Computers and Electronics in Agriculture 213 (2023) 108185

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

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

Alberto Sassu [a], Jacopo Motta [b], Alessandro Deidda [a], Luca Ghiani [a], Alberto Carlevaro [b,c], Giovanni Garibotto [b], Filippo Gambella [a,d,*]

[a] Department of agricultural Sciences, University of Sassari, Viale Italia 39 a, 07100 Sassari, Italy
[b] Aitek SpA, Funded Research Department, Via della Crocetta 15, 16122 Genova, Italy
[c] DITEN Department, University of Genoa, Via all'Opera Pia, 11 a, 16145 Genova, Italy
[d] Interdepartmental Center IA - INNOVATIVE AGRICULTURE Loc. Surigheddu, 07041 Alghero (SS), SS 127 bis, Km 28,500

ARTICLE INFO

Keywords:
Single shot detector
Multi-target object detection
Multi-temporal monitoring
Plant detection
Site-specific management
Precision agriculture

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 FPN (Feature Pyramid Network), 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. YOLOv5 achieved the best overall result. The FPN recorded a lower recall, which is desirable to achieve a minimum number of detection errors and limit the leakage of agrochemicals on false-positive targets. 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.

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 (Kriflik and Yeatman, 2005). 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 (Iriti and Vitalini, 2020). 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 (Chavarri et al., 2004).

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 (Van den Berg et al., 2020). 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 (Popp et al., 2013).

Agrochemicals are considered toxic products that affect the safety of food and ecosystems. The resulting concerns are reflected in policy initiatives, legislative regulations, and the growing demand for

* Corresponding author.
E-mail addresses: asassu@uniss.it (A. Sassu), jacopo.motta@aitek.it (J. Motta), adeidda1@uniss.it (A. Deidda), lghiani@uniss.it (L. Ghiani), alberto.carlevaro@aitek.it (A. Carlevaro), giovanni.garibotto@gmail.com (G. Garibotto), gambella@uniss.it (F. Gambella).

https://doi.org/10.1016/j.compag.2023.108185
Received 12 October 2022; Received in revised form 31 July 2023; Accepted 25 August 2023
Available online 7 September 2023

environmentally compatible management methods. Innovative smart variable-dose sprayers development for agrochemical dosage will bring a significant contribution, leading to economic and environmental benefits crucial for their implementation. Such operations, combined with decision support systems (DSS), should support farmers in applying control treatments, avoiding product waste and total energy demand, acting at the most appropriate time and place, and helping to reduce labor costs (De Bortoli et al., 2022). In particular, the application of site-specific herbicides can help reduce costs by 50 percent (GERHARDS and OEBEL, 2006). However, the high cost of precision spraying technology and the associated complexity of these systems result in farmers opting for chemical control (Gerhards et al., 2022).

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 (Lan et al., 2017). 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 (Sarri et al., 2019). Despite regulations and restrictions on aerial agrochemicals spraying, as is in Europe (European Parliament, 2009), 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 (Fadda et al., 2020; Spanu et al., 2018). Artichoke plants are attacked by several insects and pests like aphids, thrips, leaf miners, etc., which require agrochemicals application, easily deployable through UASs (Tabikha and Draz, 2022). 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 (Xue et al., 2016). 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 (Kamilaris and Prenafeta-Boldú, 2018). Previous work has 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 (Khan et al., 2021) or to determine pests' position in real-time on the orchard and plan the optimal pesticide spraying route for the agricultural UAS (Chen et al., 2021).

The Feature Pyramid Network (FPN) is a particular type of CNN but the algorithm proposed in this paper is a customized version that works as a Singles Shot Detector (SSD) (Liu et al., 2016). A CNN backbone is used to extract the informative features that will be used by the classical FPN for detection. 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 (Redmon et al., 2015) 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 (Bochkovskiy et al., 2020) 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 (Osco et al., 2021). Multi-temporal UAS imagery incorporation could significantly boost the accuracy and compensate for the low spectral resolution of RGB imagery (Feng et al., 2020). 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 (Aeberli et al., 2021).

The work aims at developing a machine learning approach based on FPN for artichoke plant detection intended for real-time UAS spraying applications and an automatic multi-temporal 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 m^2 surface in Uri (the area surveyed by the UAS was 4586 m^2 to include the field borders and avoid orthomosaics' distortion because of the reduced amount of images of that area), North-west Sardinia, Italy (Long. 8.472029, Lat. 40.623619; WGS84, EPSG 4326) at 125 m above sea level. Fig. 1 shows an overall view of the artichoke field object of the study planted on 15 July 2021.

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

2.2. UAS platform and implemented sensors

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

A RTK GNSS Reach RS+ (Emlid, Budapest, Hungary) 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 monitoring 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 15 m height AGL to obtain high quality images (232 photos for each flight). The speed was one m/s, the flight duration was 10 min 21 s, 70% side

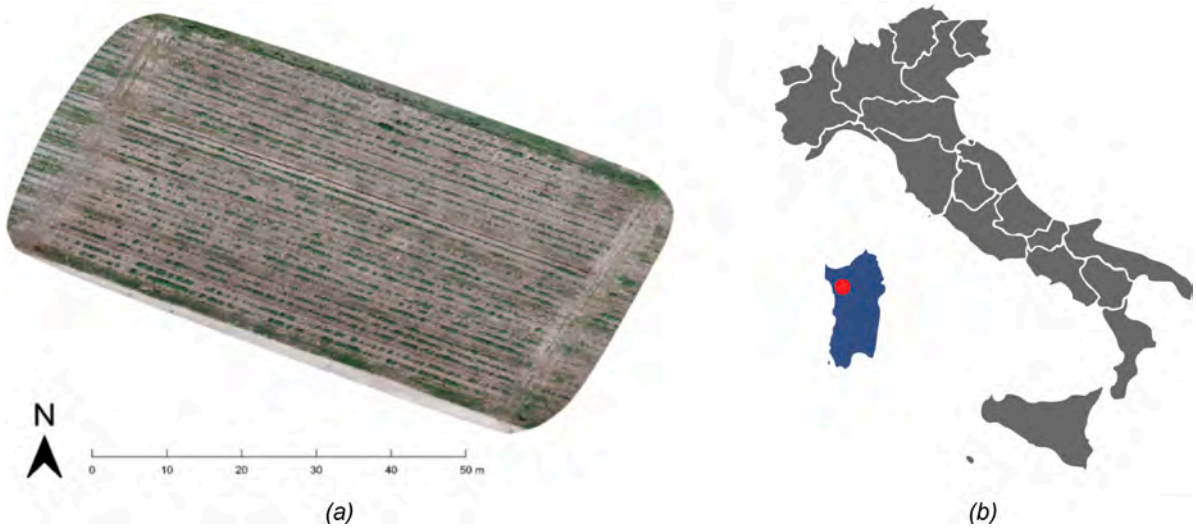

Fig. 1. The artichoke field object of the study (a) and the study site location represented by the red dot within Italy mainland in gray color and specifically in Sardinia region in blue color (b). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
UAS's flight settings and dataset's details.

Flight settings		Dataset details	
flight route length (m)	667	survey dates (n)	7
flight height AGL (m)	15	photos (n/date)	232
flight time (min.sec)	10.21	total dataset photos (n)	1624
course angle (°)	51	sensor model	DJI Phantom 4 Pro RGB CMOS 1″
take-off speed (m/s)	55	FOV (°)	84
flight speed (m/s)	59	lens's focal length (mm)	8.8/24
area (m^2)	4586	image width (px)	5472
side overlap ratio (%)	70	image height (px)	3648
frontal overlap ratio (%)	80	GSD (cm/px)	0.41

overlap ratio and 80% frontal overlap ratio, the ground sampling distance (GSD) was 0.41 cm/px. 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. Table 1 summarizes the flight settings and the dataset's details.

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 (Kang et al., 2022). 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 (Lin et al., 2017).

In the state of the art, there is a wide range of possible detection networks (Jiao et al., 2019), 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 (Fig. 2) of the object detector were made according to the following scheme:

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

Table 2
Performance measure of the FPN Detection at each individual date of the test.

Date	BBCH[1]	TP	FP	FN	Precision	Recall	F1score	SR	SP
7 Sept	14	474	11	72	0.97	0.86	0.91	0.87	0,95
14 Sept	16	533	22	53	0.96	0.91	0.93	0.91	0,96
1 Oct	35	541	39	66	0.93	0.89	0.91	0.90	0,98
15 Oct	47	544	18	57	0.96	0.90	0.93	0.94	0,96
9 Nov	51	460	32	141	0.93	0.76	0.84	0.90	0,98
3 Dec	55	450	73	125	0.86	0.78	0.82	0.92	0,98
23 Dec	59	509	71	95	0.87	0.84	0.86	0.92	0,98

[1] BBCH stages derived by (Archontoulis et al., 2010).

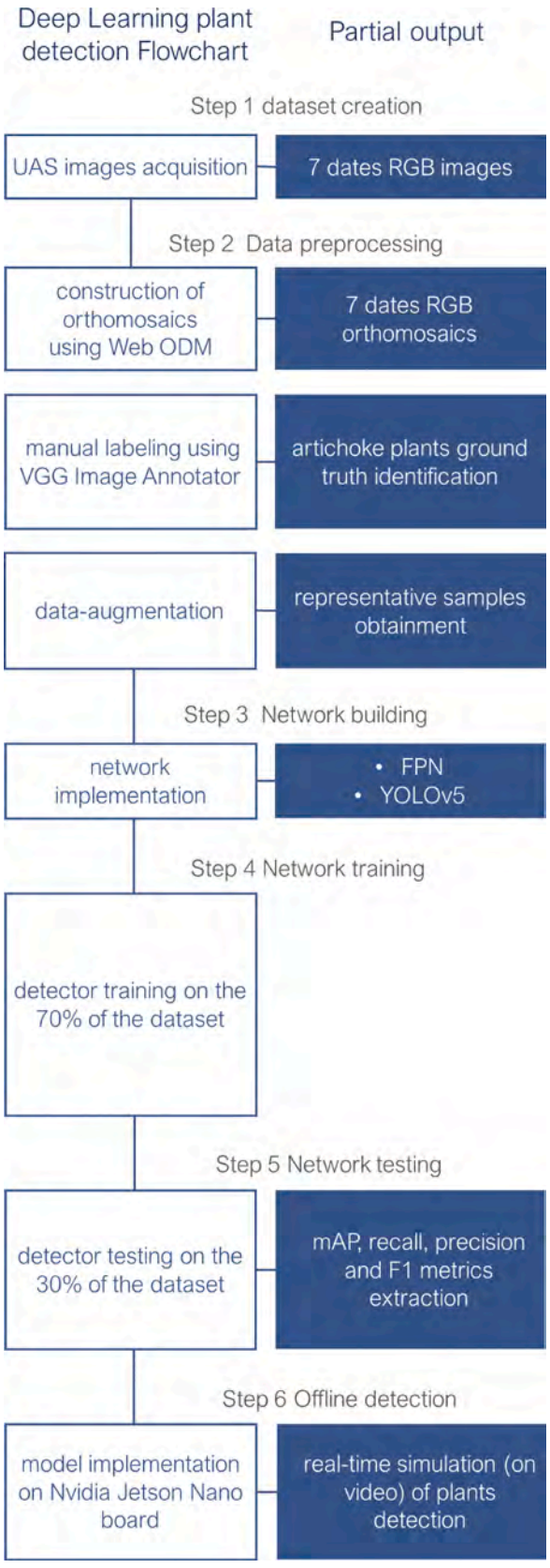

Fig. 2. The deep learning plant detection flowchart (left gray color column) and the partial output of each step (right blue color column). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

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.

2.4.1. Data processing

The preprocessing phase was common to both networks. 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 (Fig. 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.

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

As far as the implementation of the network is concerned, the input parameters are as follow:

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

Where px refers to one 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.

The approach is totally in the spirit of Meng et al., 2022, where FPN was used to improve the information power of feature extraction. Specifically, the proposed network (Fig. 4) consists of a convolutional backbone to extract the essential features, then the FPN structure is used as a neck to enhance the information and extract the features at different scales, and finally the outputs (boxiness and box coordinates) are extracted as usual with softmax and a final convolutional layer.

2.4.3. FPN training

The FPN network model has been trained from scratch, by designing and customizing all input parameters. This means that the network was initialized with random weights and biases, and then trained on a subset of the dataset of artichoke images, constructed in the preprocessing phase by 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

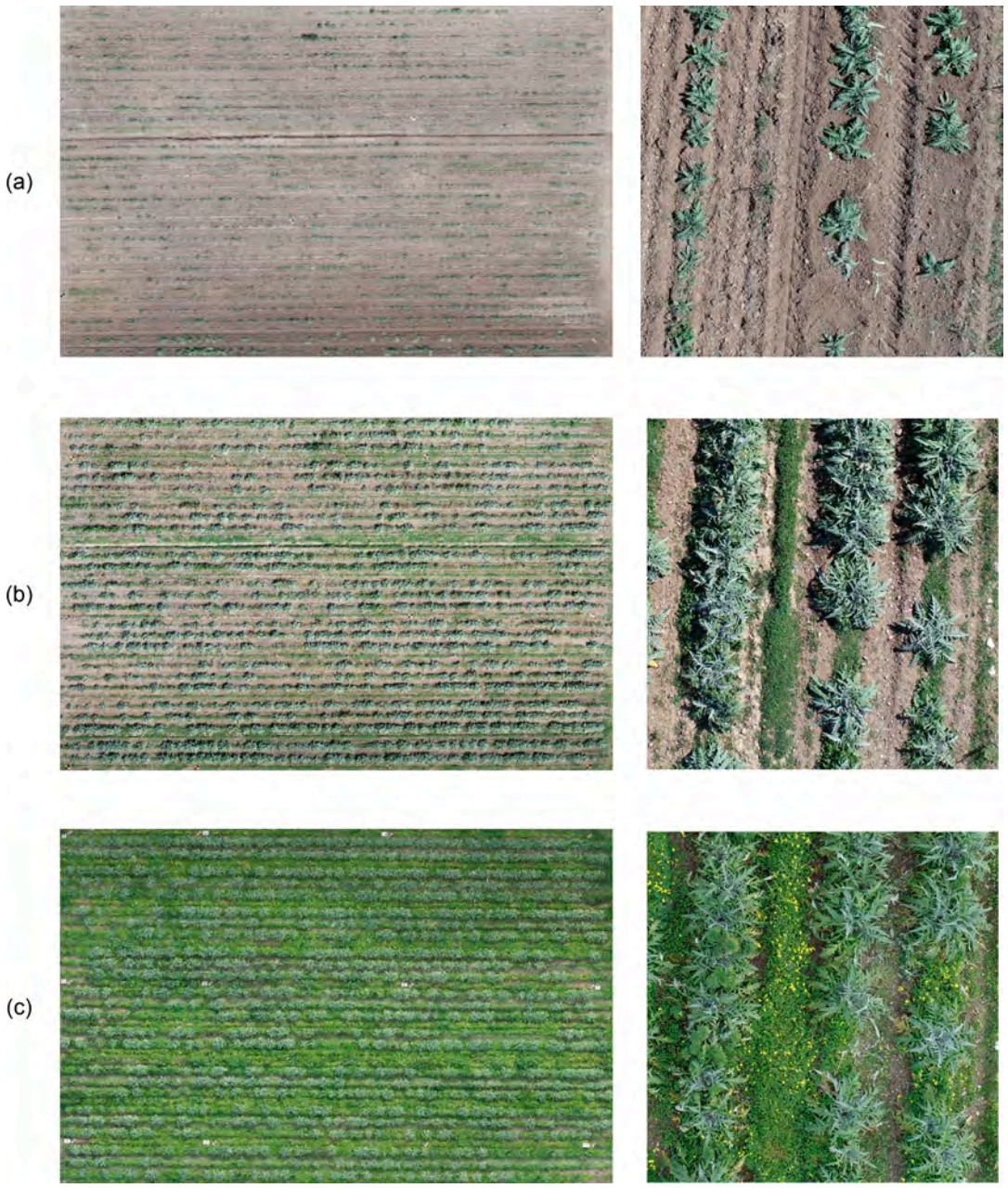

Fig. 3. RGB orthomosaics (left column) and some artichoke plants details (right column) derived by 15 m flight altitude of three different surveying dates: 7 September 2021 (a), 15 October 2021 (b), and 23 November 2021 (c).

Where IoU is the usual acronym for the Intersection over Union rate. IoU is used to evaluate the object detection performance and consists of the percentage of the ground truth bounding box covered by the prediction bounding box. 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 h to obtain satisfactory results.

2.4.4. YOLOv5 training

Transfer learning has been used to train the YOLOv5n network, the smallest version of YOLOv5. A pre-trained YOLOv5n model, which had already been trained on a large dataset of images (80 classes), was fine-tuned and re-trained on a dataset of artichoke images. Specifically, the backbone of the YOLOv5 network was used, which consists of the convolutional layers responsible of feature extractions and froze these layers. Only the last few layers of the network were trained, which are responsible for making object detections, on the artichoke dataset. In general, nano models maintain the YOLOv5 depth multiple of 0.33 but reduce the YOLOv5 width multiple from 0.50 to 0.25, resulting in about 25% fewer parameters, from 7.5 M to 1.9 M, ideal for mobile and CPU solutions.

The training phase included 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 dimension 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.

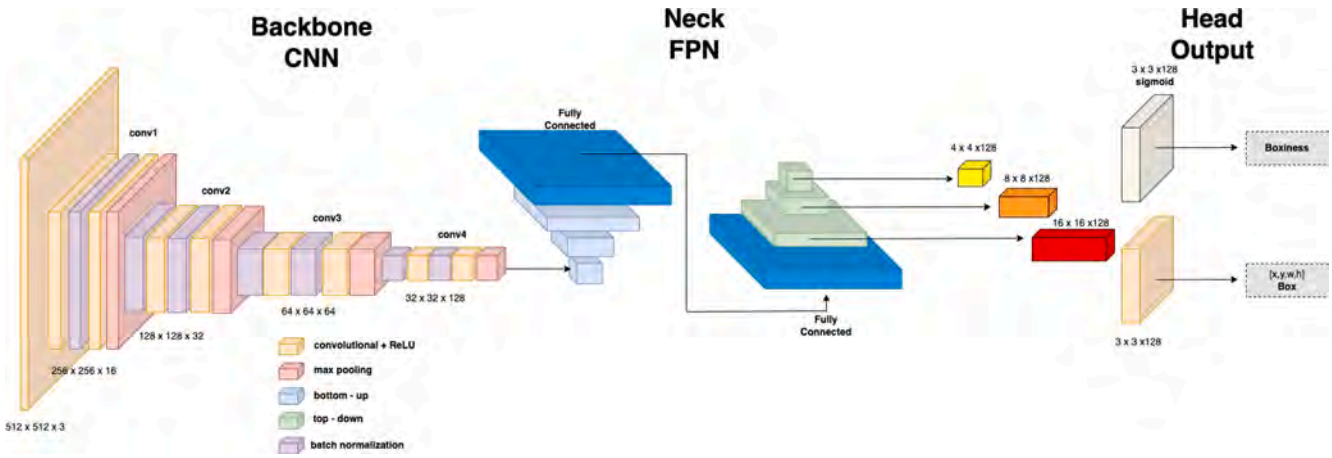

Fig. 4. The proposed classification analysis process network flow diagram.

As already 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 (Padilla et al., 2020):

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

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

$$F1score = \frac{2TP}{2TP + FP + FN} \tag{3}$$

$$surfacerecallSRI = \frac{matched\ area\ (px)}{GT\ area\ (px)} \tag{4}$$

$$surfaceprecisionSP = \frac{matched\ area}{DET\ area\ (px)} \tag{5}$$

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. To effectively analyze the detection results, it is appropriate to dwell on the fact that the target application (UAS sprayer) has two main objectives: maximizing the number of plants to be sprayed and minimizing fertilizer waste (on the ground or on weeds). Unfortunately, the classical indexes (precision and recall) are only based on the number of object-box intersections, irrespective of the actual shape and size of the plants. For this purpose, two additional indexes were introduced, namely SR (Surface-Recall) and SP (Surface-Precision), defined in equation (4) and (5), where GT_Area (px) corresponds to the number of image-pixels covered by the ground-Truth-boxes (any pixel overlap is counted only once). DET_Area (px) corresponds to the number of the image-pixels covered by the detected-boxes and Matched_Area (px) is the number of image-pixels where both GT-boxes and DET-boxes overlap each other. In some way they simulate the behaviour of the UAS-Sprayer, by predicting the expected performance of a real-time implementation.

At first, network parameters have been optimized to achieve the best possible results in the detection process. Afterwards, the designed network configurations have been applied to off-line multi-temporal monitoring and 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, since both offline processing and real-time application are dealing with the same videos and image data, whether they have been collected and stored beforehand or transmitted directly from a camera. In particular, the Nvidia Jetson Nano board (Santa Clara, California, USA) was selected, a small and powerful computer for embedded and AI IoT applications, to install the trained and tested networks for the evaluation phase.

As explained in the discussion section, in the future operational online detection phase with the UAS, only YOLOv5 will be loaded on the Jetson board: this is because YOLOv5 has proved to be a lighter network than the custom feature pyramid FPN, and it achieved better results in detection and hardware performance (power consumption, storage memory, etc.). A comparison between FPN and YOLOv5 network performance is briefly referred in 3.2 and some data on the implementation of YOLOv5 on the Jetson board are reported in Section 3.3.

2.4.7. Temporal monitoring and data integration

To correctly reconstruct the complete time history of each plant, a temporal monitoring algorithm has been developed. The first step of this process is a spatial registration of the consecutive pairs of orthomosaics (Hartley and Zisserman, 2004).

12 ground control points (GCP) (e.g., in Fig. 5) 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. Such an automatic registration is implemented in two following steps:

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

It is worthy to underline that, in this case, the homographic transformation is computed between the centers of mass of the boxes detected in two consecutive orthomosaic images. The transformation allows mapping the positions of the boxes from one image to the other, aligning them spatially. By estimating the homography, the algorithm can

Fig. 5. One of the 12 GCPs (0.5 × 0.5 m) used for image referencing.

reconstruct the complete time history of each plant by aligning and registering the detected boxes in the consecutive orthomosaic images.

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 monitoring process (Kalal et al., 2010), which, in turn, is implemented in two phases:

- monitoring 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.
- monitoring backward: when the last available orthomosaic map 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 monitoring 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.
- multiple box instances for the same ground truth, due to localization and size errors in forward–backward matching.
- new-phantom-boxes, which may appear in areas where there are no plants, often due to the presence of weeds, scattered leaves, etc.

As such, forward–backward allows the complete reconstruction of the plant development from the beginning to the end of the crop.

2.4.8. Artichoke crop field analysis

The output list of the box-plants from the temporal integration is reorganized by ordering them both in vertical and horizontal positions along the plant rows in the field. Fig. 6 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.

In a real-time application, on-board of the UAS, the best network configuration should be able to achieve a high recall rate, to cover the maximum amount of the plant targets' surface and drive the direction of the sprayer towards the actual position of the plants in the field. A high precision rate would also be desirable, to minimize the amount of spraying outside the actual crop. 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. We identified the L^2 norm (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, we calculated the distance d_i for all pairs of detected and ground-truth boxes as:

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

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 important feature output of the multitemporal analysis is 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:

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

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 monitoring algorithm, is the same for each orthomosaic ($N = 1419$ for FP network and $N = 1351$ for YOLOv5 network).

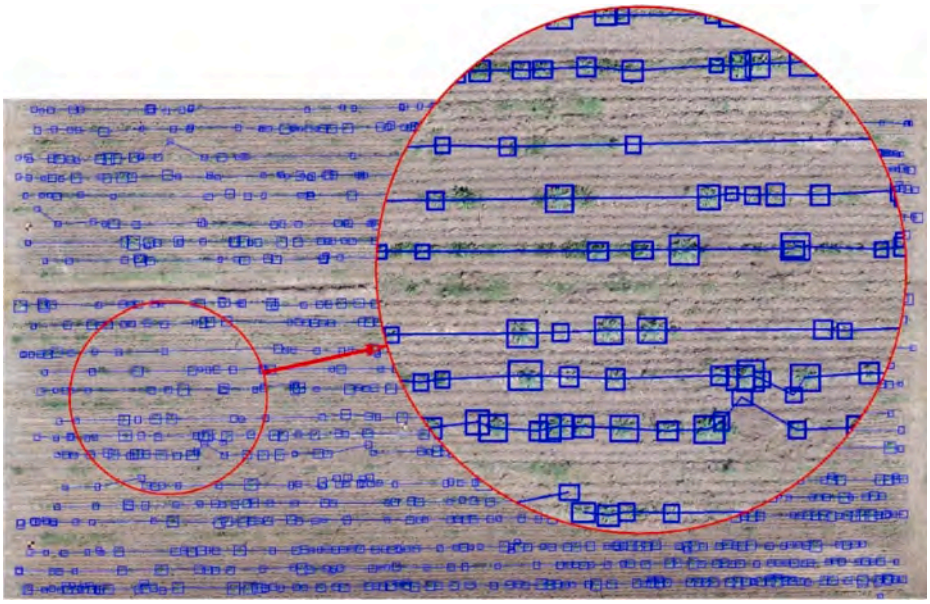

Fig. 6. A partial view of the RGB orthomosaic, with the box-plants detected and tracked over time–space (blue) and connected along each row of the crop. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3. Results

3.1. Deep learning plant detection

3.1.1. FPN performance analysis

During the evaluation phase the FPN was tested on the portion of the dataset not used for training. Table 2 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. Detection performance was also tested with different values of IoU, from 0.5 to 0.25. Of course, a slight improvement of both recall and precision was found, but always below 0.5%. Hence the entire performance test was carried on with the standard threshold IoU = 0.5. It is worth noting that surface indexes (SR and SP) are independent of the IoU value. Table 2 and Table 3 compare the performance measures obtained before and after the Forward-Backward monitoring process.

Table 2 shows how precision is slowly decreasing with time, in accordance with the crop development process. In fact artichoke plants reach their maximum growth around December, and as the plants become larger, they often interfere each other with partial overlapping which makes it difficult to distinguish artichoke plants in the field. Such an irregular distribution of the plant-boxes is also affecting the behavior of the recall index. The lower values measured in the very early stage of the test (7 Sept.), are due to the presence, in the ground-truth list, of very small target-plants which are often barely visible also by the human eye.

On the other hand, the two surface indexes SR and SP are much more stable all along the full development period. They better describe the network performance within the context of UAS-spraying application and demonstrate that more than 90% of the plant foliage can be reached by the spraying process, with less than 5% of the spraying product falling outside of the actual crop field. Forward-backward tracking allows to improve the recall index, by recovering some of the plant-boxes which were missed by the network at a certain stage of the development, but inevitably it also determines a reduction of the precision measure, due to the generation of multiple hypotheses, for very close or partial overlapping plants. Table 3 (for network FPN) and Table 5 (for YOLOv5) confirm this trend but do not contribute to determine the best network, since performance after tracking is almost identical for both.

Table 4 shows 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 one px = 0.5 cm.

3.1.2. YOLOv5 performance analysis

The same performance evaluation carried out for the custom FPN was also repeated on the YOLOv5 network for each orthomosaic of the dataset. The results, in Table 5, demonstrate a high level of precision, which is confirmed also by the SP index. On the contrary the recall index is slightly low during the early stages of the plant growth. Anyway, the SR index, reporting the actual ground plant coverage, is high (greater than 90%) and stable for the whole analysis period.

Table 3
Performance measures of the FPN network after multi-temporal monitoring.

Date	BBCH	TP	FP	FN	Precision	Recall	F1score	SR	SP
7 Sept	14	487	146	59	0.76	0.89	0.83	0.89	0.88
14 Sept	16	531	95	55	0.84	0.91	0.88	0.91	0.92
1 Oct	35	555	102	52	0.85	0.91	0.88	0.90	0.96
15 Oct	47	550	82	51	0.87	0.92	0.89	0.94	0.96
9 Nov	51	517	124	84	0.8	0.86	0.83	0.91	0.98
3 Dec	55	499	121	76	0.81	0.87	0.84	0.93	0.97
23 Dec	59	519	106	85	0.83	0.86	0.85	0.92	0.97

Table 4
Statistics of the deviation between the centers of the predicted boxes and those of ground truth.

Date	BBCH	Min (cm)	Max (cm)	Mean (cm)	Mode (cm)	Median (cm)	Std (cm)
7 Sept	14	0.00	21.82	4.89	1.11	4.30	3.24
14 Sept	16	0.00	30.56	6.17	3.53	4.74	4.99
1 Oct	35	0.50	40.05	8.99	1.80	6.40	7.40
15 Oct	47	0.50	44.77	8.10	2.50	6.51	6.35
9 Nov	51	0.70	47.16	14.11	4.03	12.06	9.73
3 Dec	55	0.70	51.24	12.30	4.03	9.92	9.02
23 Dec	59	0.00	56.64	12.08	1.11	9.48	9.51

Table 5
Performance measures of the YOLOv5 at each individual date of the test.

Date	BBCH	TP	FP	FN	Precision	Recall	F1 score	SR	SP
7 Sept	14	437	1	109	0.99	0.8	0.88	0.91	0,96
14 Sept	16	469	2	117	0.99	0.8	0.88	0.91	0,97
1 Oct	35	534	3	73	0.99	0.88	0.93	0.94	0,98
15 Oct	47	543	7	58	0.98	0.9	0.94	0.94	0,97
9 Nov	51	522	7	79	0.98	0.86	0.92	0.95	0,99
3 Dec	55	494	5	81	0.99	0.85	0.92	0.95	0,98
23 Dec	59	509	3	95	0.99	0.84	0.91	0.95	0,98

Table 6
Performance measures of the YOLOv5 network designed to achieve the lower number of detection errors.

Date	BBCH	TP	FP	FN	Precision	Recall	F1 score	SR	SP
7 Sept	14	475	127	71	0.78	0.87	0.82	0.96	0.80
14 Sept	16	492	105	94	0.82	0.84	0.83	0.96	0.85
1 Oct	35	567	56	40	0.91	0.93	0.92	0.96	0.94
15 Oct	47	554	46	47	0.92	0.92	0.92	0.96	0.95
9 Nov	51	557	51	44	0.91	0.92	0.92	0.95	0.98
3 Dec	55	530	61	45	0.89	0.92	0.9	0.95	0.97
23 Dec	59	550	44	54	0.92	0.91	0.91	0.95	0.97

Even in this case the detection performance was measured after multi-temporal forward–backward integration,

to receive confirmation, in Table 6, of a greater coverage of the foliar plant at the expense of a slightly lower precision. These performances are very suitable for multi temporal analysis of the crop, to faithfully reproduce the growth development of each plant in the field.

Finally, Table 7 shows the deviation between the centres of the predicted and ground truth boxes for the output of the YOLOv5 network. The reported values of the minimum (Min), maximum (Max), mean (Mean), mode (Mode), median (Median) and standard deviation (Std) demonstrate that the detected boxes are fairly well overlapping the ground truth, and this is an indication of the goodness of the network prediction.

Fig. 7. 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).

3.2. FPN and YOLOv5 comparison

Fig. 7 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 Fig. 7a). Both network models show appropriate performance for the problem at hand, and the area indexes (SR and SP) demonstrate a high leaf area coverage of the plants and a minimum amount of detection area outside them.

Anyway, YOLOv5 achieves a greater level of precision as compared to FPN whose training data included also a significant amount of small object-plants. On the contrary, YOLOv5 has been trained on wider and more diversified datasets, with the tendency to generalize to larger object shapes.

The statistics in Table 7 also indicate that YOLOv5 performs better in the detection of artichoke plants. In particular, the columns of mean and standard deviation show that the YOLOv5 model is more robust and was not affected by plant growth over the months.

As a last metric to compare the two networks, Fig. 8 reports the mAP index for each orthomosaic.

There is no doubt that YOLOV5 performs better than the custom FPN, as highlighted by mAP results showed in Fig. 8. The reasons for such a performance disparity can be attributed to various factors. Firstly, the model architecture plays a crucial role as YOLOv5 and FPN have different underlying architectures that can impact their performance. While FPN achieves lower performance results, it is also less complex than Yolov5, with a different number of parameters to be trained. Secondly, training strategies are significant factors to consider. Finding the optimal hyperparameters (learning rate, optimizers, train-test split, etc.) can be a challenging task that requires a considerable amount of time and energy. YOLOv5 is a state-of-the-art technology, and it is assumed that the hyperparameters have been properly tuned. Thirdly, post-processing steps applied to the model's output can have an impact on the final detection results. Techniques such as non-maximum suppression to remove duplicate detections, thresholding for confidence scores, and bounding box refinement can vary in implementation and affect the performance disparity between YOLOv5 and FPN if not consistently applied. In summary, the performance disparity can be attributed to several factors, including data availability, choice of optimal hyperparameters, and post-processing considerations, but what is worthy to note is that custom training of neural networks (like the FPN in question) can result in comparable performance with state-of-the-art techonology.

Table 7
Statistics of the deviation between the centers of the predicted boxes and those of ground truth for the YOLOv5.

Date	BBCH	min	max	mean	mode	median	std
7 Sept	14	0.00	16.86	2.54	1.11	2.61	2.15
14 Sept	16	0.00	13.41	2.83	1.11	2.12	2.19
1 Oct	35	0.00	33.63	3.94	1.11	2.69	3.84
15 Oct	47	0.00	34.05	4.66	1.11	3.35	4.16
9 Nov	51	0.00	59.03	5.70	1.58	3.60	5.88
3 Dec	55	0.00	45.02	7.05	1.58	4.52	7.24
23 Dec	59	0.00	41.50	5.84	2.23	4.03	5.55

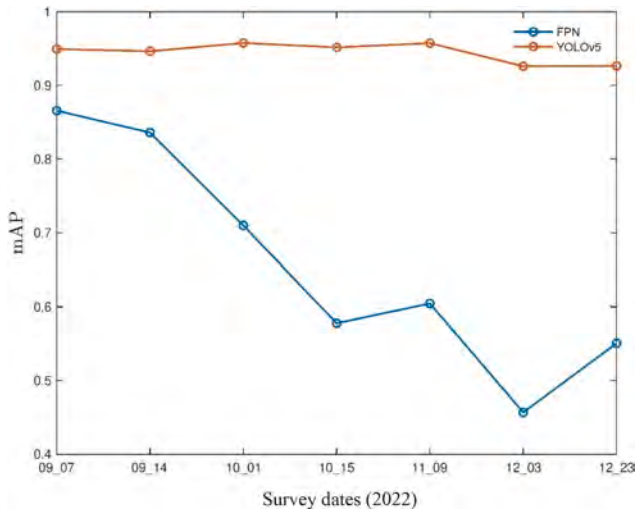

Fig. 8. The mean Average precision index for the two networks across all datasets.

3.3. On-board network detection

As mentioned in subsection 2.4.6, during the next operational phase only the YOLOv5 network will be mounted on the Nvdia Jetson board. Table 8 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 two of the four CPUs, and MAXN mode, which uses all available power up to a maximum of 15W. Fig. 9 shows the number of FPS evaluated by the board in the two different power modes. The computed statistics are composed by different parameters: CPU usage, GPU usage, RAM usage, temperature, and power consumption, each of them divided in two different power modes: energy consumption mode (5 W mode) and maximum consumption mode (MAXN mode). Specifically:

- The "Min" and "Max" columns show the minimum and maximum values for each parameter during the detection process.
- The "Std" columns show the standard deviation for each parameter, which gives an idea of how much the values fluctuate during the detection process.
- The"CPUs" rows show the percentage usage of the four CPUs on the board.
- The "GPU" row shows the percentage usage of the GPU on the board.
- The "RAM" row shows the usage of the RAM on the board.
- The "TempCPU" and "TempGPU" rows show the temperature of the CPUs and GPU, respectively.
- The "PowerAvg" row shows the average power consumption during the detection process.

It is clear from the table that YOLOv5 model uses most of the available resources of the Jetson Nanon board, with the GPU usage being the highest at 99%. The standard deviation values are relatively low, indicating that the performance of the model is consistent. The temperature of the CPU and GPU are relatively high, with the hottest value reached by the GPU in MAXN mode at 38.5 degrees. The power consumption is also relatively high with the highest consumptions reached in the MAXN mode.

Fig. 9 shows the number of FPS evaluated by the board in the two different power modes.

3.4. Multi-temporal analysis

The graphics in Fig. 10 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.

Table 8
Table of statistics of the Nvdia Jetson Nano board when running the trained YOLOv5 network.

	5 W 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	2,067,900	2,101,352	2,087,900	13,757	2,084,768	2,104,596	2,099,000	59,371
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

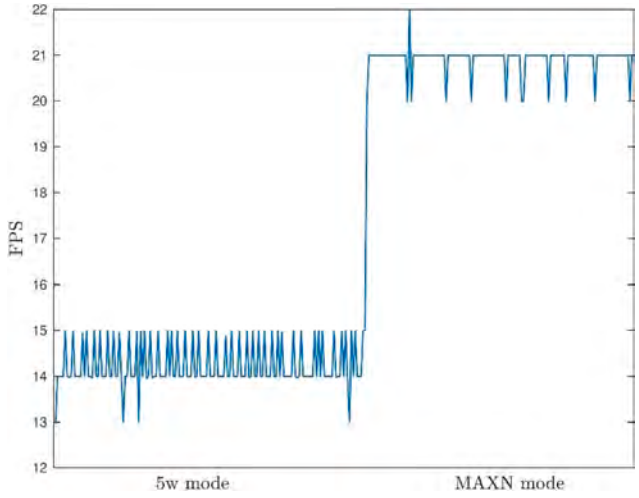

Fig. 9. FPS in the different power modes. It is evident that by using the total power of the Nvidia Jetson board, the performance in terms of FPS increases dramatically.

To address the issue of the uniform or uneven distribution of the growing rate, it is possible to display a heat-map (as in Fig. 11) where the spatial distribution of the growing index value for each tracked plant, is shown (according to the color scale) over the full crop field.

Fig. 12 shows the temporal evolution of two sample artichoke plants (detected with the YOLOv5 network, the behavior with the FPN being quite 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.

Another important result is reported in Table 9. It shows the occupancy rate (i.e., the percentage of the field occupied by the detected boxes), and the average size of the detected boxes measured for each available set. Not surprisingly, the occupancy rate increases with the growth of the artichoke plants and reaches its highest value in December (maximum plant growth). A similar observation can be made for the average plant size. If we compare FPN and YOLOv5, we can see that the results are almost the same. This is certainly due to the similar performance of the two network models, but it is mainly due to multi-temporal integration, which, through forward–backward propagation, tends to recover all missing plants and provide a reliable representation of plant evolution throughout the whole growing season.

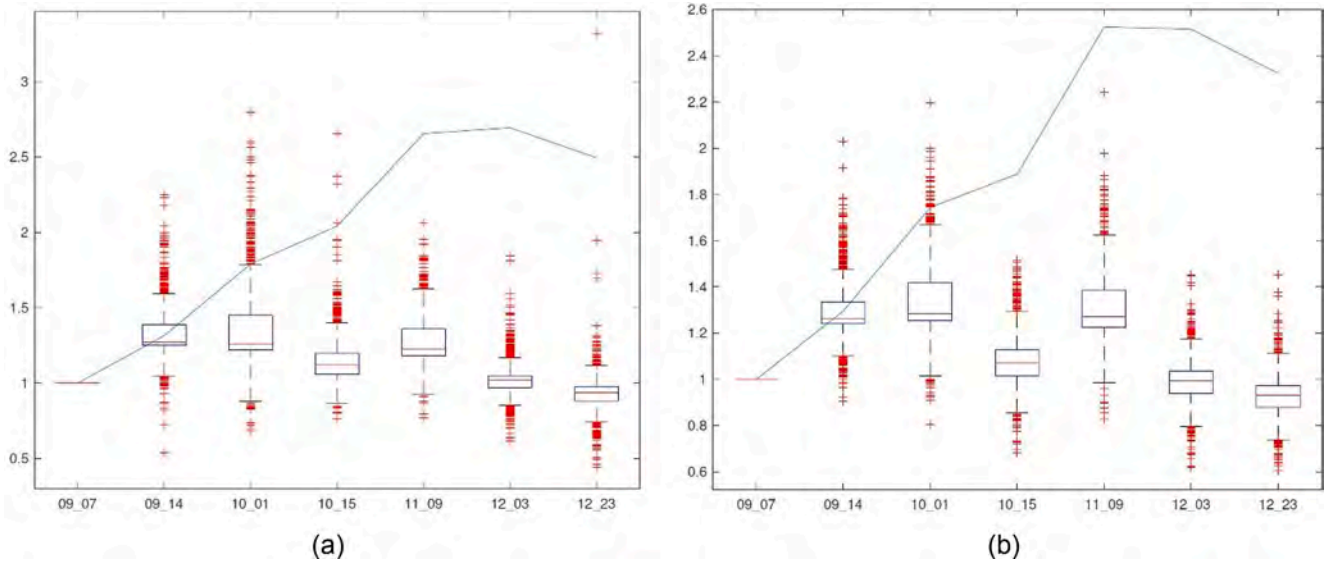

Fig. 10. Graphic plots of the growing index (GI) evolution for the FPN (a) and the YOLOv5 (b) networks. The whisker plot shows the range of variations between two consecutive dates of the experiment. The GI is measured as the ratio between the sizes of the bounding box collected at the different times of the experiment. The blue line represents the cumulative average index. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

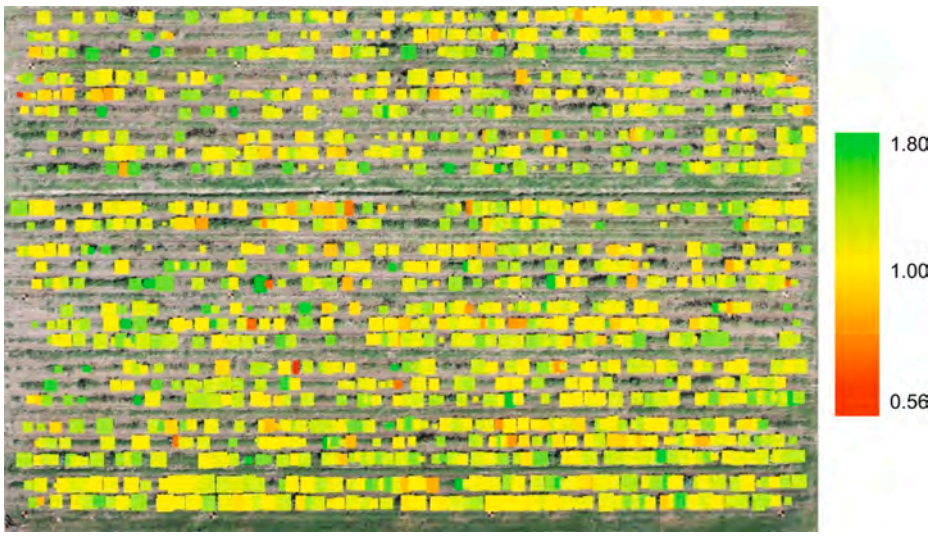

Fig. 11. The heat-map of the GI (from 1 October to 15 October of the dataset). The color scale (from red to green) is shown with the range values (from a minimum of 0.55 to the maximum of 1.8). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

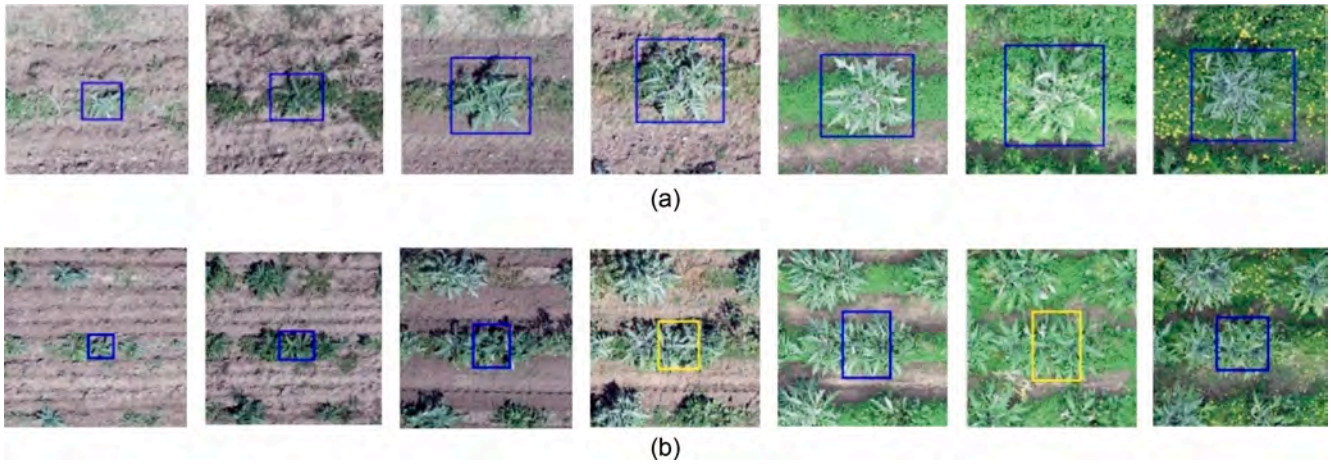

Fig. 12. History-map for each individual artichoke plant, over the full data-set; (a) simple case of an isolated plant; (b) more common case of multiple plants and their growing process (yellow boxes correspond to prediction results of the monitoring process). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 9
Occupancy rate and average size of the box-plants along the rows during the development of the crop for each network.

		FPN		YOLOv5	
Date	BBCH	Occupancy rate (%)	Avg. plant size (cm)	Occupancy rate (%)	Avg. plant size (cm)
7 Sept	14	30.03	53	28.37	57
14 Sept	16	40.38	65	39.31	71
1 Oct	35	52.34	85	52.36	93
15 Oct	47	57.90	95	56.78	97
9 Nov	51	65.59	122	66.52	127
3 Dec	55	66.06	119	66.21	125
23 Dec	59	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 (Fig. 11). A similar trend is also visible for the measure of precision which is over 90% in the early dates.

Fig. 13. 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).

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 (Fig. 13).

Results obtained on frailejones plants (family of Asteraceae) by a novel Scale Sequence Residual U-Net deep learning method showed Precision, Recall, and F1 results similar to the one reported in this work. It is worth noting that the obtained results refer to a spontaneous plant with geometric and colors (higher contrast between plants and soil background) features different from the monitored artichoke plants, which could have positively influenced plant detection. The work by (Fan et al., 2018) illustrates a three stages plant detection Convolutional Neural Network (CNN) algorithm from UAS images. The proposed methodology relies on tobacco plants identification inside candidate regions, previously identified in the first stage through image morphological operations and watershed segmentation techniques. This procedure allowed the algorithm to focus on specific areas of interest over the rows, excluding the inter-row and the surrounding soil and spontaneous plants influence, differently from the methodology proposed in this work which operates a detection on the entire field on a multi-temporal scale, and with inter-rows covered by spontaneous plants. The bare soil, high contrast, and the color difference between the target plants and the soils are crucial aspects of plant detection, as reported by (Etienne et al., 2021). The YOLOv3-based weed detection system implementation helped detect monocot and dicot weeds within corn and soybean fields, especially when the emergent crops were of similar color and size. The real-time peanut counting model proposed by (Lin et al., 2022) based on a video analysis by an improved YOLOv5 algorithm showed a 98.08% precision, with a seedling detection five times faster than the one obtained by the operator. A similar on the fly approach by (James and Bradshaw, 2020) showed invasive plant detection feasibility in the field with a commercially available drone integrated with a deep learning model and its applicability to other plant species.

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 we may obtain a complete history of the evolution of the plant starting from the first observation acquired. 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 monitoring 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 9. 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 seven 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 Fig. 6, 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 (Fig. 12), 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.

From the performance analysis carried out in section 3, it has been proved that the two considered network models (FPN and YOLOv5) satisfy the general requirements of our application. In both cases the surface indexes (SR and SP) demonstrate a high leaf area coverage of the plants and a minimum amount of detection area outside them. Anyway, we concluded YOLOv5 to be the preferred solution, especially in terms of higher precision. Moreover, considering a possible on-board real-time implementation, significant advantages were found using YOLOv5 instead of FPN, as shown in Fig. 8. This real-time implementation, even if characterized by a preliminary complex and time-consuming process of data processing and training, represents a fundamental means to optimize operations of UAS agrochemicals spraying. The entire proposed procedure is a first step towards the development of 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 (Carlevaro and Mongelli, 2022) and counterfactual explanation (Carlevaro et al., 2022), 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 images through the Nvidia Jetson Nano board, and showing comparable results with the YOLOv5 network. Tests showed a marked mAP disparity between the two networks, delineating YOLOv5 as consistently performing better throughout the growing season. The proposed automatic multitemporal monitoring 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.

CRediT authorship contribution statement

Alberto Sassu: Methodology, Validation, Investigation, Writing – original draft, Writing – review & editing. **Jacopo Motta:** Methodology, Software, Data curation. **Alessandro Deidda:** Methodology, Validation, Writing – review & editing, Visualization. **Luca Ghiani:** Methodology, Software. **Alberto Carlevaro:** Methodology, Software, Investigation, Writing – original draft. **Giovanni Garibotto:** Methodology, Data curation. **Filippo Gambella:** Investigation, Writing – original draft, Writing – review & editing, Resources, Supervision, Project administration, Funding acquisition, Methodology, Validation.

Declaration of Competing Interest

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

Data availability

The data that has been used is confidential.

Acknowledgement

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 COMP4-DRONES [grant number 826610] and ATLANTIDE project - Advanced Technologies for LANds management and Tools for Innovative Development of an EcoSustainable agriculture, Interdepartmental Center IA - INNOVATIVE AGRICULTURE Loc. Surigheddu, 07041 Alghero (SS), SS 127 bis, Km 28,500 - CUP: J88D20000070002.

References

Aeberli, A., Johansen, K., Robson, A., Lamb, D.W., Phinn, S., 2021. Detection of Banana Plants Using Multi-Temporal Multispectral UAV Imagery. Remote Sens. (Basel) 13. https://doi.org/10.3390/rs13112123.

Archontoulis, S.V., Struik, P.C., Vos, J., Danalatos, N.G., 2010. Phenological growth stages of Cynara cardunculus: codification and description according to the BBCH scale. Ann. Appl. Biol. 156, 253–270. https://doi.org/10.1111/j.1744-7348.2009.00384.x.

Bochkovskiy, A., Wang, C.-Y., Liao, H.-Y.M., 2020. YOLOv4: Optimal Speed and Accuracy of Object Detection. https://doi.org/10.48550/ARXIV.2004.10934.

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

Carlevaro, A., Mongelli, M., 2022. A New SVDD Approach to Reliable and Explainable AI. IEEE Intell. Syst. 37, 55–68. https://doi.org/10.1109/MIS.2021.3123669.

Chavarri, M.J., Herrera, A., Ariño, A., 2004. Pesticide residues in field-sprayed and processed fruits and vegetables. J. Sci. Food Agric. 84, 1253–1259. https://doi.org/10.1002/jsfa.1791.

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

De Bortoli, L., Marsi, S., Marinello, F., Carrato, S., Ramponi, G., Gallina, P., 2022. Structure from Linear Motion (SfLM): An On-the-Go Canopy Profiling System Based on Off-the-Shelf RGB Cameras for Effective Sprayers Control. Agronomy. https://doi.org/10.3390/agronomy12061276.

Etienne, A., Ahmad, A., Aggarwal, V., Saraswat, D., 2021. Deep Learning-Based Object Detection System for Identifying Weeds Using UAS Imagery. Remote Sens. (Basel) 13. https://doi.org/10.3390/rs13245182.

European Parliament, C. of the E., 2009. 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 1–16.

Fadda, A., Virdis, A., Barberis, A., Ledda, L., Melito, S., 2020. Impact of different photoperiodic treatments on "Spinoso Sardo" globe artichoke (Cynara cardunculus L. var. scolymus Fiori) head traits and elementary composition. Acta Hortic. 1284, 131–136. https://doi.org/10.17660/ActaHortic.2020.1284.17.

Fan, Z., Lu, J., Gong, M., Xie, H., Goodman, E.D., 2018. Automatic Tobacco Plant Detection in UAV Images via Deep Neural Networks. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 11, 876–887. https://doi.org/10.1109/JSTARS.2018.2793849.

Feng, Q., Yang, J., Liu, Y., Ou, C., Zhu, D., Niu, B., Liu, J., Li, B., 2020. Multi-Temporal Unmanned Aerial Vehicle Remote Sensing for Vegetable Mapping Using an Attention-Based Recurrent Convolutional Neural Network. Remote Sens. (Basel) 12. https://doi.org/10.3390/rs12101668.

Gerhards, R., Oebel, H., 2006. Practical experiences with a system for site-specific weed control in arable crops using real-time image analysis and GPS-controlled patch spraying. Weed Res. 46, 185–193. https://doi.org/10.1111/j.1365-3180.2006.00504.x.

Gerhards, R., Andújar Sanchez, D., Hamouz, P., Peteinatos, G.G., Christensen, S., Fernandez-Quintanilla, C., 2022. Advances in site-specific weed management in agriculture—A review. Weed Res. 62, 123–133. https://doi.org/10.1111/wre.12526.

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

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

James, K., Bradshaw, K., 2020. Detecting plant species in the field with deep learning and drone technology. Methods Ecol. Evol. 11, 1509–1519. https://doi.org/10.1111/2041-210X.13473.

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

Kalal, Z., Mikolajczyk, K., Matas, J., 2010. Forward-backward error: Automatic detection of tracking failures. In: 2010 20th international conference on pattern recognition. IEEE, pp. 2756–2759. https://doi.org/10.1109/ICPR.2010.675.

Kamilaris, A., Prenafeta-Boldú, F.X., 2018. Deep learning in agriculture: A survey. Comput. Electron. Agric. 147, 70–90. https://doi.org/10.1016/j.compag.2018.02.016.

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

Khan, S., Tufail, M., Khan, M.T., Khan, Z.A., Iqbal, J., Wasim, A., 2021. Real-time recognition of spraying area for UAV sprayers using a deep learning approach. PLoS One 16, 1–17. https://doi.org/10.1371/journal.pone.0249436.

Kriflik, L.S., Yeatman, H., 2005. Food scares and sustainability: A consumer perspective. Health Risk Soc. 7, 11–24. https://doi.org/10.1080/13698570500042439.

Lan, Y.B., Chen, S.D., Fritz, B.K., 2017. Current status and future trends of precision agricultural aviation technologies. Int. J. Agric. Biol. Eng. 10, 1–17. https://doi.org/10.3965/j.ijabe.20171003.3088.

Lin, Y., Chen, T., Liu, S., Cai, Y., Shi, H., Zheng, D., Lan, Y., Yue, X., Zhang, L., 2022. Quick and accurate monitoring peanut seedlings emergence rate through UAV video and deep learning. Computers and Electronics in Agriculture 197, 106938. https://doi.org/10.1016/j.compag.2022.106938.

Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S., 2017. Feature Pyramid Networks for Object Detection. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp. 936–944. https://doi.org/10.1109/CVPR.2017.106.

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

Meng, J., Jiang, P., Wang, J., Wang, K., 2022. A MobileNet-SSD Model with FPN for Waste Detection. J. Electr. Eng. Technol. 17, 1425–1431. https://doi.org/10.1007/s42835-021-00960-w.

Osco, L.P., Junior, J.M., Ramos, A.P.M., de Castro Jorge, L.A., Fatholahi, S.N., de Andrade Silva, J., Matsubara, E.T., Pistori, H., Gonçalves, W.N., Li, J., 2021. A review on deep learning in UAV remote sensing. Int. J. Appl. Earth Obs. Geoinf. 102, 102456 https://doi.org/10.1016/j.jag.2021.102456.

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

Popp, J., Pető, K., Nagy, J., 2013. Pesticide productivity and food security. A review. Agron. Sustain. Dev. 33 https://doi.org/10.1007/s13593-012-0105-x.

Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look once: Unified, real-time object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 779–788. https://doi.org/10.48550/ARXIV.1506.02640.

Sarri, D., Martelloni, L., Rimediotti, M., Lisci, R., Lombardo, S., Vieri, M., 2019. Testing a multi-rotor unmanned aerial vehicle for spray application in high slope terraced vineyard. J. Agric. Eng. 50, 38–47. https://doi.org/10.4081/jae.2019.853.

Spanu, E., Deligios, P.A., Azara, E., Delogu, G., Ledda, L., 2018. Effects of alternative cropping systems on globe artichoke qualitative traits. J. Sci. Food Agric. 98, 1079–1087. https://doi.org/10.1002/jsfa.8558.

Tabikha, R.M., Draz, A.K., 2022. Population Dynamics of Capitophorus Elaeagni (Hemiptera: Aphididae) and Its Associated Predators on Artichoke Plants in El-Behera. Alex. Sci. Exch. J. 43, 187–197. https://doi.org/10.21608/asejaiqjsae.2022.230544.

Van den Berg, H., Gu, B., Grenier, B., Kohlschmid, E., Al-Eryani, S., da Silva Bezerra, H.S., Nagpal, B.N., Chanda, E., Gasimov, E., Velayudhan, R., Yadav, R.S., 2020. Pesticide lifecycle management in agriculture and public health: Where are the gaps? Sci. Total Environ. 742, 140598 https://doi.org/10.1016/j.scitotenv.2020.140598.

Xue, X., Lan, Y., Sun, Z., Chang, C., Hoffmann, W.C., 2016. Develop an unmanned aerial vehicle based automatic aerial spraying system. Comput. Electron. Agric. 128, 58–66. https://doi.org/10.1016/j.compag.2016.07.022.