Enhancing visual autonomous navigation in row-based crops with effective synthetic data generation

Precision Agriculture
https://doi.org/10.1007/s11119-024-10157-6

Enhancing visual autonomous navigation in row-based crops with effective synthetic data generation

Mauro Martini[1] · Marco Ambrosio[1] · Alessandro Navone[1] · Brenno Tuberga[1] · Marcello Chiaberge[1]

Accepted: 3 June 2024

Abstract

Introduction Service robotics is recently enhancing precision agriculture enabling many automated processes based on efficient autonomous navigation solutions. However, data generation and in-field validation campaigns hinder the progress of large-scale autonomous platforms. Simulated environments and deep visual perception are spreading as successful tools to speed up the development of robust navigation with low-cost RGB-D cameras.

Materials and methods In this context, the contribution of this work resides in a complete framework to fully exploit synthetic data for a robust visual control of mobile robots. A wide realistic multi-crops dataset is accurately generated to train deep semantic segmentation networks and enabling robust performance in challenging real-world conditions. An automatic parametric approach enables an easy customization of virtual field geometry and features for a fast reliable evaluation of navigation algorithms.

Results and conclusion The high quality of the generated synthetic dataset is demonstrated by an extensive experimentation with real crops images and benchmarking the resulting robot navigation both in virtual and real fields with relevant metrics.

Keywords Service robotics · Autonomous navigation · Synthetic data · Deep semantic segmentation

✉ Mauro Martini
mauro.martini@polito.it

Marco Ambrosio
Marco.Ambrosio@polito.it

Alessandro Navone
Alessandro.Navone@polito.it

Brenno Tuberga
Brenno.Tuberga@polito.it

Marcello Chiaberge
Marcello.Chiaberge@polito.it

[1] Department of Electronics and Telecommunications, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Turin, Italy

Published online: 11 June 2024

Introduction

In recent years, the need for sustainable and efficient agriculture has become fundamental to meet food demand. A report from the Food and Agriculture Organization (FAO) (The Future of Food and Agriculture, 2022) forecasts that the demand for agricultural products will grow by about one-third of the actual demand by 2050. Its trends forecast that the world's population will grow by around 10 billion people by that year, leading to an expected increase in world food demand between 36 and 56% (Van Dijk et al., 2021). A modest global economic growth scenario in low- and middle-income countries would lead to a diet transition from a cereal-based one towards a higher consumption of vegetables, fruit, and meat. According to the FAO report, meeting the increased demands with the currently adopted farming technologies would likely lead to ecologically unsustainable due to greenhouse gas emissions, deforestation, and land degradation. In addition, due to climate changes, intense pesticide and fertilizer utilization is needed (Shafi et al., 2019), leading to soil pollution and water contamination, affecting nature and human health (Singh et al., 2018).

Precision agriculture has emerged as a valid alternative to traditional farming technologies in the last two decades. Four fundamental requirements can be associated with this technology, such as increasing productivity, allocating resources reasonably, adapting to climate change, and reducing food waste (Zhai et al., 2020). Precision agriculture is often led by trending and latest technology, such as mobile robotics, Artificial Intelligence (AI), the Internet of Things (IoT), and computer vision (Yépez-Ponce et al., 2023) which allow enhanced data collection and real-time monitoring of the crops, improving decision-making with the sake of reducing the overall environmental impact. Megeto et al. (2021).

Autonomous mobile systems, joined with AI tools and Deep Learning (DL), can provide a significant enhancement and competitive advancement in several agricultural tasks, reducing human labor and improving operational safety. Indeed, recent studies investigated the possibilities brought by new technologies joined together to design elaborate solutions for different tasks such as harvesting (Droukas et al., 2023), spraying (Deshmukh et al., 2021), yield estimation (Maheswari et al., 2021; Mazzia et al., 2020), disease detection (Ferentinos, 2018; Shruthi et al., 2019) and crop detection and monitoring (Comba et al., 2019; Vidović et al., 2016). Various DL strategies have been proposed to handle the autonomous navigation problem by overcoming localization restrictions in row crop settings, often integrating waypoint generation (Salvetti et al., 2023) with plant segmentation methods (Aghi et al., 2021) or local traversal path estimation (Liu et al., 2023) for intra-row control. Other ways include creating a navigation policy and training Deep Reinforcement Learning agents (Zhu & Zhang, 2021) to guide the robot along the rows, as well as directly mapping images to velocity commands (Martini et al., 2022; Wang et al., 2023). Semantic segmentation is one of the most widely used perceptual approaches among all the deep learning solutions developed (Ren et al., 2020); it is applied to agricultural images (Luo et al., 2023), and is used to identify objects on various scales: individual fruits or branches (Kestur et al., 2019; Peng et al., 2020), crop rows (Aghi et al., 2021), and entire fields (Raei et al., 2022; Song et al., 2020).

One essential component required to carry out each of these jobs is a robust and accurate navigation system. However, the weather, illumination, uneven terrain, and plants

present a number of specific problems for autonomous navigation in agricultural scenarios. GNSS sensors, such as receivers with RTK corrections, are frequently used to achieve localization in agricultural fields (Thuilot et al., 2002). Nevertheless, severe environmental factors like wide canopies may reduce the dependability of GNSS sensors inside rows with dense vegetation, particularly in the spring and summer (Kabir et al., 2016), making GPS-free approaches competitive even with the recent improvements in GPS receiver precision.

In this research, we have focused on proving the effectiveness of novel instruments for producing artificial intelligence in simulation scenarios. The main objective is to demonstrate how these tools may be used to create reliable Deep Learning (DL)-based autonomous navigation algorithms that are tailored to the needs of Unmanned Ground Vehicles (UGVs) navigating through row-based crops with the least amount of employment on real images. Moreover, virtual environments can serve as a common tool for validating and benchmarking task-specific methodologies and solutions.

The focus on creating synthetic data originates from the necessity of having sizable and varied datasets in order to efficiently train and optimize deep learning models. Through modeling different situations in farming environments, we hope to build an extensive artificial dataset that reflects various conditions encountered in row crops. This method reduces the limitations related to data collecting in real-world agricultural situations by speeding up the training process and lowering reliance on hand-labeled images. A contiguous amount of datasets (Barth et al., 2018; Häni et al., 2020) are increasingly been collected and, sometimes, shared by researchers, with a great focus on weed (Su et al., 2021) and fruit (Maheswari et al., 2021) detection. However, no similar datasets exist nor are available in the precision agriculture literature and community, although data are fundamental elements to drive the development of precise low-cost navigation algorithms for the realization of any automatic agricultural task.

The specific contributions of this study can summarized in:

- An extensive, realistic, multi-crop synthetic dataset for row plants binary segmentation;
- A procedural tool to automatically generate simulation fields in Gazebo according to geometrical parameters specified by the user for enhancing fast validation of algorithms in a virtual framework;
- A common benchmark for visual autonomous navigation of robotic platforms in row-based crops.

Figure 1 illustrates the complete pipeline proposed in the study. The dataset and the methods presented in this work have already been used in preliminary studies that demonstrate the significance and quality of the synthetic data and approach introduced so far. Navone et al. (2023a, 2023b). The first version of the study was presented in 2023 (Martini et al., 2023; Stafford et al., 2023), here we significantly enriched the dataset and the experimental results presented. Indeed, extensive experimentation has been conducted and illustrated to validate the quality of the synthetic images for crop rows perception, together with the robustness of the derived semantic segmentation-based control to guide the robot in real-world autonomous operations.

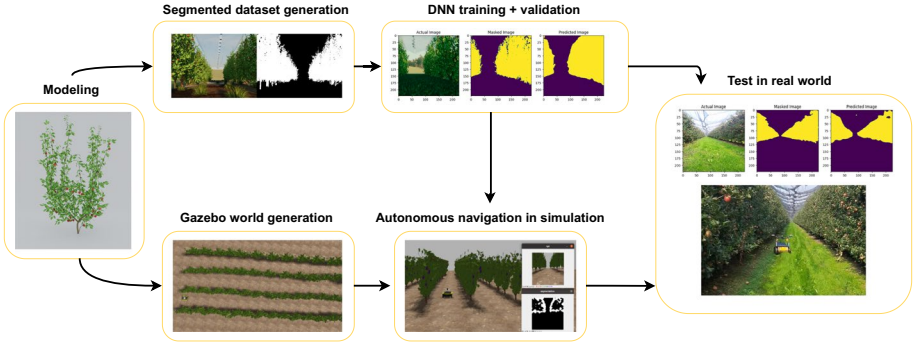


Fig. 1 An illustrative scheme describing the proposed complete pipeline for synthetic data generation and segmentation-based visual navigation of service robots in row-based crops. From the left to the right: modeling single plants is the first step, thanks to that, it is possible to generate labeled datasets for semantic segmentation and a virtual environment for testing robot navigation in an automatic way. As a final step, the developed visual navigation algorithms can be tested in the real world

Modeling and data generation

This section focuses on describing the main contributions of this work. Firstly, the necessary steps to generate the RGB and label mask images of zucchini, lettuce, chard, pear trees, and other fields with Blender,[1] a free and open-source 3D computer graphics software tool set, are illustrated. Then, the tool to procedurally generate fields according to geometrical parameters is presented and the simulation scenarios obtained in Gazebo[2] (Koenig et al., 2004), an open-source 3D robotics simulator, are used to test navigation algorithms.

Dataset generation for semantic segmentation

The first step to generate a realistic synthetic dataset for segmentation is a detailed plant model. The 3D plant models have been developed in Blender using real plant textures and standard dimensions as references. The height of the crop is considered a fundamental factor since it determines the level of obstruction of the GNSS signal, as well as the control strategy and the machine size to adopt. To consider the widest range of crops, three main categories have been identified: low crops, such as lettuce and chard, which have a height of 20–25 cm; medium crops, like zucchini, that reach 60–100 cm; and tall crops, from vineyard to fruit trees, which can reach 1.0–2.5 m and may cause GNSS signal obstruction. Some examples of 3D plant models are shown in Fig. 2. A significant factor for successfully testing navigation algorithms in simulated fields is the model of the terrain. Ground irregularity causes the rover to drift or to get stuck. Hence it requires special counteractions in developing a trajectory control strategy. For realistic terrain modeling, a plane is subdivided into multiple polygons in Blender. Then, the polygon vertices are randomly moved, respecting real proportions, to get a bumpy, irregular surface.

[1] https://www.blender.org/

[2] https://classic.gazebosim.org/

Fig. 2 Detailed example of synthetic 3D crop models realized to build the Dataset. Lavender (top) and vines (bottom) on the left. On the right lettuce (top) and autumn tree (bottom)

Fig. 3 Sample images of fields generated in Blender for a realistic rendering. The top row shows vineyards and the bottom row fruit trees during different seasons

Moreover, an additional sky model is added to the scene to have realistic visual features for background and illumination. Multiple models are employed to simulate different

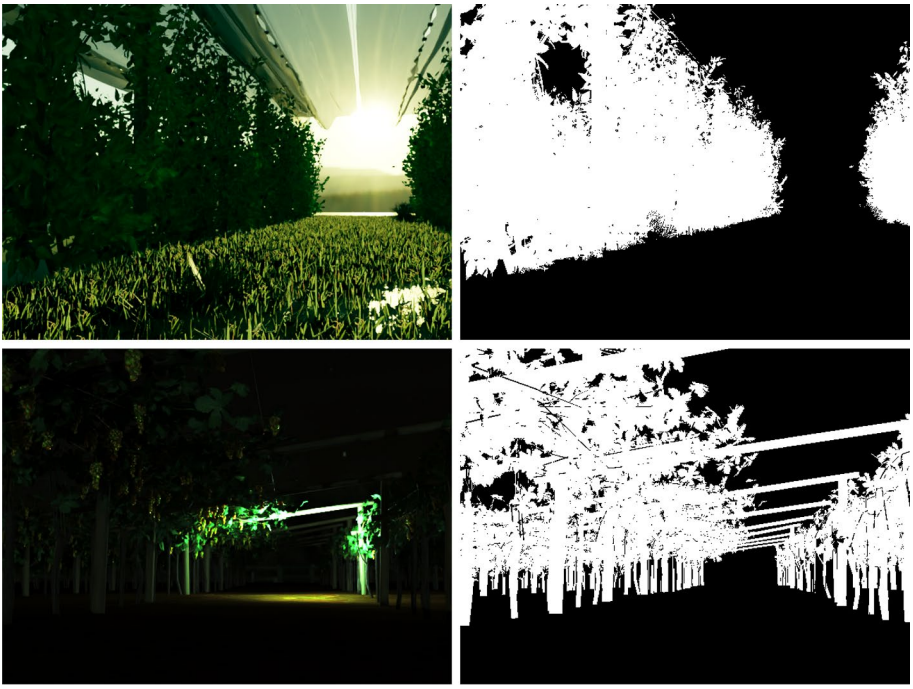


Fig. 4 Sample images of RGB and segmentation masks generated in Blender for an accurate automatic labeling process. The upper row shows an apple trees field with a cloth in backlighting conditions and its associate binary segmentation mask. The bottom row shows a pergola vineyard during dark hours with a spot light and its complex ground truth mask

Table 1 Detailed properties of each crop dataset

Crop	Samples	Type	Category	Height [m]
Lettuce	4800	Synthetic	Low	0.22
Chard	4800	Synthetic	Low	0.25
Lavender	5260	Synthetic	Low	0.3
Zucchini	19200	Synthetic	Medium	0.6
Cotton	4800	Synthetic	Medium	0.6
Vineyard	6165	Synthetic	Medium/tall	1.5–2.5
Pergola Vineyard	4800	Synthetic	Tall	3.2
Apple Tree	9600	Synthetic	Tall	2.7
Pear Tree	5205	Synthetic	Tall	3.0
Generic Tree 1	4800	Synthetic	Tall	4.5
Generic Tree 2	2785	Synthetic	Tall	4.5
Lavender	90	Real	Low	0.3
Vineyard	500	Real	Medium/Tall	1.5–2.5
Apple Tree	210	Real	Tall	2.5
Pear Tree	140	Real	Tall	3.0
Miscellaneous	100	Real	Any	Any

The section on the top reports the synthetic crops datasets generated in simulation, while the section on the bottom the real-world ones

times of the day, weather, and lighting conditions, to boost the generalization property of the segmentation neural network and avoid unexpected failures. Some examples are shown in Fig. 3. Once the environmental conditions are set in the field model, a dataset of RGB images and the associate binary segmentation masks can be generated by exploiting the Blender Python scripting functionality, which automatically divides the plants from the rest of the image. Figure 4 reports two examples of RGB and associated binary segmentation masks generated through Blender, otherwise impossible to label by hand at such level of accuracy. In this work, an extended version of the dataset is presented, composed of image samples for a wide range of crops, reported in detail in Table 1. Each dataset presents at least four sub-datasets that differ in terms of appearance and atmospheric conditions. Cloudy and sunny skies, diverse lighting, terrains, and shadows are extensively changed and combined to generate the widest offer of realistic field environments. For each sub-dataset, the camera pose, both position and orientation, has been changed to acquire diversified image samples along the whole field. Moreover, the combinations of the elements described above are further exaggerated in some smaller sub-datasets by, for example, increasing the intensity of the scene lights, using more unusual sky patterns, and increasing the presence of features on the ground. In addition to this, color correction is also performed on the entire image, increasing or decreasing specific hues to simulate more extreme, less realistic conditions. This additional process is done to increase generalization in training of the segmentation model, to achieve the most accurate gait possible even under very unusual lighting, such as in shadows or backlight conditions.

Real dataset generation

Real data have been collected for lavender, vineyard, apple and pear tree over several campaign on different fields. Thus, the images present at least two fields and environmental conditions for each crop. Nonetheless, an additional dataset of only 100 images containing miscellaneous crops, trees and high crops for the majority, have been assembled using open source low resolution images. The *Real Vineyard* dataset was originally presented in Aghi et al. (2021), but the proposed labels were coarse, hence, we re-label the samples. All real data have been labeled by a human using the *SALT* labeling tool[3] based on Segment Anything (Kirillov et al., 2023).

Lavender data are limited to a sub-set of the available images. Image were collected over three different campaign, and in the third one bad weather and field conditions drastically affect the quality of the images, proved by the fact that also human annotator faced difficulty in discerning plants from the rest. The full dataset is available at https://naspic4ser.polito.it/files/sharing/P4ZwMHt7n.

[3] https://github.com/anuragxel/salt

Procedural field generation

Algorithm 1 GenerateCropField(L_{row}, n_{rows}, n_{group}, d_{RR}, d_{rr}, d_{pp})

procedure GENERATECROPFIELD(L_{row}, n_{rows}, n_{group}, d_{RR}, d_{rr}, d_{pp})
    plantsGrid $\leftarrow$ [ ]
    terrainHeightMap $\leftarrow$ **LoadTerrainMesh**( )
    $nPlants \leftarrow$ **round**(L_{row}/d_{pp})
    $x \leftarrow 0$
    for i **from** 1 **to** n_{rows} **do**
        $y \leftarrow 0$
        for j **from** 1 **to** nPlants **do**
            $x^* \leftarrow x +$ **sample_triangular**$(-1, 1, 0) \times d_{pp}/10$
            $y^* \leftarrow y +$ **sample_triangular**$(-1, 1, 0) \times d_{rr}/10$
            $z^* \leftarrow$ **interpolate2d**(terrainHeightMap, $[x^*, y^*]$)
            append $[x^*, y^*, z^*]$ **to** plantsGrid
            $y \leftarrow y + d_{pp}$
        end for
        if i **reminder** $n_{group} = 0$ **then**
            $x \leftarrow x + d_{RR}$
        else
            $x \leftarrow x + d_{rr}$
        end if
    end for
end procedure

Testing in simulation is a fast and inexpensive way to evaluate the performance of new algorithms before engaging in complex real-world experiments. As a matter of fact, Gazebo is one of the most adopted simulators in robotics as it is extensively supported by ROS community and for this reason it is adopted also in this work. Unfortunately, modelling an entire row crop field using the Gazebo interface may result in a repetitive and time consuming process because of the large number of plants that must be placed. This method aims to streamline the process of modeling agricultural landscapes by utilizing a procedural script that generates Gazebo worlds based on a set of specified parameters and input models of individual plants and terrains. The approach outlined in Algorithm 1 relies on various key parameters such as the average row length L_{row}, total number of rows n_{rows}, row grouping n_{group}, distances between rows d_{rr} and groups d_{RR}, and plant-to-plant spacing d_{pp}, allowing for flexibility in modeling different types of row crop configurations. Figure 5 illustrates an example of crop field generated according to specific geometrical parameters. Algorithm 1 serves as the cornerstone of this method. Its output provides a grid containing positions for all plants, generated from a regular grid and then randomized using a triangular distribution. This resulting grid forms the basis for creating Gazebo worlds and Blender models, enhancing the versatility of the approach across simulation platforms. Additionally, a method is implemented to facilitate the conversion of plant models from Blender to Gazebo format, further extending the compatibility and usability of the generated models.

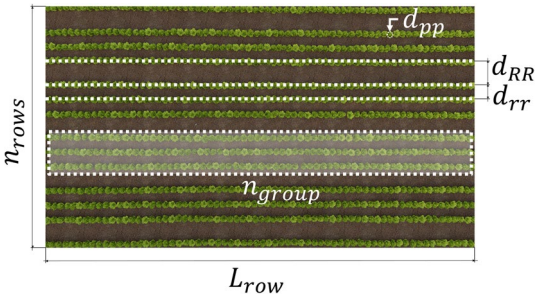


Fig. 5 Illustration of a row-based crop world realized in Gazebo according to specific geometric settings such as row-to-row distance, distance between groups of rows, and plant-to-plant distance

Table 2 Geometric features of the row-based fields realized for zucchini, lettuce, chard and pear trees

	Terrain			Plant			Field			
	Lenght	Width	$\lvert\Delta H\rvert$	Lenght	Width	Height	d_{rr}	d_{RR}	d_{pp}	Rows
Zucchini	60	38	0.20	0.82	0.90	0.60	1.80	3.60	0.70	7
Lettuce	60	25	0.25	0.38	0.34	0.22	0.70	1.40	0.40	3
Chard	60	12	0.25	0.25	0.40	0.25	0.70	1.40	0.30	3
Pears	80	45	0.30	1.40	2.20	3.20	5.00	5.00	2.20	1

$\lvert\Delta H\rvert$ indicates the maximum difference in height in the terrain. All the values are expressed in meters except for the number of rows

For accessibility and reproducibility, the code implementing this method is available on GitHub at the specified URL: https://github.com/PIC4SeR/AutomaticRowCropGenerator.

Simulation environments for autonomous navigation

The complete field models are translated into Gazebo worlds to easily evaluate autonomous navigation algorithms in simulation. Since the scope of the virtual framework is navigation, a simulation with complex visual meshes may result in an unnecessarily high computational cost. Hence, Blender models for plants and terrain are simplified and exported in Object format (.obj). This process has been conducted to generate the meshes for creating Gazebo worlds of zucchini, lettuce, chard, pear trees, and vineyards. Table 2 shows the resulting geometric features of the generated fields. Terrain, plant, and field geometry are considered field's the most relevant descriptive factors. Moreover, other important realistic features have been embedded in the fields, such as small obstacles like fallen branches or stones, terrain slope $\lvert\Delta H\rvert$, and physical elements such as stones, mud, uncultivated and uncultivated grass, snow. The complete collection of models and worlds realized for Gazebo is available at https://naspic4ser.polito.it/files/sharing/LeqCGJYp6.

Methods

The following section presents a methodology to validate the quality of synthetic data produced. First, we employ the produced synthetic RGB dataset with relative segmentation masks to train an efficient neural network that can segment target plants in RGB images.

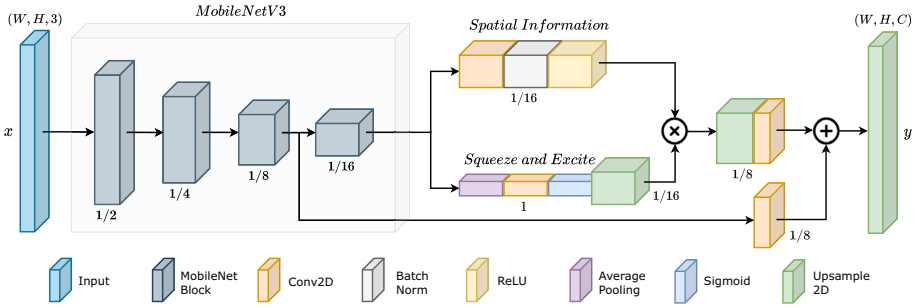


Fig. 6 The Deep Neural Network used for the study. It presents a MobileNetV3 backbone and LR-ASPP head (Howard et al., 2019). Below each block, spatial scaling factor of the features compared to the input size is reported

This initial experiment aims to prove that a synthetic dataset can generalize well over real RGB images. In the second part of the experimentation, we use a simple segmentation-based control algorithm to navigate robots through row crops, both in simulation using Gazebo, and in real-world fields. Consequently, we establish a new benchmark with relevant metrics.

Semantic segmentation DNN

The adopted neural network architecture was inspired by a previous work on crop segmentation in real scenarios (Aghi et al., 2021). Its overall architecture is depicted in Fig. 6 and its main advantage is to exploit rich contextual information from the image at a reduced computational cost.

The first stage of the network consists of a MobileNetV3 backbone (Howard et al., 2019) to extract the visual feature of the input image in an efficient way. It consists of a sequence of Inverted Residual blocks (Sandler et al., 2018) with Squeeze-and-Excitation attention sub-modules (Hu et al., 2018). They progressively reduce the spatial dimensions of the input image incrementing the number of channel features.

It is followed by an improved and reduced version of the Atrous Spatial Pyramid Pooling module (R-ASPP), namely Lite R-ASPP (LR-ASPP) (Chen et al., 2017), which upscales the extracted features through two parallel branches. The first applies a Squeeze-and-Excite sub-module to the last layer of the backbone, which reduces the spatial dimension by 1/16. A channel attention weight matrix is computed and multiplied by the unpooled features before being upsampled and passed through a convolutional layer to adjust the number of channels C to the output segmentation map. The second branch mixes lower-level and higher-level patterns in the data in the upsampling stage by taking features from an earlier stage of the backbone, which reduces the spatial dimension by 1/8, and adds them to the output of the first branch.

The input of the network has a dimensionality equal to $W \times H \times 3$, while the segmented output is equal to $W \times H$.

Moreover, since this works focuses on plants rows semantic segmentation only, a sigmoid function is used to scale the output values of the neural network between 0 and 1.

The DNN is trained using the standard cross-entropy loss between the predicted segmentation mask and the ground-truth label y:

$$L_{\mathrm{CE}}(y,\hat{y}) = -\sum_{i=1}^{N} y_i \cdot log(\hat{y}_i) \tag{1}$$

which for binary segmentation becomes a simple binary cross-entropy loss.

During both validation and testing phases, the DNN performance is evaluated through an intersection over unit (IoU) metrics:

$$mIoU(\theta) = \frac{1}{N}\sum_{i=0}^{N}\left(1 - \frac{\hat{X}^i_{seg} \cap X^i_{seg}}{\hat{X}^i_{seg} \cup X^i_{seg}}\right) \tag{2}$$

where θ is the network parameters vector, $\hat{X}^i_{seg}$ is a predicted segmentation mask, and X^i_{seg} is its ground truth mask. Plants are the only target class of interest, hence, N is always equal to 1 in the computation of IoU.

Visual autonomous navigation in row-based crops

The real-time prediction of the segmentation mask from the network presented in "Semantic segmentation DNN" section is exploited for the autonomous navigation of the rover in the fields rows. The heading of the robots with respect to the row direction is obtained from the segmentation output and, then, the control commands are computed. The complete detailed description of the control algorithm is described in Aghi et al. (2021). The algorithm involves identifying the primary cluster of zeros by adding up the columns of the segmentation mask obtained. This cluster of zeros defines the free passage of the row. Once the center of the cluster is identified, its geometrical center is used to calculate the linear and velocity commands. This type of algorithm can be integrated into a comprehensive navigation framework for the entire field, which includes a waypoint generator, a global planner, and a conventional GPS-based navigation system to control the robot outside its row. Cerrato et al. (2021).

After the segmentation mask $\hat{\mathbf{X}}^i_{seg}$ is obtained from the network at the i-th time instant. To improve robustness, the segmentation masks from the last N time instants are super-imposed with a bitwise-OR operation. In order to remove possible noise coming from the segmentation and discard information coming from plants, the depth camera data is merged. First, a depth threshold d_{th} is fixed; then, segmented pixels corresponding to an object which do not respect the threshold are discarded, obtaining a segmentation mask merged with depth data $\hat{\mathbf{X}}^i_{depth}$.

Finally, the largest cluster is searched in the array obtained from the sum on the columns of $\hat{\mathbf{X}}^i_{depth}$. The output linear and angular velocity are obtained by the following formulas:

$$v_x = v_{x,max} \cdot \left[1 - \left[\frac{\left(x_c - \frac{w}{2}\right)^2}{\left(\frac{w}{2}\right)^2}\right]\right] \tag{3}$$

$$\omega_z = -\omega_{z,gain} \cdot \left(x_c - \frac{w}{2}\right) \tag{4}$$

where w is the width of the input image, and x_c is the center of the individuated largest cluster. The gain of angular velocity $\omega_{z,gain}$ is set to 3, and the maximum linear velocity $v_{x,max}$ to 1 m/s. An improved version of this algorithm has been recently presented in Navone et al. (2023a), extending the versatility of the method to trees thick canopies through a histogram

optimization approach. In this work, both the segmentation-based control methodologies can be successfully used to demonstrate the quality of the synthetic data used to train and validate the navigation system. For the sake of simplicity, experimental tests have been conducted with the first original version of the algorithm.

Experimental testing

In this section, the quality of the synthetic data generated through simulated fields is validated in separate steps. Different crops are chosen according to their availability for testing campaigns as the target of the experiments, including both low and tall plants. A particular focus has been devoted to lavender, vineyards, and fruit trees to demonstrate the significance and completeness of the study passing from synthetic data generation to real-world usage. First, the segmentation datasets are used to train and test an efficient deep neural network to segment plant rows in RGB images. This first analysis provides the necessary insights to demonstrate that our dataset allows to drastically decrease the number of real-world images labeled by hand. Second, the virtual field scenarios in Gazebo are used to test the segmentation-based control algorithm with relevant metrics and set up a new validation benchmark for visual-based navigation in row-based crops. Finally, autonomous navigation tests are performed on the field with the real robotic platform adopting the same set of metrics. This last experimental step fully demonstrates how the proposed dataset and pipeline enhance the autonomous mobility of service robots in the agricultural field.

Semantic segmentation

Lavender, vineyards, apple, and pear trees have been chosen to carry out a complete validation of the semantic segmentation dataset and navigation methods. Nonetheless, the segmentation model has been trained and validated also on other crops data, limiting the experiments to simulation tests. Following common Deep Learning practice to obtain robust models, the backbone of the DNN has been initialized with pre-trained weights on the CityScapes segmentation datasets (Cordts et al., 2016). However, no substantial difference between the models with and without the pre-trained weights have emerged at test time. Input images are resized to a resolution of 224×224 to guarantee a balanced trade-off between accuracy and computational efficiency. Significant data augmentation transformations such as horizontal flip, greyscale, brightness and contrast random variations with a probability of 0.5 have been further applied to images at training time.

The DNN has been trained for 25 epochs with the binary cross-entropy loss defined in "Semantic segmentation DNN" section, adopting the Adam optimizer with weight decay regularization (Loshchilov & Hutter, 2018). A weight decay of $1e^{-5}$ has been used in combination with a polynomial decay for the learning value, with an initial value of $5e^{-3}$ and a final value of $5e^{-5}$. The best model has been selected according to the best IoU score of predicted plant masks on the validation set.

Quantitative results and details about data are reported in Table 3 for synthetic datasets and in Table 4 for real-world images. The synthetic test sets have been generated performing a random split of 15% from total available data, and half of that quantity is further extracted from the remaining for the validation set. For vineyards, a combination of synthetic data from vineyards, pergola vineyards, and generic trees has been used. Generic tree datasets are also mixed for apple and pear trees. The resulting amount of total synthetic

images used for each crop is reported in the tables. Moreover, the network pixel-wise prediction can be associated with class 1 of the binary segmentation problem, plants, according to a confidence threshold directly applied to the output of the model. At training time, the threshold saturation value has been fixed at 0.9 to enhance the DNN prediction confidence during the learning and validation process with a more severe policy. At testing time, especially in previously unseen conditions where the prediction confidence generally decreases, a lower value of saturation threshold can improve the resulting performance of the network for navigation purposes. Uncertain predictions often contain meaningful information about the overall region where plants are contained in the image that can be fully exploited by the visual-based controller described in "Visual autonomous navigation in row-based crops" section. According to this, in Tables 3 and 4, the last column contain the value of the threshold used for testing experiment. Two different threshold values have been selected by qualitatively analyzing the resulting robot motion in the crop rows, considering different threshold values in the range [0.5–0.8]. A threshold value of 0.7 has been adopted for low crops, since a more severe output mask is required for tiny plants and narrow rows. Differently, a 0.5 threshold results to be more effective in rows of trees, where uncertain predictions are more frequent due to the presence of branches and different canopy's geometries.

IoU results on the synthetic test sets are extremely positive, despite the quite limited amount of data used for the fine-tuning of the network. Figure 7 visually shows the predicted segmentation masks on a collection of synthetic test samples for vineyards, lavender and apple trees. Realistic elements of the environment can be noticed in the images: for lavender, a dark lighting condition is reported, whilst the vineyard presents shadows on the ground and apple trees, the typical cloth used to protect the plants. Overall, the network demonstrates a brilliant response also to challenging environmental visual variations.

As expected, with real-world crop row images, the model demonstrates lower performance but an overall acceptable result. As additional experiments, only 100 real images containing miscellaneous crops, mostly generic trees, have been added to the initial synthetic set used for training and validation. This small amount of labeled real data significantly boosts the performance of the network, especially for apple and pear trees, which also rich an IoU of 0.8778 on real images. Thus, it demonstrates that following our training process, the amount of real hand-labeled data is drastically reduced to achieve remarkable results. Figure 8 shows some real-world predictions on the same set of crops, obtained with the best crop-specific models of Table 4. Two image samples per crop are reported this time. The segmentation DNN fully demonstrates its robust performance with real data, being trained with a synthetic dataset plus a maximum of 100 real images. Different lighting conditions and camera views do not represent significant challenges for the network to generalize, thanks to the wide variety of image samples present in the training dataset. Among the tested datasets, lavender proved to be the most challenging. First, small green plants are much more difficult to distinguish from the terrain, also considering that the small amount of real images used for training mostly contain generic trees. Second, weather and lighting conditions were much more adverse during the data collection campaign.

Computational efficiency plays a crucial role in every edge device application, such as robotics perception. For this reason, the resulting segmentation models are converted in TFLite[4] to enhance inference performance on the CPU. The TFLite conversion boosts the average inference speed from 22 frame-per-second (fps) to more than 55 fps on a laptop

[4] https://www.tensorflow.org/lite?hl=it

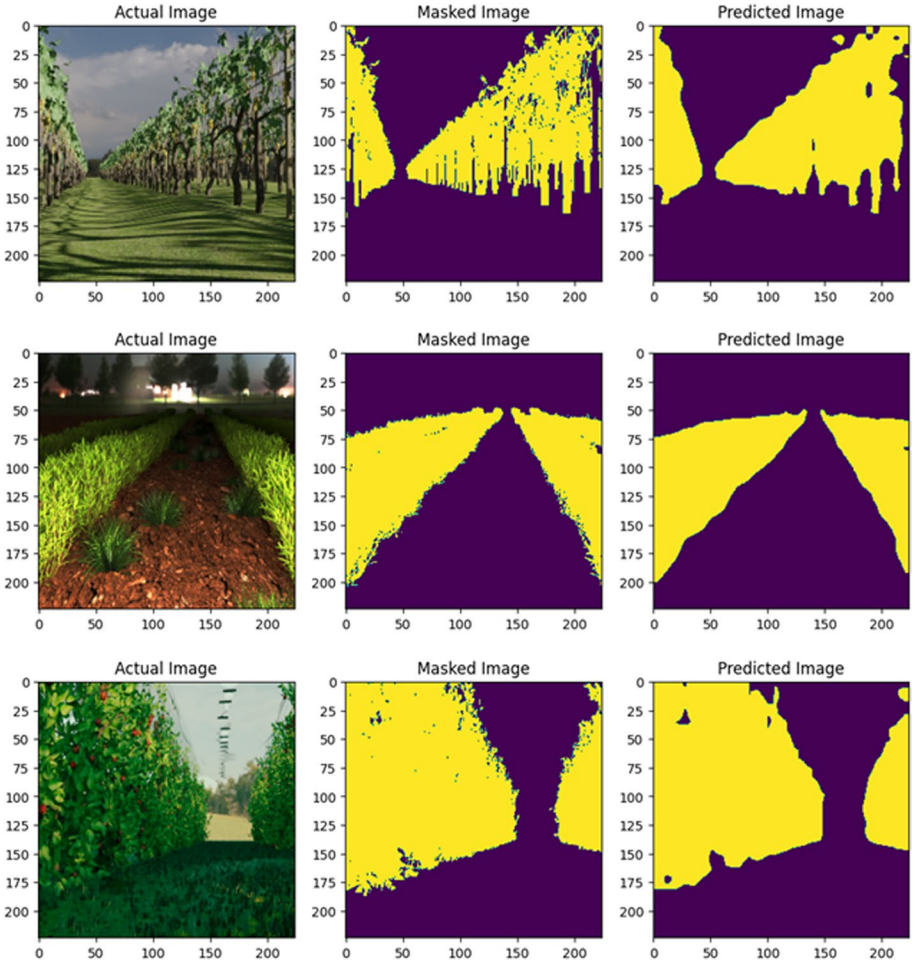


Fig. 7 Test of semantic segmentation DNN on synthetic test samples from vineyard (top), lavender (middle) and apple trees (bottom) fields. For each crop, RGB input image (left), ground truth mask (center) and the predicted mask (right) are reported

Table 3 Semantic segmentation results on synthetic test images in different crop fields

Model	Sim test IoU	Tot synth data	Threshold
Zucchini	0.8960	19200	0.7
Lettuce	0.9069	4800	0.7
Chard	0.9428	4800	0.7
Lavender	0.8812	5260	0.7
Vineyard	0.8541	13840	0.5
Apples	0.9123	15280	0.5
Pear	0.8921	7980	0.5

The threshold indicates the saturation value for binary plant segmentation used in the tests

Table 4 Semantic segmentation results on real images in different crop fields

Model	Real test IoU	Tot synth data	Real train data	Real test data	Threshold
Lavender1	0.5199	5260	0	90	0.7
Lavender2	**0.6311**	5260	100	90	0.7
Vineyard1	0.6393	13840	0	500	0.5
Vineyard2	**0.6950**	13840	100	500	0.5
Apples1	0.4575	15280	0	210	0.5
Apples2	**0.8398**	15280	100	210	0.5
Pear1	0.5120	7980	0	140	0.5
Pear2	**0.8778**	7980	100	140	0.5

For each crop, a model has been trained on synthetic data only a second one using 100 additional real images containing miscellaneous crops different from the test set

For each crop, the best model's results are indicated in bold

Intel i7 CPU, including also the normalization and resize pre-processing from camera to input resolution 224×224. The test has been performed averaging the inference of the DNNs over 1000 image samples. The conversion does not perform any pruning or quantization of model's weights, leading to no substantial difference in the accuracy performance of the DNNs. Moreover, all the models present after the conversion a size of 4MB. Considering that visual streams from the robot camera are provided at a maximum rate of 30*fps*, the obtained inference result fully satisfies the real-time requirements for visual perception.

Navigation validation in simulated environment

The segmentation masks predicted in real-time from the network are used to visually estimate the heading of the rover in the row and compute control commands accordingly. The specific segmentation-based control algorithm is described in "Visual autonomous navigation in row-based crops" section. All the tests are performed on a 20 m long path inside the row, with a maximum linear velocity of 0.5 [m/s] and a maximum angular velocity of 1.0 [rad/s]. The metrics used to evaluate the navigation in the fields are chosen to test the quality of the trajectory of the rover, and the obtained results are gathered in Table 5 The Mean Squared Error (MSE) and the Mean Absolute Error (MAE) are computed between the trajectory followed by the rover and the target one (passing at the center of the row). Moreover, to evaluate the angular oscillation of the rover, the standard deviation of the angular velocity commands and the Cumulative Heading Average (CHA) are recorded, where CHA is defined as

$$CHA = \frac{1}{T}\sum_{i=0}^{T} arctan(\frac{y_i}{x_i}) \tag{5}$$

Here, y_i and x_i represent respectively lateral and frontal deviation of a target point with respect to the rover reference frame for T temporal pose samples constituting the navigation travel. For each tested navigation scenario, the goal is a point located centrally at the end of the row. The episodic testing has been performed with the PIC4rl-gym package for automatic metrics calculation (Martini et al., 2023).

Fig. 8 Test of semantic segmentation DNN on real-world test samples from vineyard (top), lavender (middle) and apple trees (bottom) fields. For each crop, RGB input image (left), ground truth mask (center) and the predicted mask (right) are reported

The segmentation DNN shows a successful performance in all the scenarios, and the navigation task is completely accomplished. However, the lettuce field presents a higher difficulty level due to the strong irregularity of the terrain combined with stone obstacles, and a narrow space to adjust the trajectory from sudden drifts, as also emerges from the higher errors in the navigation results. Figures 9 shows a Jackal UGV (Clearpath Jackal UGV[5]) inside the Gazebo worlds.

Robot navigation test on the field

In this final experimental section, the results obtained from robot visual autonomous navigation on the field are reported and commented. The tests have been performed, according to the available crop fields in the region, in row-based fields of lavender, vineyards, apple and pear trees. More in detail, the lavender field campaign has been described in the short preliminary paper (Navone et al., 2023b), whilst other experiments have been conducted in Aglié (TO) for the vineyard and in Manta (CN) for apple and pear trees. Figure 10 shows the trajectory of the robot in representative crop rows for each of the aforementioned crops.

Table 6 reports the navigation results computed from the robot trajectories on the field, respecting the same set of significant metrics used in the simulation.

To compute Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE), it is necessary to have a target trajectory and a ground truth of the path of the robot, but obtaining these data in a real-world scenario is a complex and demanding task since sophisticated instruments are required to reach a sufficient accuracy. To overcome this fact, a position-agnostic approach has been chosen: at each time step it is computed the lateral distance from each row and the distance from the central line is obtained by performing the difference between them. The rows are identified by clustering the point clouds obtained from a highly accurate 3D LiDAR (Velodyne VLP-16), then each of the two obtained clusters is interpolated by a straight line, and the shortest distance from the robot is computed. In the plots reported in Fig. 10, the sign of the right row distance has been changed to represent the row of plants.

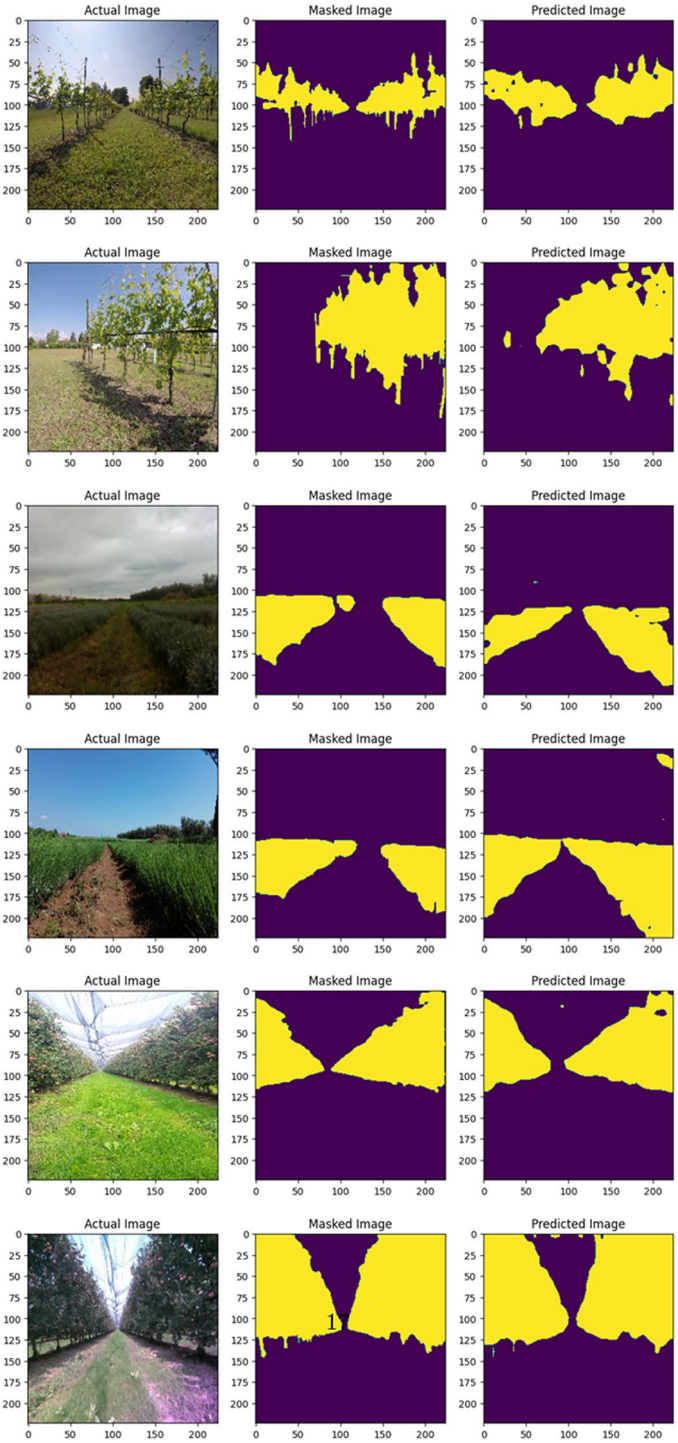
Cumulative Heading Average (CHA) is computed by using the heading provided by the high-end Attitude and Heading Reference System (AHRS) Microstrain GX5. At each time step, the difference between the average heading of the row and the heading provided by the AHRS is calculated and then averaged.

The standard deviation of the angular velocity is computed directly from the output of the control algorithm.

Results show that the robot commanded by the algorithm based on segmentation DNN is capable of traveling long distances inside a row by keeping the center with an overall maximum error of about 0.3m.

Navigation in vineyards shows consistent results, the center of the row is maintained for the whole duration of the tests with a contained error and the changes in the trajectory are always smooth. It is worth noting that the terrain of this crop at the time of the tests was pretty clean and without major unevenness. In contrast, it is noticeable that for the three

[5] https://clearpathrobotics.com/jackal-small-unmanned-ground-vehicle/


Actual Image
Masked Image
Predicted Image
Actual Image
Masked Image
Predicted Image
Actual Image
Masked Image
Predicted Image
Actual Image
Masked Image
Predicted Image
Actual Image
Masked Image
Predicted Image
Actual Image
Masked Image
Predicted Image

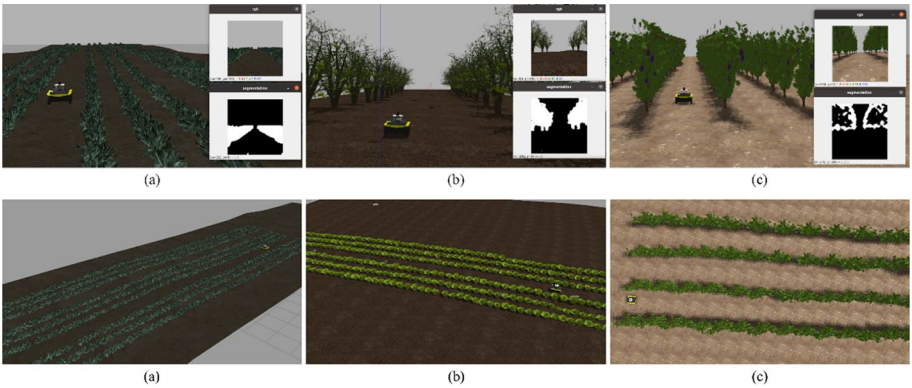


Fig. 9 Validation of segmentation-based control of the robotic platform in the realized simulated environments. The upper row shows the Jackal UGV in chard (a), pear trees (b) and vineyard (c) fields in Gazebo from the robot perspective and current image with the predicted segmentation mask. The bottom row show UGV in chard (a), lettuce (b) and vineyard (c) fields in Gazebo from above

Table 5 Results obtained from navigation tests performed on simulated crop fields in Gazebo

	CHA [rad]	MAE [m]	RMSE [m]	ω StdDev [rad/s]
Zucchini	– 0.0346	0.1167	0.1374	0.0428
Lettuce	0.0474	0.1209	0.1428	0.0224
Chard	– 0.0056	0.0145	0.02	0.0224
Pear Trees	– 0.0005	0.0058	1.9583	0.0028
Vineyard	0.0053	0.0330	0.0374	0.0255

Cumulative Heading Average (CHA), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and angular velocity Standard Deviation (ω Std Dev) are used as metrics to set up a common navigation benchmark for visual autonomous navigation in row-based crops

tests performed in orchards, sharp changes in the trajectory are often present. This behavior is the direct consequence of the presence of fruit and branches on the ground that abruptly deviate the robot from the desired central path. This condition demonstrates the effectiveness of the proposed algorithm that brings the robot back to the desired path in the row.

Conclusion

This work presents an extended synthetic dataset for binary semantic segmentation of row-based crops, together with a parametric tool to generate custom environments for fast development and evaluation of navigation algorithms. A state-of-the-art segmentation-based controller has been employed to validate the dataset and evaluate the simulated scenarios with relevant metrics. Results on both challenging synthetic and real images of crop fields demonstrated the quality of the generated data used for training. In-field autonomous navigation testing with a wheeled robotic platform further corroborates the authentic advantage of using the overall pipeline to develop low-cost reliable

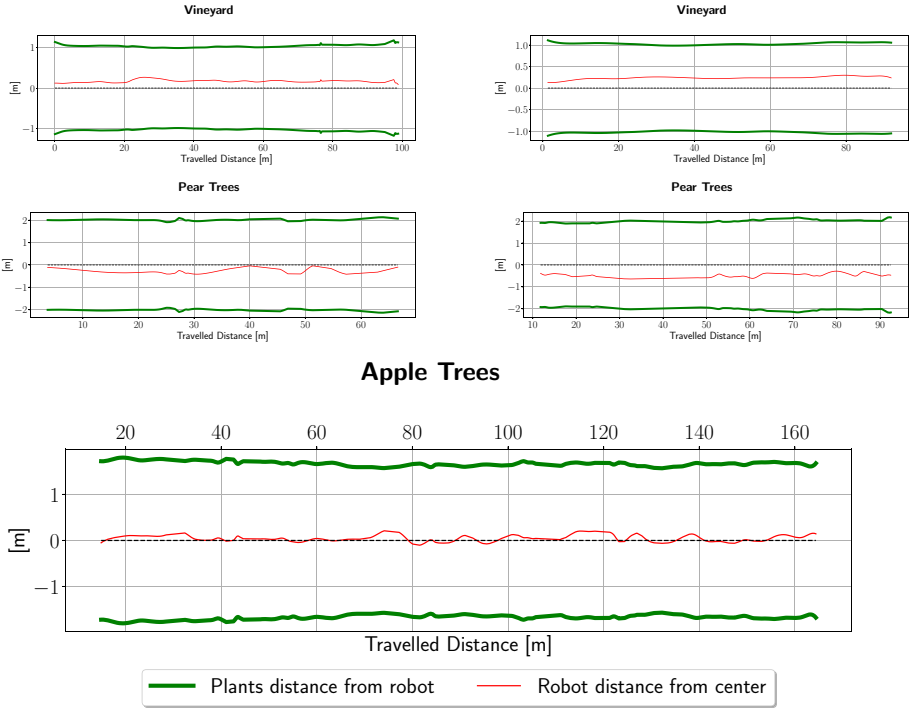


Fig. 10 Trajectories performed by the real robot in rows of vineyard, pear and apple trees. Segmentation-based visual control proves to be an effective solution to counteract small obstacles and terrain deviation from the central path

Table 6 Detailed navigation results obtained from real-world test campaign on the field

	CHA [rad]	MAE [m]	RMSE [m]	ω StdDev [rad/s]
Lavender (Navone et al., 2023b)	n/a	n/a	0.1913	n/a
Vineyard	0.0021	0.1668	0.1702	0.0225
Apple Tree	0.1071	0.0721	0.0911	0.0688
Pear Tree	0.03004	0.2843	0.2979	0.1307

visual controllers for agricultural tasks. Future work will see the extension of dataset and methods to fruits segmentation and harvesting, considering the increased challenge in modeling and rendering realistic fruit details, including diseases, for a visually guided robotic grasping or spraying.

Acknowledgements This work has been developed with the contribution of the Politecnico di Torino Interdepartmental Centre for Service Robotics (PIC4SeR https://pic4ser.polito.it). A grateful thank to *Cantina 366* (Aglié, TO) and *Agrion* (Manta, CN) for providing us with the possibility of collecting data and testing algorithms on the field.

Funding Open access funding provided by Politecnico di Torino within the CRUI-CARE Agreement.

Data availability The AgriSeg dataset used in this work is available at the webpage https://sites.google.com/view/datasetpic4ser/home-page.

Declarations

Conflict of 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.

References

Aghi, D., Cerrato, S., Mazzia, V., & Chiaberge, M. (2021). Deep semantic segmentation at the edge for autonomous navigation in vineyard rows. In *2021 IEEE/RSJ international conference on intelligent robots and systems* (*IROS*) (pp. 3421–3428). https://doi.org/10.1109/IROS51168.2021.9635969

Barth, R., Jsselmuiden, J., Hemming, J., & Van Henten, E. J. (2018). Data synthesis methods for semantic segmentation in agriculture: A Capsicum annuum dataset. *Computers and Electronics in Agriculture, 144*, 284–296.

Cerrato, S., Mazzia, V., Salvetti, F., & Chiaberge, M.: A deep learning driven algorithmic pipeline for autonomous navigation in row-based crops (2021). arXiv:2112.03816

Chen, L.-C., Papandreou, G., Schroff, F., & Adam, H.: Rethinking atrous convolution for semantic image segmentation (2017). arXiv:1706.05587

Comba, L., Biglia, A., Ricauda Aimonino, D., Barge, P., Tortia, C., & Gay, P. (2019). 2d and 3d data fusion for crop monitoring in precision agriculture. *Proceedings of the IEEE* (pp. 62–67). https://doi.org/10.1109/MetroAgriFor.2019.8909219

Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., & Schiele, B. (2016). The cityscapes dataset for semantic urban scene understanding. In *2016 IEEE conference on computer vision and pattern recognition* (*CVPR*) (pp. 3213–3223) (2016). https://doi.org/10.1109/CVPR.2016.350

Deshmukh, D., Pratihar, D. K., Deb, A. K., Ray, H., & Bhattacharyya, N. (2021). Design and development of intelligent pesticide spraying system for agricultural robot. In A. Abraham, T. Hanne, O. Castillo, N. Gandhi, T. Nogueira Rios, & T.-P. Hong (Eds.), *Hybrid Intelligent Systems* (pp. 157–170). Cham: Springer.

Droukas, L., Doulgeri, Z., Tsakiridis, N. L., Triantafyllou, D., Kleitsiotis, I., Mariolis, I., Giakoumis, D., Tzovaras, D., Kateris, D., & Bochtis, D. (2023). A survey of robotic harvesting systems and enabling technologies. *Journal of Intelligent Robotic Systems*. https://doi.org/10.1007/s10846-022-01793-z

FAO. (2022). The future of food and agriculture–Drivers and triggers for transformation. *FAO*. https://doi.org/10.4060/cc0959en

Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. *Computers and Electronics in Agriculture, 145*, 311–318. https://doi.org/10.1016/j.compag.2018.01.009

Häni, N., Roy, P., & Isler, V. (2020). Minneapple: A benchmark dataset for apple detection and segmentation. *IEEE Robotics and Automation Letters, 5*(2), 852–858.

Howard, A., Sandler, M., Chen, B., Wang, W., Chen, L., Tan, M., Chu, G., Vasudevan, V., Zhu, Y., Pang, R., Adam, H., & Le, Q. Searching for mobilenetv3. In *2019 IEEE/CVF international conference on computer vision* (*ICCV*) (pp. 1314–1324). IEEE Computer Society, Los Alamitos, CA, USA (2019). https://doi.org/10.1109/ICCV.2019.00140

Hu, J., Shen, L., & Sun, G.: Squeeze-and-excitation networks. In *2018 IEEE/CVF conference on computer vision and pattern recognition* (pp. 7132–7141) (2018). https://doi.org/10.1109/CVPR.2018.00745

Kabir, M. S. N., Song, M.-Z., Sung, N.-S., Chung, S.-O., Kim, Y.-J., Noguchi, N., & Hong, S.-J. (2016). Performance comparison of single and multi-GNSS receivers under agricultural fields in Korea. *Engineering in Agriculture, Environment and Food, 9*(1), 27–35. https://doi.org/10.1016/j.eaef.2015.09.002

Kestur, R., Meduri, A., & Narasipura, O. (2019). Mangonet: A deep semantic segmentation architecture for a method to detect and count mangoes in an open orchard. *Engineering Applications of Artificial Intelligence, 77*, 59–69.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., & Lo, W.-Y. (2023). Segment anything. arXiv:2304.02643

Koenig, N., & Howard, A. (2004). Design and use paradigms for gazebo, an open-source multi-robot simulator. In *2004 IEEE/RSJ international conference on intelligent robots and systems* (*IROS*) (IEEE Cat. No.04CH37566)(Vol. 3, pp. 2149–21543). https://doi.org/10.1109/IROS.2004.1389727

Liu, E., Monica, J., Gold, K., Cadle-Davidson, L., Combs, D., & Jiang, Y. (2023) Vision-based vineyard navigation solution with automatic annotation. arXiv:2303.14347

Loshchilov, I., & Hutter, F. (2018). Decoupled weight decay regularization. In *International conference on learning representations* (*ICLR*)

Luo, Z., Yang, W., Yuan, Y., Gou, R., & Li, X. (2023). Semantic segmentation of agricultural images: A survey. *Information Processing in Agriculture, 11*, 172–186.

Maheswari, P., Raja, P., Apolo-Apolo, O. E., & Pérez-Ruiz, M. (2021). Intelligent fruit yield estimation for orchards using deep learning based semantic segmentation techniques—A review. *Frontiers in Plant Science*. https://doi.org/10.3389/fpls.2021.684328

Martini, M., Cerrato, S., Salvetti, F., Angarano, S., & Chiaberge, M. Position-agnostic autonomous navigation in vineyards with deep reinforcement learning. In *2022 IEEE 18th international conference on automation science and engineering* (*CASE*) (pp. 477–484) (2022). https://doi.org/10.1109/CASE49997.2022.9926582

Martini, M., Eirale, A., Cerrato, S., & Chiaberge, M.: Pic4rl-gym: A ROS2 modular framework for robots autonomous navigation with deep reinforcement learning. In *2023 3rd international conference on computer, control and robotics* (*ICCCR*) (pp. 198–202) (2023). https://doi.org/10.1109/ICCCR56747.2023.10193996

Martini, M., Eirale, A., Tuberga, B., Ambrosio, M., Ostuni, A., Messina, F., Mazzara, L., & Chiaberge, M.: Enhancing navigation benchmarking and perception data generation for row-based crops in simulation (pp. 451–457). Wageningen Academic, Leiden, The Netherlands (2023). https://doi.org/10.3920/978-90-8686-947-3_56

Mazzia, V., Khaliq, A., Salvetti, F., & Chiaberge, M. (2020). Real-time apple detection system using embedded systems with hardware accelerators: An edge AI application. *IEEE Access, 8*, 9102–9114. https://doi.org/10.1109/ACCESS.2020.2964608

Megeto, G. A. S., Silva, A.G.D., Bulgarelli, R.F., Bublitz, C.F., Valente, A.C., & Costa, D.A.G.D. (2021). Artificial intelligence applications in the agriculture 4.0. *Revista Ciência Agronômica*, *51*

Navone, A., Martini, M., Ostuni, A., Angarano, S., & Chiaberge, M. (2023) Autonomous navigation in rows of trees and high crops with deep semantic segmentation. arXiv:2304.08988

Navone, A., Romanelli, F., Ambrosio, M., Martini, M., Angarano, S., & Chiaberge, M. (2023) Lavender autonomous navigation with semantic segmentation at the edge. arXiv:2309.06863

Peng, H., Xue, C., Shao, Y., Chen, K., Xiong, J., Xie, Z., & Zhang, L. (2020). Semantic segmentation of litchi branches using deeplabv3+ model. *IEEE Access, 8*, 164546–164555.

Raei, E., Asanjan, A. A., Nikoo, M. R., Sadegh, M., Pourshahabi, S., & Adamowski, J. F. (2022). A deep learning image segmentation model for agricultural irrigation system classification. *Computers and Electronics in Agriculture, 198*, 106977.

Ren, C., Kim, D.-K., & Jeong, D. (2020). A survey of deep learning in agriculture: Techniques and their applications. *Journal of Information Processing Systems, 16*(5), 1015–1033.

Salvetti, F., Angarano, S., Martini, M., Cerrato, S., & Chiaberge, M. (2023). Waypoint generation in row-based crops with deep learning and contrastive clustering. In M.-R. Amini, S. Canu, A. Fischer, T. Guns, P. Kralj Novak, & G. Tsoumakas (Eds.), *Machine Learning and Knowledge Discovery in Databases* (pp. 203–218). Cham: Springer.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.: Mobilenetv2: Inverted residuals and linear bottlenecks. In *2018 IEEE/CVF conference on computer vision and pattern recognition* (*CVPR*) (pp. 4510–4520). IEEE Computer Society, Los Alamitos, CA, USA (2018). https://doi.org/10.1109/CVPR.2018.00474

Sankhla, M.S., Kumari, M., Sharma, K., Kushwah, R., & Kumar, R. (2018). Water contamination through pesticide & their toxic effect on human health. *International Journal for Research in Applied Science and Engineering Technology*. https://doi.org/10.22214/ijraset.2018.1146

Shafi, U., Mumtaz, R., García-Nieto, J., Hassan, S. A., Zaidi, S. A. R., & Iqbal, N. (2019). Precision agriculture techniques and practices: From considerations to applications. *Sensors*. https://doi.org/10.3390/s19173796

Shruthi, U., Nagaveni, V., & Raghavendra, B.K. (2019). A review on machine learning classification techniques for plant disease detection. In *2019 5th international conference on advanced computing & communication systems* (*ICACCS*) (pp. 281–284). https://doi.org/10.1109/ICACCS.2019.8728415

Song, Z., Zhang, Z., Yang, S., Ding, D., & Ning, J. (2020). Identifying sunflower lodging based on image fusion and deep semantic segmentation with UAV remote sensing imaging. *Computers and Electronics in Agriculture, 179*, 105812.

Stafford, J.V.: Precision Agriculture '23. Wageningen Academic, Leiden, The Netherlands (2023). https://doi.org/10.3920/978-90-8686-947-3

Su, D., Kong, H., Qiao, Y., & Sukkarieh, S. (2021). Data augmentation for deep learning based semantic segmentation and crop-weed classification in agricultural robotics. *Computers and Electronics in Agriculture, 190*, 106418.

Thuilot, B., Cariou, C., Martinet, P., & Berducat, M. (2002). Automatic guidance of a farm tractor relying on a single CP-DGPS. *Autonomous Robots, 13*(1), 53–71. https://doi.org/10.1023/A:1015678121948

Van Dijk, M., Morley, T., Rau, M. L., & Saghai, Y. (2021). A meta-analysis of projected global food demand and population at risk of hunger for the period 2010–2050. *Nature Food, 2*(7), 494–501. https://doi.org/10.1038/s43016-021-00322-9

Vidović, I., Cupec, R., & Hocenski, Ž. (2016). Crop row detection by global energy minimization. *Pattern Recognition, 55*, 68–86.

Wang, Y., He, Z., Cao, D., Ma, L., Li, K., Jia, L., & Cui, Y. (2023). Coverage path planning for kiwifruit picking robots based on deep reinforcement learning. *Computers and Electronics in Agriculture, 205*, 107593.

Yépez-Ponce, D. F., Salcedo, J. V., Rosero-Montalvo, P. D., & Sanchis, J. (2023). Mobile robotics in smart farming: Current trends and applications. *Frontiers in Artificial Intelligence*. https://doi.org/10.3389/frai.2023.1213330

Zhai, Z., Martínez, J. F., Beltran, V., & Martínez, N. L. (2020). Decision support systems for agriculture 4.0 Survey and challenges. *Computers and Electronics in Agriculture, 170*, 105256. https://doi.org/10.1016/j.compag.2020.105256

Zhu, K., & Zhang, T. (2021). Deep reinforcement learning based mobile robot navigation: A review. *Tsinghua Science and Technology, 26*(5), 674–691.