**ALMA MATER STUDIORUM – UNIVERSITÀ DI BOLOGNA**
**CESENA CAMPUS**

Department of Computer Science and Engineering - DISI
Second-cycle Degree in Digital Transformation Management
Class: LM-91

# Artificial Intelligence for Digitalization in Agriculture: Considerations for the Development of a Fruit Detection System

Graduation thesis in
Machine Learning

*Supervisor*
**Dott. Guido Borghi**

*Candidate*
**Micol Cozzarizza**

5th Session
Academic Year 2022-2023

# Abstract

The adoption of digital solutions is gradually diffusing also in the realm of agriculture, due to the valuable contributions that innovative technologies can bring to a distressed sector. Among these, the application of Artificial Intelligence based fruit detection systems is receiving increasing interest, given the reliance that many technological agricultural applications have on detection tasks to execute their functions, as well as the usefulness such solutions can have in improving several activities: once they track down fruits on a tree, they are able to provide for a quality analysis of the fruits, thus rendering information over maturity level or presence of diseases, for yield estimates ahead of time or for the implementation of intelligent robots able to automatically collect fruits or perform agrochemicals spraying. Nonetheless, the development of an AI based fruit detection system is a non-trivial process since it requires many accurate and pondered considerations over intricate technological aspects relating to data requirements, feature extraction, existing models, necessary hardware configurations, as well as over the socio-economic context. Through an analysis of these elements based on relevant literature, the present elaborate aims to provide therefore a comprehensive understanding of the broader implications that arise during the conception, design, and integration phases of AI technologies for fruit detection tasks, encouraging the necessity of an holistic perspective for informed decision-making processes that could actually result beneficial for agricultural practices.

# Contents

# Chapter 1

# Introduction

According to the Food and Agriculture Organization of the United Nations, agriculture benefits society in numerous ways, foremost because it produces the food to nourish populations but also because it is a sector that worldwide provides jobs and livelihoods to almost a billion people (employing 27% of the global workforce [1]), thus sharing prosperity and helping reducing poverty [2]. In addition, agriculture plays an important role in the economic growth of a country - accounting for 4% of the global gross domestic product, but up to 30% of the GDP in some least developing countries [1]– and is fundamental to the existence of a variety of other activities linked to the sector.

Nonetheless, pressure on agricultural systems is growing due to the presence of numerous challenges. Firstly, the continuous expansion of the world population (expected to grow up to 9.7 billion by 2050 [3]), if on the on hand requires the necessity of a considerable increase in food production (namely by at least 70%, according to FAO [4]), on the other hand it implies the gradual reduction of the areas of cultivated land due to rapid urbanization. In addition to this, the agricultural sector is facing a labor shortage and, according to the Organization for Economic Co-operation and Development (OECD), the world's agricultural workforce is expected to register a yearly decline of 2% by 2030 [5]. The causes of this decline can be related mainly to an ageing of the farmer population not complemented with a simultaneous generational change of skilled workers, more and more scarce, which can be attributable to a preference for employment opportunities in more rewarding sectors [6]. As a consequence, labor costs for farm enterprises are on the rise. To top it all, food production remains extremely vulnerable to the effects of unpredictable weather conditions caused by climate change, and the lack of sufficient water resources.

In such context, digital agriculture has arisen as a scientific field geared towards developing innovative solutions for dealing with these challenges: by complementing traditional agricultural management methods with cutting-edge technologies pertaining to the Internet of Things, Machine and Deep Learning, Computer vision, Robotics, Cloud and Edge Computing, Blockchain, a revolutionary impact on various agricultural activities is achievable. For example, by combining technologies related to the Internet of Things and Machine Learning, it is possible to collect and analyze vast amount of valuable data, thus providing exhaustive and worthwhile insights of the agricultural environment. As a consequence, more effective and efficient decision-making processes are enabled and activities related to the management of crops, the estimation of their yield, their harvesting, the detection of diseases, the management of irrigation eventually improved. Or again, the implementation of autonomous robots equipped with computer vision capabilities can replace human labor for time-consuming and strenuous activities such as harvesting or weed and disease control. The ultimate effects of implementing innovative technologies in agriculture are therefore obvious: an improvement of the agricultural production and management activities and a generalized increase of the productivity and the profitability that can be overall achieved.

Among the different applications of the above-mentioned technologies in agriculture, the development of AI-based systems for the detection of on-tree fruits is receiving increasing interest. The reasons for this can be ascribable to the reliance that many technological agricultural applications have on detection tasks in order to execute their functions, as well as to the usefulness such solutions can have in improving several agricultural activities. In particular, common applications of fruit detection in agriculture relate to computer vision systems developed for unmanned aerial or ground vehicles or smartphones to quickly and automatically spot and count the fruits a crop presents. Once the detection phase is done, such systems usually provide for an analysis of the collected data, which can have different scopes. On the one side, from the appearance of the detected fruits some quality information can be derived, such as maturity level, size or presence of diseases thus assisting farmers in making relevant decisions with respect to harvesting time, fertilization, use of pesticides. The maturity level can indeed help farmers understand a fruit's growth state, which allows them to make efficient decisions related to harvesting and post harvesting activities. The detection of presence of diseases enables instead to promptly inform the farmers, so that corrective actions to avoid crop losses can be proactively taken. In addition to this, AI based detection systems can provide counting of the identified fruits, on the basis of which an early prediction of the orchard's produce is enabled: by knowing such information, farmers can implement actions to support crops growth for achieving an optimal production, as well as efficiently plan post harvesting activities for an effective allocations of labor, storage and transport resources. The relevance of this is even more evident considering that, commonly, fruit yield estimation has been carried out by manual counting, which has proven to lead to low precision results, high costs and higher time requirements for estimation. On the other side, fruit detection technologies have been vastly applied also for the automation of fruit picking activities. This is because harvesting is considered one of the most tedious, time consuming and labor intensive agricultural activities, which is even more problematic given the labor shortage the system is facing and the high cost of human resources: using intelligent robots equipped with computer vision technologies to detect and pick products in an autonomous way can help in alleviating such problems.

Nonetheless, the development of an AI-enabled fruit detection system is non trivial due to many technical, economic and social considerations that have to be addressed and are instrumental in the integration of technologies into agricultural operations. A complete understanding of the application domain specificities is indeed of fundamental importance to develop appropriate solutions that deliver accurate, reliable, efficient, and context-aware fruit detection capabilities. To provide for this, careful and exhaustive considerations of many different intricate aspects related to data collection and annotation, feature extraction, model selection and evaluation, hardware architecture are to be made.

This elaborate, through a review of the relevant literature, aims to provide valuable information for the development of an AI enabled fruit detection system by analyzing relevant aspects with respect to necessary data, existing algorithms and required hardware. In this sense, after providing in Chapter 2 a general overview of the different applications of Artificial Intelligence in agriculture (with a specific focus on fruit detection), in Chapter 3 the main implications with respect to the necessary data when building an AI system are presented. In Chapter 4 existing Machine Learning and Deep Learning models are investigated and comparatively analyzed. In Chapter 5 possible platforms and hardware considerations are outlined. In Chapter 6, an analysis of the main technical and socio-economic challenges that the development of an AI based fruit detection system pose is presented. The final chapter summarizes the conclusions on the above aspects.

# Chapter 2

# Digitalization of agriculture through AI

## 2.1 Agricultural activities enhanced by AI systems

The digitalization of agriculture has been gaining momentum given the urgency of optimizing agricultural practices, increasing the agricultural productivity and providing for more sustainable, informed, effective activities owing to the numerous issues the sector is facing such as increasing population demanding more food production, labor shortage, adverse climate and scarce water resources, among others. In particular, Artificial Intelligence is playing a key role in driving the current digital transformation of agriculture: Machine and Deep Learning algorithms have been vastly applied, in combination with other technologies pertaining to IoT, Big Data, Robotics, for the development of intelligent systems of valuable applicability in many different areas and to support many relevant agricultural activities by automating and speeding up redundant processes like crop harvesting and monitoring, as well as providing accurate knowledge throughout efficient predictive and prescriptive analyses of the vast amount of agricultural data that is possible to collect.

As described by [7], AI methods have been mainly applied for developing:

- autonomous agricultural robots to support activities like fruit picking, weed management, crop spraying and monitoring, thus replacing tedious, repetitive manual labor and improving overall productivity;

- agricultural decision support systems: thanks to AI based systems able to process vast amount of remotely sensed data, more effective and efficient decision-making processes are possible with respect to crop management, water management, harvesting and post harvesting activities;

- (mobile)expert systems: mobile/embedded devices can be used by farmers to rapidly identify crop diseases, to detect on tree products and get relevant statistics about them, as well as to monitor the overall field by being able to process satellite images

- predictive analytics: together with Big Data technologies, AI enabled agriculture predictive analytics can forecast crop yields, weather conditions and market conditions, providing to farmers useful insight on the production course so that resource optimization and proactive beneficial actions can be implemented.

Some of these applications heavily rely on object detection technologies. In light of this, the most relevant aspects characterizing the task of object detection, declined in the example of fruit detection, such as applications and main obstacles, will be further examined in the next section. On the other hand, an analysis of the main agricultural activities concerned by the application of the aforementioned AI based solutions is presented below, specifically relating to crop monitoring, yield prediction, disease and weed detection, irrigation management, soil management, harvesting and crop quality.

As final considerations, existing AI and computer vision technologies have proven efficient in addressing traditional monitoring, forecasting and planning activities, hence reducing costs, effort, time, inaccuracies and inefficiencies in the system. However, despite some real case applications, an extended adoption of such technologies is still to be realized and some of the causes of this have been analyzed by and ascribed to: lack of knowledge of the farmers, gap researchers-farmers, cost related issues [8] [9]. In addition to this, most of the results are still in the experimental phase, especially because complete solutions to some limitations due to the variability and complexity of agricultural environments haven't been found yet.

### 2.1.1 Yield Prediction

Yield prediction is regarded as both a crucial but also a challenging task. It is a task of paramount importance since a correct estimation of a crop's production can help farmers take more informed and effective decisions related to orchard management activities, as well as labor, stock and transportation allocations and planning. Nonetheless, these predictions are affected by numerous factors such as environment weather patterns, soil type, management practices, crop genotypic and phenotypic characteristics, and all their interactions [10]. Since the crop yield will be highly dependent on an optimal combination of these variables, their prediction and comprehension is extremely valuable: for example, undetected anomalies might have relevant negative effects on the crop production (e.g. huge crop losses) with all the consequent implications this creates. Considering this, accurate Machine Learning and Deep Learning models can provide valuable support.

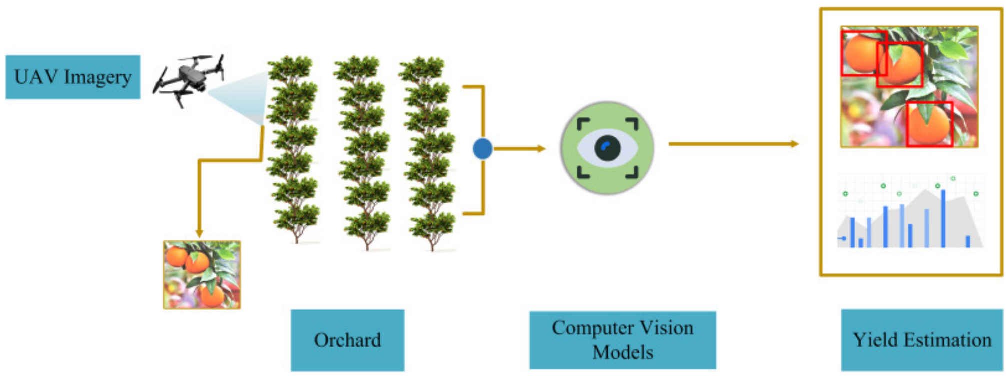


Figure 2.1: Orchard yield estimation enabled by the use of an unmanned ground vehicle for image capturing and computer vision techniques for image processing and analysis[11].

By analyzing historical and real time data from sources like remote sensors, satellites, unmanned vehicles (Figure 2.1), AI based systems are able to both foresee the factors that will affect the crop yield (like possible occurrence of disease), enabling farmers to proactively take corrective actions[12], but also to predict the final produce of a specific crop, its ideal harvest time and required inputs (e.g amount of water or fertilizers) [13]. Therefore, relying on such innovative predictive models can have beneficial outcomes on several agriculture management activities, improving decision making processes and maximizing productivity and profits.

### 2.1.2 Disease and Weed Detection

Among other activities needed to safeguard a crop health, of relevance are the one aimed at preventing and controlling the dissemination of diseases on fruits and plants, as well as the spreading of weeds, which is detrimental for the agricultural produce since contend to obtain water, minerals and other nutrients found in the soil: minimizing the presence of diseases and weeds is crucial in impeding significant crop damages that eventually translates in huge production and economical losses for farmers [10].

Currently, traditional inspection methods rely on expert farmers frequently examining crop fields on the lookout for any hint that could signal the presence of diseases and on manual or mechanical control of weeds, making such activities laborious, expensive, time-consuming and not always sufficiently accurate[14]. To cope with such constraints, ad hoc AI systems combining computer vision technologies, pattern recognition [15], machine learning and exhaustive agricultural measures, have been developed and proved efficient in detecting diseases and malicious weeds in a more precise and prompt way [16]. Once the identification step is done, the smart system is also able to provide an instant diagnosis and to suggest both urgent and preventive actions [13].

In addition, for the containment of diseases and weeds, pesticides and herbicides are usually used. Nonetheless, they are costly [12] and if not correctly quantified in their usage or if applied uniformly in the field although the presence of weeds or diseases being uneven, they can cause serious damages to the surrounding environment, waters, animals [14].

Solutions to such problems can be found again in digital agriculture: relying on IoT sensors, robotic/aerial vehicles (Figure 2.2) and AI methods, some smart systems have been developed, with the ability to distinguish weeds from crops or to localize diseased crops, and thus realize both an accurate spraying of pesticides and herbicides only to the specific areas needed or an autonomous exacerbation of weeds [13].

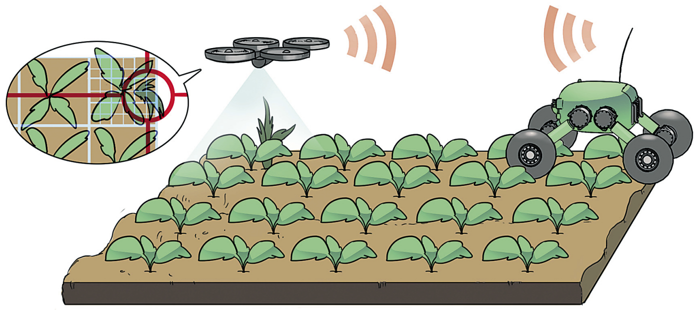

Figure 2.2: Weed detection example: an UAV examines a field, collecting data about crop density and weed pressure and coordinating and sharing information with a UGV that is used for targeted intervention and data analysis [17].

### 2.1.3 Crop monitoring

Monitoring the health status of a crop during its entire growth process is an important agricultural task for obtaining high quality abundant yields: a constant flow of information pertaining to parameters necessary to determine a crop health status, allow to uncover issues like nutrient deficiencies, water shortage, presence of diseases ahead of time. This enables agronomists to proactively adopt corrective actions, aimed at creating an optimal growth environment for the crop, so that the likelihood of obtaining a productive harvests is increased. As many other agricultural activities, also the monitoring of crops traditionally relies mainly on human labor, which is fallacious since based on subjective judgment and is not timely nor enough accurate [16]. Fortunately, advancements in computer vision and deep learning inspired the creation of solutions to perform analysis of crop images and provide therefore real time and extensive information on crops status (Figure 2.3): for example, such systems have proved able in detecting subtle changes in crops due to malnutrition or presence of disease beforehand human monitoring, as well as in providing constant and more reliable, timely information [16].

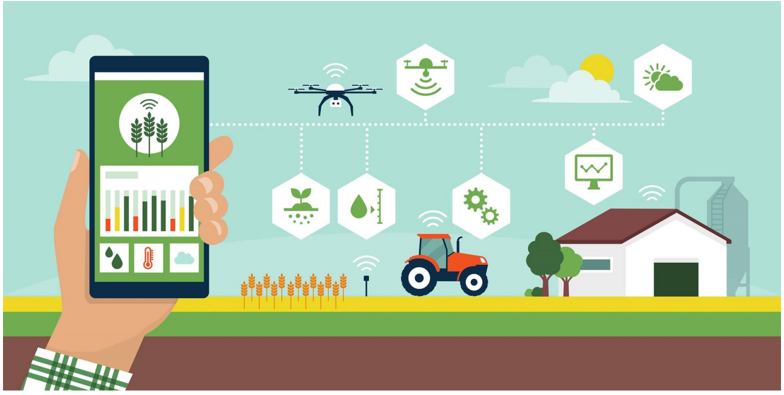

Figure 2.3: The combination of computer vision technologies, remote sensing information and robotics enables the real-time monitoring of a crop's health and of its environment

### 2.1.4 Soil Management

Knowledge related to soil properties can be very valuable in agriculture since allows for the understanding the occurrence of issues related to aspects like land degradation, unbalanced nutrients in soil due to fertilizers overuse and soil erosion that play a crucial role in a crop productivity [10]. Traditional assessment of soil properties is based on soil sampling and laboratory analysis, which usually are quite expensive, time consuming and require qualified experts [10]. In contrast, alternatively or complementarily relying on innovative technologies can be beneficial in numerous ways: the use of computer vision for collecting useful data from soil images and of IoT remote sensing for monitoring relevant parameters (temperature, moisture, electrical conductivity, pH value and nutrient content [12]) in real time, combined with AI based methods for analysis of the different data collected, allow the development of reliable cheap solutions for supporting soil management activities, eventually helping farmers in obtaining larger and better yields while minimizing the time and cost of such agricultural practices [13]. For example, computer vision based technologies have been applied and proved useful in the characterization of soil properties like soil texture, meaning the percent contribution of sand, silt and clay (Figure 2.4) [11]. Soil texture is a foundational element in supporting the ecosystem and greatly affects other soil physical properties and processes like infiltration, water holding capacity and drainage, aeration, proneness to erosion, pH buffering capacity, which all ultimately affect the agro ecosystem productivity [11].

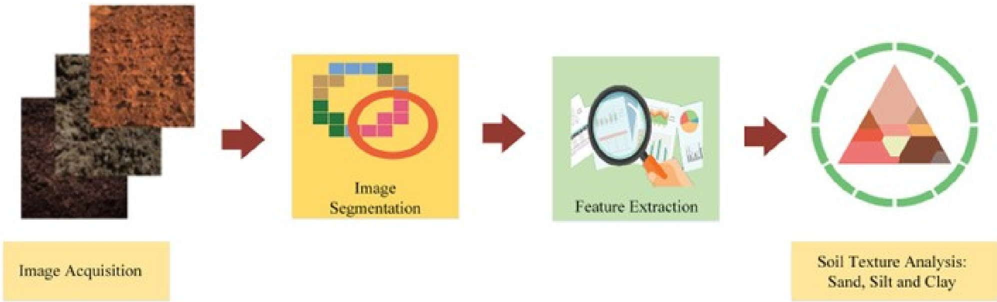


Figure 2.4: Computer vision-based soil image analysis: soil images are collected with cameras and submitted to simple computer programs to classify and categorise them considering portion of sand, silt, clay [11].

### 2.1.5 Irrigation Water Management

As stated in [18], agriculture is the sector that globally consumes available fresh water the most, since the growth and nourishment of plants and crops largely rely on it. Nonetheless, water resources are more and more scarce due to both natural factors and climate changes but also inaccurate watering practices, all which implies that interventions aimed at preserving water availability are urgent. More sustainable and efficient approaches to watering rely on a precise knowledge of the amount of water a crop requires, as well as on implementing a variable irrigation on the basis of specific requirement different zones can present, so that to avoid an unnecessary uniform irrigation on the entire field [10]. In addition, in order to provide an efficient scheduling and management of irrigation, variables like precipitation data, evaporation data, weather forecast, soil moisture data, crop growth status can result very helpful [15]. Using remote sensors, all such information can be collected and used for analyzing soil moisture and crop status. Integrating then AI methods, predictions and better decisions on the basis of the obtained values can be suggested: for example, many smart irrigation systems able to control actuators (e.g sprinkler) in an autonomous way according to the received data [13] have been developed (Figure 2.5), as well as computer vision based robots that, after analyzing and comprehending a specific situation, can perform autonomous precise irrigation only where and in the quantity needed[11].

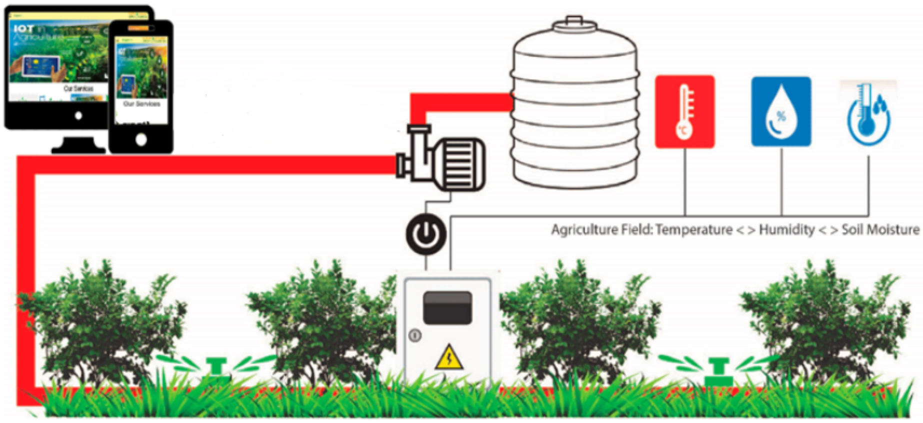


Figure 2.5: Smart irrigation system: IoT sensors gather in-field data, accessible through online applications, and are able to control irrigation pumps, which get automatically turned on and off depending on the soil moisture data that have been sensed [19]

.

### 2.1.6 Harvesting

Once a crop has finalized the growth stage and reached maturity, the agricultural produce has to be collected. Manual harvesting is still the main operating method used, which turns out to be tedious, time-consuming and costly, given the scarce availability of qualified workers [13]. To overcome such constraints, technologies pertaining to smart sensors, robotics, AI and computer vision have been considered for making harvesting processes smarter. Among others, many researches have focused on the development of AI enabled robots that automate and speed up the picking processes. They work thanks to computer vision technologies that allow the identification and localization of fruit/vegetables ready to be harvested, and then instruct a robotic arm to carefully gather the target, which is ultimately placed in a dedicated bin [13] (Figure 2.6). In addition, by leveraging Machine Learning predictions based on remotely sensed data, smart harvesting systems can provide agronomist with both yield forecasts and useful insights on crops produce, helping them in achieving the ideal harvest and supporting them in efficiently managing and planning pre harvesting, harvesting and post harvesting activities and resources [15].

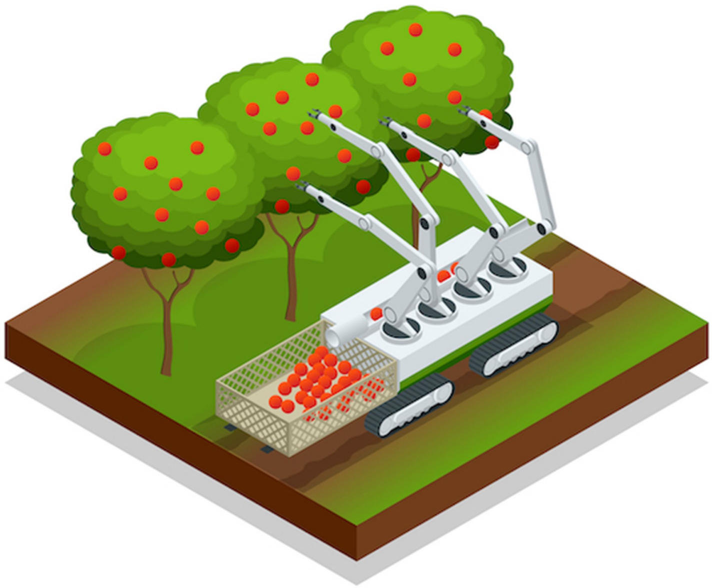

Figure 2.6: An example of selective harvesting robot: once the fruits that are ready to be harvested are localized through the use of sensors and computer vision techniques, the robotic arms carefully pluck them with the help of a picking gripper, and places them in the appropriate basket.

### 2.1.7 Crop quality

The quality of a crop produce, in terms of shape, color, size, plays a fundamental role in the determination of the market price at which the products can be sold and in the customer satisfaction, with the corresponding implications these factors have on a farmer's final economical gain. As seen in the above sub-sections, also in this case AI algorithms combined with imaging and computer vision technologies can perform automatic and meticulous quality inspections, evaluation, grading and sorting of agricultural products, thus avoiding the high cost and low efficiency of traditional manual operations usually applied for such activities [10].

## 2.2 AI for fruit detection: applications and problems

Together with image classification and segmentation, object detection constitutes one of the primary techniques computer vision is based on, which refers to all those AI algorithms to instruct a machine in extracting valuable information from images or videos [20]. Object detection approaches are useful in this sense since they can, through an accurate analysis of the image, localize and subsequently classify identified objects of interest, surrounding them by bounding boxes. Due to their vast applicability, there is nowadays a growing interest towards object detection technologies in numerous different applications domains, one of which is agriculture. Specifically, the topic of fruit detection has received increasing attention in the past decade owing to the relevance such technologies can have in improving numerous fruit production and management activities, especially those requiring vision as a sensor [21].

First off, fruit detection technologies have been implemented over aerial/ground vehicles or smartphones applications that, after capturing the image of a crop/tree through specific cameras, can quickly track down the present fruits or count them. Once the fruits are detected, an analysis phase usually takes place. This means the technology can provide exhaustive information on the fruit status with respect to, for example, the maturity level, its size or the presence of diseases, as well as yield estimates based on the counted number of fruits, all of which constitute valuable knowledge for a farmer for the following reasons. To start with, by knowing a culture is affected by pests or diseases, corrective actions can be proactively taken to avoid the infection propagating in such extended quantities that causes significant crop losses. These actions can even be implemented as additional functionalities of the intelligent robot used for detection, so that the machine vision technologies they are equipped with allows them to first detect and examine the fruits and then precisely and autonomously spray pesticides when and where needed. In addition to this, fruit detection technologies can be devised to analyze a fruit's appearance, from which valuable quality information can then be derived: the maturity level can signal their readiness for harvesting time and so help in planning related activities, while size or color results crucial especially for some crops, since the market recognition and thus the corresponding obtainable profitability much relies on these aspects [22]. On top of this, computer vision based fruit detection and counting enable the automatic forecasting of a crop yield, which is extremely valuable for a number of reasons. Commonly, fruit yield estimation has been performed by manual counting of the fruits in selected trees, activity that has proven to lead to low precision results (this is because usually just a small number of tree gets inspected, the amount of fruit on each tree can be highly variable and the assessment should be done at several times during crop growth [23]), high costs and higher time requirements for estimations, making subsequent decision making and planning activities a challenging task. Conversely, through computer vision applications/methodologies, fruits can be quickly detected and then counted, so that an early prediction of the final yield of an orchard is achievable. As a result, this can help farmers in taking more effective decisions with respect to resources that will be needed in pre harvesting activities for an optimal orchard management (e.g. quantities of fertilizers or agricultural chemicals to support the achievement of the desired yield), as well as decision referring to post harvesting activities like allocation of labor, storage capacity needed, transportation[16], thus ensuring a more efficient and profitable overall agricultural production [24].

Another agricultural task where fruit detection technologies can provide many advantages is fruit harvesting. This is because, traditionally, harvesting has been performed through manual labor, which nowadays is problematic firstly because if on the one side it is considered a labor-intensive activity, on the other side the increasing cost and ageing of the agricultural workforce are reducing the availability of human resources to carry out such tasks [25]. In addition, harvesting is deemed as being an extremely time consuming, costly and tedious task, with impact on a person's health too, given the uncomfortable positions to execute it[25]. In light of this, automatic harvesting systems can result very helpful in alleviating these problems: robots equipped with computer vision technologies can indeed autonomously instruct a mechanical arm to carefully pick the detected fruits and place them in a dedicated bin [13]. Autonomous harvesting and picking robots are among the most popular robotic applications in agriculture, due to the improvements in the speed and accuracy achieved in recent years and in considerations of the benefits they provide since they free farmers from having to carry out such tedious activities and thus allow them to focus on more strategic issues [26].

In this scenario, automated fruit detection results a valuable resource in improving fruit production and is indeed a research topic of increasing interest: many fruit detection frameworks have been developed over the years considering different methodologies based on feature extraction, machine learning or deep learning. Early approaches to image analysis for fruit detection have been based on manually design the methods to identify and extract distinctive features either pertaining to the image, such as corners, edges, blobs or to the object the interest, like color, shape, texture, that could signal the presence of a fruit in the image. Specifically, these fruits detection algorithms generally present a first step of generating potential region of interest, followed by the actual detection of the fruit. The selection of candidate regions can rely on threshold segmentation to quickly distinguish the fruit from its background based on the provided feature [27]. To this end, color is usually used though it can present misleading variations due to illumination or maturity level. In turn, a fruit's shape or texture appear to be more robust with respect to these elements. With respect to shape-based techniques, there exist ,for example, the Hough transform (HT) or the histogram of oriented gradients (HOG), while some common texture-based methods can referred to oriented FAST and rotated BRIEF (ORB), speeded-up robust features (SURF), scale-invariant feature transform (SIFT) and local binary patterns (LBP) [22]. Complex algorithms that combine many of these features have proven to lead to more precise outcomes. Although these simple image processing methods have been proven able to detect fruit targets [27], more accurate results have been obtained when coupled with machine learning algorithms: through such approaches, once regions of fruit objects are distinguished through the considered features, they can be categorized as either fruit or background by applying classifiers such as Support Vector Machine(SVM), K-Nearest Neighbour(KNN), AdaBoost, Random Forest [22]. Nonetheless, Machine Learning approaches presents some limitations too. Relying on manual feature extraction, such techniques have low generalization capabilities, are dependent on the quality of the selected features and accuracy improvements in the detection of fruits, when considering more challenging field environments, are difficult to achieve. As an example, Machine Learning algorithms used for fruit counting struggle with both clustered fruits, detecting them as one, and with fruits that appear to be divided in two by a branch, which will be considered as if there were two [28]. In such context, deep learning techniques have been increasingly used for deploying fruit detection tasks, given the promising results obtained in overcoming the challenges encountered by other techniques and in providing therefore more accurate outcomes. In particular, the superiority of convolutional neural networks lies in their strong self-learning ability: while previous approaches rely on hand engineered features to encode visual attributes to differentiate fruit regions from non-fruit regions, which makes them suitable only for a specific fruit and unique to the conditions under which data were captured, CNN have the capability of automatically learn also discriminative higher level image characteristics which cannot be extracted by traditional feature engineering [20].

No matter the fruit detection approach, the complexity and variability of outdoor orchard environments hamper high levels of accuracy and the performance of the selected detection system [20]. First of all, outdoor environments are subject to variable lightning conditions depending on the time of the day and the current weather, so that the light intensity is different in sunny and cloudy days. As a consequence, the fruit image can present oversaturated areas because of direct sunlight while capturing it, shadowed parts due to light occlusion and variable brightness depending on the light intensity [20]. This result to be problematic for a machine vision algorithm since if the fruits are not equally enlightened, and thus for some the color appears to be different, these will be ignored in the detection phase. Similarly, detection tasks became challenging in all those cases where the target fruit is of similar color with respect to its background or the leaves and branches of the canopy/plant, since this makes more difficult to identify them on the basis of color information. For some type of fruits this could happen, for example, at the beginning of their growing stage since their color is more similar to the foliage, changing mostly when the fruit has reached maturity [20]. Another problem comes from dense canopies/vegetation where the fruits in the considered image appear hardly visible since occluded, overlapping or clustered, leading to misclassification and miscounting [21].

In an effort to overcome or at least reduce the effect such constraints have on fruit detection performance, researchers have introduced some optimization strategies. A first group of optimization approaches aims at collecting higher quality datasets, where the number of images presenting the above-described problematic characteristics is minimal. Choosing the most opportune visual camera or sensor considering its imaging characteristic and applicability to the project in question can help in this sense, as well as trying to make unstructured environments more structured, for example by planting orchard in a more standard way, by accurately pruning a crop's branches and leaves and by using controllable light sources when shooting for collecting sampling images [20]. Others optimization strategies promote methods for postprocessing the collected dataset. Dataset processing techniques rely on augmentation of the available data or creation of synthetic data with the aim of enriching the quantity and variety of available dataset: data augmentation consist in a series of transformations applied to an image such as rotation, translation, mirroring, scaling, cropping, remapping into another color space, while synthetic data are data created by a computer that present similar characteristics to the real one. Additional optimization strategies carried out after the collection of necessary images, which are more robust but more technically challenging, rely on methods for improving the considered fruit detection model (for example introducing correction factors [27] or developing more powerful feature extraction [20]), as well as designing it according to the peculiarities of the dataset, selected sensor and requirements of the detection task [20].

# Chapter 3

# Data for an AI fruit detection system

## 3.1 Data acquisition

The collection of data, as suggested by a typical Machine Learning pipeline, is the starting point for the development of any Artificial Intelligence system and its paramount role in the system's final accuracy and reliability cannot be neglected: as suggested by Andrew Ng [29], one of the most globally recognized leader in Artificial Intelligence, around 80% of machine learning is data preparation and most of the accuracy improvements that can be achieved rely on high quality data. Specifically, for any ML or DL system to be effective and trustworthy, it is indispensable that the training process is initiated with huge quantities (e.g. hundreds of them for ML models, thousands for DL models) of diverse and high-quality data. Using a large dataset improves the model's capacity to learn and recognize patterns, while diversity provides the model with generalization capabilities when applied to new or unseen images [30]. In addition, data-centric attributes such as image resolution and object appearance, complexity, size and similarity play a very important role in the final accuracy of the detection task [31]. These aspects are especially true when considering the implementation of an on-tree fruit detection model: in addition to being of high quality and substantial, agricultural datasets must be indeed as much representative as possible of the variability that open-field orchards present, in order to allow an effective training of the selected model and thus make it robust enough to be able to work under different environment conditions [32]. In view of this, it is clear how a meticulous approach to data collection, encompassing representative samples and relevant features that comprehensively encapsulate the context complexity and object diversity expected, ensures the ability of a model to make accurate predictions and avoid underfitting/overfitting scenarios, granting the overall robustness and reliability of the entire ML or DL pipeline [33]. Nonetheless, data collection is a laborious task to which careful attention and sufficient dedication should be paid.

The acquisition of image data can happen in two main ways, namely by using publicly available datasets or by building a custom dataset. Both approaches present some advantages and disadvantaged that will be analyzed in the following dedicated paragraphs.

### 3.1.1 Public dataset

The advantages of relying on publicly available datasets can be attributed to their accessibility and convenience, reducing the effort and costs for data collection, preparation and annotation processes. In the research environment, it is a widespread practice to release and make available to others datasets that have been collected throughout experimental studies, to both foster the implementation of new models or to verify already developed one and possibly propose better solutions. Indeed, many vast benchmark datasets for computer vision tasks have been released like ImageNet [34], PASCAL VOC [35], COCO [36], ILSVRC [37], Open Images V4 [38], Fruit-360 dataset [39], allowing for successful improvements in object detection and encouraging the development of new architectures [32]. Although they provide for fruit datasets, the images are usually of individual fruits collected in simple scenes that can be mainly used for fruit detection within specific, structured environment: while they might be helpful in general situations, they have proved to perform poorly for models applied in open orchards, since to complete fruit detection for selected agricultural tasks, they necessitate to be trained on datasets containing more specialized images and images collected under more realistic open field settings [32]. Such necessities and the growing interest and development of computer vision applications for agriculture have favored an increase in the number of available public image dataset for task related to smart agriculture [32]. Specifically, as reviewed by [32], some usable, annotated datasets for the task of fruit detection are:

- DeepFruits [34]: around 42–170 RGB images of various resolution are provided for fruits like sweet pepper, rock melon, apple, mango, orange and strawberry with corresponding bounding box annotations of fruits
- Orchard Fruit [40]: provides 1120 apple images (size $308 \times 202$ pixels), 1964 mango images (size $500 \times 500$ pixels) and 620 almond fruits images(size $308 \times 202$ pixels). Circular and pixel-level fruit annotations are provided for apples, while rectangular bounding box annotations for mangoes and almonds
- Date Fruit [41]: consist of a first subset of 8079 color images (size $224 \times 224$ pixels) labelled into different classes according to fruit variety, maturity and harvesting decision; another subset contains the images, videos, and weight measurements of date brunches
- K Fuji RGB-DS [42]: 967 multimodal images (size $512 \times 424$ pixels) and bounding box as fruit annotations for a total of 12,839 apples
- MangoNet [43]: provides 49 high-resolution ($4000 \times 3000$ pixels) color images in jpg format, collected under natural illumination conditions. To note, in the model training phase images will have to be cropped into smaller patches to meet computation memory issues [32]
- MangoYOLO [44]: consist of 1730 images (size $612 \times 512$ pixels) collected at night with artificial lighting, for better image contrasts
- WSU apple dataset [45]: consist of 2298 RGB images of various resolution of apple trees acquired from multiple growth seasons and fruit varieties
- Fuji-SfM [46]: provides 288 RGB apple image (size $1024 \times 1024$ pixels) collected using a handheld color camera in the natural orchard conditions
- LFuji-air dataset [47]: provides annotations of 1353 apples from 11 fruit trees

- MinneApple [48]: two sets of images under natural illumination conditions are provided, one dedicated to fruit detection task with 982 images (size 1280 × 720 pixels) and another for counting tasks with of 64595, 2875 and 3395 images for training, validation and test respectively (Figure 3.1)

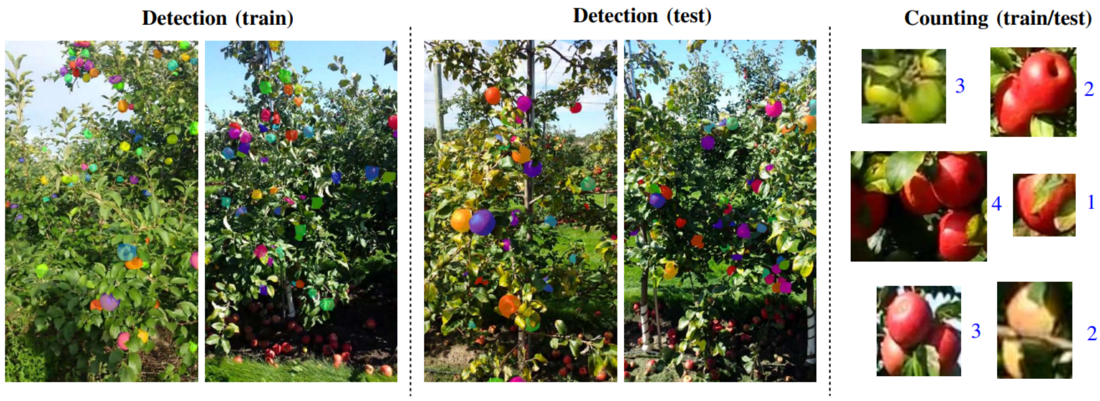


Figure 3.1: Examples of MinneApple's dataset annotated images for the detection, segmentation and counting tasks. The detection/segmentation datasets are annotated with object instance masks, while the counting dataset contains image patches and a corresponding ground truth count [48].

Nonetheless, it is clear that the field of fruit detection still suffers from the lack of openly accessible image data, a problem that gets magnified for some type of fruits for which no public dedicated dataset has been yet made available. To note, there exists datasets that have been collected by research institutes to conduct their own experiments on the topic of fruit detection, but they have not been made available to the public or permission for their use has to be requested. As a result, in some cases relying on public dataset for image acquisition is not viable option, and thus a customized data collection approach will be needed.

Some other disadvantages can be pointed out. First of all, a system developed considering public datasets might turn out to be less accurate, since the data it has been trained on are not specific and representative of the intrinsic characteristic of the field where it will be used. Furthermore, being collected by others, such datasets will have to be inspected before use in order to understand if are applicable to the specific case considered, to check the quality of the image data and of their annotations (which might even result inadequate) and to eventually consider some data cleaning procedures. Additionally, it might result that the considered public dataset lacks in representing the environmental diversity that open orchards present, which is necessary for precise in field fruit detection, as well as a sufficiently high number of images (the training of deep learning models, for example, require thousands of data for the development of a performing system). Indeed, since image collection and annotation are costly and time consuming, often the available datasets present too few images (compromising the model capacity to recognize patterns) and are not sufficiently diverse (affecting the model generalization capabilities and robustness)[32].

### 3.1.2 New dataset

Building a custom dataset is a valuable option for data acquisition when it is a viable solution, since it implies the creation of a dataset of images tailored to the specific characteristics of the orchard. Indeed, the creation of custom dataset allows to better control the quality of both the captured images and their annotations, as well as to ensure that the images in the dataset exhaustively represent the environment variability, thus allowing the development of a more accurate and robust fruit detection model, able to perform well under mutable conditions which are typical in open fields (change in illumination, differences due to maturity stage, occlusions).

However, acquiring a new dataset is extremely time consuming, expensive and laborious for many reasons [49]. First of all, specific and usually expensive equipment, like cameras with good resolutions (e.g. DSLR cameras [40]), and experts are needed for the acquisition of huge numbers of orchard images that have to be sufficiently representative of various environmental conditions (varying illumination, clustered or occluded fruits) and of different fruit growth stages, while ultimately being of high quality. The importance of images of higher resolution, containing more feature information for the model to learn, has been emphasized since it can ensure higher training accuracy and can be used to achieve so even in those cases where the amount of available images is limited, as a strategy to increase accuracy through increasing the input image resolution [31].

Once images are collected, they need to be annotated. Generally, image annotation refers to the process of semantically describing the content of the data through some labels that, depending on the use case, can be numerical or categorical. In the task of fruit detection, it mainly refers to defining the regions of interest that capture the occurrence of fruits and assigning some context information, such as the class of fruits present in the image [32]. The annotation process, though time and cost intensive, is of vital importance since the more meticulous the annotations are, the better the system can learn, ultimately delivering more accurate and robust performances in the detection of fruits even in complex orchards. There exist different ways to carry out the annotation of data. A first approach is manual annotation, which is the prevalent way to label data in computer vision task [32]. The main advantage of manual annotation is that it allows higher level of accuracy and control over the annotation process, so that objects are precisely marked and labels result more rigorous and suited for the specific use case considered. To support it, there exist open-source software tools such as labelImg, VGG Image Annotator(VIA), COCO Annotator, CVAT, LabelMe, ImageTagger, imglab, OpenLabeling, VoTT [32]. In Figure 3.2, an example of the manual labelling process of fruits with the help of the open-source tool labelImg is shown[50]. Despite these tools, manual annotation of images is extremely time-consuming and expensive, given the human resources required for it, and might not always be applicable either, especially when the number of images to be annotated is too huge (thousands of images) and of high resolution.

Figure 3.2: The manual labelling process used in [50] to identify where the fruits were located in each image from the training dataset with the help of LabelImg tool to label object bounding boxes. On the left, the annotation process for the test images is shown, on the right the different types of fruits labelled.

In such cases a solution can be found in the use of commercial platforms that offer annotation through crowdsourcing, such as Amazon Mechanical Tunk for example [51]. This requires less time and has been vastly applied for annotation of large-scale image datasets. Nonetheless, annotations are performed by non-experts and thus additional quality checks are required [52].

Another approach to data annotation is automatic labelling, which involves leveraging developed computer algorithms to assist or perform the annotation process with little to no manual intervention [52][53]. Automatic Image Labelling(AIA) technologies aim indeed at training a learning model on the basis of a given image, and on using such model to automatically assign semantic labels: even though such approach has proven to be highly efficient, for more complex images such as those relating to the agricultural environments it has not always produced satisfactory results [32]. Nonetheless, researches on the topic are increasing and technological advancements are allowing the implementation of improved models, to provide more applicability and efficiency in the agricultural context [53][33][54][55][56].

## 3.2 Data processing and data augmentation

Performing object detection within agricultural environments is intricate, given the numerous complications open fields pose with respect to variable lightning conditions and complex orchards structures that can thus compromise the detection accuracy of an AI system. Since neither feature extraction combined with machine learning nor deep learning models have proven fully capable in dealing with such challenges, accurate image pre-processing techniques can provide valuable additional support in this sense and thus strengthen the model final performance's accuracy [57]. Data pre-processing is composed of a series of procedures to refine, standardize and enhance the image data with the aim of making the considered dataset more consistent and better representative of meaningful patterns that can be learnt by the model. There exist in the literature many image pre-processing techniques [58]. Resizing is useful for ensuring a consistent resolution that is suitable for the selected model and the memory constraints, while cropping may be applied to focus on relevant regions of interest. Normalizing pixel values to have them within a standard range can help in reducing the impact of variations in lighting conditions across different images. Image enhancement provides adjustments of contrast, brightness, or sharpness, thus helping to accentuate important features in the images. For example, [57] implemented some color correction techniques like Adaptive Histogram Equalization and Histogram equalization and proved to have obtained higher performances than not applying such techniques. Additionally, geometric transformations can be considered in the preprocessing phase, which refer to the so called data augmentation techniques.

Data augmentation consist of generating variations of the considered images in an artificial way, through the application of a series of transformations, with the end goal of improving the dataset size and variability and thus supporting a model's capability to generalize to different scenarios [30]. Most common augmentation techniques are (Figure 3.3): image or bounding boxes rotation (so that the model can learn in different orientations [57]); scaling, to change an image's size so that the model can recognize object at different scales; cropping to add variety; flipping to create mirror copies of an image thus exposing the model to more information [59]. Other techniques refer to random occlusion augmentation, photometric augmentation (to change the pixel intensity values instead of pixel positions) or deep learning-based methods for data augmentation (e.g. Neural Style Transfer, Adversial training and Generative Adversial Networks - GAN) [59][30].

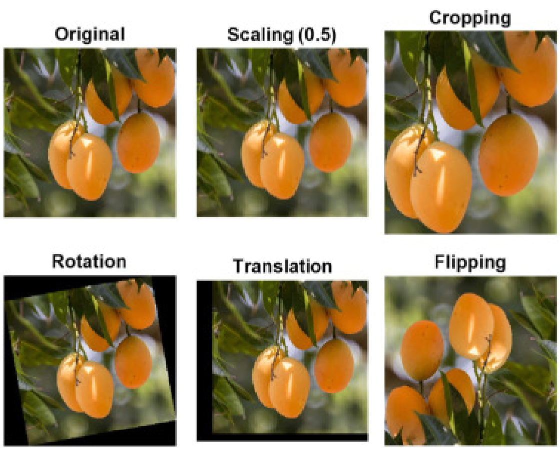


Figure 3.3: Data augmentation techniques: the image shows some geometric transformation methods used in image processing and computer vision to manipulate and modify images [60].

Data augmentation techniques have been widely considered especially for the implementation of deep learning models. This is because, in order to be able to identify sufficient complex patterns and thus produce reliable and accurate outcomes, the training phase of a DL model has to be build upon huge amount of diverse data, the collection of which is challenging due to manual effort, costs and in some cases it might even be infeasible [30]. Specifically, in [31], the experimental study conducted proves that a training dataset of 2500 objects is sufficient for most single class fruit detection. In such context, data augmentation techniques have been used to compensate in situations of scarce data.

# Chapter 4

# Models for an AI fruit detection system

The model of an AI based system represents its fundamental core since embodies the mechanisms through which it can first learn distinctive patterns and discriminative features present in the provided images, and then generalize this knowledge on new data. Given the relevant outcomes that implementing automated fruit detection system can produce in the agricultural sector, it has been a research topic of increasing interest in the past decades: many fruit detection frameworks have been developed over the years considering different methodologies based on feature extraction, machine learning or, most recently, deep learning. Selecting the appropriate model is a crucial step in developing an AI-based fruit detection system suitable for the specific use case considered and requires careful considerations on aspects such as the complexity of the task, the size of the dataset, computational resources, desired level of accuracy, necessary investments. In the literature, there exist a vast number of research papers that describe specific implementation of ML or DL techniques for the task of fruit detection considering these elements, which can be advantageous for benchmarking the proposed solution and adapt it to the case considered. A general overview of the main characteristics, existing algorithms and advantages and disadvantages of machine learning and deep learning-based models is presented in the following paragraphs.

## 4.1 Machine Learning models

### 4.1.1 General considerations

Early approaches to image processing and analysis for fruit detection have been based on manually designing procedures to identify and extract distinctive features pertaining to the fruit objects in an image, such as color, shape, texture, corners, edges, blobs. Specifically, such features are usually used as the basic element for segmentation algorithms (e.g. watershed algorithm [61]) to quickly distinguish the fruit from its background and make decisions about whether a certain region of an image corresponds to a fruit or not [27]. Color based techniques extract color features through Color Histogram, Color Set, Color Moment, and Color Coherence Vector. They showcase great results when the fruit color is notably different than its background and are less dependent on image size. However misleading variations of a fruit's color due to illumination condition or maturity level can undermine the efficacy of such approaches. In turn, techniques based on shape features, among which Hough transform (HT), Circular Hough Transform (CHT) or the histogram of oriented gradients (HOG) can be mentioned, result to be less affected by these elements, being useful also in those cases where fruit and background have similar colors and when the fruit shape differs with respect to the surrounding leaves and branches [22]. In addition, texture-based methods like oriented FAST and rotated BRIEF (ORB), speeded-up robust features (SURF), scale-invariant feature transform (SIFT) and local binary patterns (LBP) can provide valuable support in distinguishing fruits of the same color as the background since leverage on the smoother surface that fruits present, which is also invariant to changes in lighting conditions, though might not provide accurate results in case of clustering, overlapping and occlusion of fruits [22]. To overcome the many different limitations these approaches present, methods based on a combination of more of these features have proved able to improve the accuracy and robustness of the fruit detection task [62]. Although these simple image processing methods have been proved able to detect target fruits [28], their capability in doing so is limited to more structured environments and thus might not result sufficiently able to deal with more complex and variable scenes affected by many non-controllable factors (changing in lighting conditions, occluded and clustered orchards) like agricultural fields are.

An analysis of fruit detection based on image processing methods has been carried out within the ONIT Spa company. Specifically, a complete literature review of existing approaches to image processing and feature extraction has been first conducted and then complemented with trial implementations of some of the identified methods through Python scripts and leveraging the OpenCv and Scikit-Image libraries. The considerations that can be derived from the results obtained from the conducted investigations relate firstly to the complexity that relying exclusively on these methodologies for performing fruit detection tasks implies. Secondly, the findings highlight the inadequacy of these methodologies in addressing detection tasks within open-field agricultural environments given the diversity and variability that characterize them and the incapacity of these methods to capture all these features. As such, the necessity of considering alternative approaches or complementing these methodologies with other techniques for more effective solutions able to perform in intricate open fields is evident.

With the advent of Machine Learning the possibility of implementing more accurate, efficient and robust approaches to fruit object detection is achievable. Machine Learning is a subfield of AI and refers to those algorithms able to learn from the data they are exposed to without being explicitly programmed to do so, and then to generalize such knowledge over some other new unseen data. Such capability allows then these systems to be able to automatically improve their performance as they are exposed to more data over time. In the context of image data, given their high dimensionality, the training of the model is preceded by an additional step of feature extraction which involves transforming raw data into a more emblematic and compact representation. This is necessary in order to extract the most relevant characteristics of object in an image that could represent it in a discriminative way, thus allowing the algorithm to process simpler but still valuable data. Indeed, the first step of ML-based fruit detection algorithms is distinguishing target objects by extracting specific features (through the methodologies above analyzed) pertaining to their color, shape, texture, size. This allows to identify and describe potential regions of interest within the image, that then will be classified as either fruit or background. For this detection step, which implies multiple object classification and localization, traditional machine learning classifiers can be employed such as Support Vector Machine, k-Nearest Neighbour, Decision Trees or ensemble classifiers like AdaBoost and Random Forest [22].

The SVM algorithm performs classification by finding the optimal hyperplane, meaning the decision surface most able to maximally separate different classes in a high-dimensional feature space thus can correctly partition the training set. The hyperplane is determined by identifying support vectors, which are the data points that lie closest to the decision boundary and are therefore more difficult to get classified. Generally, techniques for detection of fruits rely on SVM algorithm since the classification problem is simplified, they appear able to handle variations in orchard conditions and resulted better performing than other algorithms [22]. On the other hand, they are sensitive to the tuning of parameters and accuracy is high only for binary classification tasks while lower for multi-classification problems.

The KNN clustering algorithm achieves classification by assigning to feature vectors a class label based on the majority class of its k-nearest neighbors. Thus, it result useful when local patterns and spatial distribution of the fruits are relevant and is advantageous since doesn't make assumptions about input data, can provide high classification accuracy and works well with small sized datasets. Nonetheless, the algorithm's detection accuracy can be affected by complex environment and lighting conditions and by suboptimal values of the parameter K, which can lead to overfitting or underfitting of the model. In addition, the computational effort and time required can be considerable, since proportional to the dataset's size, as well as the memory occupation [62].

The AdaBoost algorithm is a multi classifier approach based on boosting, through which a stronger classifier is build by combining several weak classifiers trained on the same set, where the mistakes of previous classifiers are learned by their successors. The advantage is that by combining in cascade several classifiers and emphasizing misclassified instances, higher accuracy in detection can usually be achieved. On the other hand, since the weight of difficult samples are exponentially increased, the training might result biased towards such difficult samples and thus the algorithm is sensitive to noisy data. In addition, since many classifiers are combined, the final complexity of the model is increased and so also the computational effort and time required for the training [62].

The Decision Tree algorithm performs a tree-like classification starting from a root node that defines a good split feature able to separate different classes as much as possible. The process is recursively applied to each node up until the leaf node is reached and the class label it stores is the result of the decision. Therefore, they provide high interpretability since the whole decision process is based on human readable sequence of decision rules that can be showed at different stages and considering the values of feasible solutions, allowing to consider the different affecting factors. The disadvantages are that they are prone to overfitting, can be biased towards datasets that result imbalanced since present some features with more values and struggle at dealing with strong feature correlations there could be between data. To overcome such limitations, more decision trees could be combined and their capabilities extended to build Random Forest algorithms: their aggregation approaches allows for a better capability in dealing with more diverse and complex dataset and thus more accurate and generalized results and adaptability to the changing conditions in open field orchards [22].

A model's learning through such algorithms is referred to as supervised learning since they are applicable in those cases where the data provided in input are annotated. There exist also the option of making the system learn by feeding it unannotated data, so called unsupervised learning. An example ML algorithm to do so is K-mean algorithm which can automatically partition a dataset into k distinct, non-overlapping clusters, and assign input data to the cluster with the nearest mean [22]. Though there exist some application examples of unsupervised algorithms, usually they obtain worse result and complicate the overall development process. Indeed, k means algorithm are sensitive to the random initial choice of k centroids which is non trivial and has a relevant impact on the final classification performance. In addition, they might result unable to provide accurate results in the presence of complex orchards with overlapping, occluded fruits or fruits of similar color than the background since in these cases the fruits discriminative characteristics, on which the algorithm much rely on to perform distinction, are not clearly separable [62].

### 4.1.2 Pros and Cons

The interest in using fruit detection algorithms based on traditional machine learning techniques is dictated firstly by their interpretability, since the decision rules and factors they are based on to perform fruit detection can be understood and eventually explained to other stakeholders (e.g. farmers) to encourage transparency and trustworthiness. In addition to this, ML models can perform well with medium size datasets (hundreds of image data) and thus result helpful in those cases where acquiring and labelling huge amount data might be challenging. They also require less training time and therefore less use of resources related to computational power and memory, which can be advantageous if they are limited [22]. Nonetheless, they rely on handcrafted features, meaning that the methodology for extracting them is based on human vision insights and intuition: if on the one side this can allow to emphasize the discriminative features of fruits so that they are tailored to the case considered, on the other hand this translates in inability of the model to generalize if applied to other different contexts or in presence of complexities [22]. Manual feature extraction is also a time consuming and tedious process and same considerations hold for the correct tuning of a huge number of the model's parameters, for which domain experts might be necessary [22]. Moreover, the performance of models based on such approaches might decrease if faced with more complex tasks: ML based fruit detection algorithms have proven to struggle in capturing highly complex patterns and thus in dealing with the intrinsic variability that the open field agricultural environments present, such as variations in lighting conditions and shadowing effects, fruits exhibiting different characteristics due to maturity level, fruits similar to their background and orchards presenting occluded, overlapped, clustered fruits [22]. As an example, Machine Learning algorithms used for fruit counting struggle with both clustered fruits, detecting them as one, and with fruits that appear to be divided in two by a branch, which will be considered as if there were two [27].

## 4.2 Deep Learning models

### 4.2.1 General considerations

The development of new advanced technologies and hardware supports like Graphic Processing Units, have enabled abundant researches aimed at exploring new and more efficient methods for improving object detection tasks, considering in particular Deep Learning implementations. Deep Learning is considered a sub-field of Machine Learning and it relies on the idea of imitating, as far as possible, the hierarchical nature of biological neurons, to provide analysis and learning capabilities by creating hierarchical structured layers of fully connected artificial neurons.

In the context of computer vision and object detection, the most commonly used deep neural networks are the so-called Convolutional Neural Networks (CNN), which have been specifically designed to process images by arranging neurons in three-dimensional matrices (feature maps) and by stacking, one on top of the other, many different layers (convolutional, pooling, fully connected layers) to process input data and features at different scales thus allowing the model to learn a vast amount of complex discriminative patterns that allow for accurate detection performances [22] [63]. The main innovation CNN provides is reflected in their capability of automatically learning high level features from data, without therefore needing manual hand-crafting of features (still they can be used as a pre-processing input). This is possible thanks to the convolutional layers where, depending on the designed kernel, convolution operations are able to extract relevant image features and generate corresponding feature maps. A pooling layer is usually used to reduce the spatial size of these feature maps, through a down-sampling operations by sampling the maximum or average value in a neighborhood range. The various type of visual information are thus learned to be extracted in a hierarchical way: filters to extract more simple features such as colour, edges, lines are learned in the early layers while filters to extract more complex and semantic information, such as shape and patterns, are learned in layers placed in depth. An activation layer then uses activation functions to process the input data and the neurons in the fully connected layer are connected to all activated neurons in the layer above it, to ultimately provide the final output. In consideration of this, DL methods have proved more able to deal with the various challenges that unstructured environments like open orchards present, thus providing more precise, fast, and robust fruit detection and recognition: their promising capability of achieving excellent performances have made them the method that recently has been prevalently considered for the development of various agricultural applications [64].

CNN for object detection are generally composed of two main structures: backbone and head. The backbone usually uses the first layers of the CNN as feature extractor to encode the extracted data into feature maps. Then, the head structure uses the feature maps provided by the backbone to predict the object locations and their class. Depending on the head structure, object detection networks can be classified in two main categories: single-stage detection methods based on regression (e.g. YOLO and SSD), and two-stage detection methods based on candidate regions (e.g. Fast and Faster R-CNN) [22]. Generally, due to their ability to propose more accurate fruit locations, two-stage fruit detection methods have shown higher accuracy levels than single-stage methods, which indeed struggle especially in presence for small sized fruits. On the other hand, two-stage fruit detection methods are slower and computationally more intensive, compared to the simpler one-stage methods [64].

Two-stage networks were the first type of CNN developed and applied in the field of fruit detection. As their name suggests, their structure is composed of two modules. In the first stage, proposals about the regions that could contain fruits are made on the basis of the feature maps produced by the convolutional layers. Each proposal consists of a fixed-size bounding box and a probability score of containing a target fruit. The N regions with the highest score are then chosen as final region of interest proposals. In the next step, a separate CNN classifies the identified regions as either fruit or background and performs bounding box regression [22][64]. Among the most commonly used two-stage CNN for fruit detection, Fast and Faster R CNN can be mentioned.

Fast R-CNN have been developed with the goal of improving the detection speed of precursor Region-based CNN. They performs feature extraction on the whole image and region proposals on the final feature map. The novelty is the implementation of a Region of Interest (RoI) pooling layer which takes feature maps and regions of interest as inputs and provides the corresponding features for each region, allowing to extract them from all regions of interest in one pass only (Figure 4.1). While this improves the detection speed, the necessity of extracting regions of fruit and providing them to the detection model is limiting. In addition, since relying on Selective Search for generating region proposal hampered further speed improvements, a new architecture was proposed, namely Faster R-CNN.

Faster R-CNN replace the heuristic selective search method of Fast R-CNN with a Region Proposal Network (RPN), that consists of fully convolutional layers aimed at proposing candidate regions from the feature maps generated by the backbone network, by filtering them on the basis of a score that indicates their probability of containing the object (Figure 4.1). The identified regions are then passed to the following stage of detection [63]. By integrating region detection into the main neural network structure, Faster R-CNN achieves near real-time detection speed with high accuracy and generalization capability[64].

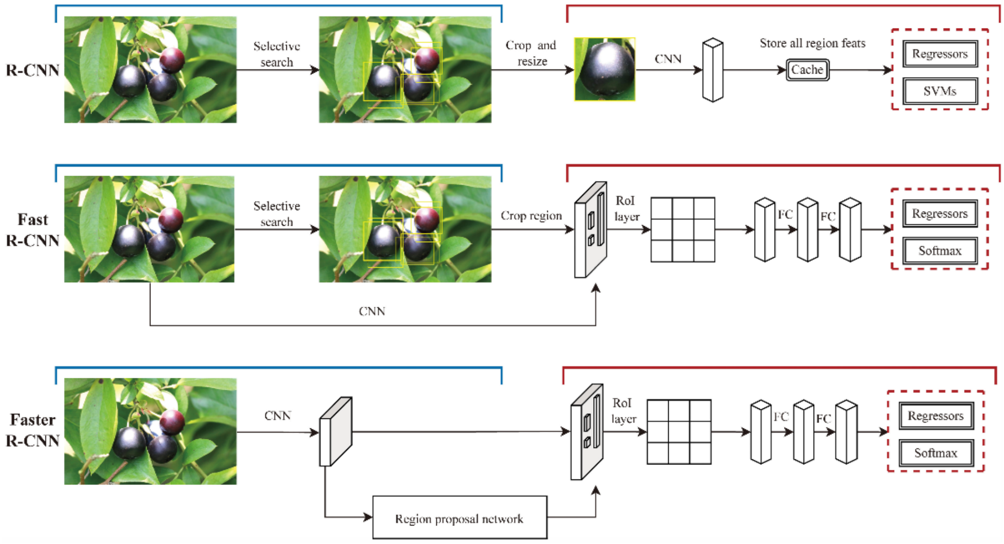


Figure 4.1: Typical R-CNN, Fast R-CNN, and Faster R-CNN frameworks for fruit detection. It can be seen the innovative introduction of the “RoI pooling layer” within the Fast R-CNN framework for improving speed detection with respect to the previous R-CNN. To provide for further improvements, in the Faster R-CNN the “RPN” is introduced, allowing the integration of region detection into the main neural network structure and thereby providing higher detection speed, accuracy and generalization capabilities [64].

To enable a further improvement in speed, one stage networks predict object class and bounding boxes simultaneously, without therefore needing a region proposal step. They define fruit detection tasks as regression problems of class confidence and bounding box locations. Single shot methods divide input fruit images into a grid of cells, extract fruit feature information through the convolutional layer, and predict object class probabilities and bounding box coordinates for each cell [64]. Most commonly used one-stage CNN for fruit detection are Single Shot Single Shot Multibox Detector and (SSD [65]) and You Only Look Once (YOLO [66]).

SSD is based on a series of convolution and pooling layers to generate multiple feature maps of different scales, for which then class probabilities and bounding box offsets are predicted (Figure 4.2). Next, top-ranked prediction boxes are outputted as the final result of fruit detection. This multiscale detection approach makes such network very effective in detecting object that might present different sizes [63]. In addition, fruit detection methods based on SSD provide high accuracy and speed. However, they present some disadvantages given that detection accuracy lowers in presence of relatively small target fruits and cases of repeated detection might occur [64].

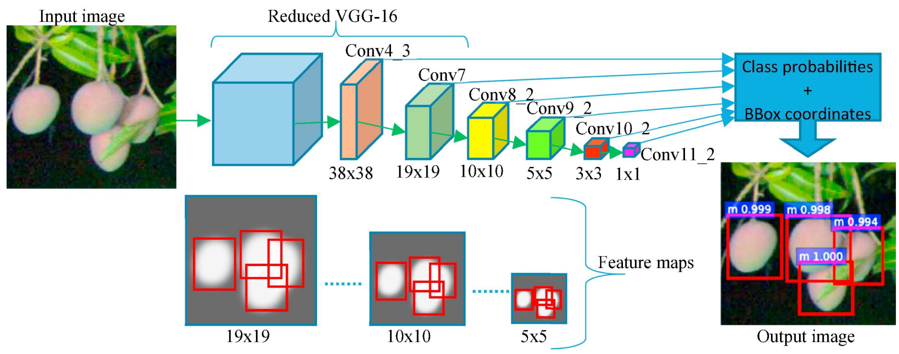


Figure 4.2: SSD object detection framework. 'Conv' refers to 'convolution layer'. Output image displays bounding box on identified ROI, classification result and probability score.[63]

YOLO is a single-shot detector where the object detection task is formulated as a single regression problem by placing the bounding box coordinates to image pixels and then assigning class probabilities. YOLO only looks at the image once to predict which objects are present and where, and does so by dividing the image into a grid where each grid predicts bounding boxes [67] (Figure 4.3). By learning to predict boxes directly from the image data, it produced many localization errors (for small objects and objects in groups) and thus further improvements in accuracy and speed have been provided over the years, leading to the development of many versions of the YOLO network each presenting different innovations [63]. YOLO is considered one of the most commonly used and advanced fruit detection algorithms due to their speed in detecting target fruits in an accurate way. Nonetheless, they do not use prior information when predicting fruit positions, which might result in a loss of fruit location accuracy, and have resulted less effective at detecting small fruits that appear in groups [64].

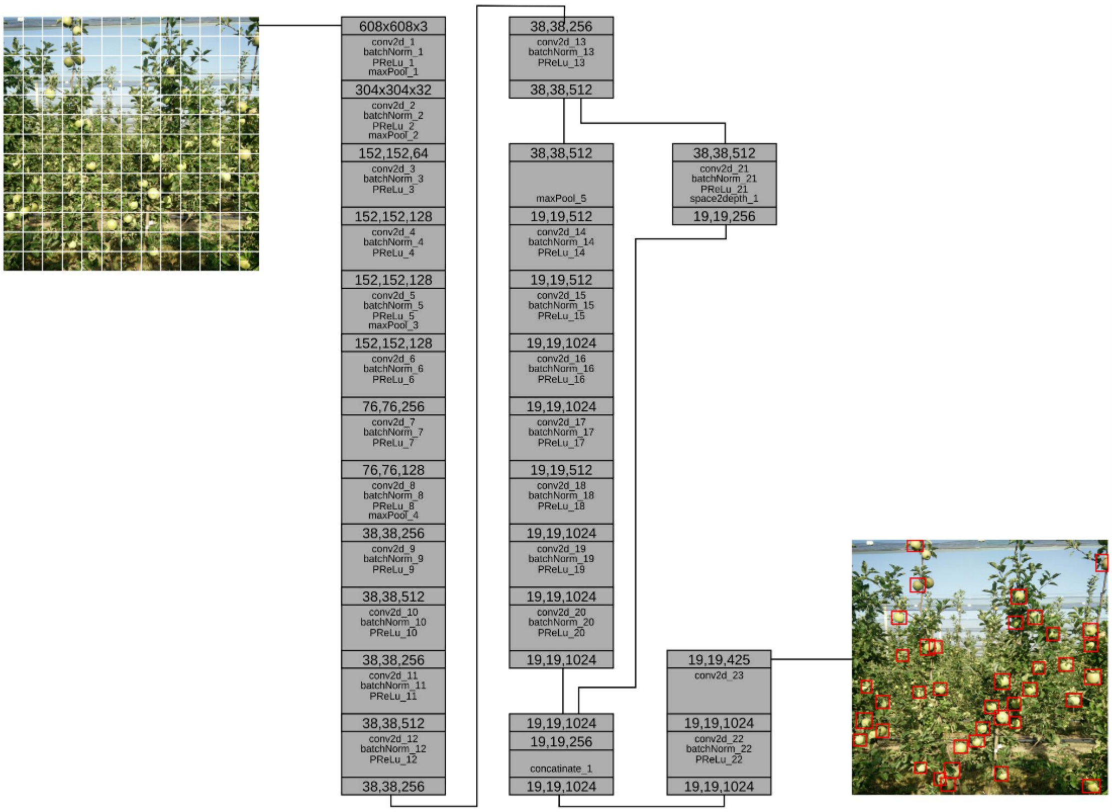


Figure 4.3: YOLOv2 framework with 24 Layers. Compared to other state-of-the art methods that treat detection, classification and region extraction as different problems, YOLO does all in one pass, hence the name [26].

### 4.2.2 Pros and Cons

The recent technological advancements have allowed the refinement of deep learning methods for object and fruit detection, providing new methods able to achieve better performances and high accuracy results even in complex unstructured environments like agricultural one are. The reason for this are to be found in the automatic feature learning process that Convolutional Neural Networks implement. In the field of image detection, extracting image features is the most critical part of the pattern recognition system, since the quality of such process directly affects the final recognition rate of the system. In CNN, feature extraction is based on a hierarchical representation of the data by means of various convolutions that enable the model to extract and learn in depth a vast amount of discriminative features of the fruit objects (in early layers simple local patterns like edges are learnt, while deeper layers capture more semantic representations of the object, like shape), thus ultimately better capturing the full complexity of the real on-field task under study. This end-to-end approach, where features are learnt directly from the raw data, results advantageous also because eliminating the need for manually engineered features implies that all the resources in terms of time, experts, costs needed to carry out such task are not anymore necessary. In addition, CNN have proved able not only to provide more accurate results in intricate environments, but are able to do so in a faster way: in some cases DL models have been specifically developed with the goal of obtaining real-time performances, like the YOLO architectures, making them suitable for applications that require rapid detection [64]. Another advantage DL models provide is that they can be developed through transfer learning approaches that implies the adoption of models that have been already trained on huge and diverse datasets (typically millions of data). By adopting these pretrained models, the knowledge thus acquired (usually weights from the earlier layers that contain information about the basic features that can be found in common objects, such as colour, shape, edges, lines) can be transferred to the model of the case, which will just require some adjustment and the fine tuning of some parameters to better optimize it and adapt its architecture to the domain considered [67]. This is a valid solution in those cases where an insufficient amount of annotated data is available or collectable, since CNN models demand thousands of them to accomplish good results and generally their superiority with respect to ML models manifests itself when vast quantities of training data are available. Nonetheless, such necessity and reliance on significant data volume can be regarded as one of the main constraints that the adoption of DL model entails: image acquisition and annotation are extremely tedious and cost, time, labor intensive processes and even more so if the necessary amount of data to collect is huge and if such data have to be heterogeneous too, so that they can exhaustively represent the variability and complexity that open field orchards present and ultimately enable the implemented models to appropriately perform. In addition, the powerfulness of deep learning architectures comes at the expense of an increased intricacy and computational cost. Their intrinsic complexity on the one hand imposes a necessary involvement of experts in deep learning and computer vision for the model's implementation and fine tuning, on the other hand has consequences over their interpretability: DL models are often considered as black boxes since their decision-making process is not always easily interpretable which can be a problem in those cases where explainability is necessary [68]. Additionally DL models, and CNN in particular, require more computational power since rely on a series of complex operations (e.g convolution, backpropagation, optimization) to allow the learning of intricate patterns through a deep hierarchical representations of huge amount of high dimensional data. To provide for this, significant computational resources and specific hardware might be needed, usually in the form of Graphic Processing Units (GPUs). In light of these considerations, the adoption of DL models demands specific expertise, hardware and software resources for a successful deployment, thus encompassing a series of expenses that might be unattainable in some cases.

## 4.3 Machine Learning vs Deep Learning

In consideration of all the aspects analyzed in the above sections, it appears evident that the choice between relying on traditional machine learning models or deep learning models when considering the development of an AI enabled fruit detection system is not straightforward: specific and appropriate considerations of numerous technical and economical aspects are needed and the final decision will involve a trade-off of various factors. Traditional machine learning models have proven effective in correctly performing within the considered agricultural environment since the extraction of relevant patterns from data relies on manual feature engineering and is thus more controllable and tailorable to the specific characteristics of the fruit images. In addition, the detection task is based upon well-established algorithms that do not require huge amount of data to efficiently train the model. As a consequence of these characteristics, ML models result advantageous when interpretability and transparency are essential, as well as in presence of limited computational resources [22]. On the other hand, deep learning models excel in automatically capturing intricate patterns and hierarchical representations from large, complex datasets, making them suitable to deal with the intrinsic variability and challenges that fruit detection applications in open field orchards are faced with, providing at the same time more accurate results. However, these improved performances require more computational resources and rely on huge amounts of annotated data for efficiently training the DL model. In light of this, the decision between the two approaches revolves around the specific requirements imposed by the considered fruit detection task, the desired level of interpretability, the practicable investments, the computational resources available or affordable and the obtainable quantity of data, since the superiority of DL approaches with respect to ML approaches manifest itself when huge quantity of training data are available(in [69], through a comparative analysis of experimental data, the authors have demonstrated that traditional machine learning provide better recognition accuracy when small-scale datasets are considered, while deep learning models show higher accuracy results on large sample datasets). In addition to the amount of available data, additional evaluations can be made considering that methods based on handcrafted features could continue to be advantageous in some cases (e.g. if the fruit presents a high contrast with respect to the background and thus can be easily distinguished), especially because of their lower use of computational and memory resources and faster training time. Nonetheless, in the existing literature there is a tendency in suggesting the adoption of deep learning approaches for fruit detection systems given the higher accuracy results they can achieve and their better generalization capabilities in detecting fruits within variable conditions (lighting changes, different versions of the same fruit, noise, background color [70]) with respect to ML models that instead struggle to provide good performances in such cases [22].

| Element | Machine Learning | Deep Learning |
|---|---|---|
| Dataset size | Medium (hundreds of data) | Large (thousands of data) |
| Computational power | Low | High |
| Required Hardware | CPU | GPU |
| Training time | Short | Long |
| Feature extraction | Manual | Automatic |
| Domain expertise | High | Low |
| Interpretability | High | Low |
| Deployment flexibility | High | Low |

Table 4.1: Traditional machine learning models vs deep learning models

# Chapter 5

# Hardware for an AI fruit detection system

When considering the development of a machine vision system for fruit detection, other important aspects to take into account are those relating to the physical platform and hardware infrastructure that are needed to deploy the considered model and thus innovating common agricultural tasks by bringing intelligent capabilities: for example, fruit detection models have been mainly considered for achieving improvements and automatization in activities such as fruit quality analysis, fruit yield estimation and fruit harvesting. To achieve so, farmers can benefit from a diverse range of solutions offered by computer vision technologies. For example, given the widespread presence of smartphone, an AI-enabled application can be developed for supporting decision-making processes: after capturing the image of an orchard, the application can be designed to detect fruits and eventually provide some analytics related to maturity level, size, presence of diseases. Or again, other computer vision based solutions have been applied to drones and robots for the automation of agricultural processes: numerous research works have focused on both the development of unmanned ground vehicles equipped with detection capabilities that allow to locate a fruit and instruct a mechanic arm to carefully pick it, providing thereby automatic harvesting, as well as on the implementation of unmanned aerial vehicles able to capture images of a fruit's tree/orchard for then providing further analysis or perform agrochemicals spraying [71]. The development of these approaches to fruit detection by using modern intelligent equipment is favored by the combination of image acquisition tools together with computer vision detection algorithms, in order to design a machine vision system that result effective in agronomic production management. An appropriate choice of such intelligent equipment is an additional guarantee to obtaining successful fruit detection performances in outdoor environments.

On the basis of the specific application considered, three main types of platforms can be identified, i.e., ground manual platforms, ground vehicle platforms and unmanned aerial vehicle platforms [27]. Whichever the type, a machine vision system generally requires a vision sensor, a processing and memory units and detection algorithms, in order to be able to analyze and comprehend complex visual data and thus enabling automation capabilities [60]. Since considerations about data and detection algorithms have been provided in the previous chapters, hereinafter the hardware aspects, namely types of platform, visual sensors and processing unit will be analyzed.

## 5.1 Platforms

A first type of employed platform for fruit detection systems can be identified in ground manual platforms which are mainly based on smartphones, given the great potential in agricultural applications they can have: the main reason for this is because they are pervasive in today's lives and can provide operability and economic advantages since are simple and cheap tools, making them appealing to agricultural producers [72]. Their suitability in performing fruit detection tasks is attributable to the recent technological advancements that enabled the development of smartphones with high resolution cameras and powerful computing capacity. In addition, due to their increased utilization popularity, improvements have been made also in the development of more applicable algorithms: lightweight DL models like YOLO and SSD have been indeed specifically designed to run on constrained devices, still providing high accuracy results [73]. An example of real-time fruit detection in orchard through smartphones is provided in [73], where the authors employed single shot multibox detector (SSD) to develop an Android APP named KiwiDetector for in field kiwifruit detection. In Figure 5.1, the interface of the app is shown: the initial interface includes an image display unit, a camera button, and an album button by which users can take or choose a kiwifruit image. The kiwifruit detection result will then be shown in the image display unit immediately.

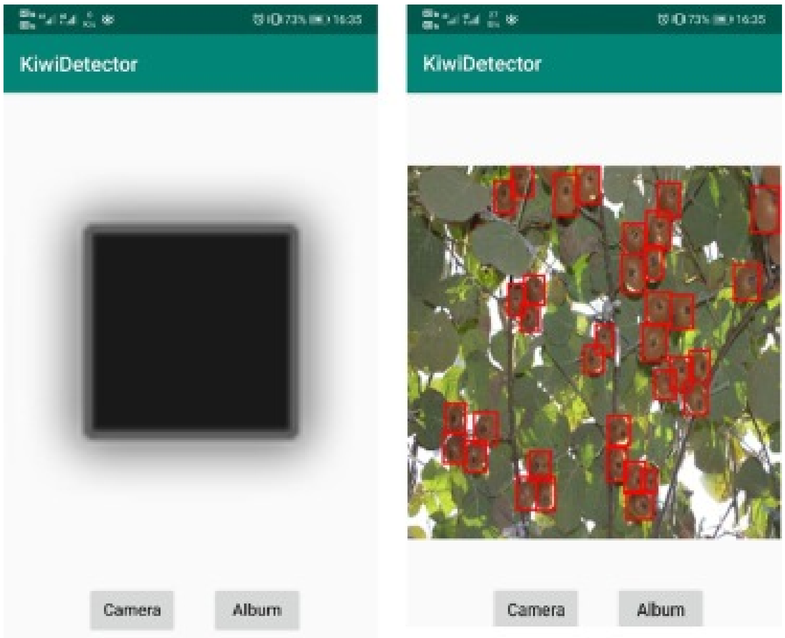


Figure 5.1: Example of ground manual platform to enable fruit detection: initial interface of the KiwiDetector app developed for Android™ smartphones (left) and its detection result display interface (right) [73].

While such platforms are more suitable for small-scale orchards, for large-scale orchards moving vehicles equipped with the necessary sensors, cameras, computers are preferable (among these, tractors are one of the most exploited machineries), for which many example applications can be found in the literature [27][74][75][44]. Nonetheless, given the shortage of agricultural workforce, employing this manually-driven vehicles is causing increasing concerns. In consideration of this, unmanned ground vehicles (UGV) have arisen as innovative solutions for performing fruit detection tasks [76][77]. These platforms implement sensors, cameras, computing capabilities and GPS modules to be able to autonomously travel around orchards to capture and analyze fruit images for rendering useful insights to the agronomist, performing accurate spraying of pesticides/fertilizers when, where and in the quantities needed or harvesting the detected products (Figure 5.2), if equipped with such additional functionalities. They allow therefore to save labor costs, prevent workers from performing laborious tasks while ensuring a longer duration of work and provide farmers with up-to-date accurate information to assist agricultural management decisions [78]. In addition, some remarks have been made on the environmental and economic benefits that automated agricultural vehicles enable when compared to conventional methods to complete various tasks, since they do not aimlessly travel through fields but rather strategically navigate to specific areas of interest by selecting the most efficient paths, thereby optimizing resource utilization [78]. Nonetheless, autonomous navigation can result challenging due to the intrinsic specificity of orchards, which can indeed be characterized by dense structures, intricate and rough terrains as well as present various obstacles, the handling of which is not trivial [78].


Figure 5.2: Typical ground vehicles examples employed for automatic harvesting activities in different orchards (plum, apple, sweet pepper, strawberry, lychee, tomato, kiwifruit) [64].

An additional type of platform that has gained popularity recently are unmanned aerial vehicles (UAV) which rely on sensor cameras, computer vision and flight control techniques to monitor large planting areas [71]. Among the most common applications, there are those of yield estimation especially in orchards that do not present occlusion problems [79][80]. This is because although UAVs platform has high efficiency in a wide range of applications, for many types of fruits there could be occlusion conditions of their trees that undermine the efficiency of top view images [27], or in case the visibility of the fruit relies more on side views, in which case a solution would be to fly the drones between rows [81]. In addition, the UAV can shoot with its unique looking-down angle and can freely adjust the imaging angle and imaging distance in the air based on the actual operation conditions, thus allowing to obtain images with more comprehensive information (Figure 5.3).

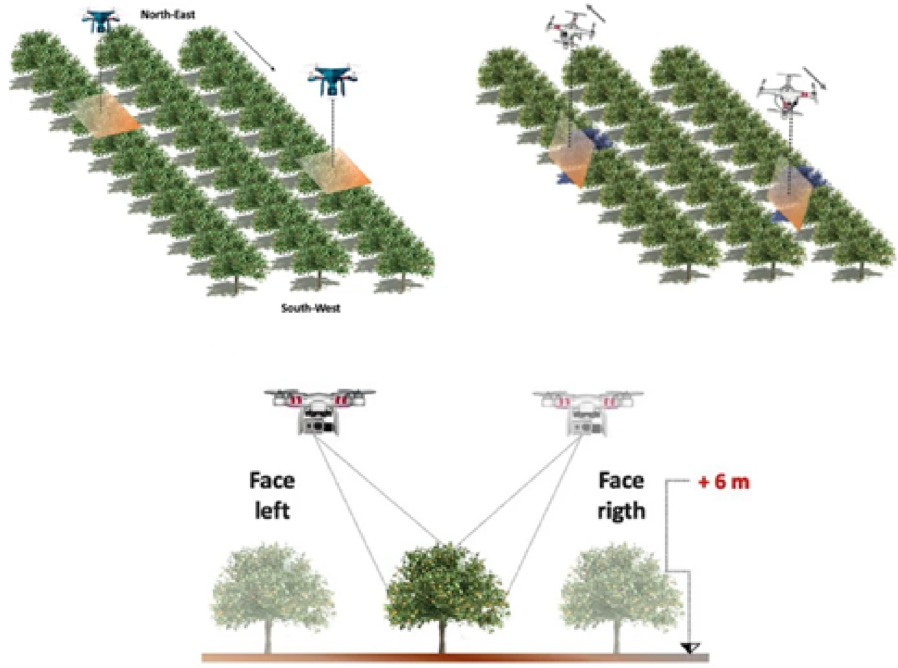


Figure 5.3: Unmanned aerial vehicles image acquisition capabilities examples, providing shooting in a looking-down angle or considering a tree's two side view [20].

Another application example is the exploitation of UAVs to perform variable rate tasks, thus allowing a precise and selective spraying of agrochemicals such as pesticides/herbicides/fertilizers [71]. The main advantages of these vehicles are their flexibility and ease of handling, their rapid implementation, their reusability and their cost-effectiveness, all of which makes them a device with significant potential for proximity-based fruit detection and orchard analysis and for providing also unique perspectives and information that would be otherwise unattainable or economically prohibitive using traditional techniques reliant on human efforts [50]. Nonetheless they present some payload and power constraints, so that some relevant technical challenges arise in consideration of mounting spraying equipment or artificial lighting for capturing quality images, as well as the limited memory and computational resources that can be implemented [71].

Choosing the optimal hardware platform for implementing an AI system for fruit detection, whether on UAVs, UGVs, or smartphones, is a complex task, involving a careful investigation of various factors: the specific application requirements, the system end goal, the targeted environment, the data collection requirements, the power and processing capabilities, the compatibility with the AI model, the budget constraints are among some of the key considerations that have to be pondered, so that a well-informed and aligned decision of the most appropriate platform can be taken.

## 5.2 Vision sensors

The vision sensor represents an important element in the considered machine vision system, since the quality of the images it is able to capture directly affects the overall system performance. However, the selection of the appropriate camera equipment is a non trivial task since many elements have to be taken into consideration for correctly choosing the sensor that best suits the needs of the desired application: task requirements, camera's sensor resolution, frame rate, image transfer rate (connectivity), and price are among the most important factors that need to be examined [78]. Generally, different type of cameras can be implemented as the vision sensors of the machine vision systems, mostly RGB, multi-spectral or stereo-vision cameras. It is important to note that due to the different types of data these sensors render, each will require the development of a different, data-specific algorithm.

RGB cameras have been the most preferred choice for fruit detection and extensively used in many agricultural vision applications given their affordability and progresses in sensing capabilities, as well as being the default build-in cameras of smartphones [60]. The camera pixel resolution, the camera filed of view and so lens focal length, camera to canopy distance and canopy height are all relevant decision factors, since the quality of the captured images depends on the their capability to deal with lighting conditions, size of object and distance from objects [78] and thus the system ability to tolerate the challenges that open fields pose. For example, as reported in [82], when using a 5 MP camera with 2.2 by 2.2 um pixels at a canopy distance of 2 m, equipped with a lens with 6 mm focal length to achieve a field of view for imaging a 4 m high canopy, a 100 mm diameter fruit will be represented by an image 12 pixels in width. In addition to this, while a camera capturing a snapshot of the crop from the side is appropriate for most fruits type, for some crops different orientations might be needed (e.g for kiwifruits cultivations, given their canopy structure, the camera should instead look upwards to the hanging fruits [82]) so that also the camera orientation is a factor to take into consideration.

Multispectral cameras, which are able to capture images in multiple bands of the electromagnetic spectrum, have also been employed for the task of fruit detection. The main reason for this is that they can exhibit better results than conventional color images captured by RGB cameras, since are able to provide additional information on the reflectance properties of fruits object, thus allowing to distinguish them even if present similar colors. Nonetheless, they are more expensive and usually not affordable for practical use [78].

While RGB cameras can only provide two-dimensional information about the scene, which in some application cases might be insufficient, stereovision cameras allow to monitor also the dimension of orchards by using two or more lenses to capture images from slightly different perspectives to provide depth perception. Specifically, orchard depth and height information, leaf shape and area are rendered to allow three-dimensional information on the canopy structure, which is especially useful for some monitoring activities [60].

## 5.3 Processor

Once an image as been captured through the provided sensor, it has to be fed to the implemented DL or ML model so that it can be analyzed and the specific results the system has been designed for can be delivered.

Providing UAVs, UGVs, or smartphones with sufficient computational power depends on various factors, and there are limits to the computational resources that can be embedded in these devices. For example, unmanned vehicles often have size and weight limitations, and specifically aerial vehicles that require lightweight and agile designs. Or again, while these platforms operate on limited battery power, powerful processors required by some models tend to consume more energy and generate heat, thus impacting the device's operational duration and requiring a way to dissipate heat, which is challenging to implement in compact devices. In addition, more powerful processors are undoubtedly more expensive, so that the overall cost of the system is increased. On top of this, some applications might require real-time detection tasks, which strictly relies on powerful processors mounted on the device. It is therefore clear that meeting all these requirements without compromising aspects related to weight, power, costs is a complex task.

For implementing fruit detection computation capabilities over UGV, UAV, smartphones two main ways can be considered: on-board processing and off-board processing, meaning transmitting data to an external computer for analysis. The choice between these options depends on the specific requirements of the fruit detection task (e.g need of real-time performances), the available computational resources of the platform, the desired level of autonomy.

On-board processing implies that the platform considered, either smartphone, UAV, UGV, is itself equipped with a sufficiently powerful processor capable of running the ML or DL model to perform detection and analysis tasks. One of the main advantages of this approach is that it enables (near)real-time detection, especially because there is no need of constant communication and transmission of image data to an external source, so that the latencies that usually occur in such process are minimized and the autonomy of the platform is increased. This is mostly useful for those applications that require tempestive decisions, like UGV for instances, if for example they have been devised to autonomous spraying of pesticides or autonomous picking of fruits. Nonetheless, there are limits on the computational power that can be implemented over such platforms given their energy supply and physical constraints. This is especially problematic in case the model implemented relies on deep learning methods, since they are data-intensive complex algorithms that require huge processing power and thus specialized hardware to run, usually a GPU, which is power consuming and quite expensive [61]. A way to accommodate this could be to optimize the considered DL model, by taking into account the size, speed and memory constraints the platform pose (e.g. YOLO algorithms have been specifically developed for providing more lightweight computation processes). Another mode of operation would be to consider implementing off-board processing: the vehicle/smartphone simply captures the image of fruits and this data is then transmitted to an external computer or cloud server where the computation and analysis required will be performed. The advantage in this case is the possibility to rely on stronger external servers to perform power-consuming tasks, which is especially useful for computationally intensive operations like running DL models could be. This approach enables also easier maintenance, auto-scaling, load-balancing and centralized analysis and storage of data, which facilitate collaboration and management [61]. Nonetheless, this solution makes the system dependent on stable and reliable connectivity for transferring the data, so if it is lacking latencies in data transmission, processing and results rendering are created and thus real-time performances limited[61]. In addition, privacy issues and the higher costs for processing huge amounts of data through such paradigm are additional concerns.

The choice between on-board and off-board processing involves therefore a trade-off between real-time responsiveness and computational capabilities so that in scenarios where immediate decision-making is critical, on-board processing is favored while for applications where latency is less critical, off-board processing can leverage on more powerful computing resources. Hybrid approaches that balance the advantages of on-board and off-board processing can also be considered: an initial image processing is performed on the considered platform for rapid decision-making and a more in-depth analysis is conducted off-board for detailed insights. Moreover, technological advancements and the availability of edge computing solutions can encourage improvement in the computational capabilities of the considered devices: lighter models such as YOLO or SSD algorithms can efficiently run on smartphone devices, while embedded systems like NVIDIA Jetson [83] or specialized processors designed for edge AI [84] are promising in providing a balance between computational power and energy, thus resulting suitable for on-board processing in UAVs and UGVs.

# Chapter 6

# Challenges and tentative solutions

The implementation of an AI fruit detection system for supporting agricultural tasks such as fruit quality analysis, yield prediction, automatic harvesting can be hindered by many challenges that derive from inherent technicalities and more general socio-economic factors that characterize the adoption of new technologies within the agricultural context. Addressing these challenges requires a collaborative effort involving researchers, policymakers, technology developers, and farmers to ensure the successful integration of innovative systems into agricultural practices with the aim of increasing efficiency, productivity and profitability. An analysis of these challenges will be provided in the following paragraphs.

## 6.1 Technical challenges

The technical challenges that the development of an AI based fruit detection system impose can be traced back firstly to the necessity of robust algorithms able to deal at best with the complexities that open field agricultural environments impose. In actual orchards, fruits are densely distributed and usually overlapped by each other or by the ubiquitous branches and leaves that can create occlusions, while some type of fruits present clustered structure or are very small in size [49]. In addition to these, also variations in fruit appearance can occur due to either differences in the degree of maturity or to the unstructured lighting conditions of open fields. These indeed change dynamically during the day and weather conditions, thus producing changes in the brightness of the fruit image, oversaturated or shadowed areas [20]. Furthermore, the outdoor orchard environment itself can present an intricate structure and thus interfere with the performance of the machine vision equipment, making it difficult to achieve the desired effect and causing poor recognition outcomes. Concrete examples of these aspects are shown in Figure 6.1 .

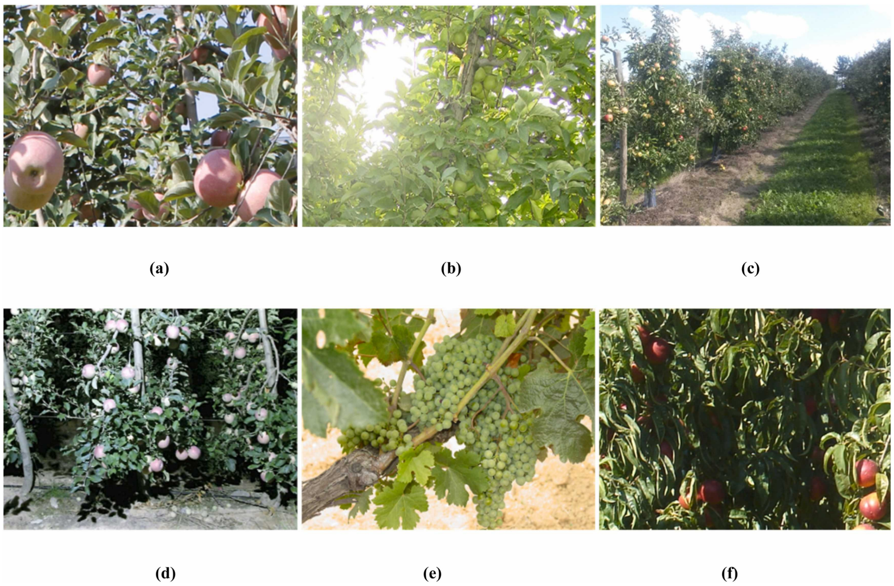


Figure 6.1: Examples of fruit detection challenges that unstructured open field environments present: a) shadowed fruits, b) fruits of similar color than their background and brightness, c) slope in an apple tree orchard, d) night time conditions e) cluster of grapes, f) fruits occlusion and shading [22].

It is clear that all these elements complicate the fruit detection task, compromising the performance and the results' accuracy that the system can obtain and imposing therefore the necessity of implementing sophisticated and robust models able to handle these challenges. To provide for this, some solutions have been proposed, for example fine-tuning feature extraction tasks, though the greatest improvements and capability in dealing with intricate environments' problems have been showed by implementing DL models [28], while considering also effective optimization strategies [20]. Nonetheless, even current improved Deep Learning network architectures still have certain limitations when dealing with highly intricate scenes, and providing such capability implies better feature extraction and information inference processes, which ultimately have direct consequences on the complexity of the model architecture [20][64]. Indeed, more complex systems imply the necessity of sufficient computational power to both train and then deploy the model, consideration especially true for data-intensive DL approaches and even more when real time performances are required, so that a tradeoff between accuracy and computational complexity has to be evaluated to provide acceptable results[28]. Complex architectures ultimately impose higher expenses in terms of needed hardware, software and experts to manage them.

Another type of technical complexities can be imputed to data availability, since data is the most fundamental requirement to build accurate AI systems: to train models that are robust enough to deal with the complexities open field orchards present, representative and substantial dataset with high quality images are a necessity. Nonetheless, on the one side publicly available dataset are scarce, of poor quality or insufficiently characteristic [85], on the other side the collection and the annotation of huge amount and diverse data, such as open field fruit images, is a tedious and labor-, time- and cost-intensive process. To alleviate the difficulties that dataset generation imposes, some solutions have been proposed. One of these is data augmentation, which relies on applying to the dataset's images several transformation techniques to increase its variability and thus size, being the latter an important concern especially when considering the training of DL models, since thousands of data are required to develop accurate systems [28]. To increase the amount of available image data, also the option of generating synthetic data is conceivable [86]. Other ways to deal with dataset generation and specifically data annotation processes that have been gaining recent interest in research, are unsupervised or semi-supervised DL approaches where only a part of data gets annotated, as well as the development of applicable automatic annotation approaches [49]. In light of the analyzed aspects, it transpires how data collection processes are not so straightforward and imply meticulous considerations of numerous aspects, especially given the importance appropriate data have in influencing the final learning capabilities and performances of the developed system.

In addition to what described above, further technical challenges arise when considering the development of an AI fruit detection system for constrained devices such as smartphones. Smartphones are increasingly appearing as highly attractive tools for many agricultural applications, given their widespread adoption and integration into everyday life, which makes them readily accessible to farmers too: they represent indeed a simple, portable and cost-effective tool that could be leveraged for instilling intelligent capabilities within agricultural tasks. Nonetheless, smartphones, and more in general mobile devices, are characterized by limited computational power and battery capacity, thus implying a series of further necessary considerations to balance quality of the prediction, speed and energy efficiency. Running sophisticated algorithms or providing (near)real-time processing for immediate feedback are resource intensive tasks that on the one hand require high computational power, and on the other hand lead to increased battery consumption. In consideration of this, an optimization of the AI model that would be ran on the device is needed to provide efficiency, which can be achieved considering more lightweight models(e.g. MobileNet [87], YOLO [66], SSD [65]), available hardware acceleration options, leveraging cloud computing for heavy processing tasks, minimizing processing time or exploring low-power modes for background processing [64]. Mobile devices present also limited memory resources, constricting the size of models and datasets that can be loaded. Solutions to this would be to rely on compression techniques or strategies like model quantization in order to reduce memory requirements, or else turn to cloud-based solutions for storing and retrieving data when needed [64]. Nonetheless, reliance on network connectivity for either model updates or cloud-based processing and storing creates dependencies and communication latencies, which are additional problems to take into consideration when counting on mobile devices. Therefore, depending on the specific application requirements (for example if real time performances are essential), providing sufficient offline capabilities and intermittent or low-bandwidth connections might be preferable [73].

On more general terms, the deployment itself of an AI based fruit detection system is also one of the most challenging phases due to a lack of deployment expertise, dependencies on third-party libraries, size of the considered model and hardware limitations of the deployment platform, such as those encountered with mobile devices. In addition, the deployment stage is often slowed down by the process of algorithm selection that usually relies on trial-and-error techniques, given that there exist many possible applicable models and figuring out the most suitable one requires either random selection or comparing multiple algorithms to then choose the one providing the best results [85].

Tackling the technical challenges that implementing an AI-based fruit detection system implies therefore careful considerations over many aspects and requires accurate strategies that encompass model optimization, platform-specific adaptations and continuous refinement to ensure a seamless integration and a sustained effectiveness of the system within the agricultural context considered.

## 6.2 Socio-economic challenges

Despite the many benefits that technological advancements applied to agriculture have been proven able to generate, specifically in the form of improved decision-making processes and thus overall increase in efficiency, productivity and profitability, there is still reluctance in their adoption all over the sector. The reasons of this have been analyzed by many researches and can be ascribed to a number of aspects [8].

First off, there is a widespread skepticism over the actual benefits that the adoption of technological innovations to automatize agricultural tasks can factually bring. This can be in consequence of a farmer's education, since less educated persons have resulted being usually less confident and less inclined towards the use of technologies, as well as to not having the sufficient knowledge to understand their benefits or to manage them [8]. In consideration of this, the use of prohibitively scientific terms during the promotion of technological solutions for agriculture have been highlighted as additional hampering factors, together with the lack of training and support provided at purchase or in case of malfunctions: without these, the potential user does not have sufficient skills and capabilities to implement the considered innovations [8]. In addition, economic barriers can be deemed among the factors that discourage the adoption of technologies, since there is a common perception that the results obtainable with their implementation are not worthwhile enough to justify the high investments in hardware, software and training that they impose. The economic gain is an important decision factor since, as for any other business, farmers aim at maximizing production and increasing profits too [9].

Proposed solutions to overcome these barriers pertain firstly to providing farmers with transparent cost-benefit analyses where both costs associated with adopting technologies are clearly outlined and long-term gains and return on investment well emphasized. In addition, an increased commitment of governments can be of valuable importance, by providing adequate policies and financial supports that eventually make the transition more economically viable and thus encourage the propagation of innovations in agriculture [8]. With respect to cultural accessibility, it should be approached by rethinking the way in which smart agriculture technologies are presented and marketed. A first way of operation could be the institution of adequate awareness campaigns to educate farmers about the practical benefits obtainable through implementing new technologies, including increased efficiency, reduced labor, improved yields and profits, eventually showcasing success stories and case studies of farmers who have successfully implemented innovative approaches to agricultural tasks. In addition, extensive training services to equip farmers with the necessary skills and knowledge to use and integrate technologies into their daily practices are of fundamental importance, as well as long-term support mechanisms to assist post-implementation with the aim of addressing any challenges and ensuring a smooth transition to innovation [8].

Incentivizing the adoption of new technological solutions and creating trustworthiness over the obtainable benefits depend also upon providing tailored solutions that fit the specific needs and challenges an agricultural industry might present, by investing in research and development to create technologies that directly address these, by fostering continuous improvement and by promoting user-friendly and intuitive technologies that are easy for farmers to understand and integrate into their existing practices. All these rely on an exhaustive domain understanding, which is non trivial. Indeed, another aspect that enhances socio-economic barriers is the cooperative gap there exist between farmers and AI researchers, whose collaboration and synergies are instead of fundamental importance for a successful implementation of AI solutions in agriculture. This is because, even though farmers are presented with a lot of difficulties during the agricultural production processes, AI researchers do not have enough agricultural knowledge and thus are not aware of them and do not know how they can actually be solved by technologies. Providing a synergic linkage between all the actors involved so that relevant knowledge is shared therebetween and thus the agricultural domain characteristics and implications are better comprehended, will pave the way for the implementation of efficient and beneficial intelligent solutions [7]. A support in this sense can also be found in the formation of interdisciplinary figures that are provided with extensive knowledge related to both agricultural and technological aspects, able therefore to better fill in the existing gaps still hampering a faster application of innovative advancements in agriculture [7].

Addressing the hesitancy of farmers in adopting new technologies involves therefore the implementation of strategies that acknowledge and mitigate their concerns while showcasing the tangible benefits that their integration can bring to numerous agricultural tasks.

# Chapter 7

# Conclusions

As a result of to the ongoing technological advancements, the contemporary landscape is experiencing an unprecedented surge in digitalization, which is profoundly impacting various industries and thereby transforming traditional business models, enhancing efficiency and fostering innovation. While the spread of digital technologies has been more prominent in some sectors, its pace and scale in agriculture have accelerated only in recent years. The reasons for this can be related to an increased awareness over the crucial necessity of ground-breaking interventions and changes in traditional farming practices, in order for the agricultural sector to be able to cope with the difficulties that it is experiencing: increasing food demand due to growing population, labor shortages, increasing costs, climate changes, scarce resources, market pressure and volatility are all factors currently burdening the farming industry. In such context, agriculture is therefore faced with the imperative to keep pace with digitalization trends, given the support and benefits that incorporating advanced technologies can bring to agricultural practices, ultimately increasing operational efficiency, improving decision making processes, improving the quality and quantity of products and yields, increasing profitability, alleviating manual labor and contributing to promote more sustainable and environment friendly practices. Nonetheless, many intricate technological, social, economic challenges hamper the pervasive adoption of technologies in the farming industry. Considering the example of implementing AI systems in the agriculture, many of these clearly emerge.

AI can be considered as one of the most important and impactful technologies that can be leveraged to automate and provide a more efficient management of agricultural tasks, thus creating productivity and profitability. Among the many different applications of AI, those requiring the automatic detection of on-tree fruits are rising increasing interest, given the innovation progresses that have been made in the machine vision field and since it is a task that result beneficial in numerous ways. First off, AI-based fruit detection technologies are capable of locating fruits on a tree and of analyzing them, thus providing useful information like maturity level or size which can help in better planning harvesting and marketing activities. Or again, by analyzing the fruit appearance, it is possible to detect the presence of diseases: this is of valuable importance to provide timely corrective actions and thus avoid detrimental crop losses. In addition, the intelligent detection of on tree fruits, complemented with counting capabilities, allows farmers to estimate yields beforehand, enabling constructive decision-making processes for harvesting and post harvesting activities, specifically with respect to orchard management, labor allocation, storage and transport resources that can be therefore organized in the most appropriate and efficient way. In many cases, fruit detection capabilities have been also infused within unmanned ground vehicles to enable for the automatic picking and collection of mature fruits, or over unmanned aerial vehicles to performs agrochemical spraying or crop monitoring in an autonomous way. The ultimate effects of implementing innovative fruit detection technologies in agriculture are therefore clear: they allow to collect and analyze vast amount of valuable data, thus providing insights that are useful for improving activities such as the management of orchards, the estimation of their yield, their harvesting, the detection of diseases, as well as allow to achieve increased overall productivity and profitability given that many tedious and time-, labor-, cost-intensive agricultural activities can be automatized. However, the implementation of an AI based fruit detection system, and more in general of whichever AI system, is a non-trivial process since it requires accurate and pondered considerations over aspects relating to data, feature extraction, model, hardware.

Data are the most important element in a Machine Learning pipeline: as suggested by Andrew Ng [29], one of the most globally recognized leader in Artificial Intelligence, around 80% of machine learning is data preparation and most of the accuracy improvements that can be achieved rely on high quality data. Nonetheless, data collection is an extremely laborious process, which could partially be facilitated by exploiting publicly available datasets, but they are quite scarce especially for more niche tasks like fruit detection and the quality of both data and their annotations will have to be checked and might not work for the specific case considered. The creation of a custom dataset, on the other hand, enable a better control over the quality of the images and their labels, that will result tailored to the specific characteristics of the orchard, thus ensuring the dataset exhaustively represent the environment variability and so allowing the development of a more accurate and robust fruit detection model. Though, the process of acquiring a new dataset is burdensome since huge amount of high quality image data are required, especially for Deep Learning models that rely on thousands of them to be properly trained and thus able to generalize over other data. In addition, the collected fruit images will have to be annotated which is an extremely time consuming, costly and labor intensive process, given the elaborated data that open field fruit images represent. Selecting the appropriate model is a crucial step in developing an AI-based fruit detection system suitable for the specific use case considered and requires careful evaluations over aspects such as the complexity of the task, the size of the dataset, the computational resources, the desired level of accuracy and the necessary investments.

Considerations over the model that best suits the application domain needs are also not straightforward and imply a careful trade-off between many aspects. DL models have surged recently as the most common used ones, mainly due to the absence of manual feature engineering (an intricate, laborious process for which expert technical knowledge is needed) and the higher accuracy result they are able to provide with respect to ML models. Nonetheless, the higher performances are strictly correlated with the amount of data available: DL models manifest their superiority in presence of huge quantity of training data (thousands of them), the collection of which, as stated above, is complicated. In addition, while in ML models the decision process in more straightforward, DL models have been defined as black boxes since characterized by deep hierarchies of numerous interconnected convolutional, pooling, activation layers so that the steps that brought to a specific output are harder to grasp. The complexities of data-intensive DL algorithms have also consequences in terms of higher training time and computational power required, which imply the necessity of additional investments in specific hardware solutions to support these, usually in the form of a GPU.

The platform over which implementing an AI-based fruit detection system needs to be properly analyzed too. Specifically, mobile and autonomous devices such as smartphones, drones and robots are increasingly being considered for achieving improvements and automatization in many agricultural activities. As for the hardware part, the main components that allow for a machine vision system are vision sensors and processing and memory units to perform the task it has been devised to. There exist different typologies of all these components and the choice of the most appropriate one relies, among others, on aspects pertaining to the specific task requirements, the platform intrinsic characteristics and the available budget.

In addition to the analyzed technical challenges, the adoption of an AI system and more in general of technologies in agriculture, is hampered by some socio-economic factors that cannot be neglected. In particular, researches have investigated how farmers with lower education levels are less confident and less willing towards the adoption of technologies, given that they lack the skills and knowledge to understand or implement them. They also have difficulties in recognizing the potential advantages that innovative solutions could actually offer to their daily activities, being hesitant towards the high initial investments that technologies require and the long-term benefits they can produce.

In light of all these considerations, it is clear how the digitalization of agriculture, if on the one side is an urgent matter for enabling the sector to persevere in an increasing pressuring situation that many issues are creating, on the other side it is a challenging and laborious process that requires careful contemplations of many intricate aspects. In this sense, effective management practices can play a crucial role in navigating the complexities that the development of intelligent systems implies and in fostering a constructive collaboration between all the involved stakeholders to provide synergic approaches for exhaustively comprehend the domain characteristics and implications, and thereby develop accurate solutions that could be effectively beneficial in improving agricultural tasks.

# Bibliography


[1] Food and Agriculture Organization of the United Nations. *FAO. 2022. World Food and Agriculture – Statistical Yearbook 2022. Rome.*

[2] Mane E. Gurbuzer L.Y. Caivano G. Piedrahita N. Schneider K. Azhar N. Benali M. Chaudhary N. Rivera R. Ambikapathi R. Davis, B. and P. 2023. Winters. *Estimating global and country-level employment in agrifood systems. FAO Statistics Working Paper Series, No. 23-34. Rome, FAO.*

[3] *United Nations Department of Economic and Social Affairs, Population Division (2022). World Population Prospects 2022: Summary of Results. UN DESA/POP/2022/TR/NO. 3.*

[4] C. Le Mouël and A. Forslund. *How can we feed the world in 2050? A review of the responses from global scenario studies*, volume 44. 05 2017.

[5] B et al. Schuh. *Research for AGRI Committee – The EU farming employment: current challenges and future prospects, European Parliament, Policy Department for Structural and Cohesion Policies, Brussels.* 2019.

[6] Federica Consentino, Gabriella Vindigni, Daniela Spina, Clara Monaco, and Iuri Peri. *An Agricultural Career through the Lens of Young People*, volume 15. 2023.

[7] Ye Liu, Xiaoyuan Ma, Lei Shu, Gerhard Petrus Hancke, and Adnan M. Abu-Mahfouz. *From Industry 4.0 to Agriculture 4.0: Current Status, Enabling Technologies, and Research Challenges*, volume 17. 2021.

[8] Federica Caffaro and Eugenio Cavallo. *The Effects of Individual Variables, Farming System Characteristics and Perceived Barriers on Actual Use of*

*Smart Farming Technologies: Evidence from the Piedmont Region, North-western Italy*, volume 9. 2019.

[9] Thomas B. Long, Vincent Blok, and Ingrid Coninx. *Barriers to the adoption and diffusion of technological innovations for climate-smart agriculture in Europe: evidence from the Netherlands, France, Switzerland and Italy*, volume 112. 2016.

[10] Lefteris Benos, Aristotelis C. Tagarakis, Georgios Dolias, Remigio Berruto, Dimitrios Kateris, and Dionysis Bochtis. *Machine Learning in Agriculture: A Comprehensive Updated Review*, volume 21. 2021.

[11] V.G. Dhanya, A. Subeesh, N.L. Kushwaha, Dinesh Kumar Vishwakarma, T. Nagesh Kumar, G. Ritika, and A.N. Singh. *Deep learning based computer vision approaches for smart agricultural applications*, volume 6. 2022.

[12] Abhinav Sharma, Arpit Jain, Prateek Gupta, and Vinay Chowdary. *Machine Learning Applications for Precision Agriculture: A Comprehensive Review*, volume 9. 2021.

[13] Sara Oleiro Araújo, Ricardo Silva Peres, José Barata, Fernando Lidon, and José Cochicho Ramalho. *Characterising the Agriculture 4.0 Landscape—Emerging Trends, Challenges and Opportunities*, volume 11. 2021.

[14] Yanfen Li, Hanxiang Wang, L. Minh Dang, Abolghasem Sadeghi-Niaraki, and Hyeonjoon Moon. *Crop pest recognition in natural scenes using convolutional neural networks*, volume 169. 2020.

[15] Rohit Sharma, Sachin S. Kamble, Angappa Gunasekaran, Vikas Kumar, and Anil Kumar. *A systematic literature review on machine learning applications for sustainable agriculture supply chain performance*, volume 119. 2020.

[16] Hongkun Tian, Tianhai Wang, Yadong Liu, Xi Qiao, and Yanzhou Li. Computer vision technology in agricultural automation —a review. *Information Processing in Agriculture*, 7(1):1–19, 2020.

[17] Alberto Pretto, Stéphanie Aravecchia, Wolfram Burgard, Nived Chebrolu, Christian Dornhege, Tillmann Falck, Freya Fleckenstein, Alessandra Fontenla,

Marco Imperoli, Raghav Khanna, Frank Liebisch, Philipp Lottes, Andres Milioto, Daniele Nardi, Sandro Nardi, Johannes Pfeifer, Marija Popovic, Ciro Potena, Cedric Pradalier, and Juan Nieto. *Building an Aerial-Ground Robotics System for Precision Farming.* 11 2019.

[18] Jasmine Neupane and Wenxuan Guo. *Agronomic Basis and Strategies for Precision Water Management: A Review*, volume 9. 2019.

[19] Md. Bayazid Rahman, Joy Dhon Chakma, Abdul Momin, Shahidul Islam, Md Ashraf Uddin, Md Aminul Islam, and Sunil Aryal. Smart crop cultivation system using automated agriculture monitoring environment in the context of bangladesh agriculture. *Sensors*, 23(20), 2023.

[20] Qiu J. Zhang Y. et al. Tang, Y. *Optimization strategies of fruit detection to overcome the challenge of unstructured background in field orchard environment: a review. Precision Agric 24, 1183–1219.* 2023.

[21] D. Brio, Valentin Tassile, Sergio Bramardi, Dario Fernández, and Pablo Reeb. *Apple (Malus domestica) and pear (Pyrus communis) yield prediction after tree image analysis*, volume 55. 12 2023.

[22] Juan C. Miranda, Jordi Gené-Mola, Manuela Zude-Sasse, Nikos Tsoulias, Alexandre Escolà, Jaume Arnó, Joan R. Rosell-Polo, Ricardo Sanz-Cortiella, José A. Martínez-Casasnovas, and Eduard Gregorio. *Fruit sizing using AI: A review of methods and challenges*, volume 206. 2023.

[23] Walsh-K.B. Wang Z. et al. Koirala, A. *Deep learning for real-time fruit detection and orchard fruit load estimation: benchmarking of 'MangoYOLO'. Precision Agric 20, 1107–1135.* 2019.

[24] Wenli Zhang, Jiaqi Wang, Yuxin Liu, Kaizhen Chen, Huibin Li, Yulin Duan, Wenbin Wu, Yun Shi, and Wei Guo. *Deep-learning-based in-field citrus fruit detection and tracking*, volume 9. 02 2022.

[25] Hanwen Kang and Chao Chen. *Fruit Detection and Segmentation for Apple Harvesting Using Visual Sensor in Orchards*, volume 19. 2019.

[26] Kushtrim Bresilla, Giulio Demetrio Perulli, Alexandra Boini, Brunella Morandi, Luca Corelli Grappadelli, and Luigi Manfrini. *Single-Shot Convolution Neural Networks for Real-Time Fruit Detection Within the Tree*, volume 10. 2019.

[27] Leilei He, Wentai Fang, Guanao Zhao, Zhenchao Wu, Longsheng Fu, Rui Li, Yaqoob Majeed, and Jaspreet Dhupia. *Fruit yield prediction and estimation in orchards: A state-of-the-art comprehensive review for both direct and indirect methods*, volume 195. 2022.

[28] Prabhakar Maheswari, Purushothaman Raja, Orly Enrique Apolo-Apolo, and Manuel Pérez-Ruiz. *Intelligent Fruit Yield Estimation for Orchards Using Deep Learning Based Semantic Segmentation Techniques—A Review*, volume 12. 2021.

[29] A. Ng. *A chat with andrew on mlops: From model-centric to data-centric ai.* 2021.

[30] Bai J. Al-Sabaawi A. et al. Alzubaidi, L. *A survey on deep learning tools dealing with data scarcity: definitions, challenges, solutions, tips, and applications. J Big Data 10, 46.* 2023.

[31] Xu (Annie) Wang, Julie Tang, and Mark Whitty. *Data-centric analysis of on-tree fruit detection: Experiments with deep learning*, volume 194. 2022.

[32] Yuzhen Lu and Sierra Young. *A survey of public datasets for computer vision tasks in precision agriculture*, volume 178. 2020.

[33] María Culman, Stephanie Delalieux, Bart Beusen, and Ben Somers. *Automatic labeling to overcome the limitations of deep learning in applications with insufficient training data: A case study on fruit detection in pear orchards*, volume 213. 2023.

[34] Inkyu Sa, Zongyuan Ge, Feras Dayoub, Ben Upcroft, Tristan Perez, and Chris McCool. *DeepFruits: A Fruit Detection System Using Deep Neural Networks*, volume 16. 2016.

[35] Van Gool L.-Williams C.K.I. et al. Everingham, M. *The PASCAL Visual Object Classes (VOC) Challenge. Int J Comput Vis 88, 303–338*. 2010.

[36] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. *Microsoft COCO: Common Objects in Context*. Springer International Publishing, Cham, 2014.

[37] Deng J. Su-H. et al. Russakovsky, O. *ImageNet Large Scale Visual Recognition Challenge. Int J Comput Vis 115, 211–252*. 2015.

[38] Rom H. Alldrin-N. et al. Kuznetsova, A. *The Open Images Dataset V4. Int J Comput Vis 128, 1956–1981*. 2020.

[39] Horea Mureşan and Mihai Oltean. *Fruit recognition from images using deep learning*, volume 10. 2018.

[40] Suchet Bargoti and James Underwood. *Deep fruit detection in orchards*. 2017.

[41] Hamdi Altaheri, Mansour Alsulaiman, Ghulam Muhammad, Syed Umar Amin, Mohamed Bencherif, and Mohamed Mekhtiche. *Date fruit dataset for intelligent harvesting*, volume 26. Elsevier, 2019.

[42] Jordi Gené-Mola, Verónica Vilaplana, Joan R Rosell-Polo, Josep-Ramon Morros, Javier Ruiz-Hidalgo, and Eduard Gregorio. Kfuji rgb-ds database: Fuji apple multi-modal images for fruit detection with color, depth and range-corrected ir data. *Data in brief*, 25:104289, 2019.

[43] Ramesh Kestur, Avadesh Meduri, and Omkar Narasipura. *MangoNet: A deep semantic segmentation architecture for a method to detect and count mangoes in an open orchard*, volume 77. Elsevier, 2019.

[44] Anand Koirala, KB Walsh, Zhenglin Wang, and C McCarthy. *Deep learning for real-time fruit detection and orchard fruit load estimation: Benchmarking of 'MangoYOLO'*, volume 20. Springer, 2019.

[45] Santosh Bhusal, Manoj Karkee, and Qin Zhang. Apple dataset benchmark from orchard environment in modern fruiting wall. 2019.

[46] Jordi Gené-Mola, Ricardo Sanz-Cortiella, Joan R Rosell-Polo, Josep-Ramon Morros, Javier Ruiz-Hidalgo, Verónica Vilaplana, and Eduard Gregorio. Fruit detection and 3d location using instance segmentation neural networks and structure-from-motion photogrammetry. *Computers and Electronics in Agriculture*, 169:105165, 2020.

[47] Jordi Gené-Mola, Eduard Gregorio, Fernando Auat Cheein, Javier Guevara, Jordi Llorens, Ricardo Sanz-Cortiella, Alexandre Escola, and Joan R Rosell-Polo. Lfuji-air dataset: Annotated 3d lidar point clouds of fuji apple trees for fruit detection scanned under different forced air flow conditions. *Data in brief*, 29:105248, 2020.

[48] Nicolai Häni, Pravakar Roy, and Volkan Isler. Minneapple: a benchmark dataset for apple detection and segmentation. *IEEE Robotics and Automation Letters*, 5(2):852–858, 2020.

[49] Qian Zhang, Yeqi Liu, Chuanyang Gong, Yingyi Chen, and Huihui Yu. *Applications of Deep Learning for Dense Scenes Analysis in Agriculture: A Review*, volume 20. 2020.

[50] O.E. Apolo-Apolo, J. Martínez-Guanter, G. Egea, P. Raja, and M. Pérez-Ruiz. *Deep learning techniques for estimation of the yield and size of citrus fruits using a UAV*, volume 115. 2020.

[51] Michael Buhrmester, Tracy Kwang, and Samuel D Gosling. Amazon's mechanical turk: A new source of inexpensive, yet high-quality, data? *Perspectives on psychological science*, 6(1):3–5, 2011.

[52] P.K. Bhagat and P. Choudhary. *Image annotation: Then and now*, volume 80. 2018.

[53] Qimin Cheng, Qian Zhang, Peng Fu, Conghuan Tu, and Sen Li. *A survey and analysis on automatic image annotation*, volume 79. 2018.

[54] Normaisharah Mamat, Mohd Fauzi Othman, Rawad Abdoulghafor, Samir Brahim Belhaouari, Normahira Mamat, and Shamsul Faisal Mohd Hussein. *Advanced Technology in Agriculture Industry by Implementing Image*

*Annotation Technique and Deep Learning Approach: A Review*, volume 12. 2022.

[55] Normaisharah Mamat, Mohd Fauzi Othman, Rawad Abdulghafor, Ali A. Alwan, and Yonis Gulzar. *Enhancing Image Annotation Technique of Fruit Classification Using a Deep Learning Approach*, volume 15. 2023.

[56] Wenli Zhang, Kaizhen Chen, Chao Zheng, Yuxin Liu, and Wei Guo. Easydam_v2: Efficient data labeling method for multishape, cross-species fruit detection. *Plant Phenomics*, 2022, 2022.

[57] David Mojaravscki and Paulo S. Graziano Magalhães. Comparative evaluation of color correction as image preprocessing for olive identification under natural light using cell phones. *AgriEngineering*, 6(1):155–170, 2024.

[58] G. Sharmila and Kavitha Rajamohan. *A Systematic Literature Review on Image Preprocessing and Feature Extraction Techniques in Precision Agriculture*. Springer Nature Singapore, Singapore, 2022.

[59] Parvinder Kaur, Baljit Singh Khehra, and Er. Bhupinder Singh Mavi. *Data Augmentation for Object Detection: A Review*. 2021.

[60] Guduru Dhanush, Narendra Khatri, Sandeep Kumar, and Praveen Kumar Shukla. *A comprehensive review of machine vision systems and artificial intelligence algorithms for the detection and harvesting of agricultural produce*, volume 21. 2023.

[61] Ulzii-Orshikh Dorj, Malrey Lee, and Sang seok Yun. *An yield estimation in citrus orchards via fruit detection and counting using image processing*, volume 140. 2017.

[62] Feng Xiao, Haibin Wang, Yaoxiang Li, Ying Cao, Xiaomeng Lv, and Guangfei Xu. *Object Detection and Recognition Techniques Based on Digital Image Processing and Traditional Machine Learning for Fruit and Vegetable Harvesting Robots: An Overview and Review*, volume 13. 2023.

[63] Anand Koirala, Kerry B. Walsh, Zhenglin Wang, and Cheryl McCarthy. Deep learning – method overview and review of use for fruit detection and yield estimation. *Computers and Electronics in Agriculture*, 162:219–234, 2019.

[64] Feng Xiao, Haibin Wang, Yueqin Xu, and Ruiqing Zhang. *Fruit Detection and Recognition Based on Deep Learning for Automatic Harvesting: An Overview and Review*, volume 13. 2023.

[65] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg. *SSD: Single Shot MultiBox Detector*. Springer International Publishing, 2016.

[66] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. *You Only Look Once: Unified, Real-Time Object Detection*. 2016.

[67] Sebastián Espinoza, Cristhian Aguilera, Luis Rojas, and Pedro G. Campos. *Analysis of Fruit Images With Deep Learning: A Systematic Literature Review and Future Directions*, volume 12. 2024.

[68] Astone Owino. *"Challenges of Computer Vision Adoption in the Kenyan Agricultural Sector and How to Solve Them: A General Perspective", Advances in Agriculture, vol. 2023, Article ID 1530629, 9 pages*. 2023.

[69] Pin Wang, En Fan, and Peng Wang. *Comparative analysis of image classification algorithms based on traditional machine learning and deep learning*, volume 141. 2021.

[70] Huijun L. Edan Y. Farjon, G. *Deep-learning-based counting methods, datasets, and applications in agriculture: a review. Precision Agric 24, 1683–1711*. 2023.

[71] Agüera P. Agüera J. et al. Martinez-Guanter, J. *Spray and economics assessment of a UAV-based ultra-low-volume application in olive and citrus orchards. Precision Agric 21, 226–243*. 2020.

[72] Xiaoming Wu Meixiang Chen Yan'an Wang Jianping Qian1, Bin Xing. *A smartphone-based apple yield estimation application using imaging, Sci. Agric. v.75, n.4, p.273-280*. 2018.

[73] Zhongxian Zhou, Zhenzhen Song, Longsheng Fu, Fangfang Gao, Rui Li, and Yongjie Cui. *Real-time kiwifruit detection in orchard using deep learning on Android™ smartphones for yield estimation*, volume 179. 2020.

[74] Mohamed Lamine Mekhalfi, Carlo Nicolò, Ivan Ianniello, Federico Calamita, Rino Goller, Maurizio Barazzuol, and Farid Melgani. *Vision System for Automatic On-Tree Kiwifruit Counting and Yield Estimation*, volume 20. 2020.

[75] Jordi Gené-Mola, Eduard Gregorio, Fernando Auat Cheein, Javier Guevara, Jordi Llorens, Ricardo Sanz-Cortiella, Alexandre Escolà, and Joan R. Rosell-Polo. Fruit detection, yield prediction and canopy geometric characterization using lidar with forced air flow. *Computers and Electronics in Agriculture*, 168:105121, 2020.

[76] Salvador Gutiérrez, Alexander Wendel, and James Underwood. *Ground based hyperspectral imaging for extensive mango yield estimation*, volume 157. 2019.

[77] Xu Liu, Steven W Chen, Chenhao Liu, Shreyas S Shivakumar, Jnaneshwar Das, Camillo J Taylor, James Underwood, and Vijay Kumar. *Monocular camera based fruit counting and mapping with semantic data association*, volume 4. IEEE, 2019.

[78] Efthimia Mavridou, Eleni Vrochidou, George A. Papakostas, Theodore Pachidis, and Vassilis G. Kaburlasos. *Machine Vision Systems in Precision Agriculture for Crop Farming*, volume 5. 2019.

[79] Yang Chen, Won Suk Lee, Hao Gan, Natalia Peres, Clyde Fraisse, Yanchao Zhang, and Yong He. *Strawberry Yield Prediction Based on a Deep Neural Network Using High-Resolution Aerial Orthoimages*, volume 11. 2019.

[80] Aharon Kalantar, Yael Edan, Amit Gur, and Iftach Klapp. A deep learning system for single and overall weight estimation of melons using unmanned aerial vehicle images. *Computers and Electronics in Agriculture*, 178:105748, 2020.

[81] Orly Enrique Apolo-Apolo, Manuel Pérez-Ruiz, Jorge Martínez-Guanter, and João Valente. *A Cloud-Based Environment for Generating Yield Estimation Maps From Apple Orchards Using UAV Imagery and a Deep Learning Technique*, volume 11. 2020.

[82] Nicholas Todd Anderson, Kerry Brian Walsh, and Dvoralai Wulfsohn. *Technologies for Forecasting Tree Fruit Load and Harvest Timing—From Ground, Sky and Time*, volume 11. 2021.

[83] Xin Feng, Youni Jiang, Xuejiao Yang, Ming Du, and Xin Li. *Computer vision algorithms and hardware implementations: A survey*, volume 69. 2019.

[84] Wenli Zhang, Yuxin Liu, Kaizhen Chen, Huibin Li, Yulin Duan, Wenbin Wu, Yun Shi, and Wei Guo. Lightweight fruit-detection algorithm for edge computing applications. *Frontiers in Plant Science*, 12, 2021.

[85] Vishal Meshram, Kailas Patil, Vidula Meshram, Dinesh Hanchate, and S.D. Ramkteke. *Machine learning in agriculture domain: A state-of-art survey*, volume 1. 2021.

[86] Wei Zhang, Ting Yao, Shiai Zhu, and Abdulmotaleb El Saddik. *Deep Learning–Based Multimedia Analytics: A Review*, volume 15. Association for Computing Machinery, New York, NY, USA, jan 2019.

[87] Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. *MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications*. 2017.

[88] Junyang Chen, Hui Liu, Yating Zhang, Daike Zhang, Hongkun Ouyang, and Xiaoyan Chen. *A Multiscale Lightweight and Efficient Model Based on YOLOv7: Applied to Citrus Orchard*, volume 11. 2022.