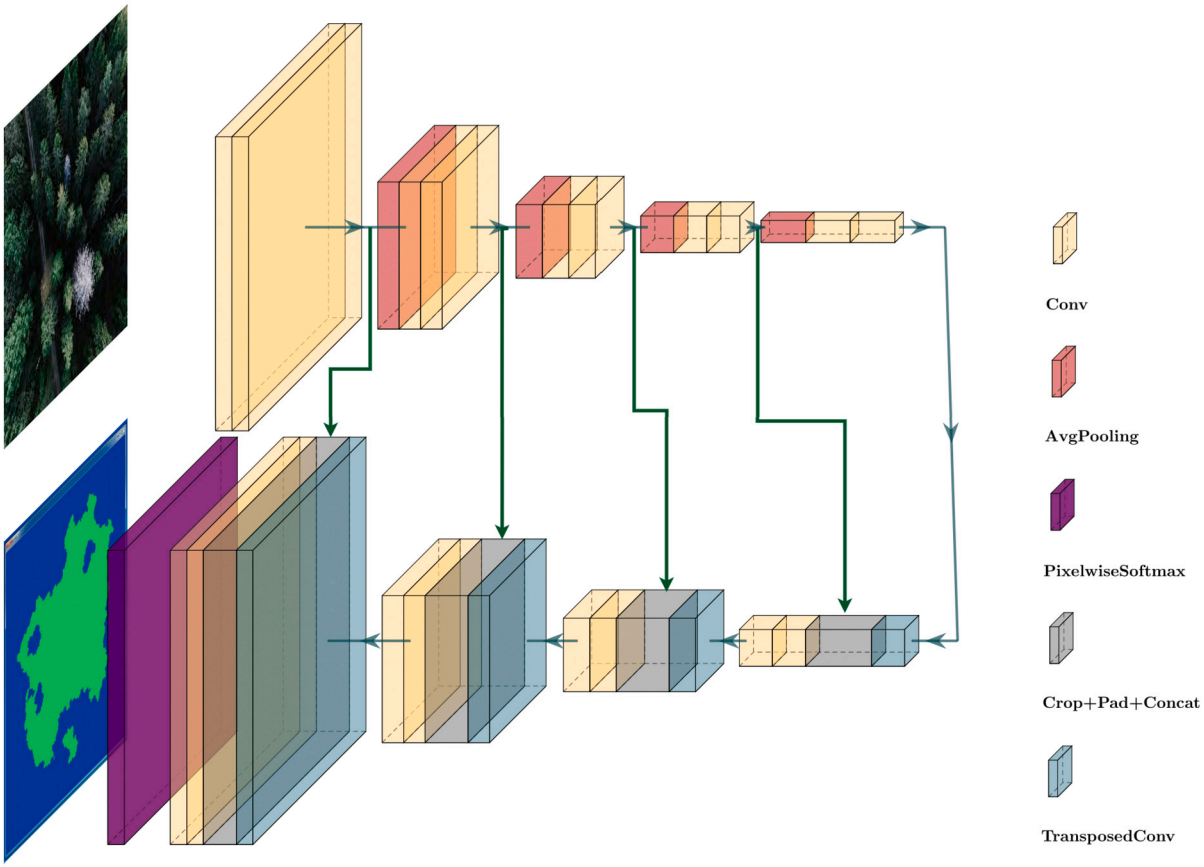


Fig. 10. Lightweight aerial SegNet proposed by Behera et al. Reprocessing of the image in the original article [26].

(whose diagram is shown in Fig. 13), a CNN network based on the classic LeNet network.

DefoNet performs better than all the other solutions and it is 6 to 8 times faster to train compared to ResNet and VGG. In [52] Pandey and Jain propose an architecture called Conjugated Dense Convolutional Neural Network (CD-CNN) for the classification of crop patches acquired by UAV. Authors claim that this solution achieves a strong distinguishing capability between the 5 proposed crop types. However, it is not possible

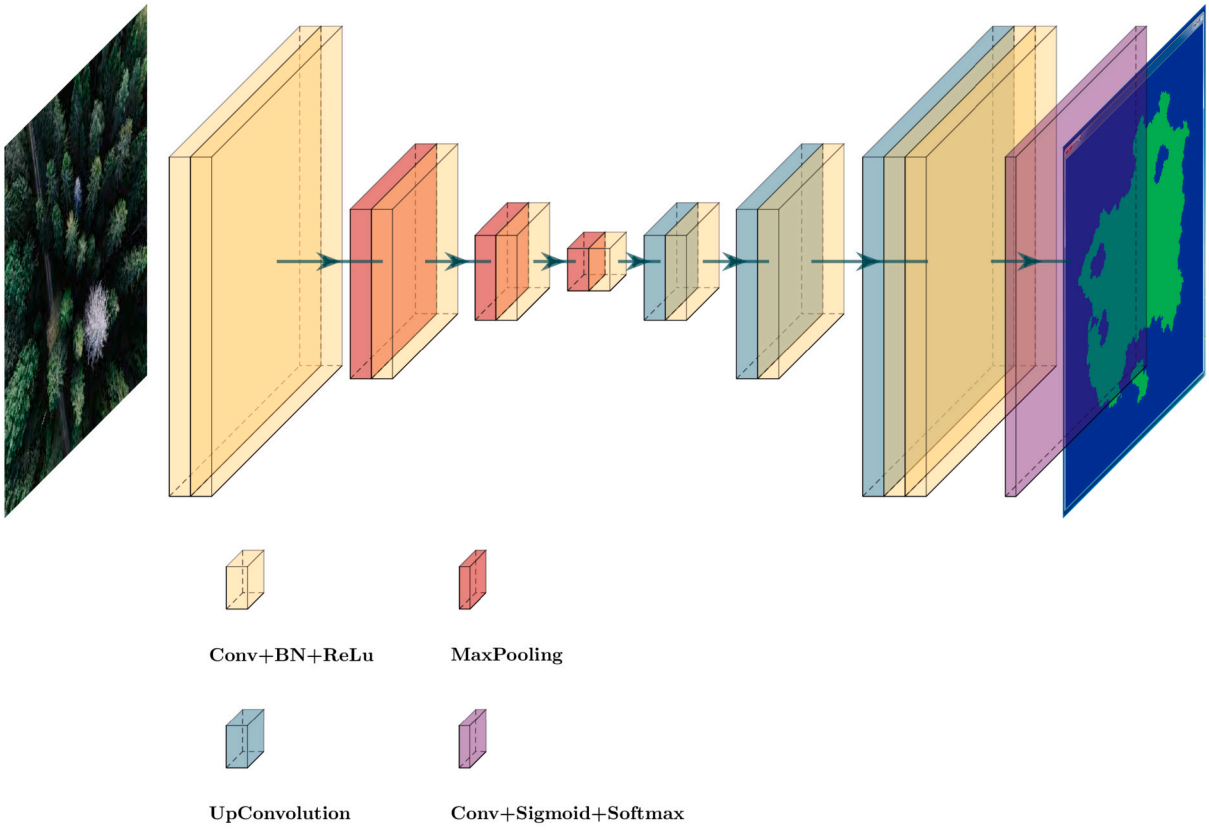


Fig. 11. Mango Tree net architecture. Reprocessing of the image in the original article [13]. In the first network training, the last softmax outputs a pixel-wise value associated with the presence of trees (2 classes, tree or background). In the retraining, it outputs a probability for each of the 3 classes: tree - canopy overlap - background.

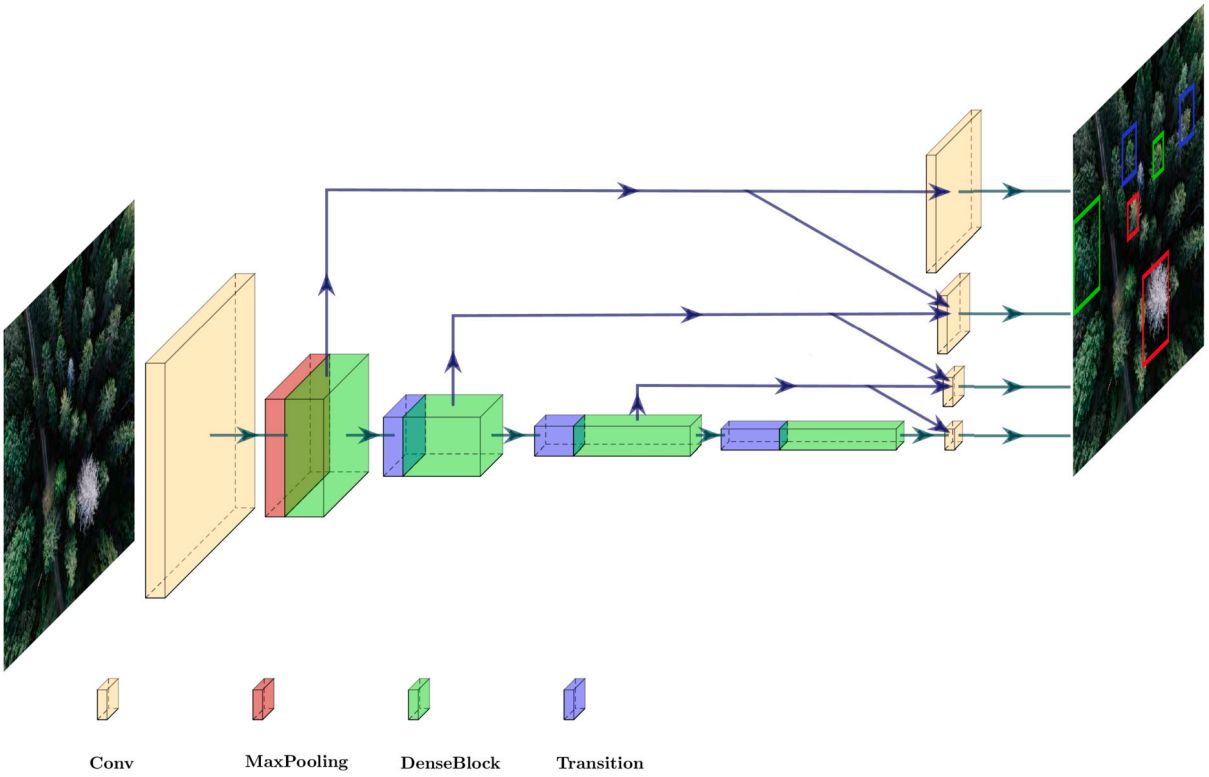


Fig. 12. Improved Yolo V3 Tiny. Reprocessing of the image in the original article [49]. The structure of dense blocks is non-trivial, and it is described in the original article.

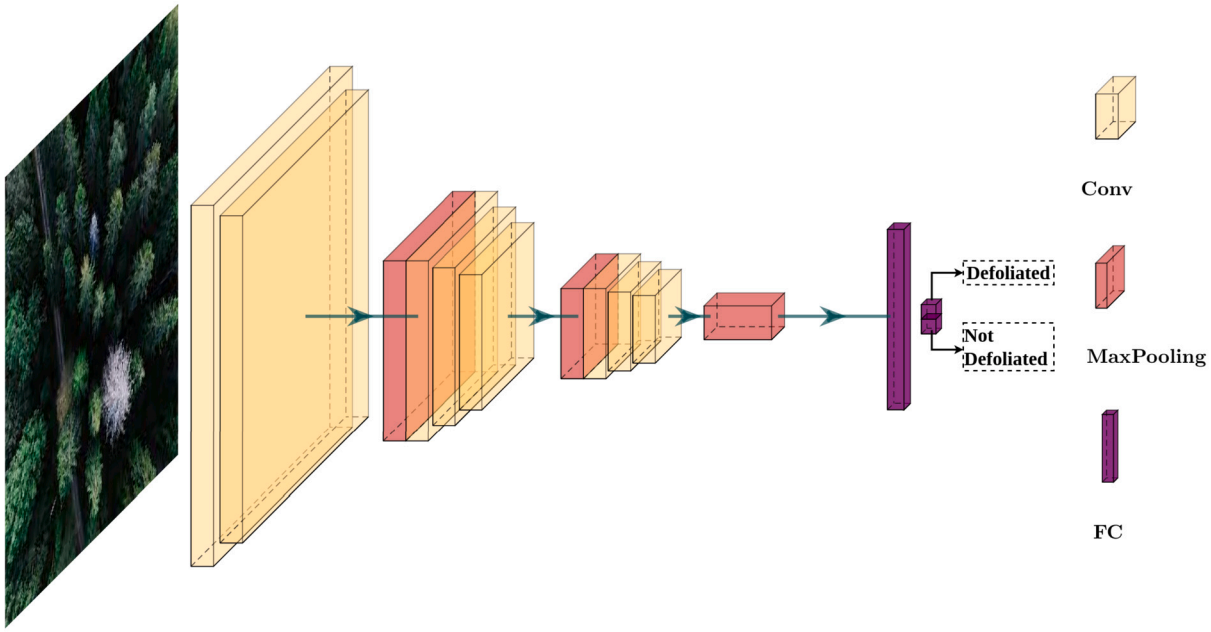


Fig. 13. DefoNet schema. Reprocessing of the image in the original article [40].

to estimate the number of parameters in this model, and the authors do not provide any information on the complexity of the architecture. The CD-CNN uses the following components in addition to the standard convolutions:

- Dense blocks: it is a module widely used in CNN that connects all layers (with matching feature-map sizes) directly with each other. It was originally proposed by Huang et al. [68],
- Conversion blocks: this block is placed between two dense blocks and it consists of a convolutional layer and a 2×2 max pooling operation,
- Composite activation function blocks: this block performs a batch normalization, a convolution and applies the SL-ReLu activation function, originally proposed by Wang et al. [69].

A schema of CD-CNN is shown in Fig. 14. D. Yu et al. in Yu et al. [36] propose a neural network regression methods (called DCNN, whose schema is in Fig. 15) and compare it with several other data driven regression methods.

This network ends with a single regressor neuron whose output represents the variable they want to estimate, that is the AGB. Song et al. [50] propose many variations of the YOLOX-tiny [70] model, the best of which is the one that is called SEYOLOX-tiny. This model applies 2 main modifications to YOLOX-tiny:

- It adds 2 Squeeze-and-Excitation(SE) [71] modules to YOLOX-tiny,
- It replaces the Varifocal loss [72] with binary cross entropy loss.

The authors justify these modifications because the baseline model is not very good for detecting small objects (maize tassels are). It inherits many different modules like Spatial Pyramid Pooling and Cross Stage Partial module [73]. The model of SEYOLOX-tiny is shown in Fig. 16.

Jaimes et al. in [37] propose a multi-stage fully parameterless and unsupervised method for semantic segmentation of UAV images. They compare their method with Segnet and DeepLab, resulting in slightly lower performance. The biggest advantage thus lies in its unsupervised and parameterless nature.

4.3. Vision transformers

A final mention to models belonging to the vision transformer (ViT) family in necessary. These architectures implement the attention mechanism in image processing. Although they have comparable performance with classical convolutional models like U-Net, they are much more complex and require more data to be trained properly [74,75]. Research is still fervent in this direction, but the real potential of these methods is still being explored, they still have to prove to have a better cost-effectiveness than convolutional models. Despite this, the attention mechanism outperforms all other known methods in processing series of tokens (the biggest example is the field of natural language processing). Some hybrid tasks, for example between image and token processing (e.g., environmental change detection from image series) can be approached with attention-based methods with interesting results [76,77]. Doing this requires large amounts of images acquired at different time, which is why satellite imagery is better suited than UAVs in this context.

4.4. Benchmarking

As described in the introduction, in our literature analysis, machine learning models are used to solve intermediate computer vision tasks necessary to extract valuable information, with the final goal to solve more complex tasks. Several works focus on benchmarking and comparing different deep learning and classic methods. Lobo Torres et al. [16] propose a study in which they compared five deep fully convolutional networks: SegNet, U-Net, FC-DenseNet, and two DeepLabv3+ with different backbones. They also test the Conditional Random Field postprocessing technique to improve the overall segmentation results. Guirado et al. [14] evaluate the performance of 2 instance segmentation techniques and propose an ensemble model that fuses both of them. The first technique is the Object-based Image Analysis (OBIA), the second one is the Mask R-CNN with a ResNet-101 backbone. Their work shows that the performance of the ensemble is superior to that of the two separate models and it improves as the resolution increases (up to 25%). dos Santos et al. [28] propose an object detection system for UAV imagery for the detection of law protected tree species (*Dipteryx alata*) in urban environment. Their system relies on an instance segmentation stage, which they implement by comparing the performance of 3 different CNN: a Faster

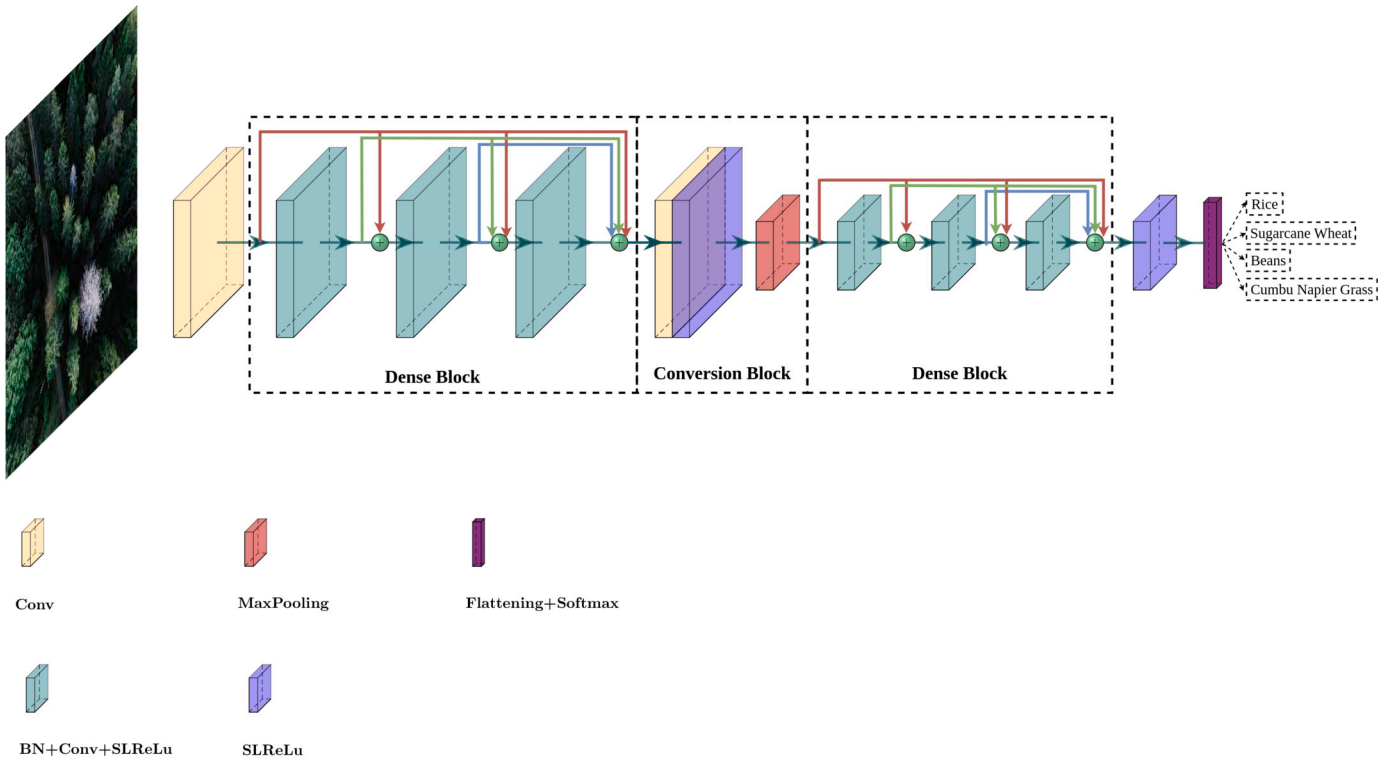


Fig. 14. CD-CNN schema. Reprocessing of the image in the original article. [52].

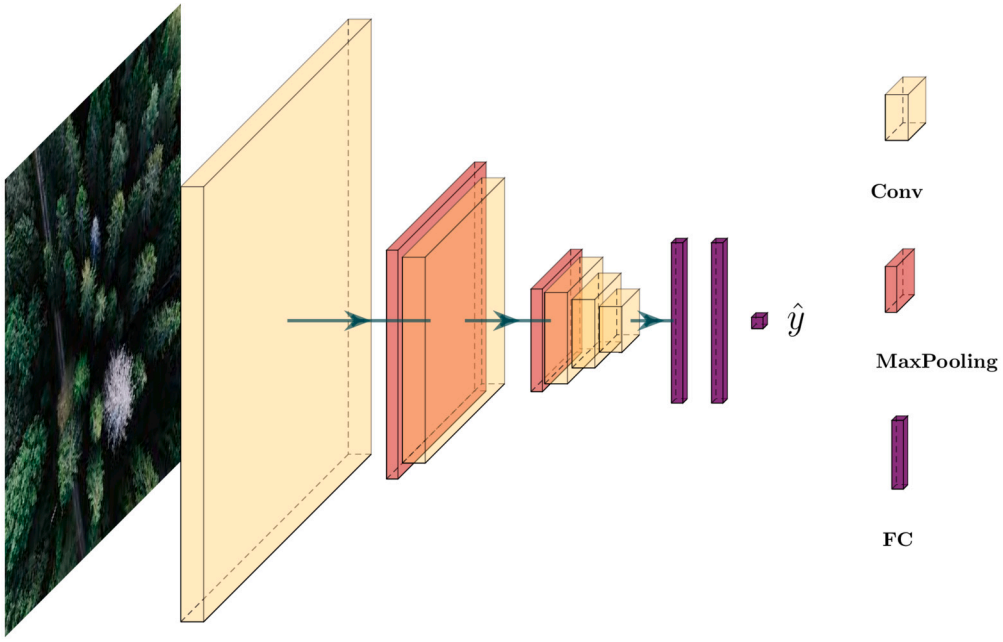


Fig. 15. Regression DCNN proposed by Yu et al. Reprocessing of the image in the original article [36].

R-CNN, a YOLOv3 and a RetinaNet. The best model in their use case was the one based on RetinaNet. Lou et al. [78] segment and estimate the width of loblolly pines crowns using deep learning techniques. They test 3 different architectures: Faster R-CNN, SSD (Single Shot Detector [79]) and YOLO v3. Onishi and Ise [41] propose a machine vision system for a 7 class CNN based classificator for tree crowns. In this article, they test 4 different standard image classification models: VGG-16, ResNet-18, ResNet-152 and AlexNet. In addition, they compare the results of these end-to-end models with an SVM-based classification with handcrafted features. Works that benchmark and compare various algorithms are grouped in the Table 4. Some of these have as their ultimate goal benchmarking itself, while others compare the proposed model with reference ones.

5. Conclusions

Image processing from remote sensing sources (specifically, UAVs) is a task that has relevance in several activities of social, academic, economic and environmental interest. The state of the art on many image

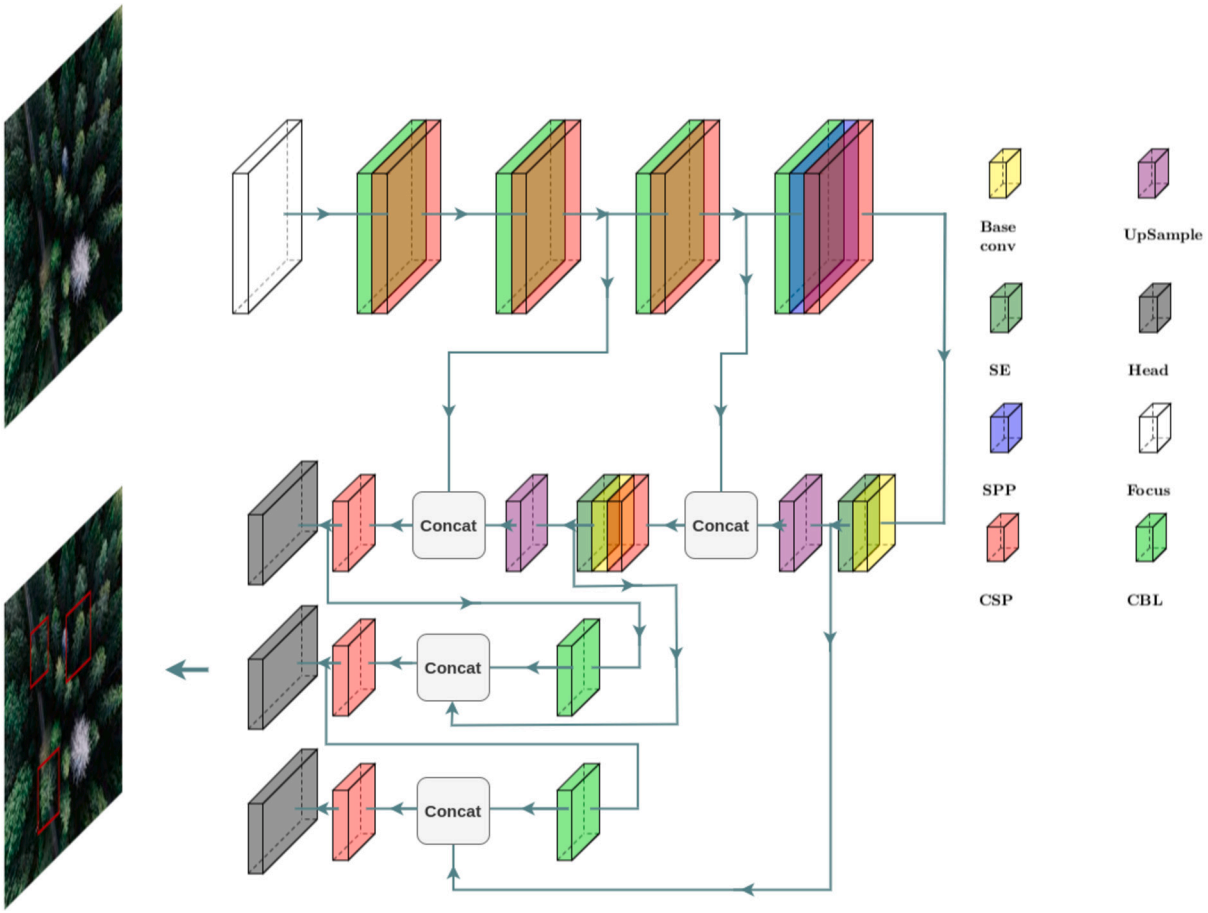


Fig. 16. SEYOLOX-tiny proposed by Song et al. [50]. CBL module performs a depthwise convolution and a standard one in parallel. Reprocessing of the image available in the original article.

Table 4
Studies that compares different architectures. Metrics abbreviations are explained in Table 6.

			Metrics					
Task	Reference	Models evaluated	P/R	F1	IoU	OA	R^2	RMSE
Semantic segmentation	Lobo Torres et al. [16]	U-Net, SegNet, DenseNet, Deeplab V3+	•	•	•	•		
	Behera et al. [26]	FCN8, FCN16, FCN32, DenseNet, SegNet, Deeplab v3, U-Net, Aerial SegNet [67], LW aerial segnet	•	•	•	•		
	Ye et al. [29]	U-Net, U^2-Net, HRNet, DeepLab v3+	•	•	•	•		
Instance segmentation - Object detection	Guirado et al. [14]	Mask R-CNN, OBIA	•	•				
	Osco et al. [30]	RetinaNet, Faster R-CNN, proposed approach	•	•				
	Song et al. [50]	YOLOX variations and proposed ones	•					
	dos Santos et al. [28]	YOLO v3, Retinanet, Fast R-CNN	•		•			
	Lou et al. [78]	Faster R-CNN, Single Shot Detector, YOLO v3	•	•		•		•
	Yu et al. [32]	Mask R-CNN, Local Maxima, MCWS	•	•				
	Junos et al. [49]	SSD, YOLO v3, YOLO v3 tiny, improved YOLO v3 tiny, Yolo v2, Faster R-CNN	•	•		•	•	•
Image Classification	Onishi and Ise [41]	AlexNet, VGG-16, ResNet-18, ResNet-512, SVM		•				
	Zhang et al. [40]	VGG-16, ResNet-50, DefoNet, classic methods	•					
	Correa Martins et al. [44]	Xception, EfficientNet, NasNetMobile	•	•				
	Pandey and Jain [52]	AlexNet, VGG16, VGG19, ResNet50, CD-CNN	•	•				
Regression	Yu et al. [36]	DCNN, RF, Multiple Linear Regression, Support Vector Machine					•	•

processing tasks is based on the neural network paradigm. For industry, smart crops monitoring can have a direct impact on the economic return to companies, as it can, for example, allow them to control (and infer) several parameters related to the amount of crop yield. In this context, neural networks allow complex correlations to be found between heterogeneous datasets from different sources and target variables, even if their behavior is unexplainable and then does not provide an explicit analytical solution. With the increasing use of data-driven systems in PA & WFM, data-intensive labeling has costs (both in terms of time and money) that are less and less sustainable. It is often necessary to rely on domain experts (which is common in many application areas), and the areas to be labeled are often huge. Some authors pointed out this is-

Table 5
Acronyms used in this paper.

Abbreviation	Extended	Introduced in
PA & WFM	Precision Agriculture & Wild Flora Monitoring	Section 1
UAV	Unmanned Aerial Vehicle	Section 1
UGV	Unmanned Ground Vehicle	Section 1
GSD	Ground Sampling Distance	Section 1
AOI	Areas Of Interest	Section 1
NIR	Near Infra Red	Section 1
SVM	Support Vector Machine	Table 2
RF	Random Forest	Table 2
MSE	Mean Square Error	Table 2
VIs	Vegetation Indexes	Table 2
NDVI	Normalized Difference Vegetation Index	Table 2
CNN	Convolution Neural Network	Table 2
ANN	Artificial Neural Network	Table 2
DNN	Deep Neural Network	Table 2
AI	Artificial Intelligence	Table 2
AGB	Above Ground Biomass	Section 2.2.3
RD	Reference Distribution	Section 2.2.3
HSV	Hue Saturation Value	Section 3.1
DSM	Digital Surface Model	Section 3.2.3
IS/OD	Instance Segmentation / Object Detection	Section 4
OBIA	Object-based Image Analysis	Section 4.4
SSD	Single Shot Detector	Section 4.4

Table 6
Metrics abbreviations.

Abbreviation	Extended name
F1	F1-Score
IoU	Intersection Over Union
K	Cohen's Kappa Coefficient
OA	Overall Accuracy
P	Precision
R	Recall
RMSE	Root Mean Square Error
RRMSE	Relative Root Mean Square Error

Table 7
Some of the indices well known in literature. Some of these were used in the works reviewed. A complete list is available at www.indexdatabase.de.

Name	Formula	Correlates with	Reference
NDVI	$\frac{NIR-Red}{NIR+Red}$	Crop health	[97]
NDWI	$\frac{Green-NIR}{Green+NIR}$	Plant water content	[98]
NDRE	$\frac{NIR-Rededge}{NIR+Rededge}$	Chlorophyll Content	[99]
GNDVI	$NIR-Green$	Chlorophyll concentration	[100]
DVI	NIR-Red	Vegetation presence	[101]
RVI	$\frac{Red}{NIR}$	High density Vegetation / biomass	[102]

sue, proposing automatic labeling methods [35] or fully unsupervised solutions [37]. Crop classification is essential to schedule and optimize farming [52] as it also helps to estimate the net yield production for each crop. There are biophysical parameters (qualitative and quantitative) whose monitoring provides direct information on the health status of plants or the productivity expected from a crop. As example, as described by Zhang et al. [40] the state of defoliation of soybean plants directly affects the productivity of the crop, which is why their system is able to categorize portions of the crop according to the severity of the defoliation. AGB is also considered an indicative parameter for productivity, as shown by Yu et al. in [36], who then tested various regression methods to estimate it. Biovolume was estimated for olive plants by Safonova et al. in [31], parameter that is correlated with oil productivity. Other parameters such as plant height, area, and canopy number are important for PA, and, although not always concurrent with industrial production, are issues addressed by several authors [13,35,78,33]. A wide interest in these techniques is also growing in the WFM domain: with climate change and global warming, the earth's ecosystem is changing ever more rapidly. The conservation status of certain plant species directly impacts the ecosystem of which these species are a part. Reliable plant monitoring systems are therefore increasingly needed. Martins et al. in [44] have analyzed with data-driven methods kettle holes, whose characteristic internal ecosystem is considered an important indicator for the health of the environment. Classification and mapping of endangered species can be successfully addressed with exposed technologies, as demonstrated by Lobo Torres et al. [16] for mapping of Dipteryx alata specimens in urban settings.

The chase to maximize standard performance scores on general-purpose reference datasets led to a huge increase in models sizes, often excessive compared to the practical problems to be solved in specific domains. This is a well-known problem in the machine learning literature, even in domains different than computer vision [80,81]. This model resizing/optimization approach, in some cases, allows to perform online inference (sometimes even training [82]) and deploy deep models in embedded devices in an Internet of Things context [83–85]. This topic opened up further horizons for academia and industry, becoming popular under the name of Tiny Machine Learning [86] (TinyML) achieving results in many tasks [87,88]. For all these reasons, in last years the research for methods to optimize and under-scale over-parametrized machine learning models is very fervent, as shown by Frankle and Carbin [89] whom proposed and empirically proven the lottery ticket hypothesis, providing a very general, versatile and effective framework that allows to prune parameters in deep neural networks. In our literature analysis, this topic emerges (it is highlighted in Table 3) and it's clear that domain-specific solutions allow parameters to be reduced at the cost of cross-domain versatility. With the emergence of foundation models/LLMs, the decision-making and planning activities have been greatly simplified, as these systems have already amply proven having cognitive abilities that improve as research progresses, being able to handle complex planning tasks and decision making reliably [90–96]. These innovations in LLM, combined with the deep learning for computer vision and the versatility/low cost/ease of use of UAV imagery technologies make the prospects towards full automation of sectors such as PA & WFM become realistic in the near future, allowing to address social and environmental challenges crucial for the entire humanity, but also generating great business opportunities typical of technological revolutions.

CRediT authorship contribution statement

Lorenzo Epifani: Writing – review & editing, Writing – original draft, Validation, Data curation, Conceptualization. **Antonio Caruso:** Writing – review & editing, Writing – original draft, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

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

Data availability

No data was used for the research described in the article.

Table 8
Summary of work reviewed grouped by core task. "Goal" column specifies the final goal of the related article. The column "#Instances" indicates the number of collected instances used for model training (objects, tiles, masks or crops, depending on the goal). The sensors column specifies which spectral channels are used by the models. The "Surface Models" column indicates whether surface models were used or not.

Task	Reference	Goal	Species	#Instances	Sensors	GSD (cm^2px)	Surface Models	Model(s)
Semantic Segmentation	Ye et al. [29]	Crowns surface and trees number estimation	Olive	751	RGB	2		U^2-Net
	Ferreira et al. [45]	Individual tree detection, species classification	A. butyracea, E. precatoria, I. detoidea		RGB	4		Deeplab V3+
	Gurumurthy et al. [13]	Crowns number estimation	Mango	297	RGB	Not specified		MangoTree Net
	Lobo Torres et al. [16]	Architectures comparison	Cumbaru		RGB	1		U-Net, SegNet, DenseNet, DeeplabV3+
	Epifani et al. [35]	Crown segmentation	Vineyard, Olive		Multi(10)	10 and 3		U-Net
	Vélez et al. [103]	Building heathmap for Botrytis Cinerea development risk	Grapevines	153	Multi (5)	1.67	•	Random forest
	Jaimes et al. [37]	Fully unsupervised vegetation segmentation	All		RGB	10		Unsupervised parameterless pipeline
Instance segmentation / Object detection	Safonova et al. [31]	Biovolume Estimation	Olive	2400	RGB, NDVI, GNDVI	3 and 13		Mask R-CNN
	Guirado et al. [14]	OBIA and Mask R-CNN benchmarking	Ziziphus Lotus		RGB	3, 10 and 50		Mask R-CNN
	Song et al. [50]	Maize tassel detection	Maize		RGB, Multi(5)			YOLOX variations
	Lou et al. [78]	Crowns size estimation	Loblolly pine		RGB	1.17 and 1.24		Faster R-CNN SSD, YOLO v3
	Hao et al. [33]	Tree crown detection and height estimation	Chinese fir	1605	RGB, NDVI	2	•	Mask R-CNN
	Junos et al. [49]	Fruits detection, network proposal	Oil palm		RGB			YOLO v2/v3/v3 tiny, Faster R-CNN, SSD
	Yu et al. [32]	Tree detection	Chinese fir	1816	Multi (5)	2	•	Mask R-CNN
	dos Santos et al. [28]	Individual tree detection, architectures comparison	Dipteryx alata	110	RGB	0.82		YOLO v3, RetinaNet, Fast R-CNN
	Osco et al. [30]	Individual tree detection and counting	Citrus	37353	Green, Red, Red edge, NIR	12.9		Network proposal
	Xiao et al. [64]	Growth monitoring across different treatments	Corn		RGB, Multi (5)	0.8 RGB, 2 Multi	•	YOLO v5, Otsu method
Image Classification	Natesan et al. [43]	Tree species classification	White pine, red pine		RGB	1, 2 and 4	•	ResNet-50
	Egli and Höpke [15]	Tree species classification for automatic observation system	Oak, beech, spruce, larch	477	RGB	0.27	•	Lightweight proposed CNN
	Onishi and Ise [41]	Tree species classification, manual vs cnn features comparison	Broad leaved tree, coniferous, evergreen broad leaved tree, C. obtuse, P. ellottii/P. taeda, strobus		RGB	5	•	AlexNet, VGG-16, ResNet 18-512, SVM
	Correa Martins et al. [44]	Kettle holes monitoring	C. riparia, C. arvense, O. aquatica, P. arundinacea, P. australis, S. alba, S. cinerea, T. latifolia, U. dioica	318 per class (after balancing)	RGB	0.9		Xception, EfficientNet, NasNetMobile
	Zhang et al. [40]	Defoliation detection	Soybean		RGB			Naive bayes, KNN, RF, SVM, Gaussian Process, VGG, ResNet, DefoNet
	Pandey and Jain [52]	Crop classification	Rice, sugarcane Wheat, beans, cumbu napier grass	2000 per class	RGB	2.7		AlexNet, VGG16, VGG19, ResNet50, CD-CNN
Regression	Fei et al. [34]	Wheat yield prediction	Wheat	180 crops	RGB, multi(5), thermal		•	Ensemble of: Cubist, SVM, DNN, ridge regression, RF
	Yu et al. [36]	AGB estimation, handcrafted features vs DCNN	Maize	57	Multi (5)		•	DCNN, multi linear regression, SVM, RF

References

[1] G. Mergos, Population and Food System Sustainability, Springer International Publishing, Cham, 2022, pp. 131–155.
[2] N. Alexandratos, World Agriculture Towards 2030/2050: the 2012 Revision, FAO, 2012.
[3] N. Cialdella, M. Jacobson, E. Penot, Economics of agroforestry: links between nature and society, Agrofor. Syst. 97 (2023) 273–277, https://doi.org/10.1007/s10457-023-00829-z, https://link.springer.com/10.1007/s10457-023-00829-z.
[4] A. Gutiérrez-Li, Feeding America: How Immigrants Sustain US Agriculture, Baker Institute, 2024.
[5] K. Ngongolo, L. Gayo, Synergistic impact of COVID-19 and climate change on agricultural resilience and food security in sub-Saharan Africa, Discov. Agric. 2 (2024) 41, https://doi.org/10.1007/s44279-024-00056-9, https://link.springer.com/10.1007/s44279-024-00056-9.
[6] R.G. Kerry, F.J.P. Montalbo, R. Das, S. Patra, G.P. Mahapatra, G.K. Maurya, V. Nayak, A.B. Jena, K.E. Ukhurebor, R.C. Jena, S. Gouda, S. Majhi, J.R. Rout, An overview of remote monitoring methods in biodiversity conservation, Environ. Sci. Pollut. Res. 29 (2022) 80179–80221, https://doi.org/10.1007/s11356-022-23242-y.
[7] C.S. Reddy, Remote sensing of biodiversity: what to measure and monitor from space to species?, Biodivers. Conserv. 30 (2021) 2617–2631, https://doi.org/10.1007/s10531-021-02216-5.
[8] F. Urbano, R. Viterbi, L. Pedrotti, E. Vettorazzo, C. Movalli, L. Corlatti, Enhancing biodiversity conservation and monitoring in protected areas through efficient data management, Environ. Monit. Assess. 196 (2023) 12, https://doi.org/10.1007/s10661-023-11851-0.
[9] P. Skrzypczyński, Path planning for an unmanned ground vehicle traversing rough terrain with unknown areas, in: R. Szewczyk, C. Zieliński, M. Kaliczyńska (Eds.), Automation 2017, Springer International Publishing, Cham, 2017, pp. 319–329.
[10] M. Rußwurm, M. Körner, Multi-temporal land cover classification with sequential recurrent encoders, ISPRS Int.l J. Geo-Inf. 7 (2018).
[11] R.C. Daudt, B.L. Saux, A. Boulch, Fully Convolutional Siamese Networks for Change Detection, 2018.
[12] D. Ienco, R. Gaetano, C. Dupaquier, P. Maurel, Land cover classification via multi-temporal spatial data by recurrent neural networks, IEEE Geosci. Remote Sens. Lett. 14 (2017).
[13] V.A. Gurumurthy, R. Kestur, O. Narasipura, Mango Tree Net – a fully convolutional network for semantic segmentation and individual crown detection of mango trees, 2019.
[14] E. Guirado, J. Blanco-Sacristán, E. Rodríguez-Caballero, S. Tabik, D. Alcaraz-Segura, J. Martínez-Valderrama, J. Cabello, Mask R-CNN and OBIA fusion improves the segmentation of scattered vegetation in very high-resolution optical sensors, Sensors 21 (2021).
[15] S. Egli, M. Höpke, Cnn-based tree species classification using high resolution rgb image data from automated uav observations, Remote Sens. 12 (2020).
[16] D. Lobo Torres, R. Queiroz Feitosa, P. Nigri Happ, L. Elena Cué La Rosa, J. Marcato Junior, J. Martins, P. Olã Bressan, W.N. Gonçalves, V. Liesenberg, Applying fully convolutional architectures for semantic segmentation of a single tree species in urban environment on high resolution UAV optical imagery, Sensors 20 (2020).
[17] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, Neural Inf. Process. Syst. (2012).
[18] C. Zhang, J.M. Kovacs, The application of small unmanned aerial systems for precision agriculture: a review, Precis. Agric. 13 (2012).
[19] Y. Lu, S. Young, A survey of public datasets for computer vision tasks in precision agriculture, Comput. Electron. Agric. 178 (2023).
[20] K.P. Seng, L. Ang, L.M. Schmidtke, S.Y. Rogiers, Computer vision and machine learning for viticulture technology, IEEE Access 6 (2018).
[21] A. Bouguettaya, H. Zarzour, A. Kechida, A.M. Taberkit, Deep learning techniques to classify agricultural crops through UAV imagery: a review, Neural Comput. Appl. 34 (2022).
[22] A. Kamilaris, F.X. Prenafeta-Boldú, Deep learning in agriculture: a survey, Comput. Electron. Agric. 147 (2018).
[23] J. Su, X. Zhu, S. Li, W.-H. Chen, AI meets UAVs: a survey on AI empowered UAV perception systems for precision agriculture, Neurocomputing 518 (2022).
[24] L.P. Osco, J. Marcato Junior, A.P. Marques Ramos, L.A. de Castro Jorge, S.N. Fatholahi, J. de Andrade Silva, E.T. Matsubara, H. Pistori, W.N. Gonçalves, J. Li, A review on deep learning in uav remote sensing, Int. J. Appl. Earth Obs. Geoinf. 102 (2021).
[25] F. Melgani, L. Bruzzone, Classification of hyperspectral remote sensing images with support vector machines, IEEE Trans. Geosci. Remote Sens. 42 (2004) 1778–1790, https://doi.org/10.1109/TGRS.2004.831865, cited by: 3495.
[26] T.K. Behera, S. Bakshi, P.K. Sa, A lightweight deep learning architecture for vegetation segmentation using UAV-captured aerial images, Sustain. Comput.: Inf. Systems 37 (2023).
[27] H. Meyer, C. Reudenbach, T. Hengl, M. Katurji, T. Nauss, Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation, Environ. Model. Softw. 101 (2018).
[28] A.A. dos Santos, J. Marcato Junior, M.S. Araújo, D.R. Di Martini, E.C.a. Tetila, H.L. Siqueira, C. Aoki, A. Eltner, E.T. Matsubara, H. Pistori, R.Q. Feitosa, V. Liesenberg, W.N. Gonçalves, Assessment of cnn-based methods for individual tree detection on images captured by rgb cameras attached to uavs, Sensors (Switzerland) 19 (2019).
[29] Z. Ye, J. Wei, Y. Lin, Q. Guo, J. Zhang, H. Zhang, H. Deng, K. Yang, Extraction of olive crown based on uav visible images and the u2-net deep learning model, Remote Sens. 14 (2022).
[30] L.P. Osco, M. d, S. de Arruda, J. Marcato Junior, N.B. da Silva, A.P.M. Ramos, E.A.S. Moryia, N.N. Imai, D.R. Pereira, J.E. Creste, E.T. Matsubara, J. Li, W.N. Gonçalves, A convolutional neural network approach for counting and geolocating citrus-trees in uav multispectral imagery, ISPRS J. Photogramm. Remote Sens. 160 (2020).
[31] A. Safonova, E. Guirado, Y. Maglinets, D. Alcaraz-Segura, S. Tabik, Olive tree biovolume from uav multi-resolution image segmentation with mask r-cnn, Sensors 21 (2021).
[32] K. Yu, Z. Hao, C.J. Post, E.A. Mikhailova, L. Lin, G. Zhao, S. Tian, J. Liu, Comparison of classical methods and mask r-CNN for automatic tree detection and mapping using UAV imagery, Remote Sens. 14 (2022).
[33] Z. Hao, L. Lin, C.J. Post, E.A. Mikhailova, M. Li, Y. Chen, K. Yu, J. Liu, Automated tree-crown and height detection in a young forest plantation using mask region-based convolutional neural network (mask r-cnn), ISPRS J. Photogramm. Remote Sens. 178 (2021).
[34] S. Fei, M.A. Hassan, Y. Xiao, X. Su, Z. Chen, Q. Cheng, F. Duan, R. Chen, Y. Ma, UAV-based multi-sensor data fusion and machine learning algorithm for yield prediction in wheat, Precis. Agric. 24 (2023).
[35] L. Epifani, V. D'Avino, A. Caruso, TEBAKA: territorial basic knowledge acquisition. An agritech project for Italy: results on self-supervised semantic segmentation, in: IEEE Symposium on Computers and Communications, ISCC 2023, Gammarth, Tunisia, July 9-12, 2023, IEEE, 2023, pp. 1116–1121.
[36] D. Yu, Y. Zha, Z. Sun, J. Li, X. Jin, W. Zhu, J. Bian, L. Ma, Y. Zeng, Z. Su, Deep convolutional neural networks for estimating maize above-ground biomass using multi-source UAV images: a comparison with traditional machine learning algorithms, Precis. Agric. 24 (2023).
[37] B.R.A. Jaimes, J.P.K. Ferreira, C.L. Castro, Unsupervised semantic segmentation of aerial images with application to UAV localization, IEEE Geosci. Remote Sens. Lett. 19 (2022).
[38] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, C.C. Loy, Y. Qiao, X. Tang, ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks, 2018.
[39] Y. Xiong, S. Guo, J. Chen, X. Deng, L. Sun, X. Zheng, W. Xu, Improved srgan for remote sensing image super-resolution across locations and sensors, Remote Sens. 12 (2020).
[40] Z. Zhang, S. Khanal, A. Raudenbush, K. Tilmon, C. Stewart, Assessing the efficacy of machine learning techniques to characterize soybean defoliation from unmanned aerial vehicles, Comput. Electron. Agric. 193 (2022).
[41] M. Onishi, T. Ise, Explainable identification and mapping of trees using uav rgb image and deep learning, Sci. Rep. 11 (2021).
[42] G. Jang, J. Kim, J.-K. Yu, H.-J. Kim, Y. Kim, D.-W. Kim, K.-H. Kim, C.W. Lee, Y.S. Chung, Review: cost-effective unmanned aerial vehicle (UAV) platform for field plant breeding application, Remote Sens. 12 (2020).
[43] S. Natesan, C. Armenakis, U. Vepakomma, Resnet-based tree species classification using uav images, in: International Archives of the Photogrammetry, Remote Sensing and Spatial, in: Information Sciences - ISPRS Archives, vol. 42, 2019.
[44] J.A. Correa Martins, J. Marcato Junior, M. Pätzig, D.A. Sant'Ana, H. Pistori, V. Liesenberg, A. Eltner, Identifying plant species in kettle holes using UAV images and deep learning techniques, Remote Sens. Ecol. Conserv. 9 (2023).
[45] M.P. Ferreira, D.R.A.d. Almeida, D.d.A. Papa, J.B.S. Minervino, H.F.P. Veras, A. Formighieri, C.A.N. Santos, M.A.D. Ferreira, E.O. Figueiredo, E.J.L. Ferreira, Individual tree detection and species classification of Amazonian palms using uav images and deep learning, For. Ecol. Manag. 475 (2020).
[46] X. Shen, A survey of Object Classification and Detection based on 2D/3D data, 2022.
[47] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation, 2015.
[48] V. Badrinarayanan, A. Kendall, R. Cipolla, Segnet: a deep convolutional encoder-decoder architecture for image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 39 (2017).
[49] M.H. Junos, A.S. Mohd Khairuddin, S. Thannirmalai, M. Dahari, Automatic detection of oil palm fruits from UAV images using an improved YOLO model, Vis. Comput. 38 (2022).
[50] C.-y. Song, F. Zhang, J.-s. Li, J.-y. Xie, C. Yang, H. Zhou, J.-x. Zhang, Detection of maize tassels for UAV remote sensing image with an improved YOLOX model, J. Integr. Agricult. 22 (2023) 1671–1683, https://doi.org/10.1016/j.jia.2022.09.021, https://linkinghub.elsevier.com/retrieve/pii/S2095311922002465.
[51] K. He, G. Gkioxari, P. Dollár, R. Girshick, R-CNN Mask, 2018.
[52] A. Pandey, K. Jain, An intelligent system for crop identification and classification from UAV images using conjugated dense convolutional neural network, Comput. Electron. Agric. 192 (2022).
[53] K. He, X. Zhang, S. Ren, J. Sun, Deep Residual Learning for Image Recognition, 2015.
[54] K. Simonyan, A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition, 2015.
[55] J.R.R. Uijlings, K.E.A. van de Sande, T. Gevers, A.W.M. Smeulders, Selective search for object recognition, Int. J. Comput. Vis. 104 (2013).
[56] R. Girshick, J. Donahue, T. Darrell, J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, 2014.

[57] R. Girshick, R-CNN Fast, 2015.
[58] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, 2016.
[59] J. Terven, D.-M. Córdova-Esparza, J.-A. Romero-González, A comprehensive review of yolo architectures in computer vision: from yolov1 to yolov8 and yolo-nas, Mach. Learn. Knowl. Extr. 5 (2023) 1680–1716, https://doi.org/10.3390/make5040083.
[60] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: unified, real-time object detection, Comput. Vis. Pattern Recognit. (2016).
[61] J. Redmon, A. Farhadi, YOLO9000: Better, Faster, Stronger, 2016.
[62] J. Redmon, A. Farhadi, YOLOv3: an Incremental Improvement, 2018.
[63] C.-Y. Wang, A. Bochkovskiy, H.-Y.M. Liao, YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors, 2022.
[64] J. Xiao, S.A. Suab, X. Chen, C.K. Singh, D. Singh, A.K. Aggarwal, A. Korom, W. Widyatmanti, T.H. Mollah, H.V.T. Minh, K.M. Khedher, R. Avtar, Enhancing assessment of corn growth performance using unmanned aerial vehicles (UAVs) and deep learning, Measurement (2023).
[65] N. Otsu, A threshold selection method from gray-level histograms, IEEE Trans. Syst. Man Cybern. 9 (1979) 62–66, https://doi.org/10.1109/TSMC.1979.4310076.
[66] U. Vepakomma, D.D. Kneeshaw, L. De Grandpré, Influence of natural and anthropogenic linear canopy openings on forest structural patterns investigated using lidar, Forests 9 (2018).
[67] T.K. Behera, S. Bakshi, P.K. Sa, Vegetation extraction from uav-based aerial images through deep learning, Comput. Electron. Agric. 198 (2022).
[68] G. Huang, Z. Liu, K.Q. Weinberger, Densely connected convolutional networks, CoRR, arXiv:1608.06993 [abs], 2016, arXiv:1608.06993.
[69] Y. Wang, Y. Li, Y. Song, X. Rong, The influence of the activation function in a convolution neural network model of facial expression recognition, Appl. Sci. 10 (2020).
[70] Z. Ge, S. Liu, F. Wang, Z. Li, J. Sun, YOLOX: exceeding YOLO series in 2021, CoRR, arXiv:2107.08430 [abs], 2021, https://arxiv.org/abs/2107.08430.
[71] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, CoRR, arXiv:1709.01507 [abs], 2017, http://arxiv.org/abs/1709.01507.
[72] H. Zhang, Y. Wang, F. Dayoub, N. Sünderhauf, Varifocalnet: an iou-aware dense object detector, CoRR, arXiv:2008.13367 [abs], 2020, https://arxiv.org/abs/2008.13367.
[73] C.-Y. Wang, H.-Y.M. Liao, I.-H. Yeh, Y.-H. Wu, P.-Y. Chen, J.-W. Hsieh, CSPNet: a new backbone that can enhance learning capability of CNN, http://arxiv.org/abs/1911.11929, arXiv:1911.11929 [cs], 2019.
[74] X. Jia, J. Bartlett, T. Zhang, W. Lu, Z. Qiu, J. Duan, U-net vs transformer: is u-net outdated in medical image registration?, http://arxiv.org/abs/2208.04939, arXiv:2208.04939 [cs, eess], 2022.
[75] M.L. Taccari, O. Ovadia, H. Wang, A. Kahana, X. Chen, P.K. Jimack, Understanding the efficacy of u-net & vision transformer for groundwater numerical modelling, CoRR, arXiv:2307.04010 [abs], 2023, https://doi.org/10.48550/arXiv.2307.04010.
[76] H. Chen, Z. Qi, Z. Shi, Remote sensing image change detection with transformers, IEEE Trans. Geosci. Remote Sens. 60 (2022) 1–14, https://doi.org/10.1109/TGRS.2021.3095166, http://arxiv.org/abs/2103.00208, arXiv:2103.00208 [cs].
[77] M. Noman, M. Fiaz, H. Cholakkal, S. Narayan, R.M. Anwer, S. Khan, F.S. Khan, Remote sensing change detection with transformers trained from scratch, http://arxiv.org/abs/2304.06710, arXiv:2304.06710 [cs], 2023.
[78] X. Lou, Y. Huang, L. Fang, S. Huang, H. Gao, L. Yang, Y. Weng, I.-K. Hung, Measuring Loblolly pine crowns with drone imagery through deep learning, J. For. Res. 33 (2022).
[79] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S.E. Reed, C. Fu, A.C. Berg, SSD: single shot multibox detector, CoRR, arXiv:1512.02325 [abs], 2015, http://arxiv.org/abs/1512.02325.
[80] Z. Li, S. Qi, Y. Li, Z. Xu, Revisiting long-term time series forecasting: an investigation on linear mapping, http://arxiv.org/abs/2305.10721, arXiv:2305.10721 [cs], 2023.
[81] G. Nalcaci, A. Özmen, G.W. Weber, Long-term load forecasting: models based on MARS, ANN and LR methods, Cent. Eur. J. Oper. Res. 27 (2019) 1033–1049, https://doi.org/10.1007/s10100-018-0531-1, http://link.springer.com/10.1007/s10100-018-0531-1.
[82] S. Disabato, M. Roveri, Incremental on-device tiny machine learning, in: Proceedings of the 2nd International Workshop on Challenges in Artificial Intelligence and Machine Learning for Internet of Things, ACM, 2020, pp. 7–13, https://dl.acm.org/doi/10.1145/3417313.3429378.
[83] A.D. Boursianis, M.S. Papadopoulou, P. Diamantoulakis, A. Liopa-Tsakalidi, P. Barouchas, G. Salahas, G. Karagiannidis, S. Wan, S.K. Goudos, Internet of things (iot) and agricultural unmanned aerial vehicles (uavs) in smart farming: a comprehensive review, Int. Things (Netherlands) 18 (2022), https://doi.org/10.1016/j.iot.2020.100187.
[84] M. Kuzman, X.d. Toro García, S. Escolar, A. Caruso, S. Chessa, J.C. López, A testbed and an experimental public dataset for energy-harvested IoT solutions, in: 2019 IEEE 17th International Conference on Industrial Informatics (INDIN), Volume 1, 2019, pp. 869–876.
[85] A. Caruso, S. Chessa, S. Escolar, F. Rincon, J.C. López, Task scheduling stabilization for solar energy harvesting Internet of Things devices, in: 2022 IEEE 27th IEEE Symposium on Computers and Communications (ISCC), Volume 2022-June, 2022, pp. 1–6, https://dl.acm.org/doi/10.1109/ISCC55528.2022.9913061.
[86] M. Roveri, Is tiny deep learning the new deep learning?, in: R. Buyya, S.M. Hernandez, R.M.R. Kovvur, T.H. Sarma (Eds.), Computational Intelligence and Data Analytics, Springer Nature Singapore, Singapore, 2023, pp. 23–39.
[87] M. Pavan, E. Ostrovan, A. Caltabiano, M. Roveri, TyBox: an automatic design and code-generation toolbox for TinyML incremental on-device learning, ACM Trans. Embed. Comput. Syst. (2023) 3604566, https://doi.org/10.1145/3604566, https://dl.acm.org/doi/10.1145/3604566.
[88] M. Pavan, A. Caltabiano, M. Roveri, TinyML for UWB-radar based presence detection, in: 2022 International Joint Conference on Neural Networks (IJCNN), IEEE, 2022, pp. 1–8, https://ieeexplore.ieee.org/document/9892925/.
[89] J. Frankle, M. Carbin, The lottery ticket hypothesis: training pruned neural networks, CoRR, arXiv:1803.03635 [abs], 2018, arXiv:1803.03635.
[90] X. Liu, H. Yu, H. Zhang, Y. Xu, X. Lei, H. Lai, Y. Gu, H. Ding, K. Men, K. Yang, S. Zhang, X. Deng, A. Zeng, Z. Du, C. Zhang, S. Shen, T. Zhang, Y. Su, H. Sun, M. Huang, Y. Dong, J. Tang, AgentBench: evaluating LLMs as agents, http://arxiv.org/abs/2308.03688, arXiv:2308.03688 [cs], 2023.
[91] H. Sun, Y. Zhuang, L. Kong, B. Dai, C. Zhang, AdaPlanner: Adaptive planning from feedback with language models, 2023.
[92] Y. Cui, S. Huang, J. Zhong, Z. Liu, Y. Wang, C. Sun, B. Li, X. Wang, A. Khajepour, DriveLLM: charting the path toward full autonomous driving with large language models, IEEE Trans. Intell. Veh. 9 (2024) 1450–1464, https://doi.org/10.1109/TIV.2023.3327715, https://ieeexplore.ieee.org/document/10297415/.
[93] S. Griewing, J. Knitza, J. Boekhoff, C. Hillen, F. Lechner, U. Wagner, M. Wallwiener, S. Kuhn, Evolution of publicly available large language models for complex decision-making in breast cancer care, Arch. Gynecol. Obstet. 310 (2024) 537–550, https://doi.org/10.1007/s00404-024-07565-4, https://link.springer.com/10.1007/s00404-024-07565-4.
[94] J.S. Park, J. O'Brien, C.J. Cai, M.R. Morris, P. Liang, M.S. Bernstein, Generative agents: interactive simulacra of human behavior, in: Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology, ACM, 2023, pp. 1–22, https://dl.acm.org/doi/10.1145/3586183.3606763.
[95] Meta Fundamental AI Research Diplomacy Team (FAIR)†, A. Bakhtin, N. Brown, E. Dinan, G. Farina, C. Flaherty, D. Fried, A. Goff, J. Gray, H. Hu, A.P. Jacob, M. Komeili, K. Konath, M. Kwon, A. Lerer, M. Lewis, A.H. Miller, S. Mitts, A. Renduchintala, S. Roller, D. Rowe, W. Shi, J. Spisak, A. Wei, D. Wu, H. Zhang, M. Zijlstra, Human-level play in the game of diplomacy by combining language models with strategic reasoning, Science 378 (2022) 1067–1074, https://doi.org/10.1126/science.ade9097, https://www.science.org/doi/10.1126/science.ade9097.
[96] W. Huang, P. Abbeel, D. Pathak, I. Mordatch, Language models as zero-shot planners: extracting actionable knowledge for embodied agents, CoRR, arXiv:2201.07207 [abs], 2022, https://arxiv.org/abs/2201.07207.
[97] S. Rouse Haas, Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation, NASA/GSFC, Type III, Final Report, 1974.
[98] B.-c. Gao, NDWI–a normalized difference water index for remote sensing of vegetation liquid water from space, Remote Sens. Environ. 58 (1996).
[99] E. Barnes, T. Clarke, S. Richards, P. Colaizzi, J. Haberland, M. Kostrzewski, P. Waller, C. Choi, E. Riley, T. Thompson, Coincident detection of crop water stress, nitrogen status, and canopy density using ground based multispectral data, in: Proceedings of the 5th International Conference on Precision Agriculture and Other Resource Management, July 16-19, 2000, Bloomington, MN USA, 6, 2000, p. 15.
[100] A.A. Gitelson, Y.J. Kaufman, M.N. Merzlyak, Use of a green channel in remote sensing of global vegetation from EOS-MODIS, Remote Sens. Environ. 58 (1996).
[101] A.D. Richardson, S.P. Duigan, G.P. Berlyn, An evaluation of noninvasive methods to estimate foliar chlorophyll content, New Phytol. (2002).
[102] C.J. Tucker, Red and photographic infrared linear combinations for monitoring vegetation, Remote Sens. Environ. (1979).
[103] S. Vélez, M. Ariza-Sentís, J. Valente, Mapping the spatial variability of Botrytis bunch rot risk in vineyards using UAV multispectral imagery, Eur. J. Agron. (2023).