Department of Agricultural Sciences
PHD COURSE IN AGRICULTURAL SCIENSES
Curriculum *"Monitoraggio e controllo degli ecosistemi agrari e forestali in ambiente mediterraneo"*

XXXV cycle

PhD dissertation

Machine learning and Unmanned Aerial Systems for crop monitoring and agrochemicals distribution optimization in orchard and horticultural systems

Dr. Alberto Sassu

Course cordinator	Prof. Severino Zara
Curriculum referent	Prof. Paola Castaldi and Prof. Filippo Gambella
Tutor	Prof Filippo Gambella
Dissertation correlator	Dott. Luca Mercenaro

2021-2022 Academic year

Acknowledgments

This work is the results of three projects goals, all of them based on developing new technologies for future agriculture.

This work is part of the activities of the project “Advanced Technologies for LANds management and Tools for Innovative Development of an EcoSustainable agriculture” (ATLANTIDE). The general objective of ATLANTIDE is the complete integration between theoretical knowledge and technologies for the definition and implementation of agricultural production models that aim at efficiently combine inputs (water, fertilizers, pesticides, energy, time) with outputs (increase efficiency, improvement of quality, reduction of production losses, reduction of resource use, reduction of land use, reduction of the ecological footprint). The team of the ATLANTIDE project is made up of three Partners with strong complementarities: Topcon Agriculture S.p.A-Private Company, Abinsula Srl-Sardinian Small-Medium Enterprise, and the Center for Innovative Agriculture, University of Sassari-Research Organization (Coordinator). This work is part of the activities of the project “ATLANTIDE” and is the topic of Work Package 7. The specific objective of WP7 Crop management of the vineyard with the combined support of precision viticulture and Artificial Intelligence (AI) technologies” is the creation of a Framework enabling the manipulation, aggregation and fusion of data from the agro-food field that allows the development from the point of view of Precision Agriculture in a European context.

This work is part of the activities of the ECSEL JU project “Comp4Drones (C4D)-Framework of Key Enabling Technologies for Safe and Autonomous Drones”. C4D started in October 2019, and it will last three years. It involves 50 partners, and it is coordinated by Indra. The Italian cluster includes 12 partners, and it is led by the University of Sassari, Department of Chemistry and Pharmacy. The main goal of the project is to provide a framework of key enabling technologies for UAS design and operation. These technologies range from application to electronic components, realized as a tightly integrated multi-vendor and compositional UAS-embedded architecture solution and a tool chain complementing the compositional architecture principles. The Italian cluster within C4D involves academic and industrial partners, aimed at developing technologies, as well as providing a use case for the assessment of those technologies. The assessment scenario is provided by the Agricultural Sciences Department of the University of Sassari and the idea is to improve precision agriculture technologies by providing more advanced observation and intervention methodologies exploiting both Unmanned Autonomous Vehicles and Unmanned Ground Vehicles.

This research was funded by PRIN: PROGETTI DI RICERCA DI RILEVANTE INTERESSE NAZIONALE-Bando 2017 “New technical and operative solutions for the use of drones in Agriculture 4.0”. The project aims to discover the potential and new

possibilities related to UAS for spraying distribution and Unmanned Ground Vehicles implementation in agricultural scenarios.

Contents

Chapter 1 – General introduction

1. Precision agriculture to feed the world

The world population will grow to 9.1 billion by the end of 2050, and food demand will increase by 70%. Due to rapid urbanization, the availability of land for agriculture will drastically decrease in the coming years, requiring new strategies to produce more with fewer inputs [1] using solutions that consider resource scarcity and farm profitability [2]. In recent years, global warming has provoked drastic changes in weather conditions. Frequent droughts and heavy rains, combined with poor farm management, are the causes of reduced food production. At the same time, agriculture is directly responsible for the emission of about 14% of global greenhouse gases and is contributing to emissions through land use changes for agricultural expansion, which accounts for a further 17% of total emissions [3]. New solutions and strategies are needed to meet the upcoming agriculture challenges and ensure sustainable food production.

Information and communication technologies (ICT), promoted by policymakers worldwide and integrated into conventional agricultural management, are helping to trigger a fourth agriculture revolution [4] through management information systems, soil sensors, accelerometers, wireless sensor networks, remote sensing technologies, web platform services, and automated guided vehicles [5,6]. ICT approaches are essential in smart agriculture, which focuses on optimizing natural resources, preserving the ecosystem and biodiversity, developing an appropriate service structure, and implementing digital technologies [7,8] for a more accurate and faster decision-making processes.

The presence of multiple sensors, some of them acquiring images, implies a management system able to fast and efficiently transform raw data into useful information for farmers. Data obtainment, storage, and elaboration still represent a weak point in agricultural digitalization. Among all technologies, the combination of Machine Learning (ML) and Unmanned Aerial Systems (UASs) appears strongly correlated in different disciplines and promises remarkable performance gains and complexity reduction. The large volume of data from UAS and robotic platforms can be processed with computer vision and ML techniques to identify and classify crop needs and deficits to execute more controlled and optimized farming practices [9]. The following sections present a short introduction to UASs and ML focusing on agriculture, and reporting potentials, limits, and relevant surveys.

2. Advances of Unmanned Aerial Systems in smart agriculture

UAS are remotely piloted aircraft platforms suitable for remote sensing data missions across multiple scenarios and scientific disciplines [10]. Able to cover large areas, providing high spatial resolution at a high temporal frequency data without impacting and

disrupting the observed ecosystems, UASs have become crucial tools for ecological monitoring, biodiversity conservation, and agriculture [11]. The last decades of technology development represented a driving force for industry growth, helping solve a series of problems related to UAS systems application. Big data, network communication, cloud computing, and Artificial Intelligence (AI) technologies bring new opportunities for UAS implementation [12].

Satellites provide multispectral and hyperspectral images on a large scale and long-term archived. They do not require operators' presence in the studied area, and in the case of open data, there are no costs associated with their acquisition. Problems such as cloud cover and revisit frequency are intrinsic to this form of sensing and will probably continue to exist [13–15]. Low to moderate-resolution satellite imagery has limitations in specific agricultural applications, such as orchard and vineyard scenarios [16]. While satellite remote sensing represents a low-resolution approach, proximal sensing and ground-based data collection offer the highest level of accuracy, but involve more time-consuming, laborious, and demanding operations. This measurement approach requires the physical presence of operators in the surveyed area and the geolocation of each sampling location to analyze the spatial variability of the studied variables. Typically, proximal level reflectance data collection is performed as ground truth data to validate the data from remote sensing sources [17,18].

Aerial remote sensing consists of two categories based on platform flight altitude. The first is the aerial sensing performed by occupied aircraft, which takes place around 500 m above ground level and involves small airplanes carrying detection and positioning systems attached to stabilization components. This technology offers higher resolution and overall data quality than satellite imagery, but the high acquisition costs and associated difficulties in acquisition and processing have limited its application in agriculture [19]. UASs, the second aerial remote sensing sub-category, represent a crucial tool for agricultural surveys and applications over the past decade. Data collection via UAS provides a non-destructive, inexpensive, and high-quality method for rapid field monitoring using low-cost platforms [20]. The ability of UAS to acquire data close to the surface and the integration of correction/calibration systems make the multi-sensor image collection of UAS unaffected by cloud cover. Different sensors cannot perform optimal data collection during a single mission flight, as their required flight parameters may vary, so individual flights for each sensor may be necessary. The ability to develop multiple data sets from a single sensor will be a great advantage, especially considering the time efficiency of UAS flights and the relatively low acquisition costs [21,22].

Although UASs are highly automatable and have achieved a high build quality level, still many safety issues exist. The presence of a trained and licensed pilot operator is mandatory due to national flight regulations [23]. The environmental conditions, the large

volumes of data, national and international legislation, the vulnerability of the ground control system to interference, and the limited flight autonomy and payload capacity represent the most common UASs' limitations.

Several UAS classifications have been suggested based on components, shape, size, and take-off weight which often determines what set of regulations it falls in. The three most common and employed UASs are the fixed wings, multirotor, and Vertical Take-Off and Landing (VTOL) [11,24]. Fixed wings, launched with specific instruments or maneuvers by the operator to take off, are not propelled vertically by any mechanism that allows them to hover. They must maintain a minimum flight speed to preserve flight altitude and to achieve the desired photos' overlaps. Able to cover large areas much more efficiently, fixed wings offer the best flight range. Multi-rotor UAS, characterized by a different number of propellers, can hover, perform agile maneuvers approaching the target area, achieve a lower flight speed, and take off and land vertically. Their ability to fly at lower altitudes and speeds allows them to obtain high-resolution images, but they cover smaller areas than fixed wings and require a longer flight time to cover the same field. VTOLs are hybrid UASs derived from fixed-wing and multi-rotor main features combination. They combine fixed wings' efficiency with the multirotor' vertical take-off and landing ability thanks to a specific system responsible for the transition and repositioning of the rotors, which increases both complexity and production costs. Such systems retain their respective advantages and overcome their limitations.

Several UAS applications can help improve knowledge, sustainability, and operational efficiency in agriculture [25].

Water represents an essential and limited resource requiring adequate improvement for wastage reduction. The intensification of food production is leading the agricultural sector to consume approximately 70% of the freshwater available worldwide [26]. Innovative solutions providing reliable and implementable information on appropriate water management systems are essential in agricultural management. Precision farming digital technologies help optimize agricultural processes and minimize environmental impact [27]. The advent of high-resolution thermal and multispectral UAS approaches has transformed and improved the accuracy of crop water stress estimations and effective irrigation scheduling by providing high-resolution spatial and temporal image in near real-time [28,29]. Remote sensing UAS combined with agrometeorological information is a vital tool for strategic management decisions providing decision-making process information and mitigating the risk of crop failure and low yields. Vegetation indices help quantify the cumulative effect of water deficit during the production season (long-term response), while indexes extracted by thermal sensors, like Crop Water Stress Index (CWSI), give an immediate response to crop water stress [30,31].

Agriculture intensification through fertilizer application is one of the main ways to improve food production. The massive and improper application of fertilizers poses an environmental risk that affects the long-term sustainability of the process and is a cost factor [32]. Precise fertilization techniques can improve food production while protecting the environment from long-term damage. The latest developments in imaging technologies have facilitated the acquisition of spatial and temporal variability as crucial information to support fertilization management and improve the cost-effectiveness and sustainability of agricultural production. Proximal Data can help obtain high-resolution data for field variability management [33]. However, proximal approaches are generally time-consuming and labor-intense to perform. Since that remote sensing imagery for crop N-status quantification and following fertilization support is not fully standardized, calibration with ground data is required for the reliability improvement of fertilizer application maps. Variable-rate application combined with soil information can help optimize fertilizer inputs, maintain appropriate yields, and reduce environmental dispersions. UAS application in management zone delineation is increasing, providing insight into how useful these tools are [34,35].

Agrochemicals pollute the environment, contaminating soil, water, crop vegetation, and non-target animals and plants. A variable proportion of crop protection products is lost through leaching, runoff, spray drift, and volatilization [36,37] In addition, agrochemical application produces dangerous effects on human health. Several studies demonstrated how acute and chronic health problems derive from the exposition to chemicals [38]. UAS can be employed for monitoring operations, checking crop health and canopy architecture, and using this information for creating prescription maps to optimize agrochemical treatments with the ground and aerial vehicles, applying the right amount where it is needed [39–41]. Agrochemical applications are generally performed through land-based or aerial technologies. Large-scale terrestrial applications are carried out by ground vehicles equipped with spraying systems that follow a fixed path in the field. Ground based treatments are time and labor-intensive and expose operators to health risks [42]. Traditional aerial spraying, typically carried out by helicopters and airplanes, is performed flying at high altitudes and speeds, increasing the risk of drift, requiring large areas for taking off and landing operations, and resulting unsuitable for small-scale farms [43–45]. In recent years, multi-rotor UASs for agrochemical applications have gained much attention. Unlike traditional aerial spraying, UAS can follow complex patterns, fly at low altitudes, adapt to hilly terrain, take off and land vertically, perform very low-volume and site-specific agrochemical applications, and avoid operator's health risks. The key factors to consider when working with spraying UASs include droplet size, application strategy, payload capacity, weather conditions, flight range, and environmental effects. Major companies are starting to develop specific nozzle types for UAS spraying, but their

characteristics are not standardized yet. Further research is needed to find the best solution to make treatments as effective as possible [46–48].

Weed management, crucial for agricultural production, which affects yield crop and quality, is generally performed using herbicides or mechanical removal. Neither mechanical methods, accused of being energy consuming, impacting soil structure, and generating erosion, nor herbicides, harmful to the environment and responsible for herbicide-resistant weed selection, can help find a definitive/long-term solution [49–51]. Weed identification and classification are crucial for herbicide control and reduction strategies. Precision agriculture provides the best technologies for field variability assessment for site-specific operations [52]. UASs, able to rapidly cover large areas and equipped with several image sensors, represent effective missing weed detection and classification tools. The most common approaches involve multispectral and hyperspectral cameras to detect and differentiate the spectral signatures of weeds [53,54].

A more recent and revolutionary approach to solving the mentioned problems involves ML image analysis [55,56]. UAS-collected information can be analyzed and post-processed using specific ML and object-based image analysis. Such processes are complex tasks that require specialized knowledge and skills to extract useful information [57]. The derived benefits will drastically boost the agriculture sector's evolution.

3. Machine learning challenges in agriculture

ML, the scientific field that allows machines to learn without being strictly programmed [58], is helping create new perspectives and opportunities to analyze and understand data in agricultural operational environments. The 'big data' produced by digital technologies in agriculture require extensive storage and processing capabilities [59]. ML algorithms help solve complex problems where human perception fails. Performance metrics, improved with experience over time, and statistical and mathematical models are used to calculate the performance of ML models and algorithms. At the end of the learning process, the trained model can be used to classify, predict or group new test data using the experience gained during the training process [60]. The ML process involves data extraction, data preparation, model building, and model deployment processes, predicting the output based on the inputs the algorithm has not been trained on before.

Supervised, unsupervised, and reinforcement learning are the three basic ML paradigms. In supervised learning, which requires the supervision of an operator, the algorithm constructs an input-output relationship based on a labeled data set and characterizes or predicts outputs of unseen inputs. The supervised classification algorithms output is used for predicting a categorical value, and the regression one for

predicting a numerical value. Unsupervised learning algorithms works with unlabeled data and discovers unknown objects by grouping similar objects, with the final goal of extracting hidden knowledge from the training data set. This approach results harder to implement than supervised learning algorithms. Reinforcement learning is another approach that learns from the environment through reward and punishment [61]. The massive increase in the size of big data affects the analysis and throughput efficiency. The core data, small but sufficient to represent the entire content, must be extracted by applying a tensor-based feature reduction model. Dimensionality reduction (DR) analysis provides a compact representation of the dataset and preserves the original data quality. Its application before a classification or regression model helps to avoid dimensionality effects. Common DR algorithms are principal component analysis, partial least squares regression, and linear discriminant analysis.

Between the most common supervised ML learning models, regression provides an output variable prediction based on known input variables. The k-means, the hierarchical technique, and the expectation-maximization technique are clustering techniques of unsupervised learning models used to find natural groupings of data (clusters). Probabilistic graphical models are applied in Bayesian Models (BM), where probability is used to represent all uncertainty within the model related to outputs and inputs within the context of Bayesian inference. The comparison of new examples with instances in the training database represent the base of Instance-Based Models (IBM). The disadvantage of these models regards the growing complexity related to the size increment of data. The Decision trees (DT) is another kind of classification or regression models shaped as a tree-like structure.

Artificial neural networks (ANNs) are inspired by the functionality of the human brain, emulating complex functions such as learning and decision-making. Divided into the traditional and deep ANN categories, they consist of several nodes arranged in several layers. One layer is dedicated to feeding the system with input data, one or more hidden layers to learning, and the last output layer is where the decision/prediction is provided. Deep ANN, better known as Deep Learning (DL) or Deep Neural Networks (DNNs) [62], is a relatively new area of ML research. It consists of computational models composed of multiple processing layers dedicated to complex representations of learned data using different levels of abstraction [63]. A common DL model is the Convolutional Neural Network (CNN), it constitutes a specific class applied to various agricultural and food production challenges, where feature maps are extracted by performing convolutions in the image domain [64]. DL models have improved agriculture and food production in many ways including yield prediction, disease detection, weed identification, crop quality quantification, and species recognition.

One of the most prominent topics in precision agriculture is yield estimation and mapping to match crop supply demand and increase productivity through crop management. DL monitoring systems provide an efficient and non-destructive method to automatically count fruits, distinguish the harvestable from the non-harvestable ones [65,66], and reduce labor requirements [67,68]. Other studies focus on yield prediction to obtain in advance information to detect problems and manage the harvesting phase [69–71].

Disease control in open fields and greenhouses is a crucial problem of modern agriculture. Agrochemicals spraying in cultivation areas, the most widely used practice for disease control, although effective has high financial and environmental costs. Residues can be found in crop products, contaminate ground water, and impact the entire eco-systems. ML, frequently used to detect and report symptoms in fruits and vegetable parts [72–75], is an integrated part of precision agriculture management, where agrochemicals input is targeted in terms of time and place [76].

Weed detection and management is a significant problem and the most relevant threat to crop production because of the high difficulty related to detection and discrimination from crops. ML algorithms combined with sensors enable the development of tools and robot weed removing, minimizing the need for herbicides [77] at a low cost and without environmental problems or side effects [78].

ML detection and classification of quality crop features can increase product prices and reduce wastage. This approach could be applied to fruit, seeds, and vegetable parts [79–81]. The main objective of the last subcategory concerns the automatic identification and classification of plant species to avoid human intervention and reduce detection and classification time [82–84].

ML has proven to be a revolutionary technique in agriculture, but despite the high level of innovation, several problems need to be solved [85]. Problems associated with sensors and ICT technologies implementation on farms and lack of information on their usage. Datasets generated by a few people obtaining images or samples in a short period and from a limited area do not reflect realistic cases. More efficient ML algorithms and scalable computational architectures could lead to faster processing and more reliable information. Furthermore, another open problem is that, since most farmers are not ML experts, there is a need to develop user-friendly systems, to solve the reported problems and to improve the applicability of this technology in daily agricultural management.

4. Objectives of the dissertation

The work aims at discovering the potential and the efficiency of UAS and ML in agriculture scenario, focusing on crop management and agrochemicals distribution optimization in orchard and horticultural cropping systems.

5. Structure of the dissertation

The dissertation includes a general introduction, three experimental chapters and a general conclusion.

Chapter 2 illustrates an operational approach to estimate individual and aggregate vineyards' canopy volume using the manual Tree-Row-Volume (TRV) and the remotely sensed Canopy Height Model (CHM) techniques, processed with MATLAB scripts, and validated through ArcGIS tools.

Chapter 3 presents the development of a grape bunch detector based on a deep convolutional neural network trained to work directly on the field in an uncontrolled environment.

Chapter 4 reports artichoke plant deep learning-based detection and georeferencing as the first step for an on-the-fly UAS spraying system and uses the gathered information to crop development monitoring in a multi-temporal approach.

References

1. How to Feed the World - 2050: High-Level Expert Forum; FAO: Rome, 2009; p. 35.
2. Amiri-Zarandi, M.; Hazrati Fard, M.; Yousefinaghani, S.; Kaviani, M.; Dara, R. A Platform Approach to Smart Farm Information Processing. Agriculture 2022, 12, doi:10.3390/agriculture12060838.
3. Seebauer, M. Whole Farm Quantification of GHG Emissions within Smallholder Farms in Developing Countries. Environmental Research Letters 2014, 9, 035006, doi:10.1088/1748-9326/9/3/035006.
4. Knierim, A.; Kernecker, M.; Erdle, K.; Kraus, T.; Borges, F.; Wurbs, A. Smart Farming Technology Innovations – Insights and Reflections from the German Smart-AKIS Hub. NJAS: Wageningen Journal of Life Sciences 2019, 90–91, 1–10, doi:10.1016/j.njas.2019.100314.
5. Sinha, B.B.; Dhanalakshmi, R. Recent Advancements and Challenges of Internet of Things in Smart Agriculture: A Survey. Future Generation Computer Systems 2022, 126, 169–184, doi:https://doi.org/10.1016/j.future.2021.08.006.
6. Reddy Maddikunta, P.K.; Hakak, S.; Alazab, M.; Bhattacharya, S.; Gadekallu, T.R.; Khan, W.Z.; Pham, Q.-V. Unmanned Aerial Vehicles in Smart Agriculture: Applications, Requirements, and Challenges. IEEE Sensors Journal 2021, 21, 17608–17619, doi:10.1109/JSEN.2021.3049471.

7. Wongchai, A.; Jenjeti, D. rao; Priyadarsini, A.I.; Deb, N.; Bhardwaj, A.; Tomar, P. Farm Monitoring and Disease Prediction by Classification Based on Deep Learning Architectures in Sustainable Agriculture. Ecological Modelling 2022, 474, 110167, doi:https://doi.org/10.1016/j.ecolmodel.2022.110167.

8. Yang, X.; Shu, L.; Chen, J.; Ferrag, M.A.; Wu, J.; Nurellari, E.; Huang, K. A Survey on Smart Agriculture: Development Modes, Technologies, and Security and Privacy Challenges. IEEE/CAA Journal of Automatica Sinica 2021, 8, 273–302, doi:10.1109/JAS.2020.1003536.

9. Bithas, P.S.; Michailidis, E.T.; Nomikos, N.; Vouyioukas, D.; Kanatas, A.G. A Survey on Machine-Learning Techniques for UAV-Based Communications. Sensors 2019, 19, doi:10.3390/s19235170.

10. Tsouros, D.C.; Bibi, S.; Sarigiannidis, P.G. A Review on UAV-Based Applications for Precision Agriculture. Information 2019, 10, doi:10.3390/info10110349.

11. Mogili, U.R.; Deepak, B.B.V.L. Review on Application of Drone Systems in Precision Agriculture. Procedia Computer Science 2018, 133, 502–509, doi:https://doi.org/10.1016/j.procs.2018.07.063.

12. Fan, B.; Li, Y.; Zhang, R.; Fu, Q. Review on the Technological Development and Application of UAV Systems. Chinese Journal of Electronics 2020, 29, 199–207, doi:https://doi.org/10.1049/cje.2019.12.006.

13. Mazzia, V.; Comba, L.; Khaliq, A.; Chiaberge, M.; Gay, P. UAV and Machine Learning Based Refinement of a Satellite-Driven Vegetation Index for Precision Agriculture. Sensors 2020, 20, doi:10.3390/s20092530.

14. Borgogno-Mondino, E.; Lessio, A.; Tarricone, L.; Novello, V.; de Palma, L. A Comparison between Multispectral Aerial and Satellite Imagery in Precision Viticulture. Precision Agriculture 2018, 19, 195–217, doi:10.1007/s11119-017-9510-0.

15. Di Gennaro, S.F.; Dainelli, R.; Palliotti, A.; Toscano, P.; Matese, A. Sentinel-2 Validation for Spatial Variability Assessment in Overhead Trellis System Viticulture Versus UAV and Agronomic Data. Remote Sensing 2019, 11, doi:10.3390/rs11212573.

16. Khaliq, A.; Comba, L.; Biglia, A.; Ricauda Aimonino, D.; Chiaberge, M.; Gay, P. Comparison of Satellite and UAV-Based Multispectral Imagery for Vineyard Variability Assessment. Remote Sensing 2019, 11, doi:10.3390/rs11040436.

17. Deery, D.; Jimenez-Berni, J.; Jones, H.; Sirault, X.; Furbank, R. Proximal Remote Sensing Buggies and Potential Applications for Field-Based Phenotyping. Agronomy 2014, 4, 349–379, doi:10.3390/agronomy4030349.

18. Fiorentini, M.; Zenobi, S.; Orsini, R. Remote and Proximal Sensing Applications for Durum Wheat Nutritional Status Detection in Mediterranean Area. Agriculture 2021, 11, doi:10.3390/agriculture11010039.

19. Matese, A.; Toscano, P.; Di Gennaro, S.F.; Genesio, L.; Vaccari, F.P.; Primicerio, J.; Belli, C.; Zaldei, A.; Bianconi, R.; Gioli, B. Intercomparison of UAV, Aircraft and Satellite Remote Sensing Platforms for Precision Viticulture. Remote Sensing 2015, 7, 2971–2990, doi:10.3390/rs70302971.

20. Deng, L.; Mao, Z.; Li, X.; Hu, Z.; Duan, F.; Yan, Y. UAV-Based Multispectral Remote Sensing for Precision Agriculture: A Comparison between Different Cameras. ISPRS Journal of Photogrammetry and Remote Sensing 2018, 146, 124–136, doi:https://doi.org/10.1016/j.isprsjprs.2018.09.008.

21. Delavarpour, N.; Koparan, C.; Nowatzki, J.; Bajwa, S.; Sun, X. A Technical Study on UAV Characteristics for Precision Agriculture Applications and Associated Practical Challenges. Remote Sensing 2021, 13, doi:10.3390/rs13061204.

22. Aslan, M.F.; Durdu, A.; Sabanci, K.; Ropelewska, E.; Gültekin, S.S. A Comprehensive Survey of the Recent Studies with UAV for Precision Agriculture in Open Fields and Greenhouses. Applied Sciences 2022, 12, doi:10.3390/app12031047.

23. Kakaletsis, E.; Symeonidis, C.; Tzelepi, M.; Mademlis, I.; Tefas, A.; Nikolaidis, N.; Pitas, I. Computer Vision for Autonomous UAV Flight Safety: An Overview and a Vision-Based Safe Landing Pipeline Example. ACM Comput. Surv. 2021, 54, doi:10.1145/3472288.

24. Rahman, M.F.F.; Fan, S.; Zhang, Y.; Chen, L. A Comparative Study on Application of Unmanned Aerial Vehicle Systems in Agriculture. Agriculture 2021, 11, doi:10.3390/agriculture11010022.

25. Radoglou-Grammatikis, P.; Sarigiannidis, P.; Lagkas, T.; Moscholios, I. A Compilation of UAV Applications for Precision Agriculture. Computer Networks 2020, 172, 107148, doi:https://doi.org/10.1016/j.comnet.2020.107148.

26. Velasco-Muñoz, J.F.; Aznar-Sánchez, J.A.; Belmonte-Ureña, L.J.; Román-Sánchez, I.M. Sustainable Water Use in Agriculture: A Review of Worldwide Research. Sustainability 2018, 10, doi:10.3390/su10041084.

27. Schieffer, J.; Dillon, C. The Economic and Environmental Impacts of Precision Agriculture and Interactions with Agro-Environmental Policy. Precision Agriculture 2015, 16, 46–61, doi:10.1007/s11119-014-9382-5.

28. Maimaitiyiming, M.; Sagan, V.; Sidike, P.; Maimaitijiang, M.; Miller, A.J.; Kwasniewski, M. Leveraging Very-High Spatial Resolution Hyperspectral and Thermal UAV Imageries for Characterizing Diurnal Indicators of Grapevine Physiology. Remote Sensing 2020, 12, doi:10.3390/rs12193216.

29. Zhang, L.; Zhang, H.; Niu, Y.; Han, W. Mapping Maize Water Stress Based on UAV Multispectral Remote Sensing. Remote Sensing 2019, 11, doi:10.3390/rs11060605.

30. Pádua, L.; Marques, P.; Adão, T.; Guimarães, N.; Sousa, A.; Peres, E.; Sousa, J.J. Vineyard Variability Analysis through UAV-Based Vigour Maps to Assess Climate Change Impacts. Agronomy 2019, 9, doi:10.3390/agronomy9100581.

31. Park, S.; Ryu, D.; Fuentes, S.; Chung, H.; Hernández-Montes, E.; O'Connell, M. Adaptive Estimation of Crop Water Stress in Nectarine and Peach Orchards Using High-Resolution Imagery from an Unmanned Aerial Vehicle (UAV). Remote Sensing 2017, 9, doi:10.3390/rs9080828.

32. Jan, P.; Calabrese, C.; Lips, M. Determinants of Nitrogen Surplus at Farm Level in Swiss Agriculture. Nutrient Cycling in Agroecosystems 2017, 109, 133–148, doi:10.1007/s10705-017-9871-9.

33. Sozzi, M.; Kayad, A.; Gobbo, S.; Cogato, A.; Sartori, L.; Marinello, F. Economic Comparison of Satellite, Plane and UAV-Acquired NDVI Images for Site-Specific Nitrogen Application: Observations from Italy. Agronomy 2021, 11, doi:10.3390/agronomy11112098.

34. Seo, B.; Lee, J.; Lee, K.-D.; Hong, S.; Kang, S. Improving Remotely-Sensed Crop Monitoring by NDVI-Based Crop Phenology Estimators for Corn and Soybeans in Iowa and Illinois, USA. Field Crops Research 2019, 238, 113–128, doi:https://doi.org/10.1016/j.fcr.2019.03.015.

35. Kayad, A.; Sozzi, M.; Gatto, S.; Whelan, B.; Sartori, L.; Marinello, F. Ten Years of Corn Yield Dynamics at Field Scale under Digital Agriculture Solutions: A Case Study from North Italy. Computers and Electronics in Agriculture 2021, 185, 106126, doi:https://doi.org/10.1016/j.compag.2021.106126.

36. Carazo-Rojas, E.; Pérez-Rojas, G.; Pérez-Villanueva, M.; Chinchilla-Soto, C.; Chin-Pampillo, J.S.; Aguilar-Mora, P.; Alpízar-Marín, M.; Masís-Mora, M.;

Rodríguez-Rodríguez, C.E.; Vryzas, Z. Pesticide Monitoring and Ecotoxicological Risk Assessment in Surface Water Bodies and Sediments of a Tropical Agro-Ecosystem. Environmental Pollution 2018, 241, 800–809, doi:https://doi.org/10.1016/j.envpol.2018.06.020.

37. Tudi, M.; Daniel Ruan, H.; Wang, L.; Lyu, J.; Sadler, R.; Connell, D.; Chu, C.; Phung, D.T. Agriculture Development, Pesticide Application and Its Impact on the Environment. International Journal of Environmental Research and Public Health 2021, 18, doi:10.3390/ijerph18031112.

38. Singh, N.S.; Sharma, R.; Parween, T.; Patanjali, P.K. Pesticide Contamination and Human Health Risk Factor. In Modern Age Environmental Problems and their Remediation; Oves, M., Zain Khan, M., M.I. Ismail, I., Eds.; Springer International Publishing: Cham, 2018; pp. 49–68 ISBN 978-3-319-64501-8.

39. Mathews, A.J.; Jensen, J.L.R. Visualizing and Quantifying Vineyard Canopy LAI Using an Unmanned Aerial Vehicle (UAV) Collected High Density Structure from Motion Point Cloud. Remote Sensing 2013, 5, 2164–2183, doi:10.3390/rs5052164.

40. Kalisperakis, I.; Stentoumis, Ch.; Grammatikopoulos, L.; Karantzalos, K. Leaf Area Index Estimation In Vineyards From UAV Hyperspectral Data, 2D Image Mosaics And 3D Canopy Surface Models. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 2015, XL-1/W4, 299–303, doi:10.5194/isprsarchives-XL-1-W4-299-2015.

41. Weiss, M.; Baret, F. Using 3D Point Clouds Derived from UAV RGB Imagery to Describe Vineyard 3D Macro-Structure. Remote Sensing 2017, 9, doi:10.3390/rs9020111.

42. Dhananjayan, V.; Ravichandran, B. Occupational Health Risk of Farmers Exposed to Pesticides in Agricultural Activities. Current Opinion in Environmental Science & Health 2018, 4, 31–37, doi:https://doi.org/10.1016/j.coesh.2018.07.005.

43. Woods, N.; Craig, I.P.; Dorr, G.; Young, B. Spray Drift of Pesticides Arising from Aerial Application in Cotton. Journal of Environmental Quality 2001, 30, 697–701, doi:https://doi.org/10.2134/jeq2001.303697x.

44. Baio, F.H.R.; Antuniassi, U.R.; Castilho, B.R.; Teodoro, P.E.; Silva, E.E. da Factors Affecting Aerial Spray Drift in the Brazilian Cerrado. PLOS ONE 2019, 14, 1–16, doi:10.1371/journal.pone.0212289.

45. Grella, M.; Gallart, M.; Marucco, P.; Balsari, P.; Gil, E. Ground Deposition and Airborne Spray Drift Assessment in Vineyard and Orchard: The Influence of Environmental Variables and Sprayer Settings. Sustainability 2017, 9, doi:10.3390/su9050728.

46. Qin, W.-C.; Qiu, B.-J.; Xue, X.-Y.; Chen, C.; Xu, Z.-F.; Zhou, Q.-Q. Droplet Deposition and Control Effect of Insecticides Sprayed with an Unmanned Aerial Vehicle against Plant Hoppers. Crop Protection 2016, 85, 79–88, doi:https://doi.org/10.1016/j.cropro.2016.03.018.

47. Lou, Z.; Xin, F.; Han, X.; Lan, Y.; Duan, T.; Fu, W. Effect of Unmanned Aerial Vehicle Flight Height on Droplet Distribution, Drift and Control of Cotton Aphids and Spider Mites. Agronomy 2018, 8, doi:10.3390/agronomy8090187.

48. Ahmad, F.; Qiu, B.; Dong, X.; Ma, J.; Huang, X.; Ahmed, S.; Chandio, F.A. Effect of Operational Parameters of UAV Sprayer on Spray Deposition Pattern in Target and Off-Target Zones during Outer Field Weed Control Application. Computers and Electronics in Agriculture 2020, 172, 105350, doi:https://doi.org/10.1016/j.compag.2020.105350.

49. Scavo, A.; Mauromicale, G. Integrated Weed Management in Herbaceous Field Crops. Agronomy 2020, 10, doi:10.3390/agronomy10040466.

50. López-Granados, F.; Torres-Sánchez, J.; Serrano-Pérez, A.; de Castro, A.I.; Mesas-Carrascosa, Fco.-J.; Peña, J.-M. Early Season Weed Mapping in Sunflower Using UAV Technology: Variability of Herbicide Treatment Maps against Weed Thresholds. Precision Agriculture 2016, 17, 183–199, doi:10.1007/s11119-015-9415-8.

51. Pavlović, D.; Vrbničanin, S.; Anđelković, A.; Božić, D.; Rajković, M.; Malidža, G. Non-Chemical Weed Control for Plant Health and Environment: Ecological Integrated Weed Management (EIWM). Agronomy 2022, 12, doi:10.3390/agronomy12051091.

52. Gerhards, R.; Andújar Sanchez, D.; Hamouz, P.; Peteinatos, G.G.; Christensen, S.; Fernandez-Quintanilla, C. Advances in Site-Specific Weed Management in Agriculture—A Review. Weed Research 2022, 62, 123–133, doi:https://doi.org/10.1111/wre.12526.

53. De Castro, A.I.; Torres-Sánchez, J.; Peña, J.M.; Jiménez-Brenes, F.M.; Csillik, O.; López-Granados, F. An Automatic Random Forest-OBIA Algorithm for Early Weed Mapping between and within Crop Rows Using UAV Imagery. Remote Sensing 2018, 10, doi:10.3390/rs10020285.

54. Peña, J.M.; Torres-Sánchez, J.; de Castro, A.I.; Kelly, M.; López-Granados, F. Weed Mapping in Early-Season Maize Fields Using Object-Based Analysis of Unmanned Aerial Vehicle (UAV) Images. PLOS ONE 2013, 8, null, doi:10.1371/journal.pone.0077151.

55. Islam, N.; Rashid, M.M.; Wibowo, S.; Xu, C.-Y.; Morshed, A.; Wasimi, S.A.; Moore, S.; Rahman, S.M. Early Weed Detection Using Image Processing and Machine Learning Techniques in an Australian Chilli Farm. Agriculture 2021, 11, doi:10.3390/agriculture11050387.

56. Bah, M.D.; Hafiane, A.; Canals, R. Deep Learning with Unsupervised Data Labeling for Weed Detection in Line Crops in UAV Images. Remote Sensing 2018, 10, doi:10.3390/rs10111690.

57. Esposito, M.; Crimaldi, M.; Cirillo, V.; Sarghini, F.; Maggio, A. Drone and Sensor Technology for Sustainable Weed Management: A Review. Chemical and Biological Technologies in Agriculture 2021, 8, 18, doi:10.1186/s40538-021-00217-8.

58. Samuel, A.L. Some Studies in Machine Learning Using the Game of Checkers. IBM Journal of Research and Development 2000, 44, 206–226, doi:10.1147/rd.441.0206.

59. Cravero, A.; Sepúlveda, S. Use and Adaptations of Machine Learning in Big Data—Applications in Real Cases in Agriculture. Electronics 2021, 10, doi:10.3390/electronics10050552.

60. Liakos, K.G.; Busato, P.; Moshou, D.; Pearson, S.; Bochtis, D. Machine Learning in Agriculture: A Review. Sensors 2018, 18, doi:10.3390/s18082674.

61. Sharma, A.; Jain, A.; Gupta, P.; Chowdary, V. Machine Learning Applications for Precision Agriculture: A Comprehensive Review. IEEE Access 2021, 9, 4843–4873, doi:10.1109/ACCESS.2020.3048415.

62. LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. Nature 2015, 521, 436–444, doi:10.1038/nature14539.

63. Kamilaris, A.; Prenafeta-Boldú, F.X. Deep Learning in Agriculture: A Survey. Computers and Electronics in Agriculture 2018, 147, 70–90, doi:https://doi.org/10.1016/j.compag.2018.02.016.

64. Kamilaris, A.; Prenafeta-Boldú, F.X. A Review of the Use of Convolutional Neural Networks in Agriculture. J. Agric. Sci. 2018, 156, 312–322, doi:10.1017/S0021859618000436.

65. Rahnemoonfar, M.; Sheppard, C. Deep Count: Fruit Counting Based on Deep Simulated Learning. Sensors 2017, 17, doi:10.3390/s17040905.

66. Yamamoto, K.; Guo, W.; Yoshioka, Y.; Ninomiya, S. On Plant Detection of Intact Tomato Fruits Using Image Analysis and Machine Learning Methods. Sensors 2014, 14, 12191–12206, doi:10.3390/s140712191.

67. Zhang, X.; He, L.; Zhang, J.; Whiting, M.D.; Karkee, M.; Zhang, Q. Determination of Key Canopy Parameters for Mass Mechanical Apple Harvesting Using Supervised Machine Learning and Principal Component Analysis (PCA). Biosystems Engineering 2020, 193, 247–263, doi:https://doi.org/10.1016/j.biosystemseng.2020.03.006.

68. Khan, H.; Esau, T.J.; Farooque, A.A.; Abbas, F. Wild Blueberry Harvesting Losses Predicted with Selective Machine Learning Algorithms. Agriculture 2022, 12, doi:10.3390/agriculture12101657.

69. Klompenburg, T. van; Kassahun, A.; Catal, C. Crop Yield Prediction Using Machine Learning: A Systematic Literature Review. Computers and Electronics in Agriculture 2020, 177, 105709, doi:https://doi.org/10.1016/j.compag.2020.105709.

70. Pantazi, X.E.; Moshou, D.; Alexandridis, T.; Whetton, R.L.; Mouazen, A.M. Wheat Yield Prediction Using Machine Learning and Advanced Sensing Techniques. Computers and Electronics in Agriculture 2016, 121, 57–65, doi:https://doi.org/10.1016/j.compag.2015.11.018.

71. Rashid, M.; Bari, B.S.; Yusup, Y.; Kamaruddin, M.A.; Khan, N. A Comprehensive Review of Crop Yield Prediction Using Machine Learning Approaches With Special Emphasis on Palm Oil Yield Prediction. IEEE Access 2021, 9, 63406–63439, doi:10.1109/ACCESS.2021.3075159.

72. Doh, B.; Zhang, D.; Shen, Y.; Hussain, F.; Doh, R.F.; Ayepah, K. Automatic Citrus Fruit Disease Detection by Phenotyping Using Machine Learning. In Proceedings of the 2019 25th International Conference on Automation and Computing (ICAC); 2019; pp. 1–5.

73. Patil, S.S.; Thorat, S.A. Early Detection of Grapes Diseases Using Machine Learning and IoT. In Proceedings of the 2016 Second International Conference on Cognitive Computing and Information Processing (CCIP); 2016; pp. 1–5.

74. Ahmed, K.; Shahidi, T.R.; Irfanul Alam, S.Md.; Momen, S. Rice Leaf Disease Detection Using Machine Learning Techniques. In Proceedings of the 2019

International Conference on Sustainable Technologies for Industry 4.0 (STI); 2019; pp. 1–5.

75. Jian, Z.; Wei, Z. Support Vector Machine for Recognition of Cucumber Leaf Diseases. In Proceedings of the 2010 2nd International Conference on Advanced Computer Control; 2010; Vol. 5, pp. 264–266.

76. Thakur, P.S.; Khanna, P.; Sheorey, T.; Ojha, A. Trends in Vision-Based Machine Learning Techniques for Plant Disease Identification: A Systematic Review. Expert Systems with Applications 2022, 208, 118117, doi:https://doi.org/10.1016/j.eswa.2022.118117.

77. Wang, A.; Zhang, W.; Wei, X. A Review on Weed Detection Using Ground-Based Machine Vision and Image Processing Techniques. Computers and Electronics in Agriculture 2019, 158, 226–240, doi:https://doi.org/10.1016/j.compag.2019.02.005.

78. Liu, B.; Bruch, R. Weed Detection for Selective Spraying: A Review. Current Robotics Reports 2020, 1, 19–26, doi:10.1007/s43154-020-00001-w.

79. Manthou, E.; Karnavas, A.; Fengou, L.-C.; Bakali, A.; Lianou, A.; Tsakanikas, P.; Nychas, G.-J.E. Spectroscopy and Imaging Technologies Coupled with Machine Learning for the Assessment of the Microbiological Spoilage Associated to Ready-to-Eat Leafy Vegetables. International Journal of Food Microbiology 2022, 361, 109458, doi:https://doi.org/10.1016/j.ijfoodmicro.2021.109458.

80. Ren, A.; Zahid, A.; Zoha, A.; Shah, S.A.; Imran, M.A.; Alomainy, A.; Abbasi, Q.H. Machine Learning Driven Approach Towards the Quality Assessment of Fresh Fruits Using Non-Invasive Sensing. IEEE Sensors Journal 2020, 20, 2075–2083, doi:10.1109/JSEN.2019.2949528.

81. Genze, N.; Bharti, R.; Grieb, M.; Schultheiss, S.J.; Grimm, D.G. Accurate Machine Learning-Based Germination Detection, Prediction and Quality Assessment of Three Grain Crops. Plant Methods 2020, 16, 157, doi:10.1186/s13007-020-00699-x.

82. Habibi, L.N.; Watanabe, T.; Matsui, T.; Tanaka, T.S.T. Machine Learning Techniques to Predict Soybean Plant Density Using UAV and Satellite-Based Remote Sensing. Remote Sensing 2021, 13, doi:10.3390/rs13132548.

83. Kartal, S.; Choudhary, S.; Masner, J.; Kholová, J.; Stočes, M.; Gattu, P.; Schwartz, S.; Kissel, E. Machine Learning-Based Plant Detection Algorithms to

Automate Counting Tasks Using 3D Canopy Scans. Sensors 2021, 21, doi:10.3390/s21238022.

84. Ghosal, S.; Zheng, B.; Chapman, S.C.; Potgieter, A.B.; Jordan, D.R.; Wang, X.; Singh, A.K.; Singh, A.; Hirafuji, M.; Ninomiya, S.; et al. A Weakly Supervised Deep Learning Framework for Sorghum Head Detection and Counting. Plant Phenomics 2019, 2019, 1525874, doi:10.34133/2019/1525874.

85. Benos, L.; Tagarakis, A.C.; Dolias, G.; Berruto, R.; Kateris, D.; Bochtis, D. Machine Learning in Agriculture: A Comprehensive Updated Review. Sensors 2021, 21, doi:10.3390/s21113758.

Chapter 2 - Integrating UASs and canopy height models in vineyard management: a time-space approach

This is the published version of the manuscript in the Remote Sensing (MDPI) journal: Sassu, A.; Ghiani, L.; Salvati, L.; Mercenaro, L.; Deidda, A.; Gambella, F. Integrating UAVs and Canopy Height Models in Vineyard Management: A Time-Space Approach. Remote Sens. 2022, 14, 130. https://doi.org/10.3390/rs14010130.

Abstract

The present study illustrates an operational approach estimating individual and aggregate vineyards' canopy volume through three years Tree-Row-Volume (TRV) measurements and remotely sensed imagery acquired with an Unmanned Aerial System (UAS) Red-Green-Blue (RGB) digital camera, processed with MATLAB scripts, and validated through ArcGIS tools. The TRV methodology was applied by sampling a different number of rows and plants per row each year to evaluate the reliability and accuracy of this technique compared with a remote approach. The empirical results indicate that the estimated tree-row-volumes derived from a UAS Canopy Height Model (CHM) are up to 50% different from those measured on the field using the routinary TRV technique in 2019, with a much higher difference in the 2016. These findings outline the importance of data integration among techniques that mix proximal and remote sensing in routine vineyards' agronomic practices, helping to reduce management costs and increase the environmental sustainability of traditional cultivation systems.

Keywords: precision viticulture; TRV; CHM; unmanned aerial systems; digital models; grapevine canopy measurement.

1. Introduction

Precision farming techniques assume the input optimization to improve production efficiency and sustainability [1,2,3,4]. A more efficient use of plant protection products and tools leads to fewer rural environments pollution loads, higher crop quality, less monetary costs, and increased production rates, positively impacting the economic and ecological sustainability of farms [5,6,7,8,9,10].

The vineyard is a heterogeneous environment where spatial monitoring techniques application for biomass development and volume characterization can be integrated into a decision support system for plant protection strategies optimization [11,12,13,14,15,16,17,18,19,20,21,22].

Farm field measurements are routinely carried out, determining the total amount of plant protection products via simplified mathematical approaches that require direct measurement of canopy height, thickness, and distance between canopies. More specifically, the Tree-Row-Volume (TRV) technique, requiring manual vineyard measurements, estimates the total plant volume by ground unit (m^3 ha^{-1}) in vineyard crown height, width, and inter-row distance [23,24,25,26]. TRV is relatively well known to grape growers and agronomists. This method has been extensively used for various purposes, including rough estimation of the adequate treatment dosage for plant protection within a specific vineyard, leading to a more comprehensive canopy

management. Other volume measurement systems such as leaf wall area (LWA) and Unit Canopy Row (UCR) were not considered in this study because they primarily referred to viticultural contexts with peculiar territorial or productive characteristics. Despite a consensus application to plant/row monitoring in largely variable agronomic and ecological field conditions in Europe as well as in other productive contexts, TRV determination is a labour-intensive and time-consuming task. To our knowledge, a specific analysis of model errors and uncertainties related to this routine approach is still lacking, despite the intense development of vineyards' canopy proximal monitoring techniques [27,28].

At the same time, a precise characterization of plants' structure was more recently performed using remote sensing tools such as image analysis techniques, stereoscopic photography, analysis of the light spectrum, and ultrasonic and optical ranging [29,30,31,32,33,34,35]. Earlier Structure from Motion (SfM) approaches provided a basic framework for 3D vineyard point clouds reconstruction, with the aim of quantifying grapevine canopy volume and Leaf Area Index (LAI) [36,37,38]. These estimations were extensively used to optimize canopy management and pest control, especially when Variable Rate Technology (VRT) was employed [39]. More recently, the acquisition of high-resolution Unmanned Aerial System (UAS) RGB imagery of the canopy has proved to be an effective tool for plant architecture estimation through the accurate and reliable digital models' computation [40,41,42,43]. The use of Ground Control Points (GCPs), located within the orchard's scene, represents an essential practice for spatial accuracy and minimization of model's errors [44,45].

Based on these premises, the present study illustrates a methodology aiming to estimate individual and aggregate vineyard canopy volume through UAS remotely sensed imagery acquired with an RGB digital camera, analysed with MATLAB 2018b scripts, and validated by ArcGIS 10.7.1 tools at various growth phase. More specifically, canopy volumes routinely measured on the field following a TRV approach were compared to remotely acquired volumes from an integrated analysis of Digital Surface Models (DSMs), Digital Terrain Models (DTMs), and Canopy High Models (CHMs) derived from SfM assessment at different time points. These results contribute to assessing the reliability of the TRV technique, verifying if the estimated volumes derived from UAS detection are comparable with, or statistically different from, the one measured in the field.

2. Materials and Methods

2.1. Study Area

The survey took place in an experimental field of 0.8 ha (Figure 1) in Usini, North-Eastern Sardinia, Italy (Lat. 40°40′10.13″; Long. 8°29′37.35″; WGS84, EPSG 4326), at 144 m above the sea level. The grapevines (*Cagnulari* cv.) were planted in a clay-loam soil with a 0.90 × 2.10 m spacing (East-West rows orientation) and trained as Vertical Shoot Position (VSP). A GNSS Leica 900 RTK receiver (Leica Geosystems) was used to record the X-Y coordinates of six GCPs for accurate georeferencing of the ortho-mosaics analysed in this study. Twenty-four additional sample points were identified over the entire surface area and geo-referenced to characterize the canopy after every UAS flight. Field measurements were run under sunny, clear sky conditions during the vine growing season in 2016 (July and August), 2017 (July), and 2019 (June), at the same phenological phase based on the specific year's weather profile.

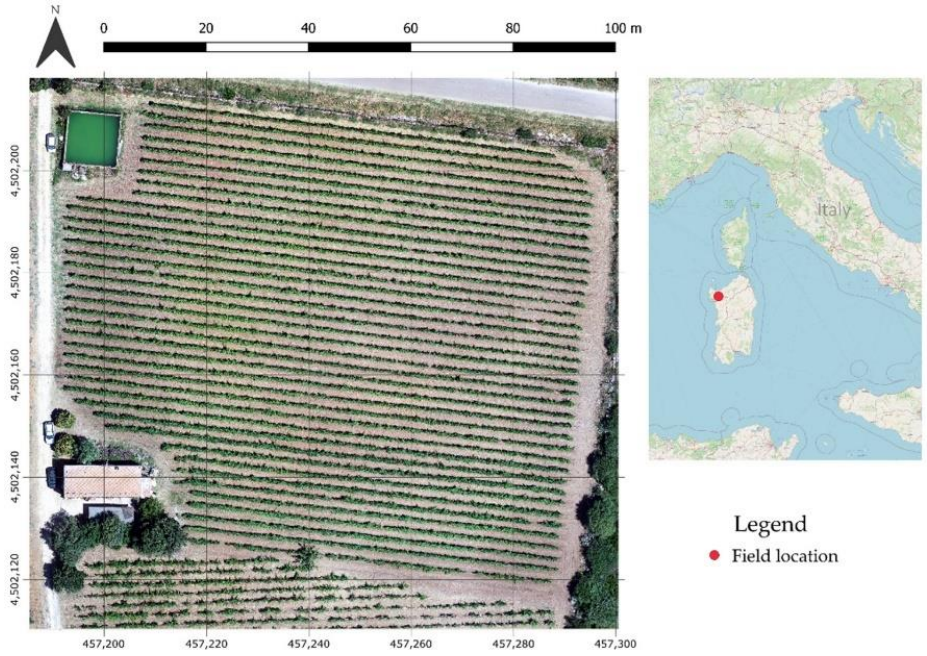


Figure 1. The experimental field located in Usini, North-Eastern Sardinia, Italy (WGS84-UTM 32N, EPSG 32632 projected coordinate system).

2.2. TRV-Based Field Measurements

In 2016, twelve measurements (every ten plants) for nine random rows were extracted, whereas 105 measurements (one for each plant of the row) were collected for four random rows used in 2017 and for six random rows in 2019. For TRV estimation, measurements were taken using a rolling tape marked on a wooden rod to facilitate height/width detection and limit the operator fatigue.

In 2016, field measurements were performed considering the highest and widest extension reached by the plant (even isolated shoots, highlighted in red color in Figure

2b) in a measurement point conventionally identified as the plant's center (Figure 2a). In 2017 and 2019, the shoots outside the uniform row shape (falling inside the blue polygon in Figure 2b) were not considered.

Since height measurements were performed considering the distance from the ground surface to the highest part of the row, also taking account of the height of the trunk, each height measurement was referred to the canopy by subtracting 0.90 m from the total value. The "cutting height" of 0.90 m (represented by the red line parallel to the ground in Figure 2a) matches the branches production zone in the cultivation system considered in this study.

(a) (b)

Figure 2. (a) The measurements of height and width rows using the wooden rod. Canopy heights were calculated subtracting 0.90 m (red line) from the total measure; (b) an exemplification of the influence of isolated shoots (in red) in height and width determination during field measurements. The blue shape represents the volume estimated in 2017 and 2019, excluding the shoots out of the row's shape considered in 2016.

TRV was calculated multiplying the average rows height and width values, by an aerial conversion factor (10,000) (1 ha = 10,000 m^2), in turn divided by the inter-row width [23] (Equation 1):

$$TRV = \frac{H \times V \times 10000}{I} \quad (1)$$

where H is the average row height (m), W is the average row thickness (m), and I is the inter-row width (m).

2.3. UAS-Based Sensing

Data acquisition in the first survey year was carried out using a customized hexacopter equipped with a CMOS sensor Canon EOS 750D of 24 megapixels resolution, sensor size 22.3 mm × 14.9 mm, focal length 50 mm f/2.8. In the following years, a DJI Phantom 4 Pro (Shenzhen, China) equipped with RGB CMOS 1" sensor of 21 megapixels resolution, Field of View (FOV) 84°, 8.8 mm/24 mm (35 mm format equivalent), f/2.8-f/11 autofocus 1 m-∞ was adopted for field measurement.

The photographic sets were acquired with a 75% front overlap and 85% side overlap at 35 m height Above Ground Level (AGL) in 2016 and 2019, and 50 m in 2017. The different sensors and the flight altitude, combined with specific elaboration processes, involved a different Ground Sampling Distance (GSD) of the digital models in the five dates. Survey information about Day of Year (DOY), Biologische Bundesanstalt, Bundessortenamt und CHemische Industrie (BBCH), and Growing Degree Day (GDD) are summarized in Table 1.

Table 1. The flight surveys details, ortho-mosaics' properties, Ground Sample Distance (GSD), vineyard Day of the Year (DOY), BBCH, and Growing Degree Day (GDD) values.

Year	Date	Orthomosaics		Vineyard phenology		
		GSD (cm)	*RMSE* (cm)	DOY	BBCH	GDD
2016	07/07	0.26	3.3	189	79	798
	02/08	0.26	3.1	215	81	1176
2017	17/07	0.90	6.4	198	79	798
	31/07	1.22	5.7	212	81	1282
2019	26/06	0.92	2.5	177	71	665

2.4. Identification of the Canopy Height Model

Standard approaches to create 3D models derived from SfM processing and CHM [44,46] adopt the absolute height of crop canopies as a target variable and define a CHM as the difference between the DSM and the DTM. In the present study, Agisoft Metashape allowed the estimation of the vineyard's soil surface elevation profile (DTM) through the classification of the dense cloud, the essential 3D model needed to obtain the 2D digital models (DSM and DTM) in the software's workflow. The specific tool named "Classify Ground Points" (located into the Dense Cloud Tools' menu) for points' classification, was used to detect the ground points and enhance the hole filling derived by the removal of the canopy points. The models' approximation derives from the inability of nadir images, taken from above, to reconstruct the lower part of the plants (Figure 3 and Figure 4a).

Starting from the resolution of each digital model (defined by the surface occupied by a single pixel), MATLAB and ArcGIS CHMs analysis extracted the covered area height information from the vineyard rows by integrating the volume of all individual pixels that

constitute the canopy. This approach generated a file containing the height values on a reference plane placed 0.90 m above the soil's surface (represented by the red line in Figure 2a). This height matches the production zone where the branches are located, contributing to remove the influence of soil surface and the stumps, allowing the measurement of the effective volume occupied by the canopy. The height information contained in the CHM is free from interference derived by background colors, shadows, or infesting plants, at least until their height exceeds the set limit of 0.90 m.

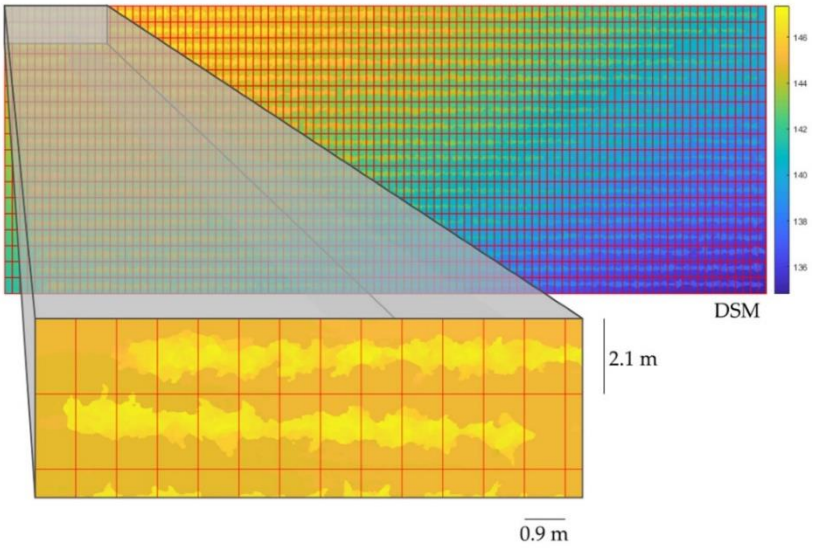


Figure 3. The Digital Surface Model (DSM) raster (colored with a pseudo-colors palette to emphasize the altitude of each pixel above sea level) employed during the MATLAB processing and divided into 0.90 m × 2.10 m blocks (as shown in the enlarged upper left portion of the ortho-image). Canopy and terrain pixels are represented with different colors in each block.

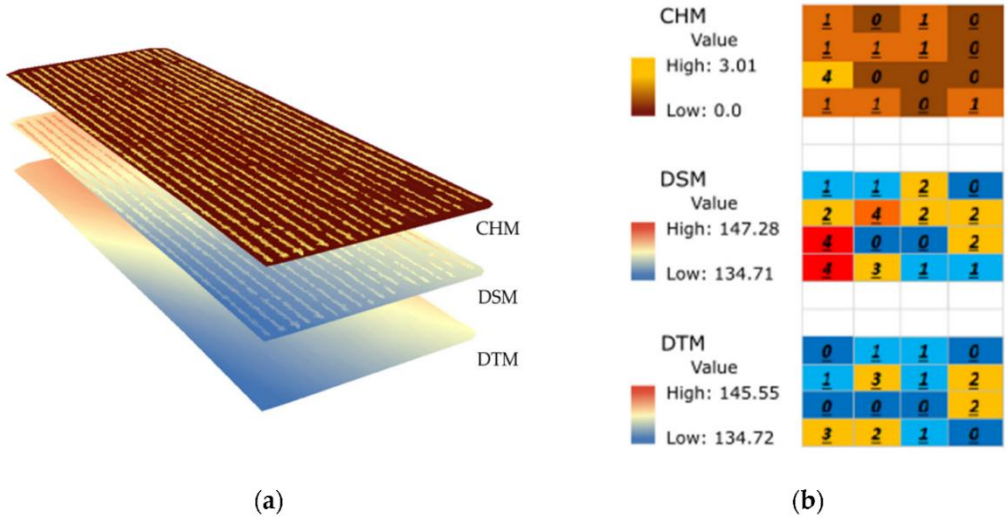


Figure 4. (a) The three-layer used to obtain canopy volumes through ArcGIS; (b) an exemplification of the Canopy Height Model (CHM) derived from the difference between Digital Surface Model (DSM) and Digital Terrain Model (DTM) rasters. Each square represents a pixel with the relative altitude above the sea level (DSM and DTM) and height above ground (CHM). The CHM height values were used to calculate the rows' volumes.

The MATLAB analysis was run solely on DSMs (Figure 3). The altitude of each row was estimated by averaging the visible soil values between each row. Measurements were performed creating a rows x columns pixel raster matrix of 0.26 cm/pixel in 2016, 1.22 cm/pixel in 2017, and 0.92 cm/pixel in 2019 (based on the Ground Sampling Distance) and dividing the raster matrix into blocks of 0.90 m × 2.10 m. The raster matrix was oriented so that the x columns and y rows were parallel and perpendicular to the vines' rows, respectively, and each block included the canopy and a small part of the ground (Figure 3). This approach allowed the reduction of the influence of vineyard slope (represented by a pseudo-color palette ranging from purple to orange color) on the canopy height measurement, providing a reference surface that consists of pixels with the same elevation. All pixel values included in the volume estimation were obtained by extracting height information above 0.90 m from the reference plan (ground) consisting of the minimum values in each block.

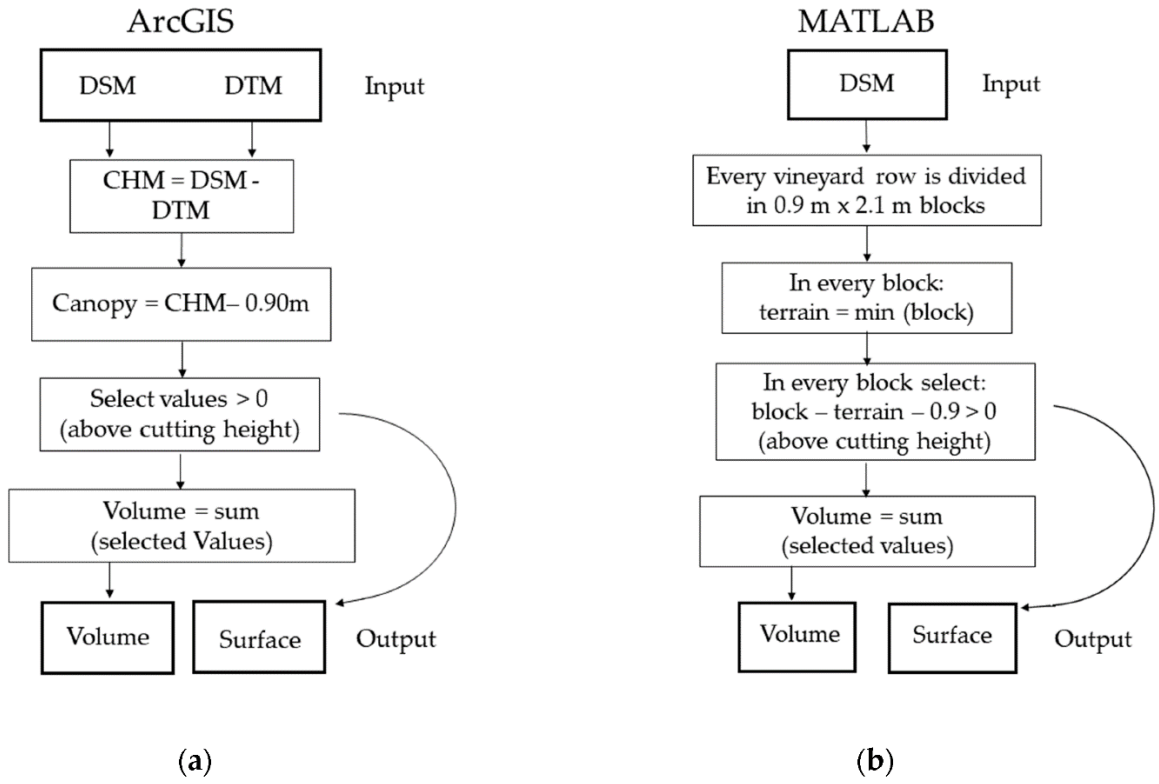


Figure 5. Flowcharts summarizing the two approaches implemented in this study, (a) the ArcGIS CHM workflow and (b) the MATLAB workflow.

ArcGIS measurements were performed by subtracting the elevation value included in the first basic input raster (DTM) from the second input raster (DSM) pixel-by-pixel value and deducting 0.90 m from each pixel height (Figure 4) using the "Raster Calculator" tool. The raster output (CHM), containing the new reference plane and the canopy, was used to calculate the volumes occupied by each row by multiplying every pixel surface to its

height without any interference related to the vineyard's slope. The Agisoft Metashape-ArcGIS CHM generation followed the same methodology of [31].

Both MATLAB and ArcGIS methods allowed an easy calculation of the Green Canopy Cover (GCC) as the percentage of field surface occupied by grapevine vegetation. The number of canopy pixels above the cutting edge of 0.90 m (as summarized in Figure 5) were counted and expressed as a per cent share in the total number of pixels in the field.

3. Results

Table 2 reports the canopy volumes determined for all grape-vine rows through field measurement and the average percentage of the GCC area [47]. The measurements were performed using the TRV technique (Equation 1) and automatic approaches based on UAS-derived digital models created using ArcGIS and MATLAB software.

Table 2. The Tree-Row-Volume (TRV) field measurements compared with the grapevine volumes calculated by ArcGIS and MATLAB, and the Green Canopy Cover (GCC), at five different dates.

Date	Field TRV(m^3/ha)	ArcGIS TRV(m^3/ha)	MATLAB TRV(m^3/ha)	MATLAB
				GCC (%)
07/07/2016	5971[1]	1991	1898	29
02/08/2016	5984[1]	1649	1580	26
17/07/2017	1271[2]	1343	1427	24
31/07/2017	1311[2]	1316	1353	30
26/06/2019	2360[3]	1550	1572	32

[1] Twelve measurements repeated for nine rows were used to extract the TRV value.
[2] One hundred measurements repeated for four rows were used to extract the TRV value.
[3] One hundred measurements repeated for six rows were used to extract the TRV value.

The intrinsic differences between the three years reflect differential plant growth and vineyard management expressed by different volume estimation through TRV, ArcGIS, and MATLAB techniques.

3.1. TRV Measurement Results

When estimating the vineyard field characteristics, the TRV technique, based on width and height rows measurements, was frequently regarded as a decent representation of the canopy structure. Canopy volumes estimation in 2016 was similar on both dates, and the same result was observed for 2017. In 2019, the estimated volume reached an intermediate value between 2016 and 2017. These differences are related to natural canopy growth differences among years and different measurement approaches. The number of rows and the measurements for each row played a crucial role in volume evaluation. A statistical analysis demonstrating the substantial instability of field estimation of canopy volumes, based on a subsample of rows, is reported in Figure 6. In this exercise, TRV was routinely re-calculated considering progressively smaller

randomly selected samples of rows in the vineyard. TRV was expressed as a percentage departure from the TRV measure derived from the full sample. Particularly, heterogeneous results of this analysis carried out at the different experimental dates for 2016, 2017, and 2019, indicate a limited robustness of the TRV estimate at decreasing sample sizes. For instance, calculating TRV on sub-samples with half number rows led to a mis-estimation above 10% of the estimated value on the whole sample. Only estimations based on a sub-sample with a high (or very high) proportion of plants, in respect of the total population, provide reliable canopy volume values. These results were consistent at different survey dates over the three study years.

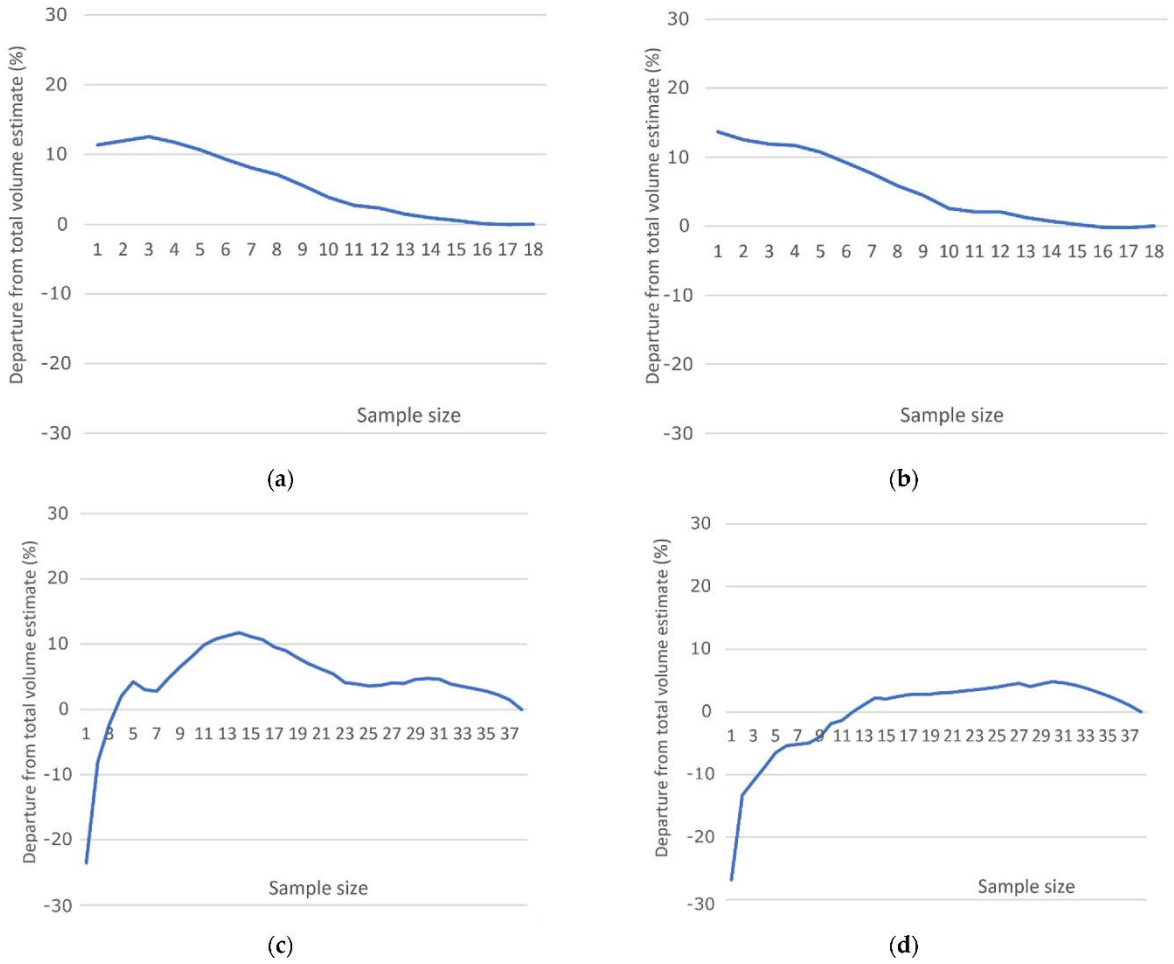


Figure 6. Statistical stability (per cent departure from whole-sample TRV estimate) of canopy volume field measurements as a function of sample size by survey date (a) 7 July 2016; (b) 2 August 2016; (c) 31 July 2017; (d) 26 June 2019.

3.2. MATLAB and ArcGIS Results

Significant differences between field and remote measurements were also observed in this study. Conversely, the remotely sensed canopy volumes estimated through MATLAB and ArcGIS procedures provided similar and consistent results for the three years of investigation. Compared to TRV results, the remote sensing data appeared

different at both dates in 2016 and slightly different in 2019. These discrepancies may confirm reliability and precision of the digital model to obtain canopy volume estimation compared with manual measurements. Considering non-parametric inference, differences in volume estimates derived from ArcGIS and MATLAB software were statistically insignificant (Mann Whitney U test, $p > 0.05$). By contrast, differences in volume estimates derived from field TRV measurement and MATLAB software (or ArcGIS software) were, in both cases, statistically significant (Mann Whitney U test, $p < 0.05$). The 2017 and 2019 field data had a slightly higher similarity with the values obtained through software elaboration. These results can be justified with (i) the different measure extraction process (in 2016 the measurements were performed considering the highest and widest extension reached by the plant, even isolated shoots) and (ii) the extension of measurements to all the plants of the investigated rows (four rows in 2017 and six in 2019). In 2019, 20 orange cards were uniformly applied in different plants over the field area to evaluate the precision of the CHM reconstruction. A comparison between the height values extracted by ArcGIS and MATLAB software, and those measured in the field, confirms the CHM model precision ($R^2 = 0.80$, $p < 0.001$, RSME = ±10.28). The GCC range of 24% and 32% obtained in July and August was similar to the values correlated with LAI obtained by Ballesteros et al. [47] on red grapevine Tempranillo cv. and Cabernet Sauvignon cv. TRV and GCC had a similar pattern of development.

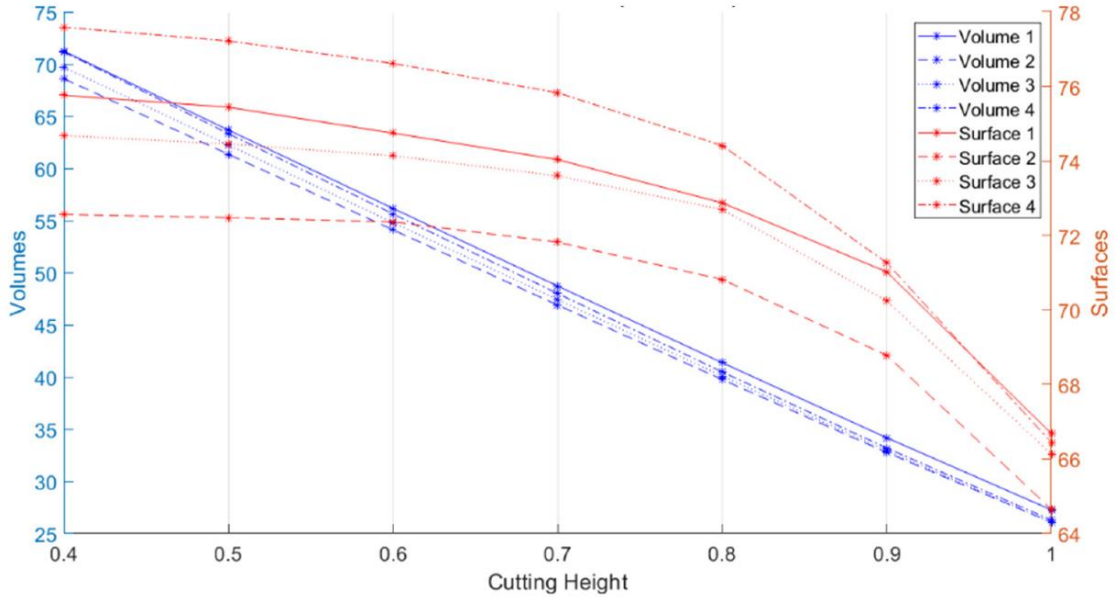


Figure 7. The absolute variation of volume (m^3) and surface area (m^2) of four vineyard rows with imposed cutting heights (MATLAB elaboration) in 2019.

Figure 7 highlights the importance of selecting the correct cutting height during the CHM elaboration to have a representative volume of the canopy. Its variation involves a constant reduction in volume and a non-linear change in surface area due to the progressive narrowing of canopy width from the bottom part to the upper. Surface and

volume variations were calculated for seven different cutting heights, ranging between 0.4 m up to 1.0 m, with 0.1 m intervals. Based on this elaboration, Figure 7 indicates how canopy volumes in June 2019 vary almost linearly, differently from the surfaces, quite limited in the early stages and decreasing afterward with the same time pattern observed for 2016 and 2017.

4. Discussion

The study aimed at estimating volume and surface variation on an experimental vineyard over three years (2016, 2017, and 2019) by correlating TRV field measurements with RGB-UAS remote sensed estimations. The combination of UAS and SfM algorithms was appropriate for vineyard volume and GCC calculation, contributing to a comprehensive description of grapevines canopy structure. These tools are of relevance since traditional methods for physiological variable estimation, routinely used by winemakers, are often time-consuming and require expensive and long-term procedures in both the field and laboratory [46].

Compared with classic methods for row volume calculation, such as the TRV, the proposed approach provided larger information, greater measurement detail, and a precision level not subjected to annual variability derived from different measurement approaches and operators' skills over the years [27]. The canopy evaluation by UAS-based imagery allows a non-destructive and standardized framework [28] avoiding (or at least reducing) the sampling error intrinsic in manual measurements [47]. The intrinsic variability of manual measurements and the related sampling errors derive from the difficulty to measure the exact height and width of the canopy, often compromised by the presence of single shoots coming out from the main volume of the row and thus affecting the overall measure [31]. Furthermore, large surface extension and a high amount of sampling points determine operator's fatigue and may lead to a high and hardly controlled level of measurement approximation [32].

Based on the empirical results of this study, UASs' high-resolution datasets provide accurate geo-referenced imagery with a high spatial and temporal resolution that is near-real-time delivered [33]. The immediate availability of data would lead farmers to consider remote sensing technology as a useful tool for timely operations [34]. Further efforts should be made to develop dedicated and user-friendly software for image analysis and automatic detection of relevant management indexes for specific agronomic practices, such as the TRV.

In this study, two simulation exercises on remotely sensed input data were proposed, respectively based on a simplified approach grounded on a user-friendly software (ArcGIS) and on a less intuitive programming scheme developed through MATLAB software. These two approaches represent different computational strategies, the former

reflecting a possible software implementation for visual integration of proximal (field) and remote sensing at the vineyard/farm scale, and the latter reflecting a generalized application for batch computation and assessment over larger spatial scales [40]. Interestingly, remote measurement techniques showed comparable results in terms of the output variables, suggesting how the use of a user-friendly approach and a more complex programming scheme does not affect the final estimation of the target variable [33]. However, MATLAB scheme allows calculation of the canopy volume directly from the DSM, thus overcoming the use of third-part software (in our case, Agisoft Metashape) to estimate differences between DSM and DTM as a preliminary step to the creation of the CHM.

Measurement heterogeneity was associated, in large part, with the difference between manual estimates and remote analysis methodologies [48,49,50,51,52]. Based on this evidence, and thanks to the intrinsic variability of the canopy structure across rows, it seems unreasonable to estimate TRV based on individual model rows or even representative plants, although this practice was frequently adopted to expedite fieldwork [37]. It has been demonstrated how whole-sample estimations are frequently biased when using a sub-sample of measures. Although earlier studies have been devoted to optimal measurement of canopy geometric features, indicating TRV techniques as a reliable reference to estimate canopy volumes [11,53], it can be assumed that the TRV scheme cannot be used as an accurate representation of geometric canopy characteristics of the whole vineyard, due to its inability to detect heterogeneity with a reduced amount of sampling [36]. For this reason, TRV measures should be considered as unrepresentative (or largely biased) in comparison with results from UAS analysis, unless TRV is performed under a huge number of measurements, which is time-consuming and unrealistic for the purposes of standard fieldwork.

In this direction, the empirical results of this study document a significant difference between the estimated volume derived from UAS techniques and TRV field measures. A precise estimation of vineyard's row volumes is crucial for (i) planning agrochemical spraying on both small and large-scale plant protection schemes, (ii) monitoring agronomic operations associated with the crop status, and (iii) preserving the quality of the final product, especially for cultivars that require specific canopy volume extension. In this perspective, the directives 2009/127/EC [54] and 2009/128/EC [55] of the European Parliament, referring to the sustainable use of pesticide, focused on the possible strategies to contain volume distribution excess doses. Assuming the crucial role in determining pesticide doses appropriate for a given vineyards, the possible overestimation of TRV based on field measures may determine an excessive use of chemicals, with negative implications for both economic management and environmental quality of farms [28].

The remote sensing methodology illustrated here allowed for an easy detection of plants (i.e., automatically scanning the whole surface relevant to measurement), evaluating apparent variations in the thickness of the canopy [34]. Moreover, this technique allowed row segmentation by using canopy pixels selected through the CHM [33]. In vineyards remote sensing is also crucial to exclude the soil and surface weeds from the analysis to avoid the mis-estimation of canopy pixel values [31]. Computer performance did not represent a limit for processing operations, needing only 2 to 4 hours for the models proposed here and resulting in significantly less effort than the manual field measurements [27]. Such a simple approach becomes more complex when images from the top and the derived digital models are not available. GCC values were comparable to the results of earlier studies reporting canopy coverage between 30% and 40% in similar trellis systems [56,57]. These results were achieved after shoot pruning and remained constant until harvest [47]. Since the BBCH values reported in Table 1 for the three survey years range from fruit set (BBCH-71) to veraison (BBCH-81), the GCC results of this study can be perfectly comparable with those mentioned above. To obtain a more reliable indicator of canopy status, this variable can be considered when estimating row volumes.

Disadvantages of UAS remote sensing applications lie in the inability to perform flights under adverse meteorological conditions, which might damage the UAS and provide unreliable data due to insufficient sunlight irradiation of the crop. Compared with other technology developed for similar purposes, UASs are relatively inexpensive with critical limitations derived from the need for specific piloting skills and knowhow to process, analyze, and to convert the acquired data into useful information for winegrowers [37]. These findings suggest the need of new technical skills able to fill the gap between winegrowers and information technologies and develop user-friendly tools to spread their use, especially in rural areas specialized in high-value wine production. The main limitation is the integration of this innovation into a complex decision support system aimed at optimizing crop management, reducing costs, operator fatigue, and the release of pollutants derived by agrochemicals' over-dosage.

The specific ArcGIS exercise developed in our study may answer this limitation, providing a simplified interface for the collection and processing of remotely sensed data and guaranteeing an easy management and integration of data recorded on the field. This technology would help farmers to control grapevine vigor and canopy growth patterns, improve support decisions for crop management, optimize pesticide and fertilizer application, and enhance yield forecasting [27]. The use of dedicated software allows for a more precise volume evaluation, analyzing the smallest details, at least in the visible parts from the top of each plant [46]. The height model resolution was

demonstrated to be comparable with results derived from more expensive sensors, such as the Lidar [26,58].

5. Conclusions

This contribution focused on the possible use of additional techniques replacing TRV field estimation with remote estimation. The empirical results of our study confirm the appropriateness of integrating (or even replacing) field measures with more precise and accurate techniques when estimating orchards structural characteristics for optimization of chemical application and other agronomic practices with a direct impact on economic costs and ecological sustainability. Although field estimation of TRV is a routinary sampling methodology still applied in many agronomic contexts, it has been demonstrated that remote applications may provide reliable and accurate tools to estimate TRV. Based on these results, extensive use of TRV is recommended, when supported by remote sensing, to better qualify errors and heterogeneities in field estimates. This is particularly important when decisions on cost estimation, agronomic practices, and sustainability issues are uniquely taken based on information derived from TRV field assessment. Proximal and remote sensing together represent promising tools in precision viticulture. In this perspective, further studies should propose new (or refined) techniques quantifying specific canopy characteristics (in addition to height and volume) from an expert interpretation of unmanned aerial vehicles' images. This specific knowledge may provide the necessary information for a comprehensive understanding of structural characteristic and functional traits of vineyards, having the final objective of enhancing together economic performances and environmental sustainability of productive farms.

Author Contributions

Conceptualization, F.G., A.S., L.S. and L.G.; methodology, F.G.; software, L.G. and L.S.; validation, A.S., L.M. and L.S.; formal analysis, A.S.; investigation, F.G., A.S., L.M. and A.D.; resources, F.G.; data curation, A.D.; writing-original draft preparation, F.G.; A.S., L.G., L.S. and L.M.; writing-review and editing, L.S. and F.G.; visualization, A.S.; supervision, F.G.; project administration, F.G.; funding acquisition, F.G. All authors have read and agreed to the published version of the manuscript.

Funding

This research was funded by "Advanced Technologies for LANds management and Tools for Innovative Development of an Eco sustainable agriculture, ATLANTIDE" CUP J88D20000070002. This project has received funding from the Regione Autonoma della

Sardegna (RAS), Italy, through the initiative called “Progetti di Ricerca e Sviluppo, da realizzare in Sardegna attraverso la collaborazione tra imprese e organismi di ricerca o infrastrutture di ricerca, secondo le modalità e le finalità previste dalla Deliberazione G.R. n. 6/27 del 31.1.2017 e in coerenza con la Smart Specialization Strategy (S3) della Regione Autonoma della Sardegna. Progetti di Ricerca e Sviluppo in collaborazione tra imprese e organismi di ricerca o infrastrutture di ricerca”. This research was also funded by “Leveraging CPS Composability & Modularity for Customized and Autonomous Civilian Drones”, H2020-ECSEL-2018-2-RIA-two-stage. This project has received funding from the ECSEL Joint Undertaking (JU) under grant agreement No. 826610. The JU receives support from the European Union’s Horizon 2020 research and innovation program and Spain, Austria, Belgium, Czech Republic, France, Italy, Latvia, Netherlands. Prof. Francesca Palumbo is also grateful to the University of Sassari that supported her research activities through the “fondo di Ateneo per la ricerca 2020”.

Acknowledgments

This work is part of the activities of the project “Advanced Technologies for LANds management and Tools for Innovative Development of an Eco Sustainable agriculture” (ATLANTIDE).

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Recanatesi, F.; Clemente, M.; Grigoriadis, E.; Ranalli, F.; Zitti, M.; Salvati, L. A fifty-year sustainability assessment of Italian agro-forest districts. Sustainability 2016, 8, 32. [Google Scholar] [CrossRef][Green Version]
2. Biasi, R.; Brunori, E.; Ferrara, C.; Salvati, L. Towards sustainable rural landscapes? A multivariate analysis of the structure of traditional tree cropping systems along a human pressure gradient in a Mediterranean region. Agrofor. Syst. 2017, 91, 1199–1217. [Google Scholar] [CrossRef]
3. Colantoni, A.; Paris, E.; Bianchini, L.; Ferri, S.; Marcantonio, V.; Carnevale, M.; Palma, A.; Civitarese, V.; Gallucci, F. Spent coffee ground characterization, pelletization test and emissions assessment in the combustion process. Sci. Rep. 2021, 11, 5119. [Google Scholar] [CrossRef] [PubMed]
4. Santesteban, L.G.; Di Gennaro, S.F.; Herrero-Langreo, A.; Miranda, C.; Royo, J.B.; Matese, A. High-resolution UAV-based thermal imaging to estimate the

instantaneous and seasonal variability of plant water status within a vineyard. Agric. Water Manag. 2017, 183, 49–59. [Google Scholar] [CrossRef]

5. Delfanti, L.; Colantoni, A.; Recanatesi, F.; Bencardino, M.; Sateriano, A.; Zambon, I.; Salvati, L. Solar plants, environmental degradation and local socioeconomic contexts: A case study in a Mediterranean country. Environ. Impact Assess. Rev. 2016, 61, 88–93. [Google Scholar] [CrossRef]
6. Bianchini, L.; Costa, P.; Dell'Omo, P.P.; Colantoni, A.; Cecchini, M.; Monarca, D. An Industrial Scale, Mechanical Process for Improving Pellet Quality and Biogas Production from Hazelnut and Olive Pruning. Energies 2021, 14, 1600. [Google Scholar] [CrossRef]
7. Cecchini, M.; Piccioni, F.; Ferri, S.; Coltrinari, G.; Bianchini, L.; Colantoni, A. Preliminary Investigation on Systems for the Preventive Diagnosis of Faults on Agricultural Operating Machines. Sensors 2021, 21, 1547. [Google Scholar] [CrossRef]
8. Tavankar, F.; Nikooy, M.; Latterini, F.; Venanzi, R.; Bianchini, L.; Picchio, R. The Effects of Soil Moisture on Harvesting Operations in Populus spp. Plantations: Specific Focus on Costs, Energy Balance and GHG Emissions. Sustainability 2021, 13, 4863. [Google Scholar] [CrossRef]
9. Borgogno Mondino, E.; Gajetti, M. Preliminary considerations about costs and potential market of remote sensing from UAV in the Italian viticulture context. Eur. J. Remote Sens. 2017, 50, 310–319. [Google Scholar] [CrossRef]
10. Vrontis, D.; Morea, D.; Basile, G.; Bonacci, I.; Mazzitelli, A. Consequences of technology and social innovation on traditional business model. Technol. Forecast. Soc. Chang. 2021, 170, 120877. [Google Scholar] [CrossRef]
11. Di Gennaro, S.F.; Matese, A. Evaluation of novel precision viticulture tool for canopy biomass estimation and missing plant detection based on 2.5D and 3D approaches using RGB images acquired by UAV platform. Plant Methods 2020, 16, 91. [Google Scholar] [CrossRef]
12. De Castro, A.I.; Jimenez-Brenes, F.M.; Torres-Sánchez, J.; Peña, J.M.; Borra-Serrano, I.; López-Granados, F. 3-D characterization of vineyards using a novel UAV imagery-based OBIA procedure for precision viticulture applications. Remote Sens. 2018, 10, 584. [Google Scholar] [CrossRef][Green Version]

13. Cinat, P.; Di Gennaro, S.F.; Berton, A.; Matese, A. Comparison of unsupervised algorithms for Vineyard Canopy segmentation from UAV multispectral images. Remote Sens. 2019, 11, 1023. [Google Scholar] [CrossRef][Green Version]

14. Di Gennaro, S.F.; Toscano, P.; Cinat, P.; Berton, A.; Matese, A. A precision viticulture UAV-based approach for early yield prediction in vineyard. In Precision Agriculture'19; Stafford, J.V., Ed.; Wageningen Academic Publishers: Wageningen, The Netherlands, 2019; pp. 370–378. [Google Scholar] [CrossRef]

15. Khaliq, A.; Comba, L.; Biglia, A.; Ricauda Aimonino, D.; Chiaberge, M.; Gay, P. Comparison of satellite and UAV-based multispectral imagery for vineyard variability assessment. Remote Sens. 2019, 11, 436. [Google Scholar] [CrossRef][Green Version]

16. Matese, A.; Di Gennaro, S.F. Practical applications of a multisensory UAV platform based on multispectral, thermal and rgb high resolution images in precision viticulture. Agriculture 2018, 8, 116. [Google Scholar] [CrossRef][Green Version]

17. Pádua, L.; Adão, T.; Sousa, A.; Peres, E.; Sousa, J.J. Individual Grapevine Analysis in a Multi-Temporal Context Using UAV-Based Multi-Sensor Imagery. Remote Sens. 2020, 12, 139. [Google Scholar] [CrossRef][Green Version]

18. Pádua, L.; Marques, P.; Adão, T.; Guimarães, N.; Sousa, A.; Peres, E.; Sousa, J.J. Vineyard variability analysis through UAV-based vigour maps to assess climate change impacts. Agronomy 2019, 9, 581. [Google Scholar] [CrossRef][Green Version]

19. Pádua, L.; Marques, P.; Hruška, J.; Adão, T.; Peres, E.; Morais, R.; Sousa, J.J. Multi-temporal vineyard monitoring through UAV-based RGB imagery. Remote Sens. 2018, 10, 1907. [Google Scholar] [CrossRef][Green Version]

20. Pichon, L.; Besqueut, G.; Tisseyre, B. A systemic approach to identify relevant information provided by UAV in precision viticulture. Adv. Anim. Biosci. 2017, 8, 823–827. [Google Scholar] [CrossRef]

21. Pichon, L.; Leroux, C.; Macombe, C.; Taylor, J.; Tisseyre, B. What relevant information can be identified by experts on unmanned aerial vehicles' visible images for precision viticulture? Precis. Agric. 2019, 20, 278–294. [Google Scholar] [CrossRef]

22. Romero, M.; Luo, Y.; Su, B.; Fuentes, S. Vineyard water status estimation using multispectral imagery from an UAV platform and machine learning algorithms for

irrigation scheduling management. Comput. Electron. Agric. 2018, 147, 109–117. [Google Scholar] [CrossRef]

23. Sutton, T.B.; Unrath, C.R. Evaluation of the tree-row-volume concept with density adjustments in relation to spray deposits in apple orchards. Plant Dis. 1984, 68, 480–484. [Google Scholar] [CrossRef]

24. Manktelow, D.W.L.; Praat, J.P. The Tree-Row-Volume Spraying System and its Potential use in New Zealand. In Proceedings of the NZ Plant Protection Conference, Lincoln, New Zealand, 18–21 August 1997; pp. 119–124. [Google Scholar] [CrossRef][Green Version]

25. Gil, E.; Escolà, A.; Rosell, J.; de Martí, S.P.; Val, L. Variable Rate Application of Plant Protection Products in Vineyard using Ultrasonic Sensors. Crop Prot. 2007, 26, 1287–1297. [Google Scholar] [CrossRef][Green Version]

26. Andújar, D.; Moreno, H.; Bengochea-Guevara, J.; De Castro, A.; Ribeiro, A. Aerial imagery or on-ground detection? An economic analysis for vineyard crops. Comput. Electron. Agric. 2019, 157, 351–358. [Google Scholar] [CrossRef]

27. Pallottino, F.; Menesatti, P.; Figorilli, S.; Antonucci, F.; Tomasone, R.; Colantoni, A.; Costa, C. Machine vision retrofit system for mechanical weed control in precision agriculture applications. Sustainability 2018, 10, 2209. [Google Scholar] [CrossRef][Green Version]

28. Fortunati, S.; Morea, D.; Mosconi, E.M. Circular economy and corporate social responsibility in the agricultural system: Cases study of the Italian agri-food industry. Agric. Econ. 2020, 66, 489–498. [Google Scholar] [CrossRef]

29. Torres-Sánchez, J.; Peña, J.M.; De Castro, A.I.; López-Granados, F. Multi-temporal mapping of the vegetation fraction in early-season wheat fields using images from UAV. Comput. Electron. Agric. 2014, 103, 104–113. [Google Scholar] [CrossRef]

30. Biasi, R.; Colantoni, A.; Ferrara, C.; Ranalli, F.; Salvati, L. In-between sprawl and fires: Long-term forest expansion and settlement dynamics at the wildland–urban interface in Rome, Italy. Int. J. Sustain. Dev. World Ecol. 2015, 22, 467–475. [Google Scholar] [CrossRef]

31. Matese, A.; Di Gennaro, S.F.; Berton, A. Assessment of a canopy height model (CHM) in a vineyard using UAV-based multispectral imaging. Int. J. Remote Sens. 2017, 38, 2150–2160. [Google Scholar] [CrossRef]

32. Matese, A.; Di Gennaro, S.F.; Santesteban, L.G. Methods to compare the spatial variability of UAV-based spectral and geometric information with ground autocorrelated data. A case of study for precision viticulture. Comput. Electron. Agric. 2019, 162, 931–940. [Google Scholar] [CrossRef]

33. Jurado, J.M.; Pádua, L.; Feito, F.R.; Sousa, J.J. Automatic Grapevine Trunk Detection on UAV-Based Point Cloud. Remote Sens. 2020, 12, 3043. [Google Scholar] [CrossRef]

34. Pijl, A.; Tosoni, M.; Roder, G.; Sofia, G.; Tarolli, P. Design of Terrace drainage networks using UAV-based high-resolution topographic data. Water 2019, 11, 814. [Google Scholar] [CrossRef][Green Version]

35. Sassu, A.; Gambella, F.; Ghiani, L.; Mercenaro, L.; Caria, M.; Pazzona, A.L. Advances in Unmanned Aerial System Remote Sensing for Precision Viticulture. Sensors 2021, 21, 956. [Google Scholar] [CrossRef] [PubMed]

36. Mathews, A.J.; Jensen, J.L.R. Visualizing and quantifying vineyard canopy LAI using an Unmanned Aerial Vehicle (UAV) collected high density structure from motion point cloud. Remote Sens. 2013, 5, 2164–2183. [Google Scholar] [CrossRef][Green Version]

37. Kalisperakis, I.; Stentoumis, C.; Grammatikopoulos, L.; Karantzalos, K. Leaf Area Index Estimation in Vineyards from UAV Hyperspectral Data, 2D Image Mosaics And 3D Canopy Surface Models. Remote Sens. Spat. Inf. Sci. 2015, XL-1/W4, 299–303. [Google Scholar] [CrossRef][Green Version]

38. Weiss, M.; Baret, F. Using 3D point clouds derived from UAV RGB imagery to describe vineyard 3D macro-structure. Remote Sens. 2017, 9, 111. [Google Scholar] [CrossRef][Green Version]

39. Gil, E.; Llorens, J.; Llop, J. Precision Viticulture: Use of New Technologies to Improve Efficiency in Spray Applications in Vineyard. In Proceedings of the Technology and Management to Increase the Efficiency in Sustainable Agricultural Systems, Rosario, Argentina, 1–4 September 2009. [Google Scholar]

40. Comba, L.; Biglia, A.; Aimonino, D.R.; Tortia, C.; Mania, E.; Guidoni, S.; Gay, P. Leaf Area Index evaluation in vineyards using 3D point clouds from UAV imagery. Precis. Agric. 2020, 21, 881–896. [Google Scholar] [CrossRef][Green Version]

41. Tsouros, D.C.; Bibi, S.; Sarigiannidis, P.G. A review on UAV-based applications for precision agriculture. Information 2019, 10, 349. [Google Scholar] [CrossRef][Green Version]

42. Tao, H.; Feng, H.; Xu, L.; Miao, M.; Long, H.; Yue, J.; Fan, L. Estimation of Crop Growth Parameters Using UAV-Based Hyperspectral Remote Sensing Data. Sensors 2020, 20, 1296. [Google Scholar] [CrossRef] [PubMed][Green Version]

43. Campos, J.; García-Ruíz, F.; Gil, E. Assessment of Vineyard Canopy Characteristics from Vigour Maps Obtained Using UAV and Satellite Imagery. Sensors 2021, 21, 2363. [Google Scholar] [CrossRef]

44. Bendig, J.; Bolten, A.; Bareth, G. UAV-based imaging for multi-temporal, very high resolution crop surface models to monitor crop growth variability. Photogramm. Fernerkund. Geoinf. 2013, 6, 551–562. [Google Scholar] [CrossRef]

45. Zarco-Tejada, P.J.; Diaz-Varela, R.; Angileri, V.; Loudjani, P. Tree height quantification using very high resolution imagery acquired from an unmanned aerial vehicle (UAV) and automatic 3D photo-reconstruction methods. Eur. J. Agron. 2014, 55, 89–99. [Google Scholar] [CrossRef]

46. Geipel, J.; Link, J.; Claupein, W. Combined spectral and spatial modeling of corn yield based on aerial images and crop surface models acquired with an unmanned aircraft system. Remote Sens. 2014, 6, 10335–10355. [Google Scholar] [CrossRef][Green Version]

47. Ballesteros, R.; Ortega, J.F.; Hernández, D.; Moreno, M.Á. Characterization of Vitis vinifera L. canopy using unmanned aerial vehicle-based remote sensing and photogrammetry techniques. Am. J. Enol. Vitic. 2015, 66, 120–129. [Google Scholar] [CrossRef]

48. Bajocco, S.; Salvati, L.; Ricotta, C. Land degradation versus fire: A spiral process? Prog. Phys. Geogr. 2011, 35, 3–18. [Google Scholar] [CrossRef]

49. Bajocco, S.; De Angelis, A.; Salvati, L. A satellite-based green index as a proxy for vegetation cover quality in a Mediterranean region. Ecol. Indic. 2012, 23, 578–587. [Google Scholar] [CrossRef]

50. Bajocco, S.; Ceccarelli, T.; Smiraglia, D.; Salvati, L.; Ricotta, C. Modeling the ecological niche of long-term land use changes: The role of biophysical factors. Ecol. Indic. 2016, 60, 231–236. [Google Scholar] [CrossRef]

51. Smiraglia, D.; Ceccarelli, T.; Bajocco, S.; Salvati, L.; Perini, L. Linking trajectories of land change, land degradation processes and ecosystem services. Environ. Res. 2016, 147, 590–600. [Google Scholar] [CrossRef] [PubMed]

52. Kazemzadeh-Zow, A.; Zanganeh Shahraki, S.; Salvati, L.; Samani, N.N. A spatial zoning approach to calibrate and validate urban growth models. Int. J. Geogr. Inf. Sci. 2017, 31, 763–782. [Google Scholar] [CrossRef]

53. Anifantis, A.S.; Camposeo, S.; Vivaldi, G.A.; Santoro, F.; Pascuzzi, S. Comparison of UAV Photogrammetry and 3D Modeling Techniques with Other Currently Used Methods for Estimation of the Tree Row Volume of a Super-High-Density Olive Orchard. Agriculture 2019, 9, 233. [Google Scholar] [CrossRef][Green Version]

54. Council of the European Union; European Parliament. Directive 2009/127/EC of the European Parliament and of the Council of 21 October 2009 amending Directive 2006/42/EC with regard to machinery for pesticide application (Text with EEA relevance). Off. J. Eur. Union 2009, L310, 29–33. [Google Scholar]

55. Council of the European Union; European Parliament. Directive 2009/128/EC of the European Parliament and Of the Council of 21 October 2009 establishing a framework for Community action to achieve the sustainable use of pesticides (Text with EEA relevance). Off. J. Eur. Union 2009, L309, 71–86. [Google Scholar]

56. Campos, I.; Nealeb, C.M.U.; Caleraa, A.; Balbontína, C.; Piquerasa, J.G. Assessing satellite-based basal crop coefficients for irrigated grapes (Vitis vinifera L.). Agric. Water Manag. 2010, 98, 45–54. [Google Scholar] [CrossRef]

57. López-Urrea, R.; Montoro, A.; Mañas, F.; López-Fuster, P.; Fereres, E. Evapotranspiration and crop coefficients from lysimeter measurements of mature Tempranillo wine grapes. Agric. Water Manag. 2012, 112, 13–20. [Google Scholar] [CrossRef]

58. Sanz, R.; Rosell, J.R.; Llorens, J.; Gil, E.; Planas, S. Relationship between tree row LIDAR-volume and leaf area density for fruit orchards and vineyards obtained with a LIDAR 3D Dynamic Measurement System. Agric. For. Meteorol. 2013, 171–172, 153–162. [Google Scholar] [CrossRef][Green Version]

Chapter 3 - In-field automatic detection of grape bunches under a totally uncontrolled environment

This is the published version of the manuscript in the Sensors (MDPI) journal: Ghiani, L.; Sassu, A.; Palumbo, F.; Mercenaro, L.; Gambella, F. In-Field Automatic Detection of Grape Bunches under a Totally Uncontrolled Environment. Sensors 2021, 21, 3908. https://doi.org/10.3390/s21113908.

Abstract

An early estimation of the exact number of fruits, flowers, and trees would help farmers to make better decisions on cultivation practices, plant disease prevention, and estimate harvest labor force size. The current practice of yield estimation based on manual counting of fruits or flowers by workers is a time consuming and expensive process and it is not feasible for large fields. Automatic yield estimation based on artificial intelligence and agricultural robots provides a viable solution in this regard. In a typical image classification process, the task is not only to specify the presence or absence of a given object on a specific location, while counting how many objects are present in the scene. The success of these tasks largely depends on the availability of a large amount of training samples. This paper presents a detector of grape bunches, based on a deep convolutional neural network trained to detect vine bunches directly on the field. Experimental results show a 91% mean average precision.

Keywords: deep learning; grape detection; object detection; precision agriculture; precision viticulture

1. Introduction

Precision agriculture evaluates spatial and temporal variability of field data through automatic collection and digitization of extensive information databases. Different types of sensors are applied to develop high-efficiency approaches to optimize input use, maximize crop production, reduce wastes, guarantee environmental sustainability, and obtain economic benefits [1,2,3,4]. These approaches apply to viticulture in terms of efficient use of inputs, such as fertilizers, water, chemicals, or organic products [5,6]. In this context, the use of machinery and labor for harvesting, pruning, or other crop management operations focuses on improving the efficiency of each plot within the vineyard [7,8,9,10,11].

Vineyards are characterized by high spatial and temporal heterogeneity and are influenced by pedo-morphological characteristics, climate, phenology, and cropping practices [12]. These variables can influence grape yields and quality, and their prediction is the main goal of precision viticulture. Farmers are encouraged to pursue the economic benefits and achieve the desired oenological results by the latest technologies combined with decision support techniques [13,14,15].

Emerging viticulture technologies are not fully developed, and several challenges still need to be addressed. While much of the work is currently promising, much effort is required to the so-called “vineyard of the future”. Viticulturists may therefore get advantage of modern tools to monitor and tailor the management of their vineyards. Data

in vineyard management include knowledge of the leaf area, fruit harvesting, yield estimation, evaluation of grapes quality, and grapevine cultivar identification [16,17]. Yield estimation is of critical importance in the wine industry. Traditionally, yield forecasts were based on manual counting of grapevine numbers, bunch numbers per vine, and manual and destructive sampling of bunches to determine their weights, berry size, and number.

When computer vision [18] and Machine Learning (ML) [19] are considered, object detection [20] represents a technique that deals with the detection of one or more categories of objects via the digitization of a given image or video. Object detection tasks can be roughly split into the object localization (where each object is located within the image) and object classification (which category each object belongs to). The location of the bounding box around each detected object will be returned in pixels as the x and y coordinates of the top left corner, and the width and height of the box.

As in many other applications of machine learning, in the last decade, Deep Learning (DL) [21] methods proved to be among the most effective in object detection [22,23]. Many different techniques have been implemented starting from R-CNN (Region Based Convolutional Neural Networks) [24], Fast R-CNN [25], and Faster R-CNN [26] up to, among many others, YOLO (You Only Look Once) [27] and Mask R-CNN [28]. Many of these techniques have been successfully applied in the agricultural field.

The aim of the work of Sa et al. [29] was to build a fruit detection system. Using transfer learning and fine-tuning techniques they were able to train a multi-modal Faster R-CNN model with a really limited number of images. They combined RGB (Red Green Blue) and NIR (Near InfraRed) information building in a reliable multi-modal system. The comparison with a Conditional Random Field with hand-crafted features method previously presented by the same team proved the validity of the approach. Bresilla et al. [30] trained a YOLO convolutional network for fruit detection and localization in images of apple and pear trees. Preliminary results were improved by some network modification, dataset augmentation, and the generation of synthetic images. The network was first trained to detect apples using apple trees images. The trained network was then “fine-tuned” with pear tree images to also detect pears. To estimate the biovolume of olive trees, Safonova et al. [31] used deep learning instance segmentation methods. They analyzed RGB images and two well-known normalized difference vegetation indexes. Several Mask R-CNN-based models were used for the segmentation of olive tree crowns and shadows to estimate the biovolume of individual trees. Fuentes et al. [32] proposed a robust DL-based detector for real-time tomato diseases and pests recognition.

Several experiments were conducted with an in-depth analysis of various deep learning architectures and feature extractors. Accuracy was further increased by data

augmentation techniques and the system was able to effectively recognize nine different types of diseases and pests. Picon et al. [33] presented several crop disease classification methods for mobile devices (Android, iOS, and Windows Phones) using a Deep Residual Neural Network with 50 layers and 224 × 224 input image size. They first extended an already existing dataset collecting leaves images in Spain and Germany. The leaves were labeled as healthy or affected by Rust, Septoria, and Tan Spot. Three different kinds of inputs were provided to the neural network: the resized full image, a leaf mask crop, and a superpixel based tile. Several data augmentation techniques were applied, and the training phase was repeated adding an artificial background to the images. Experimental results proved to be interesting with significant increases caused by the super pixel segmentation, the artificial background, and the image augmentation. To detect bunches of grapes or single berries, several methods have been proposed. Reis et al. [34] were able to detect red and white grapes, experimentally selecting a few intervals of RGB values by trial and error, collecting images during the night to limit light/brightness variations. After a sequence of iterations of the morphological dilation, the bunch regions were located and measured. Diago et al. [35] automatically estimated the number of flowers per inflorescence. The images of the inflorescences, with a uniform background of black color, were first converted from the RGB to the CIELAB color space (CIE L*a*b, where CIE stands for International Commission on Illumination in French), then segmented using thresholding based on histogram values. The elimination of local peaks (lower than a threshold) and a final post-processing filtering allowed to find and identify the brighter points corresponding to the flowers. An automatic system for shoot detection and yield estimation has been proposed by Liu et al. [36]. Images were converted from the RGB to the L*a*b color space and an Otsu thresholding technique [37] is used for the first segmentation step. An unsupervised feature selection followed by an unsupervised shoot classification using the K-means clustering algorithm leads to the shoot identification. Diago et al. [38] proposed a methodology to characterize the grapevine canopy and assess leaf area and yield through RGB images. They used the Mahalanobis distance to classify leaves (young or old), wood, grapes, and background. Font et al. [39] acquired images at night under controlled artificial illumination to simplify the grape segmentation procedure. They analyzed both the RGB and the Hue, Saturation, and Value (HSV) color spaces and segmented the images with five different methods: thresholding with the Otsu method followed by a sequence of morphological filtering; Mahalanobis distance between the three-dimensional color intensities; a Bayesian classifier; a Linear Color Model; a three-dimensional color-intensity histogram. A methodology for segmenting inflorescence grapevine flowers was presented by Aquino et al. [40]. They applied some morphological operators to the images in the HSV color spaces and a top-hat transformation to emphasize bright details. After binarization

and pyramidal decomposition, the regional peak corresponding to the inflorescence was found. An automated image analysis framework for berry size determination was proposed by Roscher et al. [41]. Working in the YIQ color space and after the detection of berry candidates with the circular Hough transform, they extracted several features from the image patches around the detected circles. Berry diameters were measured after using conditional random field to classify those patches as berry or non-berry. Another berry detection method using images converted to the CIELAB color space was proposed by Aquino et al. [42]. Images were acquired with dark cardboard placed behind the cluster. After an Otsu thresholding and some filtering, berries candidates were selected by finding regional maxima of illumination, and then six descriptors were extracted, and false positives were discarded using two different supervised-learning classifiers: Neural Network and Support Vector Machine. Liu and Whitty [43] eliminated irrelevant regions in the image by thresholding the H and V channels in the HSV color space obtaining potential bunch areas and reduced the noise by applying several morphological operations. The resulting bunches in 80 images were manually labeled as true or false and 54 different measures from RGB, HSV, and L*a*b color spaces were extracted. After applying the ReliefF algorithm and a sequential feature selection to reduce the feature dimensions, the SVM was used to train the system. Nuske et al. [44] predicted yields in vineyards through cameras and illumination mounted on a vehicle. They detect potential berry locations using a Radial Symmetry Transform and an Invariant Maximal Detector, then they use Gabor filters, a SIFT descriptor, and a Fast Retinal Descriptor to classify the detected points as grapes or not-grapes through a randomized KD-forest. To avoid double-counting of grapes between consecutive images, the grape locations were registered. A sequence of calibration measurements allows the team to predict yields with remarkable precision. That work was continued by Mirbod et al. [45] that used two algorithms (Angular invariant maximal detector and Sum of gradient estimator) for berry diameter estimation. Coviello et al. [46] introduced the Grape Berry Counting Network (GBCNet). It belongs to the family of Dilated CNNs and it is composed by ten pre-trained convolutional layers for feature extraction and by a dilated CNN for density map generation. The authors were able to estimate the number of berries in the image achieving good performances on two datasets, one with seven different varieties and one with only one variety. Finally, a more comprehensive review of computer vision, image processing, and ML techniques in viticulture has been proposed by Seng et al. [47].

Three main limitations characterize many of the works summarized in this section and schematized in table 1: the detection process is not fully automated, it is usually based on a limited amount of data (dozens or hundreds of images), and it is also based on a limited amount of grape variety (in most of the cases no more than two). Therefore, a

method applied on images acquired in a vineyard under specific conditions may not work as well in another vineyard or may not even work in the same vineyard as some of those conditions change.

Table 1. Proposed grape detection methods comparison.

Reference	Fully Automated Detection Process	Large Data Set (More Than a Thousand)	Large Grape Variety (More Than Ten)
[34]	Yes (by camera internal flash at night)	No (190 images of white grapes, 35 images of red grapes)	No (Port)
[35]	No (using a uniform background of black color)	No (90 images)	No (Tempranillo, Graciano, and Carignan)
[36]	Yes	Yes (thousands of images extracted from hundreds of videos)	No (Chardonnay and Shiraz)
[38]	Yes	No (70 images)	No (Tempranillo)
[39]	Yes (with artificial illumination at night)	No (40 images)	No (Flame Seedless)
[40]	No (capturing inflorescences facing the Sun and casting a shadow on the scene to create a homogeneous illumination)	No (40 images)	No (Airen, Albariño, Tempranillo, and Verdejo)
[41]	Yes	No (139 images)	No (Riesling, Pinot Blanc, Pinot Noir, and Dornfelder)
[42]	No (using a dark cardboard behind the cluster)	No (152 images)	Yes (Tempranillo, Semillon, Merlot, Grenache, Cabernet Sauvignon, Chenin Blanc, and Sauvignon Blanc)
[43]	Yes	No (160 images)	No (Shiraz and Cabernet Sauvignon)
[44]	Yes (with natural illumination, flash illumination, and cross-polarized flash illumination)	Yes (more than one thousand images)	No (Traminette, Riesling, Chardonnay, Petite Syrah, Pinot Noir, and Flame Seedless)
[45]	Yes	Yes (more or less 100,000 images)	No (Petite Syrah and Cabernet Sauvignon)
[47]	Yes	Yes (GrapeCS-ML dataset: more than 2000 images)	Yes (Merlot, Cabernet Sauvignon, Saint Macaire, Flame Seedless, Viognier, Ruby Seedless, Riesling, Muscat Hamburg, Purple Cornichon, Sultana, Sauvignon Blanc, Chardonnay, Shiraz, Pinot Noir)

This work	Yes	Yes (GrapeCS-ML dataset: more than 2000 images + Internal dataset: 451 images)	Yes (Merlot, Cabernet Sauvignon, Saint Macaire, Flame Seedless, Viognier, Ruby Seedless, Riesling, Muscat Hamburg, Purple Cornichon, Sultana, Sauvignon Blanc, Chardonnay, Shiraz, Pinot Noir, Vermentino, Cannonau (i.e., Granache), Cagnulari (i.e., Graciano))

One of the intentions of this paper was to overcome these problems by developing a grape detector to analyze images automatically acquired by a vehicle moving in a generic vineyard (located in an unspecified geographical area with an unknown grape variety). A detector, based on an R-CNN (Region Convolutional Neural Network), was trained and tested on the *GrapeCS-ML* dataset containing more than 2000 images of much different varieties described in the next section. An internal dataset was used to further test the framework on different grape varieties and under different environmental conditions.

2. Materials and Methods

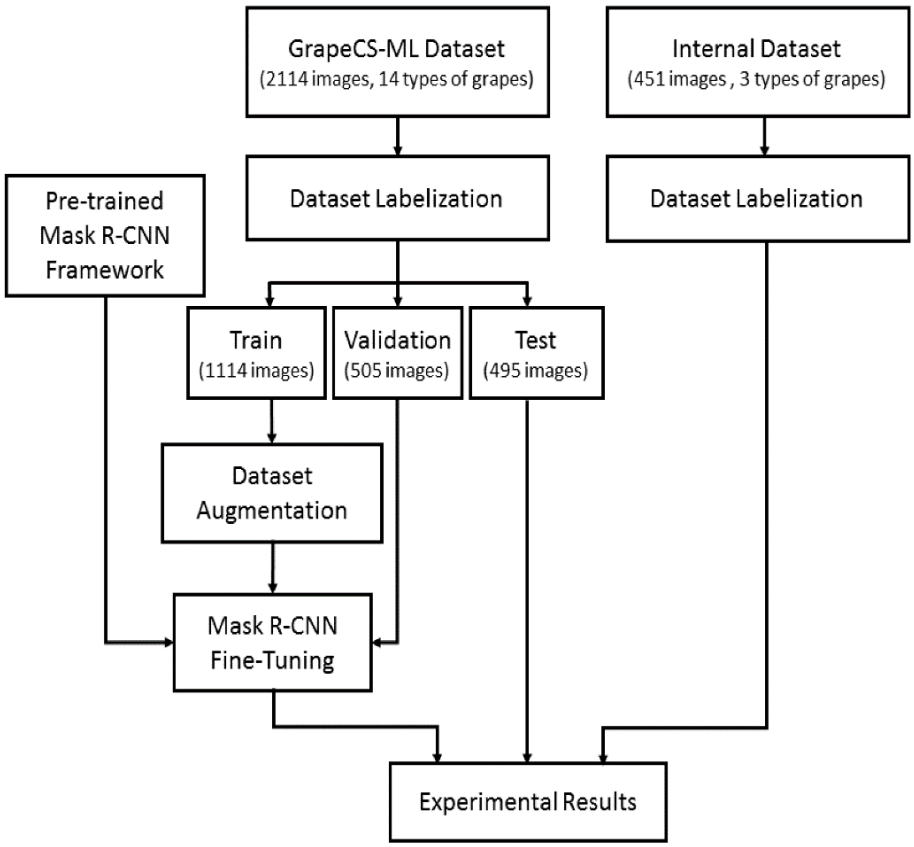


Figure 1. Detailed workflow of the proposed methodology. After the labeling, the data set is divided in train, validation, and test. A pre-trained Mask R-CNN framework is fine-tuned using the augmented train set and the validation set. The experimental results are obtained by applying the detector to both the test set and our internal dataset.

In this section, is fully described the proposed methodology summarized in Figure 1. The GrapeCS-ML dataset was labeled and divided in train, validation, and test subsets. Augmentation techniques [48] were applied to the training subset. A pre-trained Mask R-CNN framework was fine-tuned using the train and validation subsets, and from the trained network and the test subset were obtained the experimental results.

2.1. Dataset

The main difficulty in applying ML techniques in the agronomic field is the availability of useful data for training and testing. In 2018 the Charles Sturt University released the freely downloadable zip file GrapeCS-ML dataset [47], containing more than 2000 images of 15 grape varieties at different stages of development and collected in three Australian vineyards. The images are divided into five subsets:

- Set 1: *Merlot* cv. bunches, taken in seven rounds from the period January to April 2017.
- Set 2: Designed for research on berry and bunch volume and color as the grapes mature, featuring *Merlot*, *Cabernet Sauvignon*, *Saint Macaire*, *Flame* Seedless, *Viognier*, *Ruby* Seedless, *Riesling*, *Muscat Hamburg*, *Purple Cornichon*, *Sultana*, *Sauvignon Blanc*, and *Chardonnay* cvs.
- Set 3: Subsets for two cultivars (*Cabernet Sauvignon* and *Shiraz*) taken at dates close to maturity.
- Set 4: Subsets of images for two cultivars (*Pinot Noir* and *Merlot*) taken at dates close to maturity, focusing on color changes with the onset of ripening.
- Set 5: *Sauvignon Blanc* cv. bunches taken on three different dates. Each image also contains a hand-segmented region defining the boundaries of the grape bunch to serve as the ground truth for evaluating computer vision techniques such as image segmentation.

Although several subfolders contain some data such as the grape variety and the date of acquisition, a meaningful information is missing: the ground truth, i.e., the position of the bunches inside the different images. Therefore, the smallest Bounding Boxes were hand-drew around every bunch of grapes for each image uing the “Image Labeler” app (Figure 2) available in Matlab. As shown in Figure 2, the app enables the user to define a set of class labels (in our case just one class named “grape”) to draw a rectangle that is the Region of Interest (RoI) around each selected object and to label that ground truth as belonging to one of the previously defined classes.

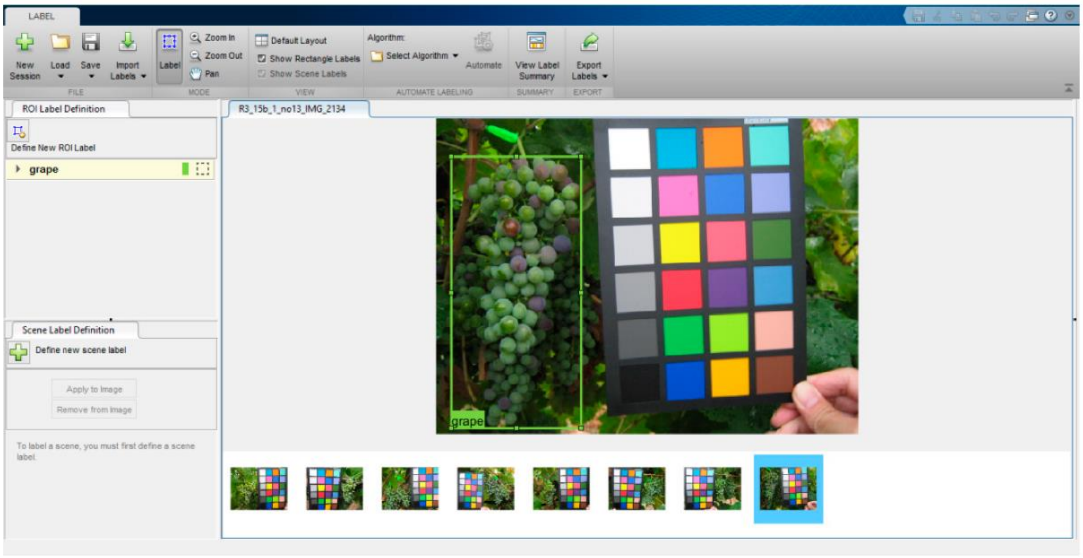


Figure 2. MATLAB Image Labeler used in the labeling process. For each image the smallest bounding box was hand drawn around every bunch of grapes.

A color reference or a volume reference is present in most of the images (a few examples are shown in Figure 3), but this information was ignored to obtain a fully automated detection process.

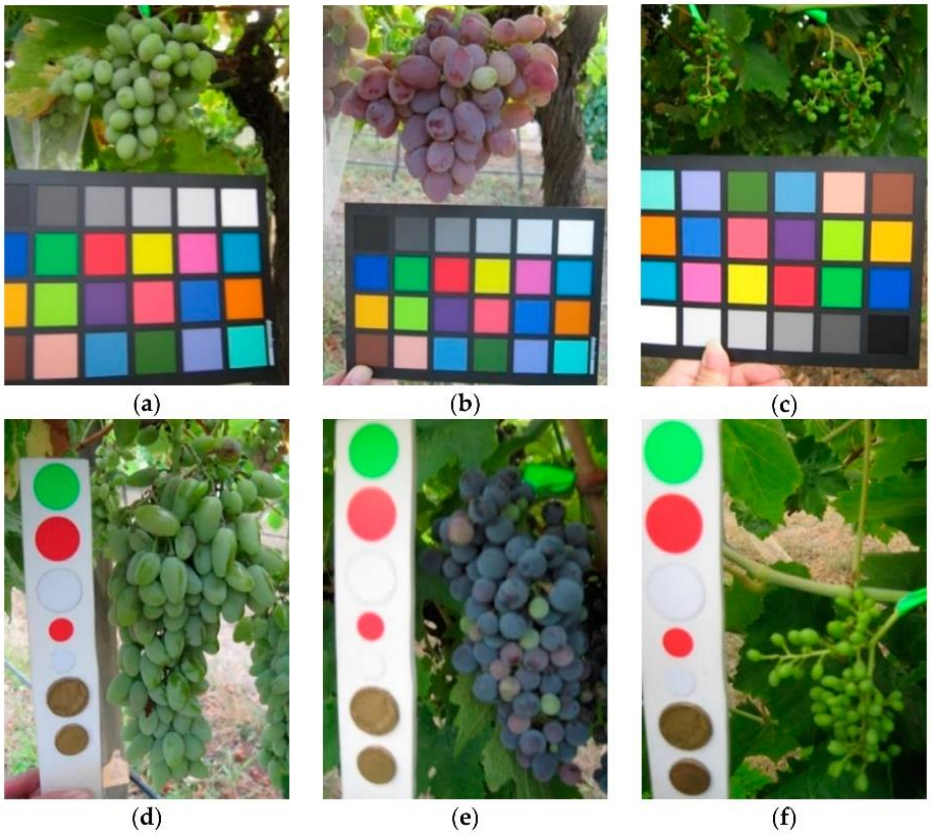


Figure 3. Samples images from GrapeCS-ML dataset 2: (a–c) include a color reference; (d–f) contain a volume reference.

During the last 15 years, thousands of digital images of bunches were collected at the Department of Agricultural Sciences, University of Sassari (a few examples are presented in Figure 4).

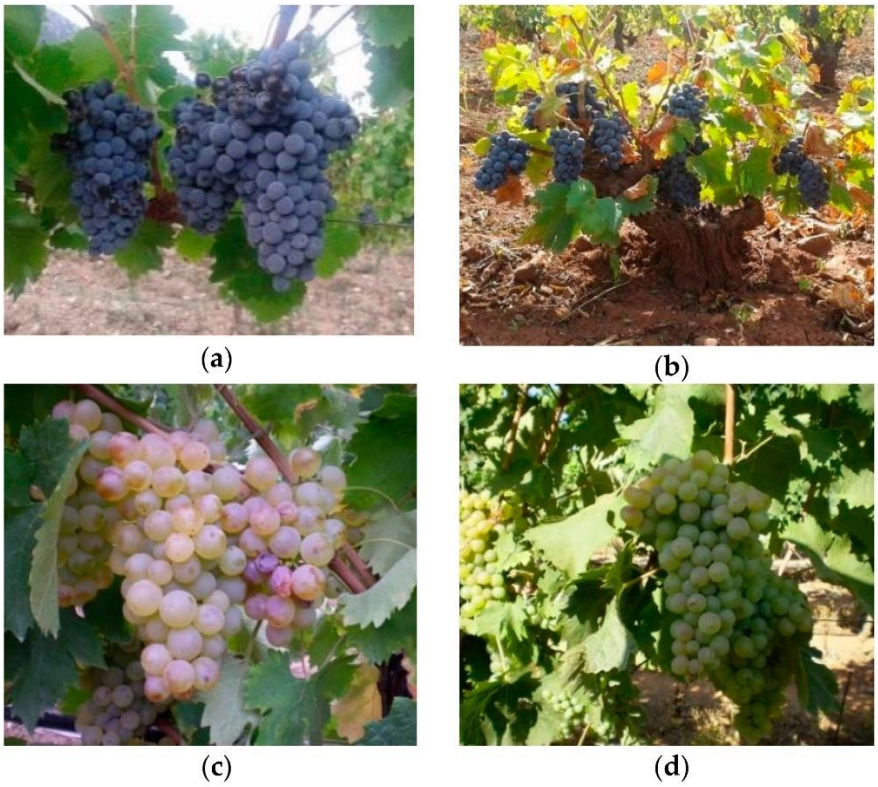


Figure 4. Samples images from our internal dataset: (a) cv. Cannonau; (b) cv. Cagnulari; (c,d) cv. Vermentino with different stage of maturation.

While all the GrapeCS-ML images of different grape varieties were collected in Australian vineyards, the ones in our dataset were collected all around in Sardinia Island (Italy), literally on the other side of the world. The number of available images were in the thousands, and they were acquired all around several Sardinian vineyards. Some contained the entire vineyard, others in perspective the space between two rows or an entire row imaged from one end. The purpose of our work was to train a detector able to analyze images automatically acquired by a vehicle moving between the vine rows. Therefore, photos acquired between the rows at about one meter from the leaf wall were only selected. A total of 451 images were selected to further test the trained network. It is worth emphasizing the importance of testing the system on a dataset that contains images like those used for the test. Moreover, it would be even more important to ascertain the ability of the system to provide good detection results on images very different from those present in the training set. In fact, while in the former case, a well performing detector on a specific vineyard could be obtained, in the latter the result would be a “universal” detector able to work anywhere.

2.2. Mask R-CNN Framework for Grape Detection

Given its performance on several well-known object detection benchmark datasets [22,23], It has been chosen to train our system with the Mask R-CNN method [28]. The Python implementation used in this work is freely downloadable from https://github.com/matterport/Mask_RCNN (accessed on 3 June 2021) [49].

The Mask R-CNN framework (Figure 5) segmentation is an extension of Faster R-CNN, and it adopts a two-stage procedure.

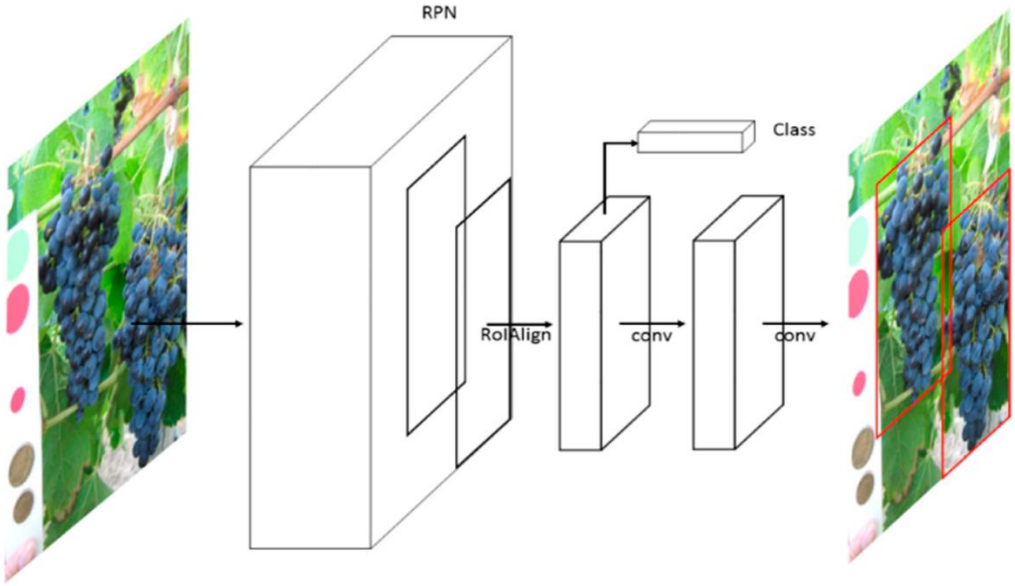


Figure 5. MaskR-CNN framework (He et al. [28]). In this two-stage procedure, the first stage, called Region Proposal Network (RPN), estimates the position of bounding boxes. The second stage performs a classification, a bounding box regression, and extracts a binary mask.

The first stage is called Region Proposal Network (RPN) and is a fully convolutional network. The RPN can be trained to predict region proposals at different scales and aspect ratios; therefore, it is used to estimate the position of bounding boxes. The second stage corrects the RoI misalignments in the RoIAlign layer and then performs in parallel a classification, a bounding box regression, and extracts a binary mask to output the bounding boxes and the segmentation masks of the selected object [28]. In this work, it was trained only the system to extract the bounding boxes values while ignoring the segmentation.

2.3. Training Procedure

It is well-known that deep learning training process requires a huge number of samples, hundreds of thousands, or even millions. In addition, training a model from scratch is tremendously expensive in terms of required computational power but also in terms of processing time. Luckily, the availability of a pre-trained model allows the execution of the so-called “fine-tuning”. In the fine-tuning process, a model trained on some huge

(millions of samples) dataset is “specialized” on different data and this further training requires much fewer resources.

In our case, the starting point was a ResNet101 (a convolutional neural network that is 101 layers deep) pre-trained on the MS COCO (Microsoft Common Objects in Context), a dataset containing hundreds of thousands of images belonging to 80 different classes [50]. Basically, a network trained to be able to detect objects belonging to the 80 different classes of the MS COCO has been retrained to specialize on the grape class. The availability of pre-trained weights for MS COCO simplify the first part of the training, since those weights can be used as a starting point to train a variation on the network. Google Colab was used, a cloud service that provides free access to computing resources including GPUs. The experiments were executed by a virtual machine with 12 GB of RAM, an Nvidia graphic card (Tesla P100-16GB), and 68 GB of disk space. It was performed fine-tuning (Goodfellow et al. [21]) using the GrapeCS-ML dataset images. The dataset was divided into a train (set 1, containing more than 1000 images), validation (set 2, containing more than 500 images), and test (sets 3, 4, and 5, containing nearly 500 images); see Table 2 for further details.

Table 2. Number of images contained in the GrapeCS-ML Dataset and in the internal dataset.

GrapeCS-ML Dataset		
Train	Set 1	1114 images
Validation	Set 2	505 images
	Set 3	204 images
Test	Set 4	242 images
	Set 5	49 images
Internal Dataset		451 images

The internal dataset collected at the University of Sassari contains 451 images from all around Sardinia. The photos collect images of clusters of the main cultivars grown on the island. Specifically, of the 451 photos, almost 200 are *Cannonau* and *Vermentino* cultivars. Every single photo represents a different biotype or clone obtained following two important experimental works on mass and clonal selection for cv. *Cannonau* [51] and varietal comparison for cv. *Vermentino* [52].

The other photos were collected mainly in collection vineyards of the University of Sassari where all the regional varieties registered in the national register of Italian vine varieties are grown [53].

Regarding the presence of different varieties, it has been pointed out the main difference with respect to similar works. The introduction of several different varieties will probably contribute to the generalization, but it is difficult to evaluate this contribution if examples of all the varieties are present in train, validation, and test at the same time. In our work

something totally different has been done, since there are notable differences, in terms of varieties, between train, validation, and test (in the Australian GrapeCS-ML dataset), and above all, a second test dataset was created with further different varieties (the Italian internal dataset).

Images dimensions in the first four sets of the GrapeCS-ML dataset are almost always 480 × 640 or 640 × 480. Conversely, images dimensions in the set 5 of GrapeCS-ML dataset and in the internal dataset vary a lot, from 480 × 640 up to 3024 × 4032 or 4608 × 3456 and many more (see Table 3). Since those sets are both used to test the system, consistent results could prove the robustness even towards considerable variations in size. To be processed by the Mask R-CNN framework, all the images are automatically resized to 1024 × 1024 pixels. The aspect ratio is preserved, so if an image is not square it is padded with zeros.

Table 3. Numerosity (in brackets) per different size of the images contained in the GrapeCS-ML dataset and in the internal dataset.

	GrapeCS-ML Dataset
Set 1	480 × 640 (1102), 640 × 480 (7), 1200 × 1600 (5)
Set 2	480 × 640 (253), 640 × 480 (198), 1200 × 1600 (28), 1600 × 1200 (26)
Set 3	480 × 640 (81), 640 × 480 (81), 1200 × 1600 (21), 1600 × 1200 (21)
Set 4	480 × 640 (35), 640 × 480 (206)
Set 5	640 × 480 (1), 3024 × 3024 (12), 3024 × 4032 (36), 3402 × 3752 (1)
Internal Dataset	360 × 640 (1), 480 × 640 (29), 640 × 480 (17), 1600 × 2128 (2), 1904 × 2528 (3), 2048 × 1536 (36), 2112 × 2816 (23), 2304 × 3072 (1), 2320 × 3088 (120), 2560 × 1536 (3), 2816 × 2112 (139), 3072 × 2304 (9), 3088 × 2320 (43), 3456 × 4608 (2), 4160 × 2340 (1), 4608 × 3456 (22)

To expand the size of the training part of the dataset, it was used a technique called “data augmentation” through which many modified versions of the images in the dataset are created by horizontally flipping, translating, and adding artificial blur and contrast (a few augmentation examples are shown in Figure 6). This technique allows to considerably extend the number of samples presented to the network during the training phase and, accordingly, to increase its detection and generalization capabilities. Moreover, variations in blurring, color, and brightness are a major problem in the field of computer vision. While other authors try to limit those variations as much as possible, on the contrary, in this work were included as many variations as possible in our training using dataset augmentation, so that the system “learns” to detect a grape bunch under as many as possible different conditions. It is worth noting that it was used only set 1 to train due to the highest numerosity; more than 1000 images which is half of the entire GrapeCS-ML Dataset. The training of the network with a single variety, which could

quickly lead to overtraining, is balanced using data augmentation and the high number of varieties present in the validation set.

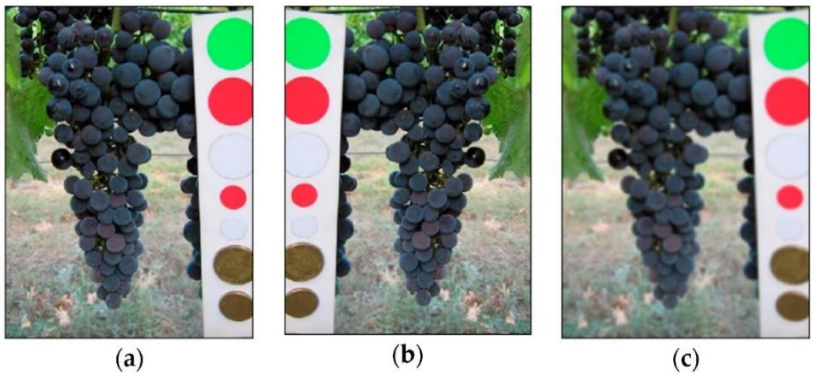


Figure 6. Examples of train dataset augmentation: (a) original image; (b) horizontal flipping; (c) image blurring.

3. Results

3.1. Performance Evaluation

Intersection over Union (IoU) measure (Equation 1) was used to evaluate the effectiveness of the proposed approach for bunches detection, which allows us to estimate the precision in the overlap between a bounding box obtained by the classifier and that defined as ground truth that is the one hand drawn during the 'labelling' process.

$$IoU = \frac{Ground\ Truth\ \cap\ Prediction}{Ground\ Truth \cup Prediction} \quad (1)$$

This measure is given by the ratio between the intersection and the union of the surfaces of the two bounding boxes (Figure 7), and it is positively evaluated if it exceeds a given threshold value (usually 0.5, but other values can also be considered [22]).

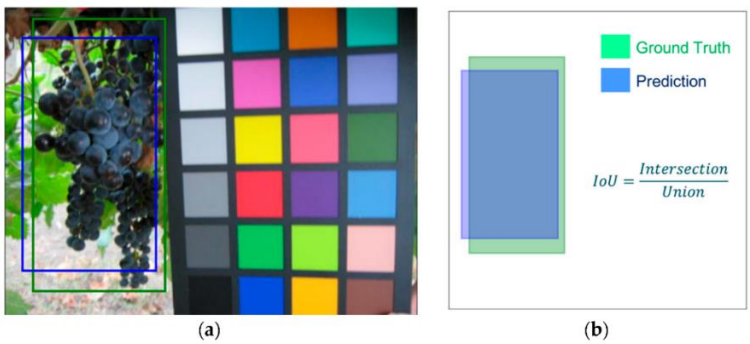


Figure 7. Evaluation of the IoU—Intersection over Union. This value is the ratio between the intersection and the union of the surfaces of the blue bounding box obtained by the classifier (Prediction) and the green one hand drawn during the 'labelling' process (Ground Truth). In (a) a sample image, in (b) a description of the calculation process.

In Figure 8, two examples of IoU are presented, one higher and the other lower than 0.5.

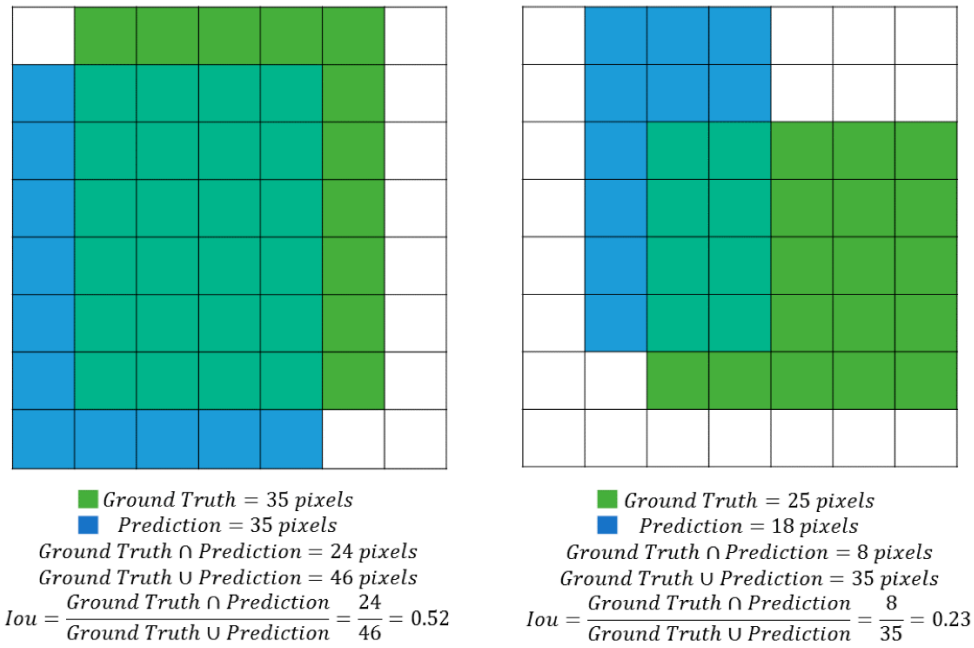


Figure 8. Two examples of IoU. In the example on the left the ratio between intersection and union of the ground truth and prediction bounding boxes is higher than 0.5 (0.52) while in the example on the right the ratio is lower (0.23).

The following values are defined:

- TP (True Positive): bounding boxes correctly detected (IoU > 50%).
- FP (False Positive): bounding boxes wrongly detected (there are no bunches or IoU < 50%).
- FN (False Negative): bounding boxes not detected where the bunches are present.

Precision (the ratio between the number of correctly detected bunches and the total number of objects detected as bunches in the image) and Recall (the ratio between the number of correctly detected bunches and the number of all the bunches present in the image) can therefore be calculated for each class as:

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

Since each bounding box is detected with a certain probability, values of this probability higher than a certain threshold represent the more probable grape's locations. As this threshold grows from 0.0 to 1.0, all possible Precision and Recall values are obtained.

These values can be used to plot, for each image, a curve with Precision as y value and Recall as x value called the Precision–Recall Curve. The most important points are those for which there is a value change for Precision or for Recall. The purpose of this procedure is the computation of the area below this curve that is called Average Precision (AP) and can be used as a measure of the detection performance on the image. In Figure 9, an example of the Precision–Recall curve obtained during our experiments is presented. The mean of all the obtained values is known as mean Average Precision (mAP) and is among the most used metrics in the field of object detection.

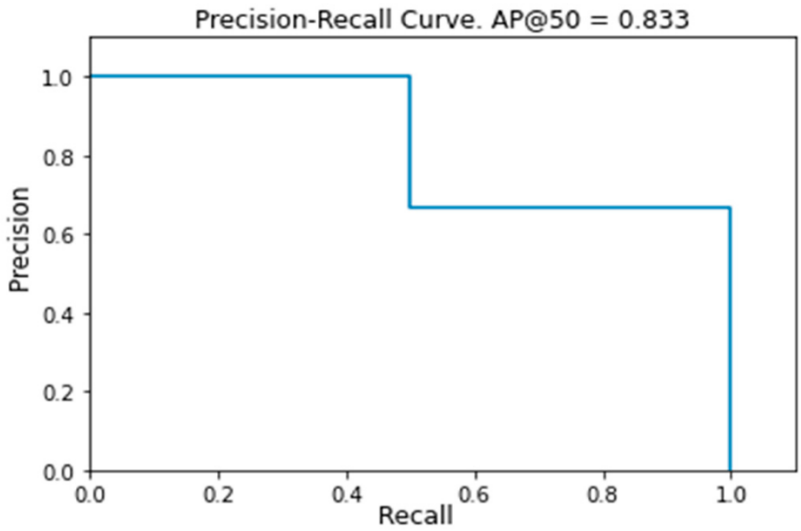


Figure 9. Example of Precision–Recall curve obtained during our experiments. The Average Precision, that is the area below the curve, has a value of 0.833. In this example there are three Precision or Recall value changes, but that number of changes could be different for each image.

3.2. Loss Function

An important step in the training of a model is the selection of a loss function to evaluate the network performances. The sum of losses obtained from the three different outputs of the Mask R-CNN framework was chosen among many possible values, as it represents the best compromise between the three different losses:

$$L = L_{cls} + L_{box} + L_{mask} \quad (4)$$

L_{cls} is the classification loss, L_{box} is the bounding-box loss, and Lmask is the mask loss as described in [28].

It is well known that during the training process, the validation loss is essential in choosing when to stop. As a matter of fact, if the training loss (evaluated on the train dataset) indicates how well the system is learning to perform the object detection on the training set (that is the already known data), the validation loss (evaluated on the

validation dataset) explains how much the system can generalize the detection capability on never seen data. Figure 10 shows the training and validation loss values obtained by our system. The number of epochs, which is the number of times the learning algorithm update the model by analyzing the entire training dataset, is used as a temporal scale.

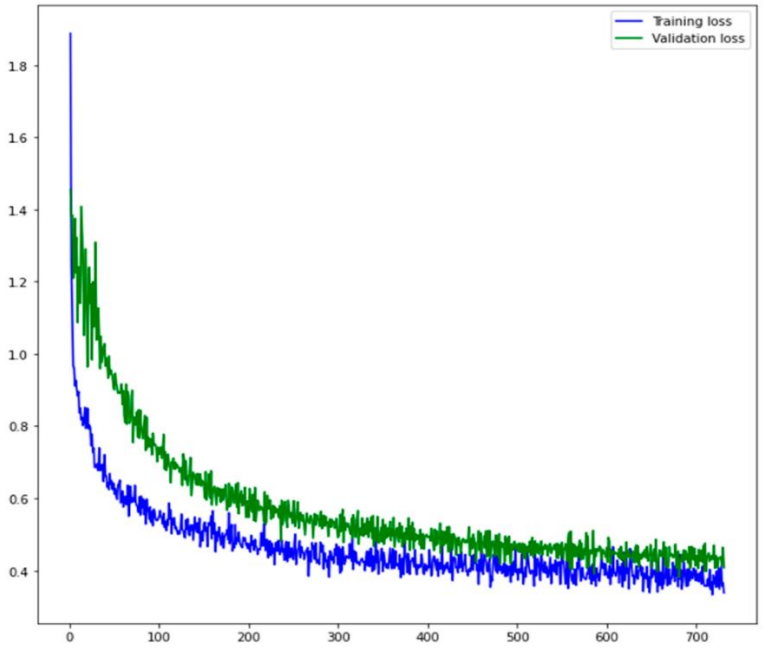


Figure 10. Training and validation loss profile over the number of epochs, which is the number of times the learning algorithm update the model by analyzing the entire training dataset. The two curves show the performance improvement on training and validation data.

3.3. Detection Results

The result obtained by applying the detector on the test samples was a mAP value of 92.78%. It means that a large majority of the bounding boxes have been correctly detected. In their works, Reis et al. [34] correctly identified 91% of white grapes and 97% of red grapes, Diago et al. [35] obtained a global Recall of 74.3% and a global Precision of 92.9% for flower detection in grapevine inflorescence, Liu and Whitty [43] detected bunches with an average accuracy of 88.0% and a recall of 91.6%, Aquino et al. [40] detected flowers with an average Precision of 0.8338 and a Recall of 0.8501, and Liu et al.'s [36] average detection performance was an Accuracy of 0.8683 and an F1 Score of 0.9004. Unfortunately, it is difficult to make a direct comparison among all those results and ours. Indeed, the evaluated metrics are not always the same and, most importantly, the experimental set-up used to retrieve the images is different (usage of different cameras and acquisition with different light conditions) as well as the vines varieties are different. Despite this, it can be observed that the obtained values are competitive with most of the works presented since, despite the different metrics, the recognition rate almost never exceeds 92%. The only case where the recognition rate seems higher is

when images are captured at night using the camera's internal flash with very little light/brightness variation [34]. A more reliable comparison can be made with the results obtained by Seng et al. [47] on the GrapeCS-ML Dataset although it is not clear which images were used for training, for validation, and for testing. By applying six different algorithms on four different color spaces, the highest classification rate they were able to achieve was 84.4% for white cultivars and 89.1% for red cultivars. In our work, there is no distinction between white and red cultivars, but with a mAP value of 92.78%, it can be claimed that our results are competitive with what is currently the state of the art.

The detailed results are shown in the second column of Table 4 (train complete, with augmentation). The validation and test values are very similar, proving the generalization capability of the system. Since the test dataset is composed of three subsets of the GrapeCS-ML Dataset, the mAP of each of them is presented. The considerable variation in the results is because the images in the three sets have very different characteristics. As shown in Figure 11, while in set 3 the bunches images are usually well defined and easy to detect, in set 4 and, even more, in set 5 there is a greater overlap between different bunches.

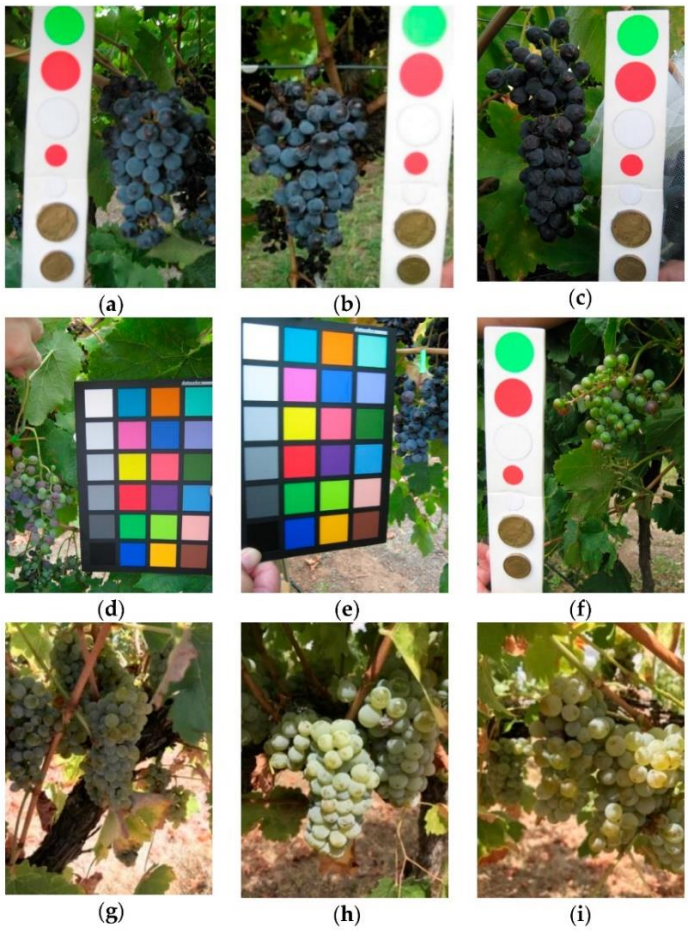


Figure 11. Images from the three GrapeCS-ML subsets included in the test: (a–c) set 3; (d–f) set 4; (g–i) set 5.

Furthermore, the prevalence of red grapes in set 3 makes the detection much easier compared to the detection of white grapes, more similar in color to the surrounding vegetation.

Table 4. Experimental results on both GrapeCS-ML and our internal dataset. The detector has been trained in three different ways: using the entire set 1 as train, with dataset augmentation; using only 10% of set 1 as a train, with and without dataset augmentation.

Dataset Name	mAP		
	Train Complete, with Augmentation	Train 10%, with Augmentation	Train 10%, without Augmentation
Validation (Set 2)	93.97%	90.95%	85.24%
Test (Set 3 + Set 4 + Set 5)	92.78%	90.98%	87.65%
Set 3	98.77%	98.69%	97.30%
Set 4	89.18%	86.70%	83.40%
Set 5	85.64%	80.07%	68.44%
Internal Dataset	89.90%	86.41%	70.75%

The system on our internal dataset was also tested to assess the generalization capability of the proposed framework. Since, concerning the GrapeCS-ML dataset, our images contain different grape varieties, different vegetation, and different colors, it would be important to replicate on our dataset results like those obtained with the original test. As shown in Table 4, it was obtained an 89.90% mAP that it is only slightly smaller than the other values.

Two more training were performed to determine the importance of the size of the dataset used for train and to determine the importance of the augmentation techniques. The described workflow was followed using only a reduced set of the original train (10% of the training images randomly selected) in one case with and in the other without the dataset augmentation. As it could have been expected, in the third and fourth column of Table 4 the mAP values decrease especially in the experiments performed without augmentation. The obtained results prove the importance of a high number of images in the train but also of the use of augmentation techniques. Most of the results show a decrease between 3% and 5%, passing from the value obtained with the complete train, to those with the reduced train, to those with the reduced train and without augmentation. Exceptions are the always very high values obtained for set 3 for which the decrease is limited to values around 1% and those on set 5 and the internal dataset which are much lower than the others. It is particularly important to highlight the different values obtained for the internal dataset: if the considerable reduction in the number of train images causes a limited reduction in performance (around 4%), the absence of augmentation leads to a drop in performance (more than 15%).

Since the overlap between bunches and the presence of smaller bunches is probably the factor that reduced the detection capability, the different sets were divided into subsets based on the number of ground truth objects in each image. The results presented in Table 5 as expected prove that the detection capability usually decreases as the number of objects in the image increases. It is worth noting that by analyzing the results from this different perspective, those that seemed evident differences between the various datasets are considerably reduced. Whatever the dataset, when there is just a single bunch in the image, the detection rate is always high.

Table 5. Experimental results on both GrapeCS-ML and our internal dataset based on the number of bunches present in the images. After each mAP value, in brackets, the number of examined images is shown.

Dataset Name	mAP (Total Number of Images)					
	1 bunch	2 bunches	3 bunches	4 bunches	5 bunches	6 bunches
Validation (Set 2)	98.85% (369)	82.61% (126)	57.22% (10)			
Test (Set 3, 4, 5)	99.75% (395)	65.41% (73)	72.59% (15)	51.72% (8)	60.00% (2)	64.63% (2)
Set 3	100.00% (195)	72.22% (9)				
Set 4	99.45% (181)	57.70% (53)	65.28% (8)			
Set 5	100.00% (19)	96.97% (11)	80.95% (7)	51.72% (8)	60.00% (2)	64.63% (2)
Internal Dataset	96.79% (218)	85.39% (166)	76.11% (46)	80.89% (17)	99.17% (4)	

4. Discussion

As stated before, the system is not error-free, since some bounding boxes are not detected at all, and others are not correctly detected, meaning that their IoU, with the ground truth, is lower than 0.5. In Figure 12, an example of correct detection on a test image is shown (the green boxes represent the ground truth, while the blue ones are the detection results) since the IoU is clearly greater than 0.5. Other examples (Figure 13), show some of the typical problems of object detection. In Figure 13a, only one out of two bounding boxes is correctly detected. In Figure 13b and 13c, the two bunches are detected but as a single element. This is one of the cases in which the error can be considered as “less severe”, since the area containing the bunches has been correctly detected. Unfortunately, when many bunches stay so close together inside the same image, they are difficult to distinguish. In Figure 13d, the picture is out of focus and only the larger of the two bunches has been correctly detected. These examples confirm the results presented in Table 5 since a single bunch is almost always correctly detected. Most of the errors are due to the presence of bunches that are too small and out of focus or to the inability to distinguish partially overlapping bunches.

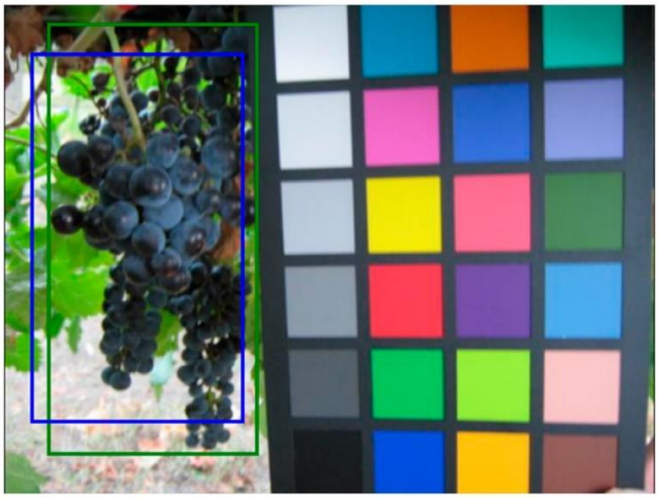


Figure 12. Example of correct detection on a test image from the GrapeCS-ML dataset. The green box represents the ground truth while the blue one is the detection results. The IoU of the two boxes is greater than 0.5.

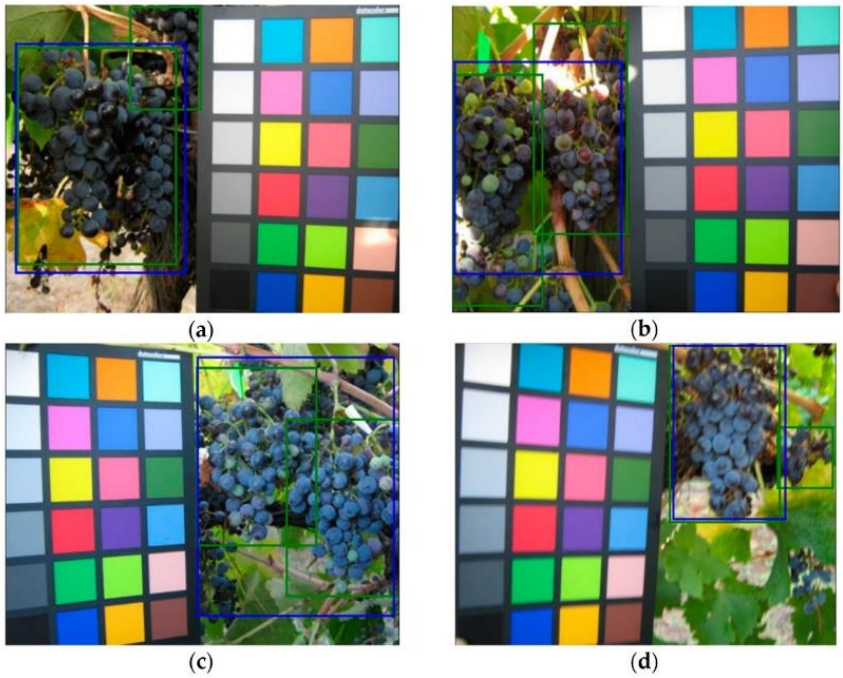


Figure 13. Examples of errors in the GrapeCS-ML dataset. The green boxes represent the ground truth while the blues ones are the detection results. In (a) only one out of two bounding boxes is correctly detected, in (b,c) the two bunches are detected but as a single element, in (d) only the larger of the two bunches is correctly detected.

Results obtained with the internal dataset can be considered excellent due to the considerable difference between the grape varieties images in this dataset, and those used to train the system. In Section 2, it was stated that our aim was the development of a grape detector able to analyze images automatically acquired in a generic vineyard. It could be claimed that those are the most important results presented in this work.

As in the previous cases, most of the errors are due to the incorrect detection of overlapping bunches (Figure 14a), others are caused by the inability to correctly detect shaded parts (upper-left box in Figure 14b). Another error is the incorrect detection of some leaves as bunches (Figure 14c) and it is probably due to the difference between the Sardinian grape varieties and those in the training dataset. In this case, the difference between the leaves in the image and all of those previously shown, belonging to the GrapeCS-ML dataset, is evident.

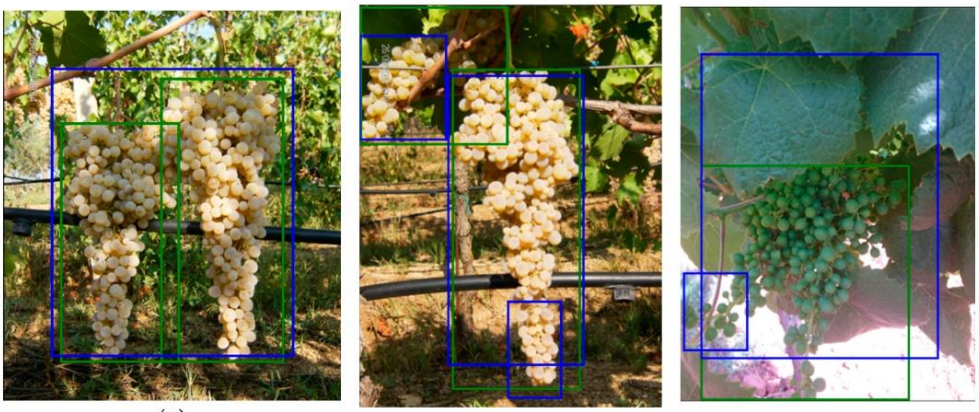


Figure 14. Examples of errors in the internal dataset. The green boxes represent the ground truth while the blues ones are the detection results. In (a) the incorrect detection of overlapping bunches, in (b) undetected shaded parts, and in (c) leaves incorrectly detected as bunches.

This type of error can be significantly reduced by training and testing the system with images collected in the same geographic area or, even better, in the same vineyard, but the focus of this work was, on the contrary, the analysis of a generalized capability of the framework. This capability is shown, for example, in the two images of the internal dataset in Figure 15, where the same bunches are depicted.

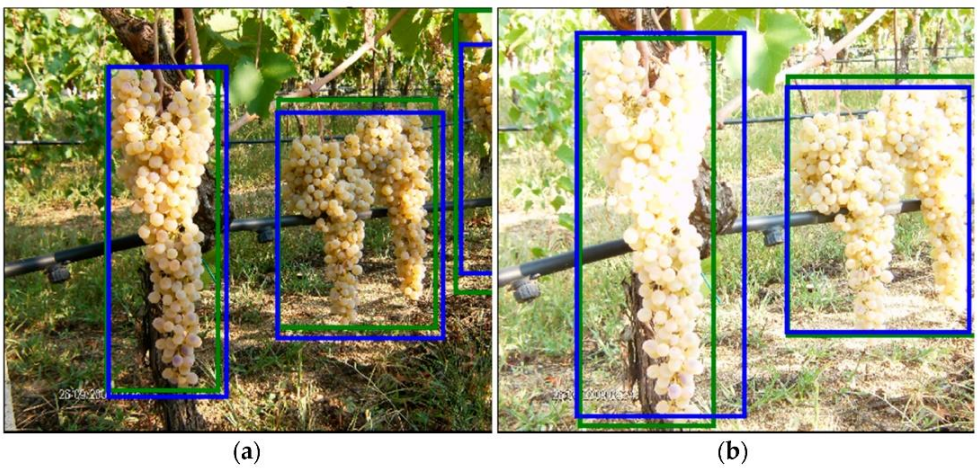


Figure 15. Example of same bunches correctly detected in two similar images. The image (a) is significantly overexposed compared to the image (b). The green boxes represent the ground truth while the blues ones are the detection results.

It is worth noting that, despite the image in Figure 15a is considerably overexposed, all the clusters have been correctly recognized as in Figure 15b. The ability to correctly detect grape bunches of varieties never seen before under uncontrolled lighting conditions is the main novelty of this work. Once again, it is worth emphasizing that, even in the most difficult cases, single bunches are almost always correctly detected as shown in Table 5.

Details of yield estimations involving traditional methods such as the lag phase method and others can be found in [54,55]. The quality assessment by visual inspection scales poorly to large vineyards and appears inaccurate due to the subjectivity of the human evaluator [56]. Moreover, these methods are expensive, inaccurate (if the yield is unevenly distributed across the vineyard), laborious, and time-consuming since they require a manual measurement of specific cluster features [57]. The precise knowledge of the number of bunches and their position in the vineyard would automate such activities.

5. Conclusions

In this paper, a detector of grape bunches based on the Mask R-CNN framework was presented. The GrapeCS-ML dataset was used to train the system and to evaluate its performances. The detector was tested on an internal dataset collected in several Sardinian vineyards during the last decade since the main goal was the training of a system capable of detecting bunches regardless of both the grape variety and its geographical location. The presented results are promising since most of the bunches were correctly detected and many of the errors were only due to the incorrect detection of two adjacent bunches as one. In fact, single bunches were usually correctly detected even in the most difficult cases, despite problems as shadowing and over exposition. It is worth noting that the importance of the presented methodology is that good results are obtained not only on the GrapeCS-ML database, which has been used to train the system but also on our internal dataset, confirming the portability to different scenarios. This is something novel at the state of the art, where methodologies are normally customized over a precise context of application and not proved to be portable. This approach is suitable to be employed, after appropriate training, in multiple scenarios of fruit detection and tracking from autonomous systems, reducing the subjectivity of the human evaluator during the visual quality assessment and optimizing monitoring operating times. Indeed, the achieved results represent valuable first results within the activities of the Comp4Drones (C4D) project. Starting from images' collection, it would be possible to more precisely monitor the development of the grapes, detect the diseases, estimate the yield in terms of quantity and quality, and predict the appropriate time for harvesting. In C4D the idea is to allow running those types of analysis both off-

line and on-line. In this regard, as future work, it is planned a new set of training with different frameworks, starting from the YOLO detector, which is known to be less precise but much faster in order of images analyzed per second with respect to the Mask R-CNN. Such a characteristic makes YOLO particularly suitable for the usage on embedded platforms, such as co-processing units acting as companion computers [58,59], which will allow advanced on-line processing on-boards of Unmanned Autonomous Ground Vehicles. Additionally, the teamwork plans also to collect a new set of images in the vineyards where field trials will be carried out in the coming years.

Author Contributions

Conceptualization, F.G.; methodology, L.G. and A.S.; software, L.G.; validation, L.G., F.G., L.M., and A.S.; formal analysis, L.G. and A.S.; investigation, F.G.; resources, L.G. and L.M.; data curation, L.G. and L.M.; writing—original draft preparation, L.G. and F.G.; writing—review and editing, F.G., L.G., A.S., and F.P.; visualization, A.S. and F.P.; supervision, F.G.; project administration, F.P.; funding acquisition, F.P. and F.G. All authors have read and agreed to the published version of the manuscript.

Funding

This research was funded by “Leveraging CPS Composability & Modularity for Customized and Autonomous Civilian Drones”, H2020-ECSEL-2018-2-RIA-two-stage. This project has received funding from the ECSEL Joint Undertaking (JU) under grant agreement No 826610”. This research was also funded by “Advanced Technologies for LANds management and Tools for Innovative Development of an EcoSustainable agriculture, ATLANTIDE”.

Data Availability Statement

The GrapeCS-ML database is openly available at doi:10.26189/5da7a8603c55c. The images of the Internal Dataset were collected by different teams and were granted to us for internal use only.

Acknowledgments

The Authors acknowledge the SME “Azienda Agricola CARPANTE”, Sassari, Sardinia, Italy, for the support and for the materials used for the experiments. This work is part of the activities of the ECSEL JU project “Comp4Drones (C4D)—Framework of Key Enabling Technologies for Safe and Autonomous Drones”.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Blackmore, S. Developing the Principles of Precision Farming. In Proceedings of the International Conference on Agropoles and Agro-Industrial Technological Parks (Agrotech 99), Barretos, Brazil, 15–19 November 1999. [Google Scholar]
2. Sudduth, K.A. Engineering technologies for precision farming. In Proceedings of the International Seminar on Agricultural Mechanization Technology for Precision Farming, Suwon, Korea, 27 May 1999; pp. 5–27. [Google Scholar]
3. Zhang, N.; Wang, M.; Wang, N. Precision agriculture—A worldwide overview. Comput. Electron. Agric. 2002, 36, 113–132. [Google Scholar] [CrossRef]
4. Lindblom, J.; Lundström, C.; Ljung, M.; Jonsson, A. Promoting sustainable intensification in precision agriculture: Review of decision support systems development and strategies. Precis. Agric. 2017, 18, 309–331. [Google Scholar] [CrossRef][Green Version]
5. Ghiani, L.; Sassu, A.; Lozano, V.; Brundu, G.; Piccirilli, D.; Gambella, F. Use of UAVs and Canopy Height Model Applied on a Time Scale in the Vineyard. In Proceedings of the International Mid-Term Conference of the Italian Association of Agricultural Engineering, Matera, Italy, 12–13 September 2019; pp. 837–844. [Google Scholar]
6. Ghiani, L.; Sassu, A.; Piccirilli, D.; Marcialis, G.L.; Gambella, F. Development of a matlab code for the evaluation of spray distribution with water-sensitive paper. In Proceedings of the International Mid-Term Conference of the Italian Association of Agricultural Engineering, Matera, Italy, 12–13 September 2019; pp. 845–853. [Google Scholar]
7. Pierce, F.J.; Nowak, P. Aspects of precision agriculture. In Advances in Agronomy; Elsevier: Amsterdam, The Netherlands, 1999; Volume 67, pp. 1–85. [Google Scholar]
8. Matese, A.; di Gennaro, S.F. Technology in precision viticulture: A state of the art review. Int. J. Wine Res. 2015, 7, 69–81. [Google Scholar] [CrossRef][Green Version]

9. Sartori, L.; Gambella, F. Comparison of mechanical and manual cane pruning operations on three varieties of grape (Cabernet Sauvignon, Merlot, and Prosecco) in Italy. Trans. ASABE 2014, 57, 701–707. [Google Scholar]

10. Guido, V.; Mercenaro, L.; Gambella, F. Application of proximal sensing in viticulture: Comparison of different berry state conditions. Chem. Eng. Trans. Open Access 2017, 58, 613–618. [Google Scholar]

11. Sassu, A.; Ghiani, L.; Pazzona, A.; Gambella, F. Development and Implementation of an Ultra-Low Volume (ULV) Spraying Equipment Installed on a Commercial UAV. In Proceedings of the International Mid-Term Conference of the Italian Association of Agricultural Engineering, Matera, Italy, 12–13 September 2019; pp. 563–571. [Google Scholar]

12. Smart, R.E. Principles of grapevine canopy microclimate manipulation with implications for yield and quality. A review. Am. J. Enol. Vitic. 1985, 36, 230–239. [Google Scholar]

13. Bramley, R.G.V.; Lamb, D.W. Making Sense of Vineyard Variability in Australia. In Proceedings of the International Symposium on Precision Viticulture, Ninth Latin American Congr. on Viticulture and Oenology; 2003; pp. 35–54. Available online: https://docplayer.net/21684245-Making-sense-of-vineyard-variability-in-australia.html (accessed on 5 June 2021).

14. Satorra, J.A.; Casasnovas, J.A.M.; Dasi, M.R.; Polo, J.R.R. Precision viticulture. Research topics, challenges and opportunities in site-specific vineyard management. Span. J. Agric. Res. 2009, 7, 779–790. [Google Scholar]

15. Sassu, A.; Gambella, F.; Ghiani, L.; Mercenaro, L.; Caria, M.; Pazzona, A.L. Advances in Unmanned Aerial System Remote Sensing for Precision Viticulture. Sensors 2021, 21, 956. [Google Scholar] [CrossRef] [PubMed]

16. Casasnovas, J.A.M.; Aymerich, X.B. Viticultura de precisión: Predicción de cosecha a partir de variables del cultivo e índices de vegetación. Rev. Teledetec. 2005, 24, 67–71. [Google Scholar]

17. Dunn, G.M.; Martin, S.R. Yield prediction from digital image analysis: A technique with potential for vineyard assessments prior to harvest. Aust. J. Grape Wine Res. 2004, 10, 196–198. [Google Scholar] [CrossRef]

18. Szeliski, R. Computer Vision: Algorithms and Applications; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2010. [Google Scholar]

19. Duda, R.O.; Hart, P.E.; Stork, D.G. Pattern Classification; John Wiley & Sons: Hoboken, NJ, USA, 2012. [Google Scholar]

20. Viola, P.; Jones, M. Robust real-time object detection. Int. J. Comput. Vis. 2001, 4, 4. [Google Scholar]

21. Goodfellow, I.; Bengio, Y.; Courville, A.; Bengio, Y. Deep Learning; MIT Press: Cambridge, MA, USA, 2016; Volume 1. [Google Scholar]

22. Zhao, Z.Q.; Zheng, P.; Xu, S.; Wu, X. Object detection with deep learning: A review. IEEE Trans. Neural Netw. Learn. Syst. 2019, 30, 3212–3232. [Google Scholar] [CrossRef][Green Version]

23. Jiao, L.; Zhang, F.; Liu, F.; Yang, S.; Li, L.; Feng, Z.; Qu, R. A survey of deep learning-based object detection. IEEE Access 2019, 7, 128837–128868. [Google Scholar] [CrossRef]

24. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587. [Google Scholar]

25. Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448. [Google Scholar]

26. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Mach. Intell. 2015, 39, 1137–1149. [Google Scholar] [CrossRef] [PubMed][Green Version]

27. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788. [Google Scholar]

28. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2961–2969. [Google Scholar]

29. Sa, I.; Ge, Z.; Dayoub, F.; Upcroft, B.; Perez, T.; McCool, E.C. Deepfruits: A fruit detection system using deep neural networks. Sensors 2016, 16, 1222. [Google Scholar] [CrossRef] [PubMed][Green Version]

30. Bresilla, K.; Perulli, G.D.; Boini, A.; Morandi, B.; Grappadelli, L.C.; Manfrini, E.L. Single-shot convolution neural networks for real-time fruit detection within the tree. Front. Plant Sci. 2019, 10, 611. [Google Scholar] [CrossRef][Green Version]

31. Safonova, A.; Guirado, E.; Maglinets, Y.; Alcaraz-Segura, D.; Tabik, E.S. Olive Tree Biovolume from UAV Multi-Resolution Image Segmentation with Mask R-CNN. Sensors 2021, 21, 1617. [Google Scholar] [CrossRef]

32. Fuentes, A.; Yoon, S.; Kim, S.C.; Park, E.D.S. A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition. Sensors 2017, 17, 2022. [Google Scholar] [CrossRef][Green Version]

33. Picon, A.; Alvarez-Gila, A.; Seitz, M.; Ortiz-Barredo, A.; Echazarra, J.; Johannes, E.A. Deep convolutional neural networks for mobile capture device-based crop disease classification in the wild. Comput. Electron. Agric. 2019, 161, 280–290. [Google Scholar] [CrossRef]

34. Reis, M.J.C.S.; Morais, R.; Peres, E.; Pereira, C.; Contente, O.; Soares, S.; Valente, A.; Baptista, J.; Ferreira, P.J.; Cruz, J.B. Automatic detection of bunches of grapes in natural environment from color images. J. Appl. Log. 2012, 10, 285–290. [Google Scholar] [CrossRef][Green Version]

35. Diago, M.P.; Sanz-Garcia, A.; Millan, B.; Blasco, J.; Tardaguila, J. Assessment of flower number per inflorescence in grapevine by image analysis under field conditions: Grapevine flower number per inflorescence by image analysis. J. Sci. Food Agric. 2014, 94, 1981–1987. [Google Scholar] [CrossRef]

36. Liu, S.; Cossell, S.; Tang, J.; Dunn, G.; Whitty, M. A computer vision system for early stage grape yield estimation based on shoot detection. Comput. Electron. Agric. 2017, 137, 88–101. [Google Scholar] [CrossRef]

37. Otsu, N. A threshold selection method from gray-level histograms. IEEE Trans. Syst. Man Cybern. 1979, 9, 62–66. [Google Scholar] [CrossRef][Green Version]

38. Diago, M.P.; Correa, C.; Millán, B.; Barreiro, P.; Valero, C.; Tardaguila, J. Grapevine Yield and Leaf Area Estimation Using Supervised Classification Methodology on RGB Images Taken under Field Conditions. Sensors 2012, 12, 16988–17006. [Google Scholar] [CrossRef][Green Version]

39. Font, D.; Tresanchez, M.; Martínez, D.; Moreno, J.; Clotet, E.; Palacín, J. Vineyard Yield Estimation Based on the Analysis of High Resolution Images Obtained with Artificial Illumination at Night. Sensors 2015, 15, 8284–8301. [Google Scholar] [CrossRef][Green Version]

40. Aquino, A.; Millan, B.; Gutiérrez, S.; Tardáguila, J. Grapevine flower estimation by applying artificial vision techniques on images with uncontrolled scene and multi-model analysis. Comput. Electron. Agric. 2015, 119, 92–104. [Google Scholar] [CrossRef]

41. Roscher, R.; Herzog, K.; Kunkel, A.; Kicherer, A.; Töpfer, R.; Förstner, W. Automated image analysis framework for high-throughput determination of grapevine berry sizes using conditional random fields. Comput. Electron. Agric. 2014, 100, 148–158. [Google Scholar] [CrossRef][Green Version]

42. Aquino, A.; Diago, M.P.; Millán, B.; Tardáguila, J. A new methodology for estimating the grapevine-berry number per cluster using image analysis. Biosyst. Eng. 2017, 156, 80–95. [Google Scholar] [CrossRef]

43. Liu, S.; Whitty, M. Automatic grape bunch detection in vineyards with an SVM classifier. J. Appl. Log. 2015, 13, 643–653. [Google Scholar] [CrossRef]

44. Nuske, S.; Wilshusen, K.; Achar, S.; Yoder, L.; Narasimhan, S.; Singh, S. Automated visual yield estimation in vineyards. J. Field Robot. 2014, 31, 837–860. [Google Scholar] [CrossRef]

45. Mirbod, O.; Yoder, L.; Nuske, S. Automated Measurement of Berry Size in Images. IFAC-PapersOnLine 2016, 49, 79–84. [Google Scholar] [CrossRef]

46. Coviello, L.; Cristoforetti, M.; Jurman, G.; Furlanello, C. GBCNet: In-Field Grape Berries Counting for Yield Estimation by Dilated CNNs. Appl. Sci. 2020, 10, 4870. [Google Scholar] [CrossRef]

47. Seng, K.P.; Ang, L.-M.; Schmidtke, L.M.; Rogiers, S.Y. Computer Vision and Machine Learning for Viticulture Technology. IEEE Access 2018, 6, 67494–67510. [Google Scholar] [CrossRef]

48. Shorten, C.; Khoshgoftaar, T.M. A survey on image data augmentation for deep learning. J. Big Data 2019, 6, 60. [Google Scholar] [CrossRef]

49. Abdulla, W. Mask R-CNN for Object Detection and Instance Segmentation on Keras and TensorFlow. Github. 2017. Available online: https://github.com/matterport/Mask_RCNN (accessed on 16 October 2020).

50. Lin, T.-Y.; Maire, M.; Belongie, S.; Bourdev, L.; Girshick, R.; Hays, J.; Perona, P.; Ramanan, D.; Zitnick, C.L.; Dollár, P. Microsoft COCO: Common Objects in Context. airXiv 2015, arXiv:1405.0312[cs]. [Google Scholar]

51. Mercenaro, L.; Usai, G.; Fadda, C.; Nieddu, G.; del Caro, A. Intra-varietal agronomical variability in Vitis vinifera L. cv. cannonau investigated by fluorescence, texture and colorimetric analysis. S. Afr. J. Enol. Vitic. 2016, 37, 67–78. [Google Scholar] [CrossRef]

52. Nieddu, G.; Chessa, I.; Mercenaro, L. Primary and secondary characterization of a Vermentino grape clones collection. In Proceedings of the 2006 First International Symposium on Environment Identities and Mediterranean Area, Corte-Ajaccio, France, 10–13 July 2006; pp. 517–521. [Google Scholar] [CrossRef][Green Version]

53. Mercenaro, L.; Oliveira, A.; Cocco, M.; Nieddu, E.G. Biodiversity of Sardinian grapevine collection: Agronomical and physiological characterization. Acta Hortic. 2017, 65–72. [Google Scholar] [CrossRef]

54. Komm, B.; Moyer, M. Vineyard Yield Estimation. 2015. Available online: https://research.libraries.wsu.edu:8443/xmlui/handle/2376/5265 (accessed on 16 October 2020).

55. Sabbatini, P.; Dami, I.; Howell, G.S. Predicting Harvest Yield in Juice and Wine Grape Vineyards. Ext. Bull. 2012, 3186, 12. [Google Scholar]

56. Nuske, S.; Achar, S.; Bates, T.; Narasimhan, S.; Singh, S. Yield estimation in vineyards by visual grape detection. In Proceedings of the 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems, San Francisco, CA, USA, 25–30 September 2011; pp. 2352–2358. [Google Scholar] [CrossRef][Green Version]

57. Palacios, F.; Diago, M.P.; Tardaguila, J. A Non-Invasive Method Based on Computer Vision for Grapevine Cluster Compactness Assessment Using a Mobile Sensing Platform under Field Conditions. Sensors 2019, 19, 3799. [Google Scholar] [CrossRef] [PubMed][Green Version]

58. Nezan, J.F.; Siret, N.; Wipliez, M.; Palumbo, F.; Raffo, L. Multi-purpose systems: A novel dataflow-based generation and mapping strategy. In Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS), Seoul, Korea, 20–23 May 2012; pp. 3073–3076. [Google Scholar]

59. Li, L.; Sau, C.; Fanni, T.; Li, J.; Viitanen, T.; Christophe, F.; Palumbo, F.; Raffo, L.; Huttunen, H.; Takala, J.; et al. An integrated hardware/software design methodology for signal processing systems. J. Syst. Archit. 2019, 93, 1–19. [Google Scholar] [CrossRef]

Chapter 4 - Artichoke deep learning detection network for site-specific agrochemicals UAS spraying

This is the under-revision version of the manuscript in the Computers and Electronics in Agriculture (Elsevier) journal: Sassu, A.; Motta, J.; Deidda, A.; Ghiani, L.; Alberto Carlevaro, A.; Garibotto, G.; Gambella, F. Artichoke Deep Learning Detection Network for Site-specific Agrochemicals UAS Spraying. Comput Electron Agric. Submited on 12 Oct. 2022.

Abstract

Input optimization is a distinguishing characteristic of Precision Agriculture approaches, helping reduce the environmental impact and costs and increase vegetable production quality. Thanks to the high automation evolution of Unmanned Aerial Systems (UAS), a new approach derived from their combination with Deep Learning techniques is leading to significant improvements in agricultural management practices. The study aims at artichoke plants detection and georeferencing as a first step for an on-the-fly, real time, UAS spraying system, and use the gathered information to monitor crop development through a multi-temporal approach. A commercial UAS, equipped with an RGB sensor, acquired images of the artichoke field located in Sardinia (Italy) during the 2021-2022 season in different crop growth stages. The Feature Pyramid Network (FPN), trained and compared with the YOLOv5 (You Only Look Once) network, showed a high detection level with an average F1 score of around 90%, and satisfactory off-line performances on the Nvidia Jetson Nano board. The multi-temporal approach influenced detection performances, with an inverse response of precision and recall metrics. The growing index trend showed a distinct value in October, peaking at the beginning of December as expected. The proposed approach contributes to designing future automatic and reliable site-specific UAS agrochemicals application and the classification of management zones.

Keywords. Single shot detector; Multi-temporal tracking; Plant detection; Site-specific management; Precision agriculture.

1. Introduction

Food sustainability and consumer protection are relevant issues today, as demonstrated by growing consumer interest in vegetable production and distribution on the market [1]. Because of the increasing food demands and the high impact of plant diseases on the global annual yield losses, chemical input in agriculture is still mandatory to protect crops against insects, pests, and fungi [2]. Agrochemical distribution is a dangerous operation with a high impact on consumers' safety and the environment. Often misapplied with considerable risks for consumers, agrochemical residues can be found in food, feed, water bodies, and non-target organisms [3].

Conventional spraying mechanization, deployed by ground machinery, is essential to reduce human and environmental harm and labor intensity. However, more effective and efficient application techniques are required to reduce the environmental impact of agrochemicals [4]. Agricultural aerial spraying by airplanes and helicopters, often considered an economical and rapid method for agrochemical application, is known for

covering large fields without any physical impact on crops or soil structure and causing high product overdose and losses due to poor distribution accuracy [5]. Unmanned aerial systems (UAS), as well as allowing the acquisition of images and data from a different perspective, have recently gained attention for pesticide spraying operations [6]. They can follow complex patterns, fly at low altitudes, adapt to different terrains, perform vertical take-offs and landings, and perform low-volume and site-specific agrochemical applications with low risks for operators' health [7]. Despite regulations and restrictions on aerial agrochemicals spraying, as is in Europe [8], multi-rotor UASs are under study for spraying applications worldwide, and they are the best candidates to replace conventional aerial vehicles.

The globe artichoke *Cynara cardunculus* L. var. *scolymus Fiori*, also known as *Spinoso Sardo*, is a Mediterranean native crop diffused in Sardinian Island (Italy) that strongly contributes to the agricultural economy of the region [9,10]. Artichoke plants are attacked by several insects and pests like aphids, thrips, leaf miners, etc., which require agrochemicals application, easily deployable through UASs [11]. UAS spraying operations planning, performed by defining target area borders, flight height Above Ground Level (AGL), speed, spray width, flow rate, etc., is easily applicable to cover crops like rice, corn, and wheat, but not to horticultural crops like artichoke. Site-specific spraying distribution, essential to reduce the amount of chemical product released on the soil surface, requires the coordinate references of each plant, obtainable by using an RTK GNSS station or indirectly by UAS images [12]. A fast and real-time approach is crucial to optimize UAS spraying operations, reduce the overall operation time, and execute accurate distributions over target plants.

In this scenario, the Deep Learning approach represents a valid and effective solution for real-time recognition and the consequent execution of a task [13]. Previous works have been carried out on the combination of UAS spraying technique and deep learning object detection in agricultural scenarios for an accurate real-time recognition system for spraying areas [14] or to determine pests' position in real-time on the orchard and plan the optimal pesticide spraying route for the agricultural UAS [15].

The Feature Pyramid Network (FPN) is a particular type of Single Shot Detector (SSD) [16]. These types of algorithms work in a single forward pass of the network, locating and classifying objects at the same time. The basic concepts of these networks imply the use of a grid that divides the image into cells responsible for detecting objects in that region of the image and the use of priors and predefined boxes responsible for detecting objects of specific sizes and shapes within a grid cell. In the FPN, it is possible to recall that the main structure of the architecture is composed of a bottom-up pathway for feature extraction and a top-down path for position detection on the image. The

combination of these two phases allows the network to detect objects of different scales with a good level of location precision in rapid training times.

In this work, the FPN performance was compared to a well know network, the YOLOv5. YOLO (You Only Look Once) is an algorithm for object detection developed in 2016 [17] based on regression: instead of selecting the part of the interest of an image, it predicts classes and bounding boxes in one run of the algorithm (for this, Once), so it belongs to the SSDs class as the custom FPN explained before. YOLOv5 is about 88% smaller than YOLOv4 (27 MB vs. 244 MB), 180% faster than YOLOv4 (140 fps vs. 50 fps), and it is roughly as accurate as YOLOv4 on the same task (0.895 mAP vs. 0.892 mAP). The main problem is that there is no official document for the YOLOv5 version, except the concept paper of YOLOv4 [18] and references therein.

Deep learning based networks are generally applied in agricultural scenarios for counting and detecting plants and plantation rows, crucial for plant health monitoring or plantation gaps identification after the seedling process [19]. Multi-temporal UAS imagery incorporation could significantly boost the accuracy and compensate for the low spectral resolution of RGB imagery [20]. Such approach could be applied to different crops for plant counting, crop health monitoring, yield estimation, and to plan optimized fertilizer, pesticide, and other input distribution within farm management [21].

The work aims at developing a machine learning approach for artichoke plant detection intended for real-time UAS spraying applications and an automatic multi-temporal tracking procedure for crop monitoring and UAS path planning development. In addition, to obtain a reliable model that can adapt to real-world applications and agricultural needs, the custom network was compared to the state-of-the-art YOLOv5 model.

2. Materials and Methods

2.1 Study site and survey date

Experiments took place in an Artichoke cultivation. (cv. *Spinoso sardo*) on a 3000 m2 surface in Uri, North-west Sardinia, Italy (Long. 8.472029, Lat. 40.623619; WGS84, EPSG 4326) at 125 m above sea level. Figure 1 shows an overall view of the artichoke field object of the study.

The surveys were performed following the phenological development of the culture with 2 weeks frequency in the first part of the growth and 1 month in the last phases, for a total of 7 days.

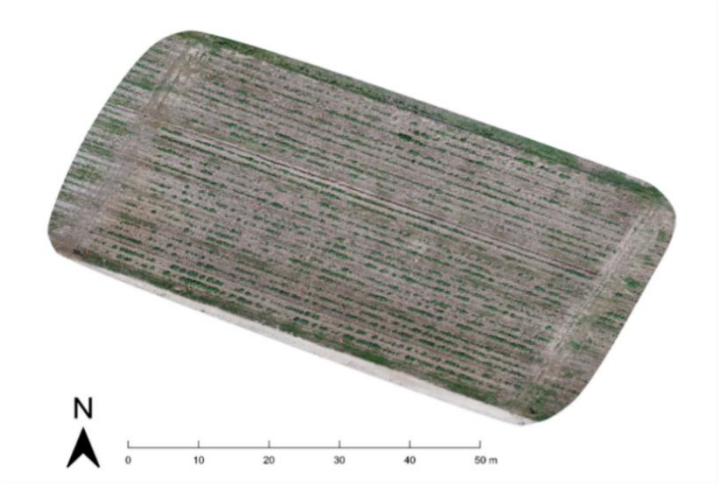


Figure 1. The artichoke field object of the study.

2.2 UAS platform and implemented sensors

Remote image acquisitions were performed by a DJI Phantom 4 Pro UAS equipped with RGB CMOS 1” sensor of 21 megapixels resolution, Field of View (FOV) 84°, 8.8 mm/24 mm (35 mm format equivalent), f/2.8-f/11 autofocus 1 m to ∞.

A RTK GNSS Reach RS+ (Emlid) connected to a NTRIP correction system was used to record the geographic coordinates of 12 Ground Control Points (GCPs), to obtain high accuracy orthomosaics and perform the temporal tracking process described in the next chapters.

2.3 UAS images acquisition campaign

During the 2021-2022 season, several images of the artichoke field in different growth stages were acquired by the DJI Phantom 4 Pro RGB sensor in nadiral position (perpendicular to the ground).

Automatic flights were performed using the android based DJI pilot app, able to guarantee the execution of standardized photos and videos acquisition by following a constant path at a specific height above ground level (AGL).

All flights for the orthomosaics creation were performed at 15m height AGL to obtain high quality images (232 photos for each flight). The speed was 1 m/s, the flight duration was 10 min 21 s, 70% side overlap ratio and 80% frontal overlap ratio. A flight of 80 m was performed to test the networks performances on a previous flight of the entire field.

The weather conditions were generally sunny and with clear sky; different lighting conditions in the photo-set were mainly due to the changing inclination of the incident radiation during the growing season.

2.4 Deep learning plant detection

Nowadays, it is increasingly easy to find predefined neural networks suitable for addressing various deep learning tasks, especially in object detection [22]. However, the

ability to build, train and test a custom neural network allows to better adapt the algorithm to the problem faced, giving more significance to the scientific work. With this aim, this section of the paper is devoted to explaining the network structure used in the detection phase, namely a custom FPN [23]. In the state of the art, there is a wide range of possible detection networks [24], each of which has its own strengths depending on the problem to be addressed. In our application scenario, which involves the use of a UAS capable of flying at different heights, the choice of the network had to consider its ability to detect objects at different scales. This is a peculiarity of FPN and the main reason this network was implemented for this work. The organization of the implementation (Figure 2) of the object detector was made according to the following scheme:

- Data preprocessing
- Network building
- Network training
  - FPN training
  - YOLOv5 training
- Network testing and performance evaluation
- Offline detection

where “offline detection” refers to the application of the trained model on a real scenario (i.e., on video and images collected by the UAS of the test area) but not in real-time.

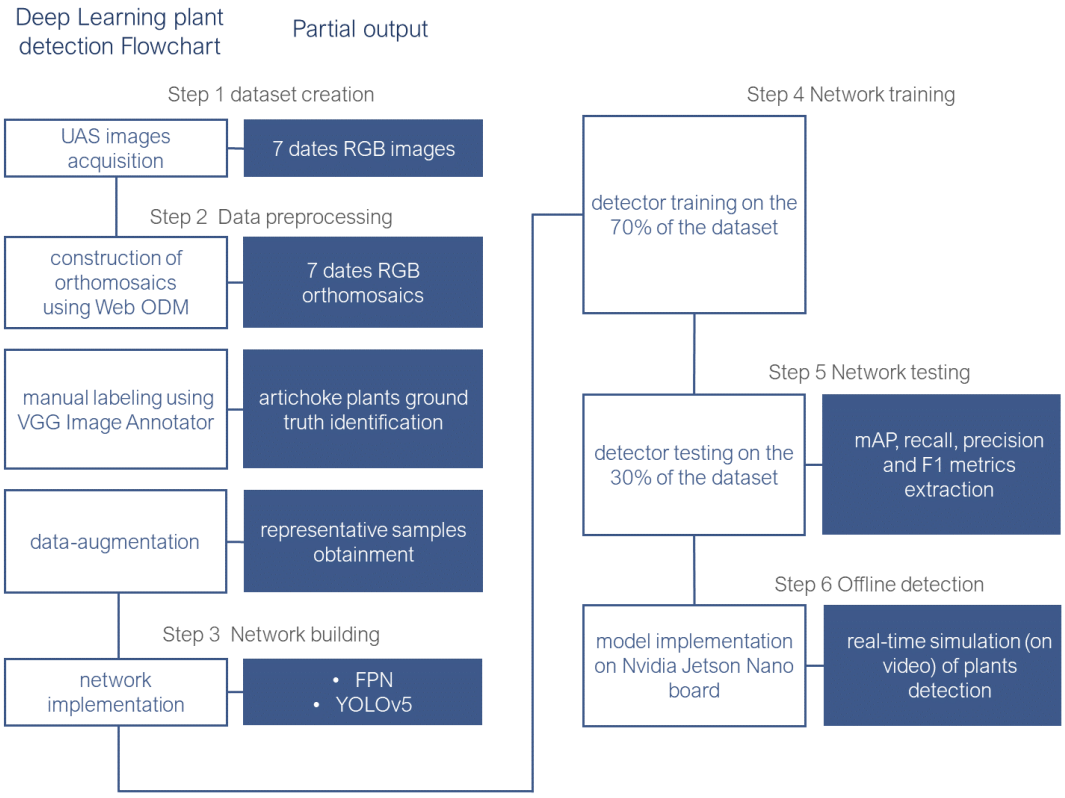


Figure 2: The deep learning plant detection flowchart (left gray color column) and the partial output of each step (right blue color column).

2.4.1 Data processing

Common to both networks was the preprocessing phase. Once the UAS acquisition phase was carried out, the images collected during the flight were merged to form a single orthorectified and high-resolution image called orthomosaic (Figure 3). The orthomosaics construction was made using OpenDroneMap, an application and API for UASs image processing capable of constructing an orthomosaics from a group of individual georeferenced images.

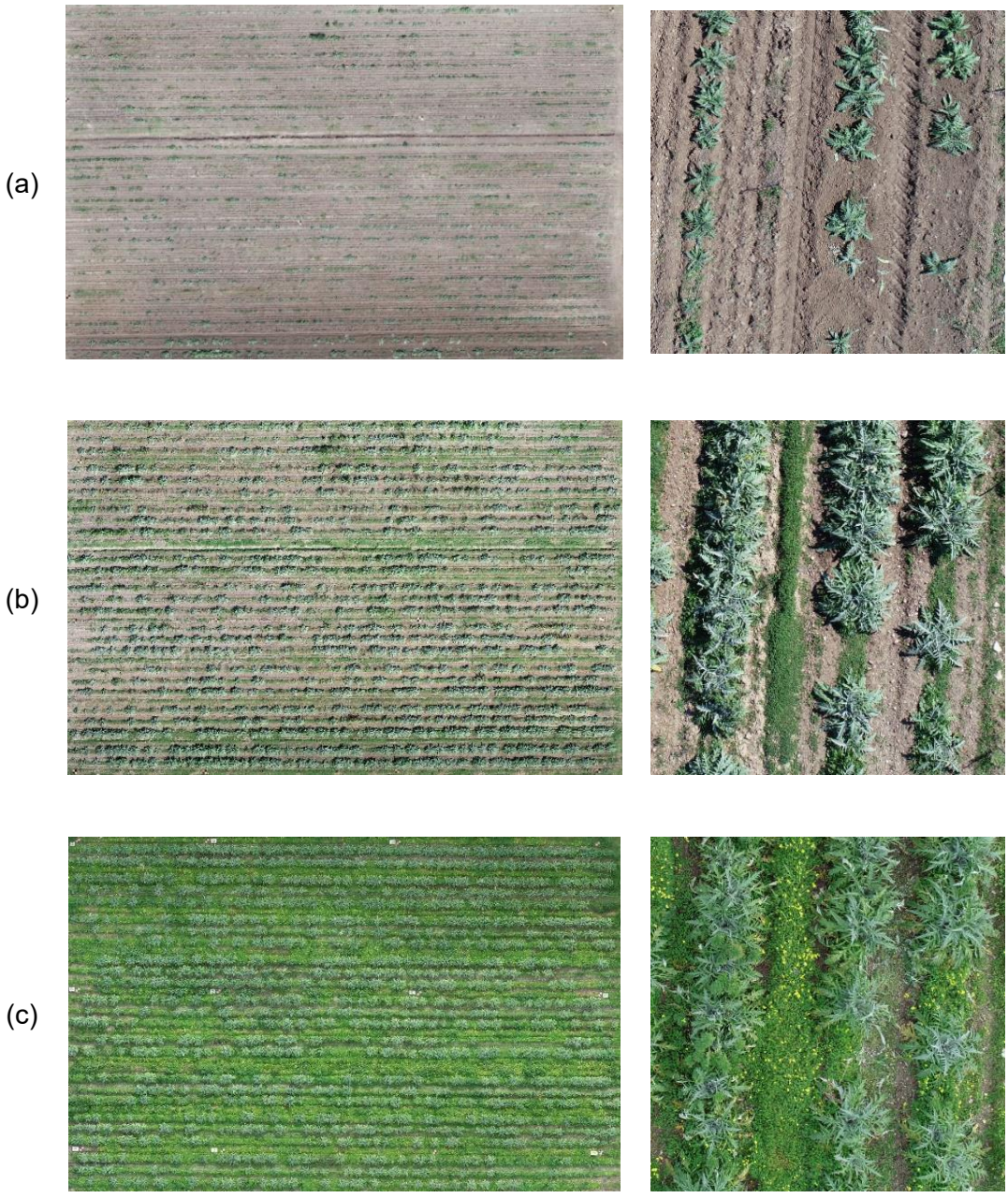


Figure 3. RGB orthomosaics (left column) and some artichoke plants details (right column) derived by 15 m flight altitude of three different surveying dates: 7 September 2021 (a), 15 October 2021 (b), and 23 November 2021 (c).

Seven orthomosaics corresponding to UAS flights in the months between September and December 2021 were generated, and an image dataset was extracted from each of them for the network training. After an initial phase of manual labeling performed using the software VGG Image Annotator (VIA) to obtain the ground truth of the data, the dataset generation was performed by randomly cropping orthomosaics and applying data-augmentation algorithms (rotation, blurring, saturation, etc.) to the obtained images to produce representative samples. The resulting dataset was then divided into training and test sets and provided as input to the detection network.

2.4.2 FPN building

Specifically, regarding the implementation of the network, the input parameters are as follow:

- Grid Sizes: (4×4, 8×8, 16×16) px^2
- Priors Sizes: (1×1, 16×9, 9×16) px^2
- Input Size: (512×512) px^2
- Total parameters: 2.8 M.

Where px refers to 1 pixel size (0.5 cm). Regarding the specifics of Python, 3.10.5 opencv-python 3.4.11.43, NumPy 1.21.2, SciPy 1.21.2, and matplotlib 3.4.3 were used.

2.4.3 FPN training

Once the network architecture was defined, the detector was trained on the datasets constructed in the preprocessing phase, dividing them into training and testing sets with a proportion of 70% and 30% respectively. The test set was used to get initial feedback on the network detection performance before applying the model for offline detection, as further explained in section 2.4.6. Recalling that the Location loss is the mismatch between the ground truth box and the predicted boundary box and that the Boxiness loss measures how confident the network is of the objectness of the computed bounding box, the specific training inputs, common to all the trainings, are as follows:

- Loss Function: Boxiness loss + Location loss.
- IoU: 50%
- Batch Size: 8
- Learning Rate: 1e-4
- Optimizer: Adam

Where IoU is the usual acronym for the Intersection over Union rate. By varying the dataset, the training time and the number of epochs are affected. Each new orthomosaic was merged with the training dataset. Therefore, the results obtained on the last dataset

are the most complete, the dataset having consisted of all previous orthomosaics. Generally, each network was trained for about 72-80 hours to obtain satisfactory results.

2.4.4 YOLOv5 training

Regarding our case of study, the smallest version of YOLOv5 (YOLOv5n) was also trained. Nano models maintain the YOLOv5 depth multiple of 0.33 but reduce the YOLOv5 width multiple from 0.50 to 0.25, resulting in ˜25% fewer parameters, from 7.5M to 1.9M, ideal for mobile and CPU solutions.

The training phase was conducted similarly to the customized FPN. Thus, there was a preprocessing phase in which data-images were obtained from the cropped orthomosaics of the various UAS flights (from September to December 2021) and then provided to the network for training. The training included 2000 epochs and the batch was 8.

2.4.5 Network performance testing and evaluation

For both networks, before applying the model for offline detection, a test was performed for each training phase to evaluate the network performance and to maximize the networks' detection capabilities.

As pointed out in the training phase of the network, 30% of the datasets were used to test the network, extracting the measures of mAP, recall, precision and F1 score that are the usual metrics adopted for the evaluation of a machine learning model [25]:

$$recall = \frac{TP}{TP + FN} \tag{1}$$

$$precision = \frac{TP}{TP + FP} \tag{2}$$

$$F1\ score = \frac{2TP}{2TP + FP + FN} \tag{3}$$

Where, as usual, TP, FP, and FN indicate the number of true positive (intended as the correctly detected artichoke plants), false positive (weeds or other objects incorrectly detected as artichokes), and false negative (the undetected artichoke plants) respectively. Recall (sometimes called sensitivity) is a measure of the detection efficiency of the network to minimize the number of missed objects. Precision is a measure of the network accuracy to achieve the minimum number of detection error. F1 score is the weighted average of Recall and Precision including both false positives and false negatives. This first testing phase proved how essential is to achieve the best

results in the offline detection phase, allowing the network parameters to be set in the most suitable way. After the network achieved the desired performance, it was applied to off-line orthomosaic analysis.

2.4.6 Offline detection

The step prior to the application of the model in a real-time scenario is the evaluation of its offline behavior. Strictly speaking, the quality of detection in the offline phase should be almost the same in the real-time application, since what the network processes are always videos and images, whether they have been collected and stored beforehand or transmitted directly from a camera. In fact, the Nvidia Jetson Nano board was selected, a small and powerful computer for embedded and AI IoT applications, to install the trained and tested networks for the evaluation phase.
In the operational phase, i.e., in online detection with the UAS, only YOLOv5 will be loaded on the Jetson: this is because YOLOv5 is a lighter network than the custom feature pyramid, and to achieve the best results in detection, hardware performance (power consumption, storage memory, etc.) must also be considered. All the performance of YOLOv5 on Jetson has been reported in the Section 3.3.

2.4.7 Temporal tracking

To correctly reconstruct the complete time history of each plant, a temporal tracking algorithm has been developed. The first step of this process is a spatial registration of the consecutive pairs of orthomosaics [26].

Figure 4. One of the 12 GCPs used for image referencing.

The ground control points (GCP) (e.g., in Figure 4) have been placed on the ground by recording their georeferenced position (longitude, latitude, altitude), according to WGS84 (EPSG 4326) geographic projection model. As such, all the centers of gravity of the DL detected boxes can be always remapped in world coordinates with an affine transformation, to support optimal UAS-sprayer mission planning before, in real-time crop operations.

The adopted solution for orthomosaics registration at different recording times is an automatic registration process, based on the information provided by the box-plants detected by the Neural Network process. Such automatic registration is implemented in two following steps:

- search for the overall translation (dx, dy) that maximizes the IoU between the boxes detected in the two consecutive orthomosaic images.
- selection of the boxes with IoU above a predetermined threshold (> 50%) and computation of the relative homographic transform between the centers of mass.

The mean square deviation of the homographic transformation is varying between 9 and 12 cm, which roughly corresponds to the actual spread of the center of gravity positions of the matched boxes, in two consecutive time frames.

The automatic registration allows to perform a spatial prediction between the coordinates of the orthomosaic at time t0 (past) with those at time t1 (next) and vice versa. It represents the basis of the time tracking process [27], which, in turn, is implemented in 2 phases:

- tracking forward: for each box detected at time t0, the best match is searched (in terms of max IoU) among all the boxes detected at time t1; if an acceptable match is not found (IoU threshold), a new hypothesis (prediction) is generated and added to the list of the boxes available at time t1, thus ensuring the propagation and continuity of the current track.
- tracking backward: when the last available orthomosaic data has been reached, the process is repeated in reverse, generating backward predictions for all the boxes that do not have yet connections with the previous stages of the crop.

The result of this tracking process is a series of complete traces, from the first orthomosaic image recorded, up to the last available, for all the box-plants that have been detected by the Neural Network. The total number of traces inevitably includes some errors that can be classified as:

- missed-box-plants, mainly due to low image contrast, interference with other elements of the scene, etc.
- localization and size errors, with low IoU values, as compared to the truth.

- new-phantom-boxes, which appear in areas where there are no plants, often due to the presence of weeds, scattered leaves, etc.

With time tracking it is possible to correct most of these errors and obtain better performance levels (recall and precision) as compared to the analysis at each individual stage of the crop.

2.4.8 Artichoke crop field analysis

The output list of the box-plants from the temporal integration is re-organized by ordering them both in vertical and horizontal positions along the plant rows in the field. Figure 5 shows a subset of the crop field (the full size is 14112 × 9072 px), with the overlap of the detected and time-tracked box-plants.

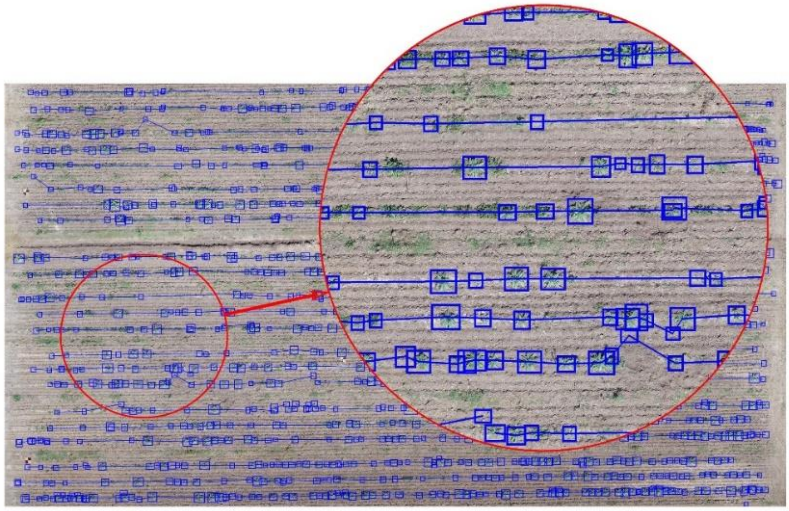


Figure 5. A partial view of the RGB orthomosaic, with the box-plants detected and tracked over time-space (blue) and connected along each row of the crop.

In a real-time application, on-board of the UAS, to drive the direction of the sprayer towards the actual position of the plants, the best network configuration should exhibit a high recall rate, to minimize the number of missed targets; some additional detection errors (lower precision) will be acceptable in this case. Given the potentially real-world application, it is worth noting how much the detected boxes differ from the ground truth: minimal variations on detection can drastically affect the UAS's spraying positioning over the target plant and the operation planning, so the deviation between the centers of the detected boxes and their respective truth boxes to apply appropriate countermeasures was estimated. The L^2 norm was identified (the Euclidean distance) as the most representative measure for estimating the distance between the centers of the boxes, since it considers deviations along all directions of the plane. Calling $c_i^D = \left(c_{x,i}^D, c_{y,i}^D\right)$ and with $c_i^T = \left(c_{x,i}^T, c_{y,i}^T\right)$ the centers of the i-th detected box and ground-truth box,

respectively, it was calculated the distance d_i for all pairs of detected and ground-truth boxes as:

$$d_i = \sqrt{(c_i^D)^2 - (c_i^T)^2} = \sqrt{(c_{x,i}^D - c_{x,i}^T)^2 + (c_{y,i}^D - c_{y,i}^T)^2} \quad (4)$$

From this data representation it is possible to achieve an automatic segmentation of each plant row with corresponding parameters (number of plants/rows, vegetation-mass index and density of the plants) which provide a clear view of the health status of the crop.
Another output of the multitemporal analysis was the Growing index related to the seasonal development of the crop size. It is computed as the ratio between the average size (width and height) of the bounding box of the plants detected at the different times of the experiment:

$$GI_i = \frac{1}{N}\sum_{i=1}^{N} \frac{w_j^i + h_j^i}{w_j^{i-1} + h_j^{i-1}}, i = 2, \dots, 7 \quad (5)$$

Where w_j^i and h_j^i are, respectively, the width and the height of the *j* bounding boxes of the orthomosaic *i* , and *N* is the number of the bounding boxes in the orthomosaic, which, thanks to the temporal tracking algorithm, is the same for each orthomosaic (*N* = 1419 for FP network and *N* = 1351 for YOLOv5 network).

3. Results

3.1 Deep learning plant detection

3.1.1 FPN

The evaluation phase of the FPN was conducted testing the network on the portion of the dataset not used for training. Table 1 shows the performances of the network in the detection of the artichoke plants. TP is the number of the boxes correctly identified by the network, with an overlap measure IoU of more than 50% over the ground truth boxes that were selected during the manual annotation process. FP is the number of boxes detected by the network which do not correspond to ground truth boxes, or the IoU is lower than the 50% threshold. FN is the number of ground truth boxes (in the annotation list) that have been missed by the detection network. Refer to Section 2.4.5 for the definition of precision, recall and F_1 score. Table 1 and Table 2 compare the performance measures obtained before and after the Forward-Backward tracking process.

Table 1. Performance measure of the FPN Detection at each individual date of the test.

Date	TP	FP	FN	Precision	Recall	F1score
09_07	474	11	72	0.977	0.868	0.919
09_14	533	22	53	0.96	0.91	0.934
10_01	541	39	66	0.933	0.891	0.912
10_15	544	18	57	0.968	0.905	0.936
11_09	460	32	141	0.935	0.765	0.842
12_03	450	73	125	0.86	0.783	0.82
12_23	509	71	95	0.878	0.843	0.86

Table 2. Performance measures of the FPN network after multi-temporal tracking.

Date	TP	FP	FN	Precision	Recall	F1score
09_07	487	146	59	0.76	0.89	0.83
09_14	531	95	55	0.84	0.91	0.88
10_01	555	102	52	0.85	0.91	0.88
10_15	550	82	51	0.87	0.92	0.89
11_09	517	124	84	0.8	0.86	0.83
12_03	499	121	76	0.81	0.87	0.84
12_23	519	106	85	0.83	0.86	0.85

In this case the DL SSD network has been designed to achieve the minimum number of detection errors (maximum precision), which adversely affects the detection rate (lower value of recall). Anyway, this choice has proved to be optimal for the overall tracking process which tends to propagate box predictions in both time directions, based on the assumption that detection is reliable and correct. Table 2 shows the results after tracking. Table 3 show the metrics extracted from d_i to evaluate the deviation between the centers of the predicted boxes and those of ground truth. Statistics were first calculated per pixel and then reported in cm with the equivalence that 1 px = 0.5 cm.

Table 3. Statistics of the deviation between the centers of the predicted boxes and those of ground truth.

Date	Min (cm)	Max (cm)	Mean (cm)	Mode (cm)	Median (cm)	Std (cm)
09_07	0.00	21.82	4.89	1.11	4.30	3.24
09_14	0.00	30.56	6.17	3.53	4.74	4.99
10_01	0.50	40.05	8.99	1.80	6.40	7.40
10_15	0.50	44.77	8.10	2.50	6.51	6.35
11_09	0.70	47.16	14.11	4.03	12.06	9.73
12_03	0.70	51.24	12.30	4.03	9.92	9.02
12_23	0.00	56.64	12.08	1.11	9.48	9.51

3.1.2 YOLOv5

The evaluation phase was conducted as for the custom FPN, and Table 4 shows the results of YOLOv5 for each dataset.

Table 4. Performance measures of the YOLOv5 at each individual date of the test.

Date	TP	FP	FN	Precision	Recall	F1 score
09_07	437	1	109	0.99	0.8	0.88
09_14	469	2	117	0.99	0.8	0.88
10_01	534	3	73	0.99	0.88	0.93
10_15	543	7	58	0.98	0.9	0.94
11_09	522	7	79	0.98	0.86	0.92
12_03	494	5	81	0.99	0.85	0.92
12_23	509	3	95	0.99	0.84	0.91

As with the previous network, a tracking process was carried out for YOLOv5n to improve the quality of the prediction. Table 5 shows the prediction results after tracking.

Table 5. Performance measures of the YOLOv5 network designed to achieve the lower number of detection errors.

Date	TP	FP	FN	Precision	Recall	F1 score
09_07	475	127	71	0.78	0.87	0.82
09_14	492	105	94	0.82	0.84	0.83
10_01	567	56	40	0.91	0.93	0.92
10_15	554	46	47	0.92	0.92	0.92
11_09	557	51	44	0.91	0.92	0.92
12_03	530	61	45	0.89	0.92	0.9
12_23	550	44	54	0.92	0.91	0.91

Table 6. Statistics of the deviation between the centers of the predicted boxes and those of ground truth for the YOLOv5.

Date	min	max	mean	mode	median	std
09_07	0.00	16.86	2.54	1.11	2.61	2.15
09_14	0.00	13.41	2.83	1.11	2.12	2.19
10_01	0.00	33.63	3.94	1.11	2.69	3.84
10_15	0.00	34.05	4.66	1.11	3.35	4.16
11_09	0.00	59.03	5.70	1.58	3.60	5.88
12_03	0.00	45.02	7.05	1.58	4.52	7.24
12_23	0.00	41.50	5.84	2.23	4.03	5.55

3.2 FPN and YOLOv5 comparison

Figure 6 reports the detection of the whole study area with the two different networks: both showed to be able to detect a high number of plants (detection of the whole orthomosaic is an onerous task for one network) but with a better prevalence of YOLOv5, which, as mentioned, tends to minimize false positives (identified by the bigger red squares in the top-right and bottom-right parts of Figure 6a).

(a) (b)

Figure 6. Detection of the entire crop area with both nets. It is possible to see in the upper right part of the image (a) how the FPN detects several false positives of the artichoke plant. In contrast, the YOLOv5 network performs approximately perfect detection of the entire field (b).

As a last metric to compare the two networks, the mAP index for each orthomosaic is reported in Figure 7.

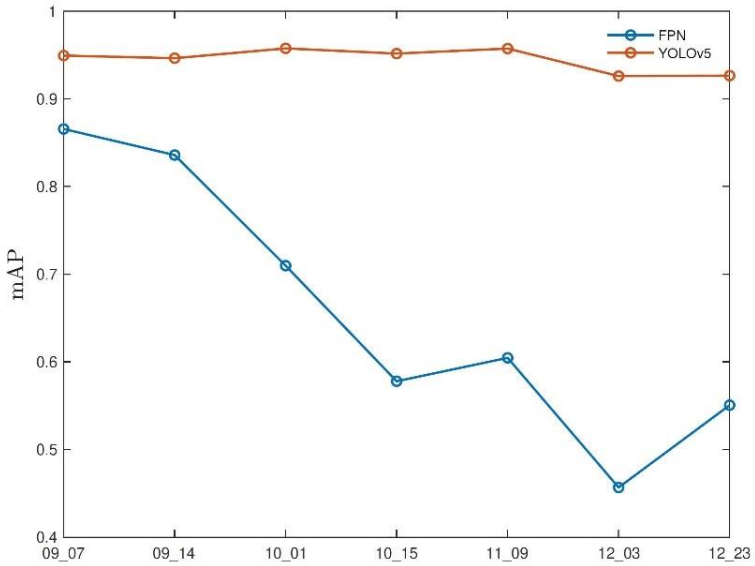


Figure 7. The mean Average precision index for the two networks across all datasets.

3.3 Offline detection

As mentioned in subsection 2.4.6, in the operational phase only the YOLOv5 network will be mounted on the Nvdia Jetson board. Table 7 shows some statistics of the Jetson board when the YOLOv5 network is running on it in the test phase. The statistics were obtained from jetson-stats, a package for monitoring and controlling NVIDIA Jetson [Orin series, Xavier series, Nano, TX1, TX2]. The performance statistics were evaluated in two different power modes of the board: 5w mode, in which the board operates at low power consumption and uses only 2 of the 4 CPUs, and MAXN mode, which uses all available power up to a maximum of 15w.

Table 7. Table of statistics of the Nvdia Jetson Nano board when running the trained YOLOv5 network.

	5W mode				**MAXN mode**			
	Min	**Max**	**Mean**	**Std**	**Min**	**Max**	**Men**	**Std**
CPU1 (%)	10.00	100.00	66.08	23.13	2	81	38.14	11.82
CPU2 (%)	12.00	100.00	66.39	22.77	1	81	36.48	10.89
CPU3 (%)	\	\	\	\	0	53	35.52	11.33
CPU4 (%)	\	\	\	\	0	54	35.98	11.58
GPU (%)	7.00	99.00	56.36	44.65	0	99	79.48	34.82
RAM	2067900	2101352	2087900	1375	2084768	2104596	2099000	59371
TempCPU (°)	28.50	33.00	30.87	1.36	32.00	38.5	35.65	1.63
TempGPU (°)	28	32.5	30.28	1.49	32	37	35.76	1.36
PowerAvg (mW)	1945	3035	2553	551.1	3435	4722	4237	383.93

Figure 8 shows the number of fps evaluated by the board in the two different power modes.

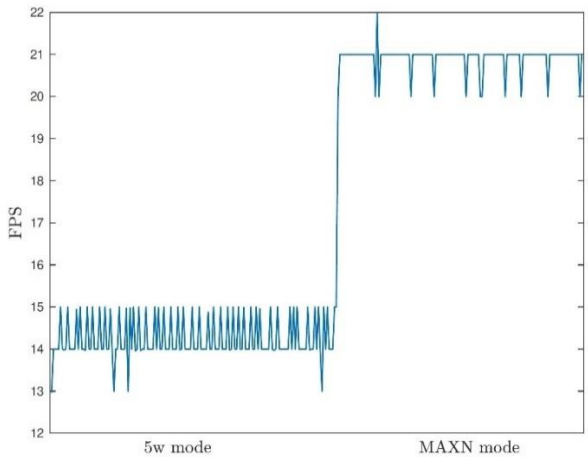


Figure 8. fps in the different power modes. It is evident that by using the total power of the Nvidia Jetson board, the performance in terms of fps increases dramatically.

3.4 Multi-temporal analysis

The graphics in Figure 9 shows the evolution of the Growing Index (GI) for each network. The figure shows the range of variations on a differential scale (between consecutive maps) as well as the cumulative value during time.

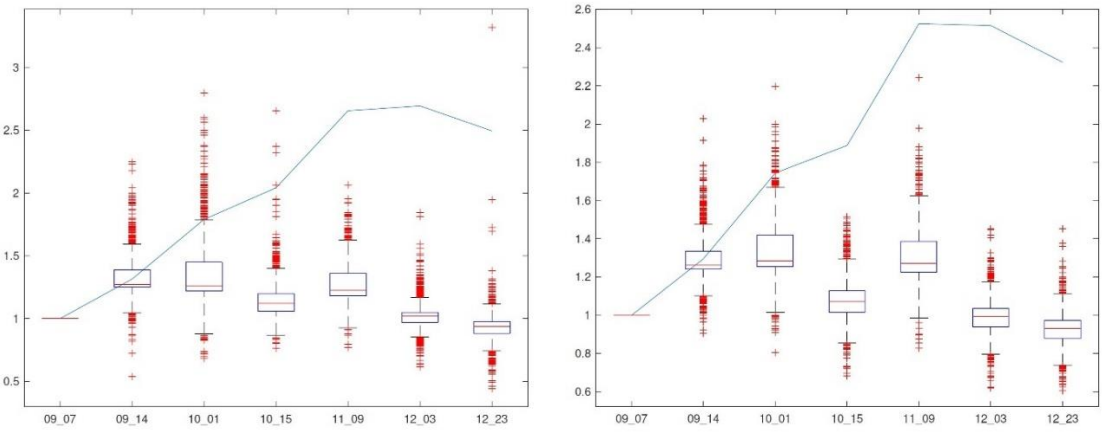


Figure 9. Graphic plots of the growing index (GI) evolution for the FPN (a) and the YOLOv5 (b) networks. The whisker plot shows the range of variations between two consecutive dates of the experiment. The GI is measured as the ratio between the sizes of the bounding box collected at the different times of the experiment. The blue line represents the cumulative average index.

To address the issue of the uniformity or uneven distribution of the growing rate, it is possible to display a heat-map (as in Figure 10) where the spatial distribution of the growing index value for each tracked plant, is shown (color scale) over the full crop field.

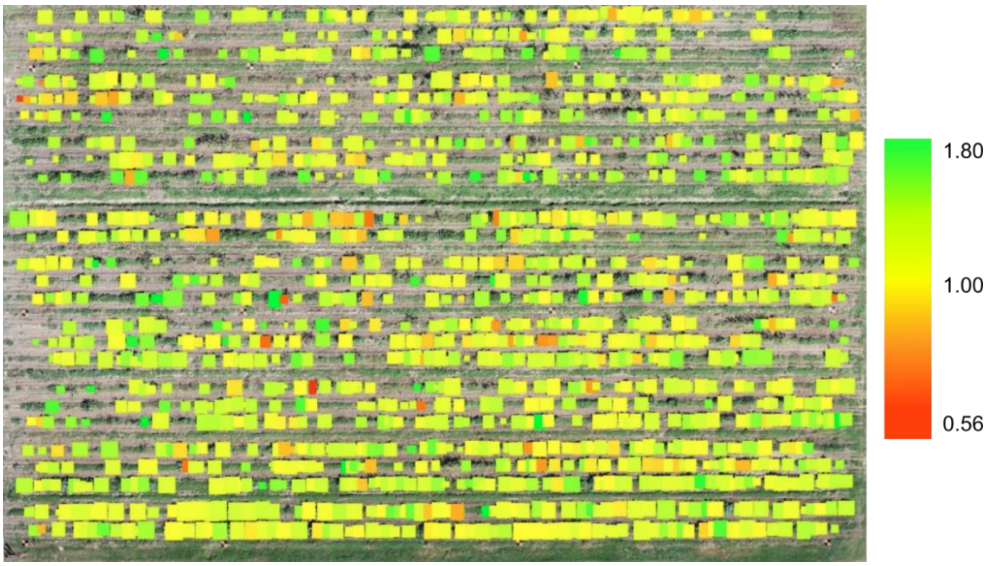
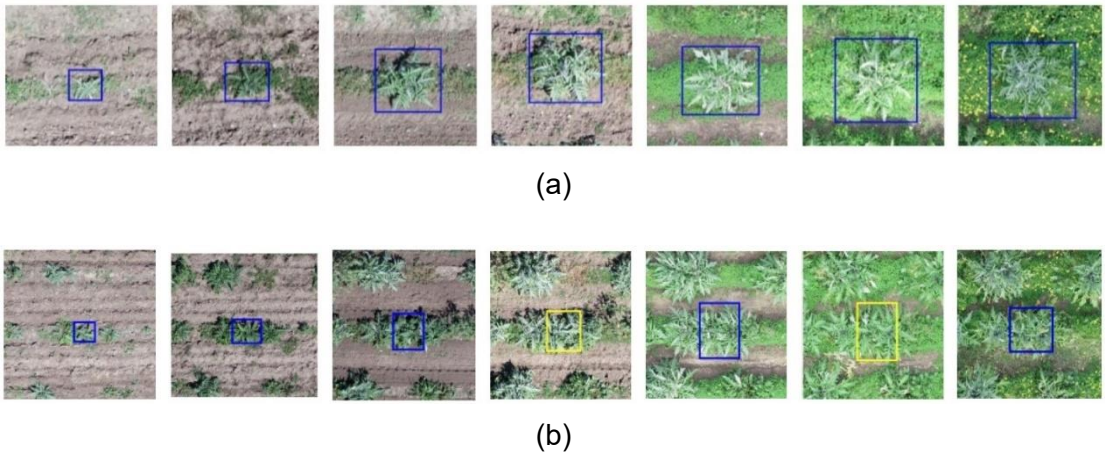


Figure 10. The heat-map of the GI (from 1 October to 15 October of the dataset). The color scale (from red to green) is shown with the range values (from a minimum of 0.55 to the maximum of 1.8).

Figure 11 shows the temporal evolution of two sample artichoke plants (detected with the YOLOv5 network, the behavior with the FPN being the same) in a simple isolated case (a) and a quite common situation (b) where the growing of nearby plants is quickly reaching a size of mutual interference and partial overlaps.



(a)

(b)

Figure 11. History-map for each individual artichoke plant, over the full dataset; (a) simple case of an isolated plant; (b) more common case of multiple plants and their growing process (yellow boxes correspond to prediction results of the tracking process).

Another result reported in Table 8, highlight the occupancy rate of the plants along the rows and its development through the growing season.

Table 8. Occupancy rate and average size of the box-plants along the rows during the development of the crop for each network.

	FPN		YOLOv5	
Date	**Occupancy rate (%)**	**Avg. plant size (cm)**	**Occupancy rate (%)**	**Avg. plant size (cm)**
09_07	30.03	53	28.37	57
09_14	40.38	65	39.31	71
10_01	52.34	85	52.36	93
10_15	57.90	95	56.78	97
11_09	65.59	122	66.52	127
12_03	66.06	119	66.21	125
12_23	61.60	107	61.95	115

4. Discussions

The results obtained on the various datasets show that the network performs satisfactorily on artichoke plants detection, irrespective of the date of the test. It is worth reminding that the same trained network was used for the whole experimental season, without any optimization for the individual datasets. This time independence is

particularly important for an industrial application of this technology in precision agriculture because it can be applied to different scenarios in a small amount of time for different crop applications.

In general, the detection rate (recall) is higher in the early period of the crop when the plants are smaller and isolated and is lower in the late period of the year. This behavior highlights the increasing difficulties to detect and distinguish plants in the last phases due to the mutual overlap of the bigger plants within the rows (Figure 10). A similar trend is also visible for the measure of precision which is over 90% in the early dates.

Missing plants tend to decrease during the experimental dates, except on 9 November 2021. The different illumination condition derived by a different angle of the incident radiation, and the presence of *Oxalis pes-caprae* L., one of the most abundant alien species in artichoke fields during the last days of winter, could have given a higher contrast helping the detection system to distinguish artichoke plants more easily from weed or other elements in the latest surveys dates (Figure 12).

(a) (b)

Figure 12. Two different date acquisitions of the same field portion. The Oxalis pes-caprae invasive plants on 23 December 2021 (b) determined a stronger contrast compared to the 9 November 2021 survey (a).

As compared to the performance detection analysis carried out individually for each stage of the crop, significant improvements can be obtained by using a multi-temporal analysis, with the aggregation of information from all stages of the crop. As such a complete history of the evolution of the plant starting from the first acquired observation could be obtained. For example, it is possible to compensate for a possible lack of detection on a certain date, thanks to the availability of new detection data, as well as to

manage the partial overlaps of neighboring plants, and correct some evident errors of localization and size.

From the availability of the complete temporal traces of each plant it is possible to obtain useful indicators on the evolution of the crop which can be used by the expert agronomist to properly plan irrigation and fertilization interventions and improve plant productivity and health. In addition to the overall estimates on the evolution of the crop (vegetative mass, growth indices), averaged over the entire observed field, some detailed spatial maps can also be provided to highlight any anomalies or non-uniformities in the different areas of the field.

The result of tracking all the detected box-plants over time increases the number of instances by filling most missing data (with box-prediction), and allows a remarkable improvement of recall, at the expense of a reduced level of precision (i.e., more candidates for nearby plants). Moreover, the spatial ordering of all plant- boxes along each individual row of the crop field allows additional measures like the occupancy rate of the plants as shown in Table 8. The number of plants is progressively increasing, with respect to the background (terrain and weed), by reaching a maximum value at the beginning of December, in accordance with the growing index. The field consists of 25 rows of plants with an initial estimated planting of 3000 samples. The number of detected plants during the 7 experimental campaigns reveals a strong reduction of plants (more than 60%) during the very early stages of development, with a progressive stabilization. The multitemporal analysis allowed the obtainment of a more efficient net that, if applied to the same crop in the next years will achieve better performance to the first, more balanced network. This approach, after the first year of image acquisition and network training, can be applied regularly for crop analysis, without the need of repeating the training process.

As stated in the materials and methods section, the detection process involved the conversion of the WGS84 georeferenced images in a XY coordinate system to easily perform the detection process. The artichoke plants detector has been developed not only to later create a real time detection system implemented on board of a spraying UAS, but also to hypothesize the future creation of a path planning system useful to define the borders of the field and the optimized route the UAS will follow, adapting the flight course to maintain the position of the nozzles over the plants. The use of the RTK positioning system is a mandatory equipment to perform such operations, but as frequently happens, these systems face low accuracy problems, especially in remote zones characterized by poor correction signal cover. Any error in the positioning related to the frequent low accuracy of GPS systems implemented on board of UASs will be solved by using the real-time plant detector system. From tables 2 and 5, it can be said that the network predictions do not deviate overly (on average) from the ground truth,

confirming the good detection performance reported in tables 1 and 4. So, apparently, the evidence does not imply taking countermeasures to align the UAS and properly control the detected artichoke, but further analysis will need to be done when the model will be tested in real scenarios. Based on the representation reported in Figure 5, it will be possible to perform path-planning optimization for spraying operations (by UAS or Unmanned Ground Vehicle).

Moreover, a full history of the vegetation process is obtained, for each individual plant (Figure 11), at the different stages of the development process. This detection system's ability will open new scenarios for plant detection, easily allowing the operator to monitor the entire field and evaluate the condition of each plant, specifically for those that show different conditions respect to the rest of the field.

Once the net is trained to detect a specific crop, a first explorative flight should be performed each year after plants' emergence to identify the exact field borders (which also match the operations limits), create an optimized flying route based on the size and positions of nozzles, and have a time zero status of the field. To maintain a low waste of agrochemicals products, intermediate monitoring flights should be performed to verify the exact number, the size of plants and, in case of missing plants, adapt the flight parameters and the required agrochemical amount to distribute.

YOLOv5's performance is marked by a low detection error (precision over 98%) but with a less accurate detection rate (recall just over 80%), which, however, rises above 90% after temporal tracking, and is different from the FPN results, which reported more balanced values of precision and recall (see tables 4 and 5). In contrast, the F1 index, which is a weighted average of precision and recall, settles at similar values for both networks. Compared with the same results from the custom FPN, in Table 1 the prediction statistics change drastically, the data being more balanced: precision drops to more standard values (from 78% to 92%) and recall increases to more acceptable values (from 84% to 93%).

The results reported in Table 6 are quite similar, but YOLOv5 performs a slightly better alignment between the detected boxes and ground truth: this may be related to the fact that, as shown in Table 4, YOLOv5 tends to have higher precision in box recognition against a lower number of detected boxes. In Figure 7 YOLOv5 is significantly better than FPN, which has a decreasing trend in performance as the months of the datasets increase. For both configurations (although in MAXN mode the performance is significantly better) the fps rate is satisfactory (between 14 and 15 fps in the 5w mode and between 20 and 21 fps for the MAXN mode), allowing this board to be used for operational and industrial purposes.

These real-time approaches, even if characterized by a preliminary complex and time-consuming process of data processing and training, are required to optimize operations

like UAS agrochemicals spraying. The entire procedure was a first approach to develop a detector system finalized to perform real-time spraying operation over horticultural plants (in this case artichoke) but mostly to identify the process workflow, highlight the potentialities and, most of all, discover the related limits. The rising application of UASs in precision agriculture scenarios relies on the optimization of operations regarding spot/site-specific input application, path planning and quick response obtainment. Future works will involve the use of possible explainable AI techniques, like safety regions [28] and counterfactual explanation [29], to improve plant detection and give more strength to the multitemporal analysis framework.

5. Conclusions

A machine learning approach for artichoke plant identification for UAS real-time spraying applications was developed. The FPN showed satisfactory detection performances in testing and offline phases, processing videos and images through the Nvidia Jetson Nano board, and showing comparable results with the YOLOv5 network. The proposed automatic multitemporal tracking and analysis procedure showed the possibility of developing a UAS path planning procedure for flight optimization, needed to execute accurate and precise agrochemicals distribution. Such procedure allowed crop monitoring over the entire season, showing important results related to the growing heterogeneity of the field. The next steps, on the strength of the encouraging obtained results, will be to incorporate the Nvidia Jetson Nano board directly on the UAS to perform real-time detection and spraying application, giving a potentially strong and significant scope to this work. Moreover, multi-temporal analysis allows the exploration of crucial information to improve detection reliability and develop an automatic procedure for crop development monitoring.

Acknowledgments

The authors thank the Sarciofo Company (Uri, Sardinia, Italy), site of the surveys during the 2021-2022 season.

Funding

This work was supported by the ECSEL JU-funded project COMP4DRONES [grant number 826610].

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Kriflik, L.S.; Yeatman, H. Food Scares and Sustainability: A Consumer Perspective. Health, Risk & Society 2005, 7, 11–24, doi:10.1080/13698570500042439.

2. Iriti, M.; Vitalini, S. Sustainable Crop Protection, Global Climate Change, Food Security and Safety—Plant Immunity at the Crossroads. Vaccines 2020, 8, doi:10.3390/vaccines8010042.

3. Chavarri, M.J.; Herrera, A.; Ariño, A. Pesticide Residues in Field-Sprayed and Processed Fruits and Vegetables. Journal of the Science of Food and Agriculture 2004, 84, 1253–1259, doi:https://doi.org/10.1002/jsfa.1791.

4. Berg, H. van den; Gu, B.; Grenier, B.; Kohlschmid, E.; Al-Eryani, S.; Bezerra, H.S. da S.; Nagpal, B.N.; Chanda, E.; Gasimov, E.; Velayudhan, R.; et al. Pesticide Lifecycle Management in Agriculture and Public Health: Where Are the Gaps? Science of The Total Environment 2020, 742, 140598, doi:https://doi.org/10.1016/j.scitotenv.2020.140598.

5. Popp, J.; Pető, K.; Nagy, J. Pesticide Productivity and Food Security. A Review. Agronomy for Sustainable Development 2013, 33, doi:10.1007/s13593-012-0105-x.

6. Lan, Y.B.; Chen, S.D.; Fritz, B.K. Current Status and Future Trends of Precision Agricultural Aviation Technologies. International Journal of Agricultural and Biological Engineering 2017, 10, 1–17, doi:https://doi.org/10.3965/j.ijabe.20171003.3088.

7. Sarri, D.; Martelloni, L.; Rimediotti, M.; Lisci, R.; Lombardo, S.; Vieri, M. Testing a Multi-Rotor Unmanned Aerial Vehicle for Spray Application in High Slope Terraced Vineyard. Journal of Agricultural Engineering 2019, 50, 38–47, doi:https://doi.org/10.4081/jae.2019.853.

8. European Parliament, C. of the E. Directive 2009/128/EC of the European Parliament and Of the Council of 21 October 2009 Establishing a Framework for Community Action to Achieve the Sustainable Use of Pesticides (Text with EEA Relevance). Official Journal of the European Union 2009, 1–16.

9. Fadda, A.; Virdis, A.; Barberis, A.; Ledda, L.; Melito, S. Impact of Different Photoperiodic Treatments on “Spinoso Sardo” Globe Artichoke (Cynara Cardunculus L. Var. Scolymus Fiori) Head Traits and Elementary Composition. Acta Horticulturae 2020, 1284, 131–136, doi:https://doi.org/10.17660/ActaHortic.2020.1284.17.

10. Spanu, E.; Deligios, P.A.; Azara, E.; Delogu, G.; Ledda, L. Effects of Alternative Cropping Systems on Globe Artichoke Qualitative Traits. Journal of the Science of Food and Agriculture 2018, 98, 1079–1087, doi:https://doi.org/10.1002/jsfa.8558.

11. Tabikha, R.M.; Draz, A.K. Population Dynamics of Capitophorus Elaeagni (Hemiptera: Aphididae) and Its Associated Predators on Artichoke Plants in El-Behera. Alexandria Science Exchange Journal 2022, 43, 187–197, doi:doi: 10.21608/asejaiqjsae.2022.230544.

12. Xue, X.; Lan, Y.; Sun, Z.; Chang, C.; Hoffmann, W.C. Develop an Unmanned Aerial Vehicle Based Automatic Aerial Spraying System. Computers and Electronics in Agriculture 2016, 128, 58–66, doi:https://doi.org/10.1016/j.compag.2016.07.022.

13. Kamilaris, A.; Prenafeta-Boldú, F.X. Deep Learning in Agriculture: A Survey. Computers and Electronics in Agriculture 2018, 147, 70–90, doi:https://doi.org/10.1016/j.compag.2018.02.016.

14. Khan, S.; Tufail, M.; Khan, M.T.; Khan, Z.A.; Iqbal, J.; Wasim, A. Real-Time Recognition of Spraying Area for UAV Sprayers Using a Deep Learning Approach. PLOS ONE 2021, 16, 1–17, doi:10.1371/journal.pone.0249436.

15. Chen, C.-J.; Huang, Y.-Y.; Li, Y.-S.; Chen, Y.-C.; Chang, C.-Y.; Huang, Y.-M. Identification of Fruit Tree Pests With Deep Learning on Embedded Drone to Achieve Accurate Pesticide Spraying. IEEE Access 2021, 9, 21986–21997, doi:10.1109/ACCESS.2021.3056082.

16. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. SSD: Single Shot MultiBox Detector. In Proceedings of the Computer Vision – ECCV 2016; Leibe, B., Matas, J., Sebe, N., Welling, M., Eds.; Springer International Publishing: Cham, 2016; pp. 21–37.

17. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection 2015.

18. Bochkovskiy, A.; Wang, C.-Y.; Liao, H.-Y.M. YOLOv4: Optimal Speed and Accuracy of Object Detection 2020.

19. Osco, L.P.; Junior, J.M.; Ramos, A.P.M.; Jorge, L.A. de C.; Fatholahi, S.N.; Silva, J. de A.; Matsubara, E.T.; Pistori, H.; Gonçalves, W.N.; Li, J. A Review on Deep Learning in UAV Remote Sensing. International Journal of Applied Earth Observation and Geoinformation 2021, 102, 102456, doi:10.1016/j.jag.2021.102456.

20. Feng, Q.; Yang, J.; Liu, Y.; Ou, C.; Zhu, D.; Niu, B.; Liu, J.; Li, B. Multi-Temporal Unmanned Aerial Vehicle Remote Sensing for Vegetable Mapping Using an

Attention-Based Recurrent Convolutional Neural Network. Remote Sensing 2020, 12, doi:10.3390/rs12101668.

21. Aeberli, A.; Johansen, K.; Robson, A.; Lamb, D.W.; Phinn, S. Detection of Banana Plants Using Multi-Temporal Multispectral UAV Imagery. Remote Sensing 2021, 13, doi:10.3390/rs13112123.

22. Kang, J.; Tariq, S.; Oh, H.; Woo, S.S. A Survey of Deep Learning-Based Object Detection Methods and Datasets for Overhead Imagery. IEEE Access 2022, 10, 20118–20134, doi:10.1109/ACCESS.2022.3149052.

23. Lin, T.-Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature Pyramid Networks for Object Detection. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017; pp. 936–944.

24. Jiao, L.; Zhang, F.; Liu, F.; Yang, S.; Li, L.; Feng, Z.; Qu, R. A Survey of Deep Learning-Based Object Detection. IEEE Access 2019, 7, 128837–128868, doi:10.1109/ACCESS.2019.2939201.

25. Padilla, R.; Netto, S.L.; Silva, E.A.B. da A Survey on Performance Metrics for Object-Detection Algorithms. 2020 International Conference on Systems, Signals and Image Processing (IWSSIP) 2020, 237–242.

26. Hartley, R.I.; Zisserman, A. Multiple View Geometry in Computer Vision; Second.; Cambridge University Press, ISBN: 0521540518, 2004;

27. Kalal, Z.; Mikolajczyk, K.; Matas, J. Forward-Backward Error: Automatic Detection of Tracking Failures. In Proceedings of the Proceedings of the 2010 20th International Conference on Pattern Recognition; IEEE Computer Society: USA, 2010; pp. 2756–2759.

28. Carlevaro, A.; Mongelli, M. A New SVDD Approach to Reliable and Explainable AI. IEEE Intelligent Systems 2022, 37, 55–68, doi:10.1109/MIS.2021.3123669.

29. Carlevaro, A.; Lenatti, M.; Paglialonga, A.; Mongelli, M. Counterfactual Building and Evaluation via EXplainable Support Vector Data Description. IEEE Access 2022, 10, 60849–60861, doi:10.1109/ACCESS.2022.3180026.

Chapter 5 – General conclusions

Concluding remarks and future perspectives

The dissertation aims to explore the potential and efficiency of unmanned aerial systems and machine learning approaches for crop management and agrochemical distribution optimization in orchard and vegetable cultivation systems. All the reported papers highlight how these technologies, even in a preliminary stage, could help solve present and future issues of contemporary agriculture. Further research is needed to optimize their application and promote wider use, overcoming the required programming and scientific skills through software development with a user-friendly configuration.

Chapter 2 focused on the possible use of alternative UAS remote sensing techniques to replace TRV for canopy estimation in a vineyard. The results confirm the appropriateness of integrating or replacing field measures with more precise and accurate techniques when estimating orchards' structural characteristics for the optimization of chemical application and other agronomic practices with an impact on costs and ecological sustainability. This knowledge may provide the necessary information for a comprehensive understanding of structural characteristics and functional vineyard traits with the final objective of enhancing economic performances and environmental sustainability of productive farms. Remote sensing represents a promising tool in precision viticulture, further studies should propose new or refined techniques for quantifying additional canopy features.

The objective of the contribution reported in Chapter 3 was to develop a system for grape bunch detection regardless of variety and geographic location. Therefore, the Deep Learning detector, trained and tested on the GrapeCS-ML dataset, was also tested on an in-house dataset collected from several Sardinian vineyards over the past decade. The obtained results are promising, as most of the grape bunches were detected correctly. Errors arise from the incorrect detection of two bunches as one. In the state-of-the-art, with customized methodologies for a precise application context, the obtained results confirm the portability, flexibility, and novelty of the proposed approach for different scenarios. This approach, applicable to fruit detection and tracking operations of autonomous systems, will help reduce human subjectivity during the visual status assessment, optimize the monitoring time, and reduce the environmental impact of derivative operations such as agrochemical distribution and harvesting.

Chapter 4 reports the results of a Feature Pyramid Network (FPN) machine learning algorithm for artichoke plant identification for real-time UAS spraying applications. The FPN showed satisfactory detection performance in the test and offline phases, processing video and images through the Nvidia Jetson Nano board and showing comparable results with the YOLOv5 network. The proposed automatic multitemporal tracking and analysis procedure enabled the monitoring of crops throughout the season, showing crucial results related to field growth heterogeneity. The next step is to

incorporate the Nvidia Jetson Nano on a UAS for real-time tracking and spray application. In addition, use multitemporal analysis to improve detection reliability and develop an automatic procedure for monitoring crop development.

The thesis first analyzed both technologies and their current application and later explored the results obtained from their combination. The results are encouraging, showing high compatibility and potential. UASs represent a promising tool already well-established in agricultural scenarios. Further improvements regard flight autonomy and payload capability increment, lowering costs (especially about the implemented sensors), automation, and more permissive flight regulations (particularly limiting, especially for aerial agrochemical distribution). Rapid technological evolution should be followed by appropriate regulation able to capture society's and companies' needs. Major improvements and changes are needed to regulate aerial distribution of agrochemicals via UAS in Europe, or at least to accelerate and facilitate trials, which are essential to explore the limits and potential of this technology. Machine learning saw considerable employment in the last decade thanks to the calculation capability improvement, the establishment of new professional profiles experienced in the topic, knowledge and methods sharing. The major ethical issues surrounding AI implementation affect its diffusion, especially when it is free to make decisions without human supervision and to access private information. This is far from a dystopic future based on machines and computers, but to make good use of such technology it is necessary to think about its regulation and possible derived risks.

The presented approaches report some of the multiple tools and technologies smart agriculture can use. Further studies and practical applications are required to test and evaluate their potential in real scenarios. In the future decades their development will be essential to reach agriculture purposes to feed the world population. Technology advance is essential, but the development of accessible technology in terms of costs and usability is crucial to enable its spreading worldwide, especially in those country characterized by higher population growth rate and urgent need of input optimization.