Bush Detection for Vision-based UGV Guidance in Blueberry Orchards: Data Set and Methods

Vladan Filipović, Dimitrije Stefanović, Nina Pajević, Željana Grbović, Nemanja Djuric, Marko Panić
BioSense Institute, Novi Sad, Serbia
{vladan.filipovic, dimitrije.stefanovic, nina.pajevic, zeljana.grbovic}@biosense.rs
nemanja@temple.edu, panic@biosense.rs

Abstract

Object detection has reached strong performance in the last decade, having seen its usage spreading to various application areas, such as medicine, transportation, sports, and others. However, one of the more underutilized areas where advanced detection methods have yet to fully fulfill their promise is in the area of agriculture, where a strong potential exists for applying learned models to achieve practical, real-world impact affecting a large number of people. In this work, we focus on this application area and consider the problem of orchard guidance for ground robots, focusing on obstacle and plant detection from RGB camera images. First, we present an overview of public data sets used to train models to detect relevant objects from camera images and other sensor inputs. Then we introduce a novel data set collected in blueberry orchards that contains camera images in various conditions and provides blueberry bushes as targets for detection. The introduced data set provides the research community with a novel task of blueberry bush detection, which was not commonly considered thus far due to the lack of relevant data sets. We describe a detailed analysis of the data set, and finally provide an experimental study with several state-of-the-art deep object detection models, that set a baseline for the performance on this novel data set. The data set is made available online, enriching the variability of the existing tasks in the field and supporting further development of smart agriculture applications.

1. Introduction

Object detection in images is a topic that has gained a lot of attention in recent years due to its widespread applications and many well-publicized successes [44]. For example, face and text detection tasks have been thoroughly explored by the community [21], and vehicle, pedestrian, and traffic sign detection are being rapidly developed due to the popularity of autonomous driving [8]. While the applications relying on object detection have become ubiquitous in many parts of human activity, the task is less explored in the field of agriculture [45]. Nevertheless, particularly in this area there exists a large potential for real-world impact through the application of advanced learned models that can help to improve yield and crop management and efficiency, thus directly affecting a large number of people that depend on agriculture for work and sustenance. This has many potential use cases, such as helping to reduce the environmental impact and improve production efficiency [45], or in forestry to improve biomass management and prevent wildfires [12], to name a few.

A large number of detection challenges in agriculture can be brought under the umbrella of small object detection that is used to address various tasks in agriculture [3, 22], such as for improving yield estimation [17] or optimizing robotic harvesting [15]. However, one of the most prominent applications in agriculture is for the task of autonomous vehicle guidance, since the detected objects can be used as landmarks for the localization of unmanned ground vehicles (UGVs) and their path planning in complex environments where obtained GPS information can be unreliable [20,24,46]. In the domain of vehicle guidance, a distinction can be made between natural ecosystems that arise without human interference and have an unorganized arrangement of features within them, such as woods and forests, and artificial ecosystems where some level of organization among features exists, such as plantations and orchards where the plants are placed into rows [38]. In the latter case, automation in viticulture has recently received significant attention [3,4,6,26], and further diversification of applications to more complex, bush-structured fruits is a step forward that can lead to the development of new agricultural solutions.

In our current work we consider this task and focus on blueberry bush detection, which is a high-value crop [28] and whose production gained substantial popularity in Europe and North America. However, due to high production costs and lack of workforce, there exists a demand for suc-

Table 1. Publicly available annotated data sets for the task of vision-based ground vehicle navigation

Ref.	Data set name	Environment	Targets	Modality	Number of images (original / augmented)
[10, 13]	ForTrunkDet	Forest (3 locations)	Tree trunk (eucalyptus and pinus)	RGB and thermal (thermal for one location)	2,895 / 24,210
[11, 14]	ForTrunkDet v2	Forest (3 locations)	Tree trunk (eucalyptus and pinus)	RGB and thermal (thermal for one location)	5,325 / 49,608
[3, 35]	VineSet - Vine Trunk Image/Annotation Dataset	Vineyard (4 locations)	Vine trunk	RGB and thermal (thermal for one location)	952 / 9,481
[1, 2, 26]	VineSet - Grape Bunch and Vine Trunk Dataset	Vineyard (4 locations)	Vine trunk and grape bunch	RGB and thermal (thermal for one location)	1,939 / 428,498 (split to patches)
[5, 6]	Humain Lab Vine Trunk Dataset	Vineyard	Vine trunk	RGB	899 / 2,786
[20]	Different Light Conditions Pear Orchard Dataset	Pear orchard (2 locations)	Tree trunk (pear)	Thermal	5,313 / 7,563
Ours	Ground-level Blueberry Orchard Dataset v1	Blueberry orchard	Blueberry bush and pole	RGB	2,000 / -

cessful deployment of autonomous unmanned ground and aerial vehicles for tasks such as parcel zoning [36], spraying of weeds, or soil analysis. Further, detection models of blueberry bushes could potentially be generalized to other bush-structured fruits, such as raspberries or blackberries, thus expanding the application of autonomous robots in precision agriculture.

Our contributions can be summarized as follows:

- We provide and describe in detail a new open-source data set for blueberry bush detection;
- We analyze the performance of the state-of-the-art detectors, and set a baseline for the bush detection task to support further studies.

2. Related work

In the following, we give an overview of methods and data sets related to vision-based UGV navigation. The relevant work covers applications of detection models in natural settings such as forests, as well as vineyards and orchards.

2.1. Overview of methods

While there exist various distance sensors that can be used for data acquisition in agriculture applications, such as lidar, infrared, or time-of-flight cameras, RGB cameras are still the most prominent robotic navigation method [46]. Thanks to the color information and ability to obtain high-resolution images at a low cost, RGB cameras have the ability to extract relevant characteristics of plants, including the species or ripeness of the detected trees [19], and allow for reliable object detection in complex environments using one of many efficient models developed in the last few years that offer fast training and inference.

Most models considered in the literature have focused on the problem of plant or plant part detection using such 2D camera-based sensor data as inputs. In [13], the authors presented a tree trunk detection benchmark between seven models on a forestry data composed of both RGB and thermal images, and showed that YOLOv4-tiny [7] achieved the best overall performance. Their work, extended in [12] with a larger data set, considered thirteen models with different input resolutions tested on four edge devices, positioned YOLOv7 [42] as the best trade-off between accuracy and detection speed among the baselines. In [24], a single RGB camera is used with the Faster R-CNN model [33] to detect tree trunks as obstacles when guiding small UAVs at low altitudes, and the performance of this navigation system has been verified in eleven successful flight tests.

In an orchard, particularly in vineyards, UGV navigation is important for precision viticulture and its automation [27]. In [3] authors present a vine trunk data comprising RGB and thermal images and compared MobileNet [18] and Inception v2 [39] detection models. The same data was used in [4], where authors investigated the impacts of the image and data size on the detection performance. Another vine trunk data set was introduced in [6] and a comparison of six popular object detectors was conducted, indicating that EfficientDet-D0 [40] is the most suitable for integration on the UGV. Other orchard-grown cultures where UGV navigation is being investigated include pears, apples, and oranges. For example, extraction of fruit row centerlines is done in [46] based on the YOLOv3 model for pear tree trunk detection. Navigation in low-light settings using a thermal camera is explored in [20], where the Faster R-CNN model [33] showed reliable performance at three different times of the day. Within apple orchards, in [37] authors proposed improvements to the YOLOv5 model [41]

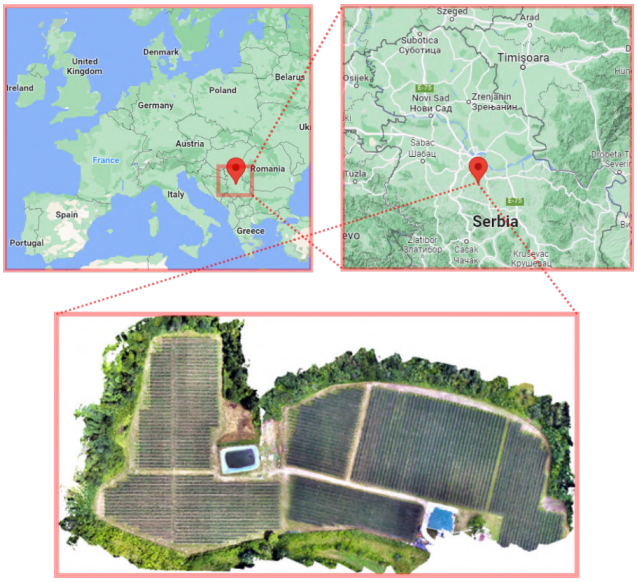


Figure 1. Geographic location of the blueberry orchard of interest

Figure 2. UGV platform used for data collection

for the detection of tree trunks and workers with the goal of path planning and obstacle detection. Going beyond deep methods, the authors of [9] developed a HOG- and SVM-based algorithm to detect tree trunks in dwarf orange orchards using a color camera and an ultrasonic sensor.

2.2. Data sets

Commonly used public data sets for comparison and benchmarking of detection models are Pascal VOC, MS-COCO, Open Image, and ImageNet [44], where these data sets mainly contain objects that are required for visual perception in a non-agricultural context. In more specific domains such as agriculture or forestry, data sets are usually purpose-made for specific tasks, limited in size and content (e.g., number of classes and scene variability). Nevertheless, the abovementioned general data sets are commonly used for pretraining of models due to the lack of data which is a common problem for agricultural use cases [25].

In the domain of UGV guidance in outdoor conditions, a large number of studies were done using small bespoke data. In Table 1 we present a brief overview of annotated publicly available data designed for the task of vision-based ground vehicle navigation, used in the aforementioned literature. References are given for both repositories on which the data sets are stored and papers with detailed data descriptions. The focus of the table is on vision-based perception, thus lidar or similar range-based data is not considered. In terms of imaging modalities RGB images are dominant, which is expected considering the arguments discussed in Section 2.1. Thermal cameras are used either on their own or alongside RGB cameras to enable operations in low-light or night-time settings. All presented data sets are small with the number of original images in the order of thousands, compared to popular general-purpose data sets with millions of images [44]. Lastly, an overview of additional forestry data sets with no annotations is given in [14].

From described literature and related data sets it is apparent that the work on vehicle guidance is mainly focused on tree trunk detection, and the performance of state-of-the-art models remains unexplored for the task of bush detection. Nevertheless, there is a growing need for bush detection for the task of precise UGV guidance in blueberry orchards. In particular, such task is important for soil monitoring and weed spraying [36], as well as in forestry where, besides the trees, perception of objects like bushes and rocks is of utmost importance [14].

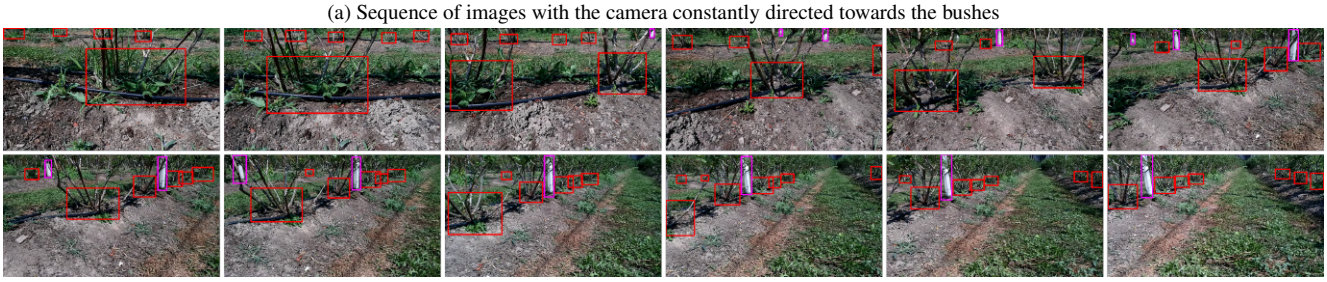
In this paper, we aim to mitigate this problem and present a data set to support a task of blueberry bush detection, that can be used for benchmarking models, pretraining of models for similar use cases, or as an extension of data sets with similar applications for the purpose of augmentation and improving generalization. The data set is publicly available online[1] and is meant to support further research into this topic. It is planned to be further extended and improved through the addition of more data and images in the years to come. In addition to introducing a novel data set, we also analyze three commonly used detection models to set a baseline for the task of blueberry bush detection.

3. Data set for blueberry bush detection

The created data set contains 2,000 RGB images of scenes from a blueberry (lat. *Vaccinium corymbosum*) orchard captured in the village of Babe, Serbia, with the exact location illustrated in Fig. 1. There are two types of annotated objects: the *bush* label corresponding to the base of the blueberry bush, and the *pole* label corresponding to hail netting poles and similar obstructing objects such as lamp posts

[1] https://doi.org/10.5281/zenodo.7813238, last accessed April 2023

(a) Sequence of images with the camera constantly directed towards the bushes



(b) Sequence of images with the camera rotating in the blueberry row

Figure 3. Examples of image sequences, with labels indicating blueberry bush (in red) and poles (in purple)

or wooden legs of bumblebee hives (distinguishing poles is important to prevent equipment damage in operations such as soil sampling and pruning). Images were captured using an RGB module of the Luxonis OAK-D device[2], mounted at a height of about 0.5 meters on our UGV platform such as the one given in Fig. 2, with the resolution of 1920×1080 pixels and stored in the lossless PNG format. The data comprises 20 image sequences of variable length. The camera was mostly directed from the row center towards the plants (i.e., the optical axis of the camera is orthogonal to the row direction), although there are sequences where the camera is rotated to look down the row, as seen in Fig. 3. The intended focus was on the base of a blueberry plant and the surrounding bank on which it grows, however, the camera direction and angle are not constant, as seen in the figure.

Images were captured on three occasions in March, May, and August of 2022, and they contain various artifacts and environmental conditions that can be expected in outdoor applications. In addition to the natural complexity of the blueberry bush shape (see Fig. 4a) and different types of obstacles (see Fig. 4b), some of the most notable sources of variability in bush appearance are lighting, high contrast, shadows, and saturation (see Fig. 4c), as well as camera occlusions by weeds and branches (see Fig. 4d). Along with the described scenes, there is a small number of outliers such as images with no labeled objects whatsoever, images from different perspectives, or images containing trash, examples of which are illustrated in Fig. 5. A deliberate effort was made to collect the images in a variety of outdoor conditions, however it should be noted that there existed several data-collecting requirements that limited some of the data variability, such as focusing on one orchard and avoiding fog, rain, or very low-light conditions.

[2]https://shop.luxonis.com/products/oak-d, last accessed April 2023

3.1. Data statistics

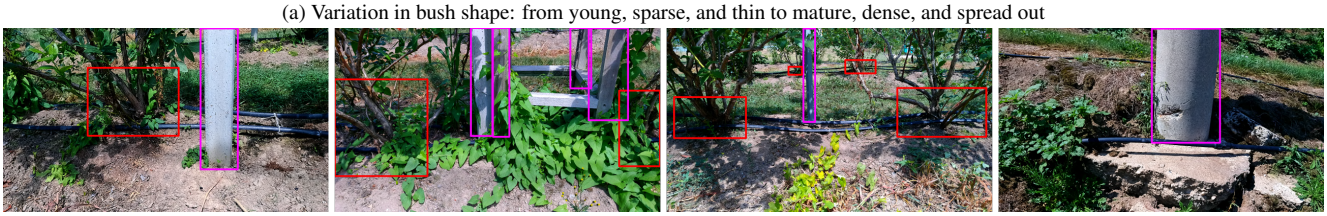
Statistics of bounding box positions and shapes are presented in Fig. 6. The distribution of objects on the x-axis is fairly uniform, with peaks at the edges of the image frame corresponding to bushes that are partially out of frame. On the y-axis, the majority of objects are located in the upper half of the frame, which is caused by a large number of objects in distant background rows, as well as the focus of most images being on the point where the bush meets the ground. In terms of the bounding box shape, poles are expectedly taller than bushes. There are two distinct modes appearing in both object distributions, with bigger bounding boxes corresponding to objects in the closest row and smaller boxes corresponding to more distant rows.

Of the total $2{,}000$ images, there are $1{,}935$ images containing at least one bush and 597 images containing at least one pole, with 593 images containing both and 61 images of orchard scenes with no annotated objects (Fig. 7). In total, there are $5{,}245$ and 833 instances of bushes and poles, resp., giving an average of 2.62 and 0.42 instances per image, respectively. The data is randomly split into train, validation, and test sets with $1{,}490$, 200, and 310 images, resp., aiming to achieve 75%, 10%, and 15% split. As the data contains 20 sequences, the split is made based on sequences rather than individual images to prevent data leakage.

3.2. Data labeling

The two classes are annotated with bounding boxes, using a semi-automated iterative procedure and Python-based

(a) Variation in bush shape: from young, sparse, and thin to mature, dense, and spread out



(b) Different types of obstacles in images: hail netting pole, bumblebee hive legs, and lamp posts

(c) Variation in contrast and shadows: from light and diffuse to dark and opaque

(d) Various occlusions: different types and quantities of weeds and branches obstructing the bush

Figure 4. Examples of images illustrating data variation and artifacts, with labels indicating blueberry bush (in red) and poles (in purple)

(a) Examples of background images with no objects present

(b) Views from irregular heights, positions, angles, and out-of-context objects

Figure 5. Examples of images illustrating irregular scenes present in the data set

LabelMe software [34]. A bush is defined as an above-ground part of a blueberry plant that is connected to the soil. An upper part of a bounding box should encompass blueberry branch splitting and leaves, and the lower part should include soil and vegetation in near proximity. The goal is to focus on the triangular shape of the bush base and encapsulate it into approximately square-shaped bounding boxes. Bushes are labeled regardless of their position in the image, distance from the camera, the possible occurrence of occlusion, or strong shadow presence, even if they are not crucial for UGV guidance, as long as they can be distinguished from the background as bushes by a human annotator. The region where the bush comes into contact with the soil is important for tasks such as the selection of soil sampling point location, or localizing a weed spraying area. On the other hand, poles are well-defined solid objects and they are annotated such that the bounding box captures the visible part of a pole.

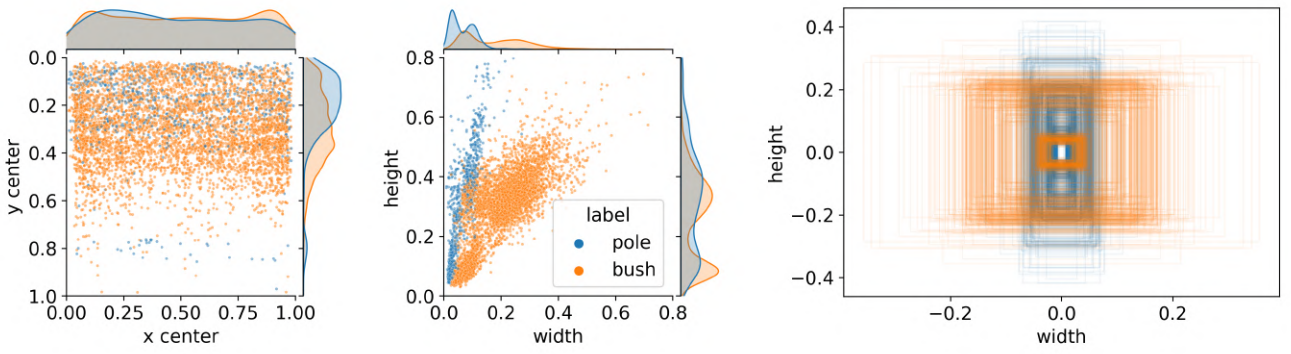


Figure 6. Label statistics: distribution of positions and bounding box dimensions, as well as shape summary (axes are normalized)

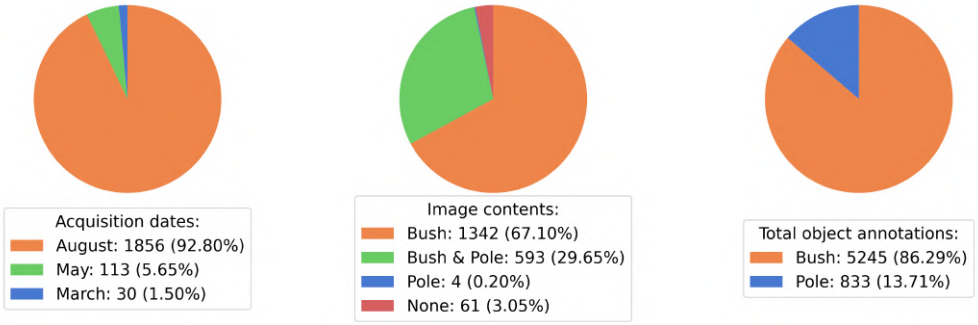


Figure 7. Data set statistics: acquisition dates, number of images containing at least one instance of a given class, and label counts

A labeling procedure is done in the following manner. First, 400 images that capture the data variability (such as various types of bushes, poles, camera directions, and occlusions) were hand-picked for manual labeling. Selected images were divided into ten subsets and each was labeled by a different graduate student annotator. An initial YOLOv5s model (discussed in more detail in Section 4) was trained using the manually-labeled data, and then run on the remaining 1,600 images. This architecture was chosen due to fast training and good empirical results obtained during the development process. Annotations obtained in such a manner were visually inspected, manually corrected, and added or removed when necessary. This initial YOLOv5s model was used only to optimize the annotation process and was discarded after the labeling process.

Since the edges of the blueberry bushes were often ambiguous, the bounding box shapes of manual annotations differed between 10 annotators (see Fig. 8), even when given identical instructions as described above. Training the initial model with these variable shapes enabled the network to better converge to a bounding box shape that is consistent and accurate. Annotations were then converted and stored in the YOLO format. Note that no predetermined augmentation was done to the data set, leaving that decision to the users of the data. Further extensions of the data set are planned to include a significantly larger amount of images, as well as additional two mono images for stereo vision and depth estimation, which will further support the development of positioning and guidance algorithms.

4. Baseline methods for bush detection

In this section we propose a baseline method for blueberry bush detection based on YOLOv5 [41] architecture, which has shown great performance in terms of detection accuracy and inference speed across multiple applications [16, 30, 43]. Being an extension of YOLOv3, the YOLOv5 model uses CSP-Darknet53 architecture as a backbone [7], optimized spatial pyramid pooling neck (SPPF, which is an optimized SPP from [32]), and YOLOv3 head. The loss that is minimized during training is composed of the following three components: classification, objectness, and location losses. The first two are formulated as binary cross-entropy, while the latter is complete intersection-over-union loss [7]. The YOLOv5 model comes in a combination of different levels of complexity (n, s, m, l, and x, corresponding to nano, small, medium, large, and extra large) and input image sizes (such as the P5 and P6 variants with the maximum resolutions of 640 and 1280, respectively).

Figure 8. Difference in annotations between 5 different annotators (final labels shown in solid lines, individual annotations in dotted lines)

Table 2. Experimental results for the considered detection models

Model	All				Bush				Pole				Param count	Latency [ms]
	P	R	mAP 50	mAP 50-95	P	R	mAP 50	mAP 50-95	P	R	mAP 50	mAP 50-95		
YOLOv5n	0.940	0.790	0.859	0.440	0.927	0.840	0.912	0.479	0.954	0.740	0.805	0.401	1.7M	96.4
YOLOv5s	0.892	0.800	0.873	0.472	0.882	0.871	0.909	0.500	0.903	0.730	0.838	0.444	7.0M	179.8
YOLOv5m	0.930	0.797	0.872	0.489	0.935	0.860	0.924	0.510	0.924	0.735	0.820	0.467	20.8M	313.2

Using a larger number of parameters results in more powerful models, however the trade-off between accuracy and inference speed makes only the lighter models viable for the considered real-time edge-based application. For that reason, we considered a nano version that has 1.7M parameters (referred to as YOLOv5n), a small one with 7.0M parameters (YOLOv5s), as well as a medium one with 20.8M parameters (YOLOv5m). The default input resolution of the models is 640×640 pixels, and we resized the collected images to match that resolution. Starting from a YOLOv5 model pre-trained on the COCO data set, we conducted training for 50 epochs using the proposed data set, setting a batch size to 32, the learning rate to 0.01, and using the Adam optimizer [23].

As no pre-determined augmentation is done to the raw data set, we relied on the default techniques implemented in YOLOv5 data loaders. During the training procedure, the original set of training images is loaded and modified with different augmentations in each epoch. Images are flipped around the vertical axes with the probability of 0.5, randomly scaled up or down by up to 50%, translated left or right by up to 10%, and their hue, saturation, and value were scaled by up to 1.5%, 70%, and 40%, respectively. Along with these manipulations, mosaic augmentation [7] was applied where a 4-image mosaic is created from the current image and three other random images to make the detection model more robust.

Following the YOLO framework, the model inference is conducted by dividing an input image into a grid of cells (set to 32×32) and predicting a set of bounding boxes for each cell according to the predefined number of anchor boxes of different sizes (set to 10×13, 16×30, 33×23, 30×61, 62×45, 59×119, 116×90, 156×198, 373×326). During the evaluation, a default matching IoU threshold of 0.6 was used. We set the detection confidence threshold value to that for which the max-F1 is reached on the validation set, resulting in a detection threshold of 0.24 in the final evaluation. Finally, Non-Maximum Suppression (NMS) [31] with IoU threshold of 0.6 is applied to model output in order to eliminate redundant detections.

5. Experimental results

In order to evaluate the considered models we calculated the following four metrics: precision (P), recall (R), mean average precision (mAP) with IoU threshold set to 0.5 (mAP50), and average mAP with a threshold in a range from 0.5 to 0.95 (mAP50-95) [29]. Metrics are calculated considering joint detections from both classes (referred to as "All"), as well as for each class separately (referred to as "Bush" and "Pole"). Inference times were computed and reported for a single image by evaluating on CPU rather than GPU, in order to better match the target edge platform where the models would be deployed.

The evaluation results are presented in Table 2. We can see that YOLOv5m achieves slightly better metrics than either YOLOv5n or YOLOv5s, although these models with fewer parameters still achieved comparable results on both classes. It should also be mentioned that YOLOv5n and YOLOv5s exhibit significantly faster inference times than YOLOv5m (nearly 4 and 2 times faster, respectively). We also note that the precision is quite large across the board, with somewhat lower recall, which we further explore in the remainder of this section.

To better understand the detection performance we focus on YOLOv5s given its good balance between accuracy and latency, and provide several illustrative qualitative examples, presented in Fig. 9. Dashed yellow and cyan bounding boxes represent blueberry bush and pole detections, respectively, while solid red and magenta bounding boxes represent their corresponding ground truth labels. Descriptions

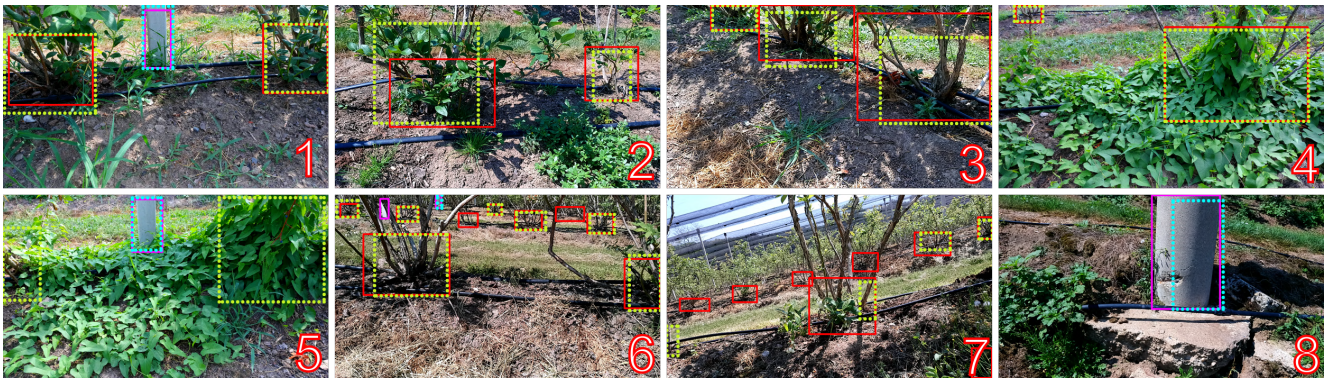


Figure 9. Examples of the detection results of YOLOv5s, detailed discussion provided in the text

of these examples are given in left-to-right top-to-bottom order. In the first image we see an example of a best-case scenario, correct detection of two bushes and one pole in a scene with a small amount of occlusion by the weeds. When it comes to the second example, two observations can be made. First, it illustrates the case where the detected bounding box deviates significantly from the label, while still managing to capture the bush quite well. Second, it shows two significantly different plants, one young and sparse and the other mature and covered by leaves, both detected by the model. Moving forward, the third image illustrates the model's ability to detect bushes at different scales, with varying distances as their locations shift down the row. The fourth image illustrates how well the model handles severe occlusions by weeds, as well as the ability to detect very distant bushes that are barely present in the image. We can see that both of these detections are nearly perfect in this example.

In the bottom row of Fig. 9 we focus on mistakes and deficiencies encountered during the analysis. The fifth image illustrates an interesting situation where the model predicted two false positive bush detections which do resemble true objects although the significant occlusion makes the example very ambiguous, and the annotators eventually decided not to label these cases. The sixth and seventh examples are images with multiple false negative objects that the model failed to detect. We can see that the missed bushes are mostly further away from the camera, or occluded by the other objects. While we are investing efforts to resolve such failure modes, we note that more fundamental improvements to the detectors are out of the scope of our current work. Instead, our focus is on providing reasonable baselines on the novel data set that are mostly relying on default settings, which can support future work on further model iterations and improvements. The eighth and final image shows an object of a pole class that is of a larger size and not present in the training set, included in the test set to help evaluate the generalization abilities of the model. We see the model detects such a large object, although the detection box does not fully match the label's true dimensions.

6. Conclusion

In our current work we explored the topic of vision-based blueberry bush detection. The literature review showed that bush detection is mostly an unexplored topic in the domain of UGV guidance and similar ground-level applications, despite its importance in areas of agriculture and forestry. It was shown that the majority of the focus in these fields is on tree trunk detection, which is reflected in detection targets of publicly available data sets, while there are no major available data sets focused on bushes. We presented a novel ground-level blueberry bush data set, annotated for the detection of bushes and obstacles. The data set is captured in a blueberry orchard, encompassing a large variability in plant characteristics, camera positions, lighting settings, and occlusions. Both the data set and the labeling procedure were described in detail. Along with the novel data, multiple variations of the state-of-the-art YOLOv5 detection model were used to set the baseline detection metrics on this data set. Results showed that trained models achieved promising evaluation metrics on the considered task, thus setting a good basis for further work on improving the performance on the task of blueberry bush detection.

7. Acknowledgements

This research has received funding from the project FLEXIGROBOTS, grant agreement No. 101017111 under European Union's Horizon 2020 research and innovation program. The work is also supported through the ANTARES project under grant agreement SGA-CSA No. 739570 under FPA No. 664387[3]. Finally, we are grateful to our colleagues Dragana Blagojević, Lazar Lemić, Dejan Pavlović, Katarina Petranović, and Mirjana Radulović for participating in the data labeling process.

[3]https://doi.org/10.3030/739570, last accessed April 2023

References

[1] André Silva Aguiar, Sandro Augusto Magalhães, Filipe Neves Dos Santos, Luis Castro, Tatiana Pinho, João Valente, Rui Martins, and José Boaventura-Cunha. Grape bunch detection at different growth stages using deep learning quantized models. *Agronomy*, 11(9):1890, 2021. 2

[2] André Silva Aguiar and Sandro Magalhães. Grape bunch and vine trunk dataset for Deep Learning object detection. https://zenodo.org/record/5139598, 2021. [Online; accessed 17-March-2023]. 2

[3] André Silva Aguiar, Nuno Namora Monteiro, Filipe Neves dos Santos, Eduardo J Solteiro Pires, Daniel Silva, Armando Jorge Sousa, and José Boaventura-Cunha. Bringing semantics to the vineyard: An approach on deep learning-based vine trunk detection. *Agriculture*, 11(2):131, 2021. 1, 2

[4] Khadijeh Alibabaei, Eduardo Assunção, Pedro D Gaspar, Vasco NGJ Soares, and João MLP Caldeira. Real-time detection of vine trunk for robot localization using deep learning models developed for edge tpu devices. *Future Internet*, 14(7):199, 2022. 1, 2

[5] Eftichia Badeka, Theofanis Kalampokas, Eleni Vrochidou, Konstantinos Tziridis, George A Papakostas, Theodore P Pachidis, and Vassilis G Kaburlasos. Humain-Lab-vine-trunk-dataset. https://github.com/humain-lab/vine-trunk, 2020. [Online; accessed 17-March-2023]. 2

[6] Eftichia Badeka, Theofanis Kalampokas, Eleni Vrochidou, Konstantinos Tziridis, George A Papakostas, Theodore P Pachidis, and Vassilis G Kaburlasos. Vision-based vineyard trunk detection and its integration into a grapes harvesting robot. *International Journal of Mechanical Engineering and Robotics Research*, 10(7):374–385, 2021. 1, 2

[7] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao. Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*, 2020. 2, 6, 7

[8] Azzedine Boukerche and Zhijun Hou. Object detection using deep learning methods in traffic scenarios. *ACM Computing Surveys (CSUR)*, 54(2):1–35, 2021. 1

[9] Xianyi Chen, Binquan Zhang, Liang Luo, et al. Multi-feature fusion tree trunk detection and orchard mobile robot localization using camera/ultrasonic sensors. *Computers and Electronics in Agriculture*, 147:91–108, 2018. 3

[10] Daniel Queirós da Silva and Filipe Neves dos Santos. ForTrunkDet - Forest dataset of visible and thermal annotated images for object detection. https://zenodo.org/record/5213825, 2021. [Online; accessed 17-March-2023]. 2

[11] Daniel Queirós da Silva and Filipe Neves dos Santos. ForTrunkDetV2 - Forest dataset of visible and thermal annotated images for object detection (augmented version). https://zenodo.org/record/7186052, 2022. [Online; accessed 17-March-2023]. 2

[12] Daniel Queirós da Silva, Filipe Neves dos Santos, Vítor Filipe, Armando Jorge Sousa, and Paulo Moura Oliveira. Edge ai-based tree trunk detection for forestry monitoring robotics. *Robotics*, 11(6):136, 2022. 1, 2

[13] Daniel Queirós da Silva, Filipe Neves Dos Santos, Armando Jorge Sousa, and Vítor Filipe. Visible and thermal image-based trunk detection with deep learning for forestry mobile robotics. *Journal of imaging*, 7(9):176, 2021. 2

[14] Daniel Queirós da Silva, Filipe Neves dos Santos, Armando Jorge Sousa, Vítor Filipe, and José Boaventura-Cunha. Unimodal and multimodal perception for forest management: review and dataset. *Computation*, 9(12):127, 2021. 2, 3

[15] Leonidas Droukas, Zoe Doulgeri, Nikolaos L Tsakiridis, Dimitra Triantafyllou, Ioannis Kleitsiotis, Ioannis Mariolis, Dimitrios Giakoumis, Dimitrios Tzovaras, Dimitrios Kateris, and Dionysis Bochtis. A survey of robotic harvesting systems and enabling technologies. *Journal of Intelligent & Robotic Systems*, 107(2):21, 2023. 1

[16] Zheng Ge, Songtao Liu, Feng Wang, Zeming Li, and Jian Sun. Yolox: Exceeding yolo series in 2021. *arXiv preprint arXiv:2107.08430*, 2021. 6

[17] Sebastian Gonzalez, Claudia Arellano, and Juan E Tapia. Deepblueberry: Quantification of blueberries in the wild using instance segmentation. *Ieee Access*, 7:105776–105788, 2019. 1

[18] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017. 2

[19] Kenta Itakura and Fumiki Hosoi. Automatic tree detection from three-dimensional images reconstructed from 360 spherical camera using yolo v2. *Remote Sensing*, 12(6):988, 2020. 2

[20] Ailian Jiang, Ryozo Noguchi, and Tofael Ahamed. Tree trunk recognition in orchard autonomous operations under different light conditions using a thermal camera and faster r-cnn. *Sensors*, 22(5):2065, 2022. 1, 2

[21] Jaskirat Kaur and Williamjeet Singh. Tools, techniques, datasets and application areas for object detection in an image: a review. *Multimedia Tools and Applications*, 81(27):38297–38351, 2022. 1

[22] Faina Khoroshevsky, Stanislav Khoroshevsky, and Aharon Bar-Hillel. Parts-per-object count in agricultural images: Solving phenotyping problems via a single deep neural network. *Remote Sensing*, 13(13):2496, 2021. 1

[23] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 7

[24] HY Lee, Hann Woei Ho, and Ye Zhou. Deep learning-based monocular obstacle avoidance for unmanned aerial vehicle navigation in tree plantations: Faster region-based convolutional neural network approach. *Journal of Intelligent & Robotic Systems*, 101:1–18, 2021. 1, 2

[25] Yuzhen Lu and Sierra Young. A survey of public datasets for computer vision tasks in precision agriculture. *Computers and Electronics in Agriculture*, 178:105760, 2020. 3

[26] Sandro Costa Magalhães, Filipe Neves dos Santos, Pedro Machado, António Paulo Moreira, and Jorge Dias. Benchmarking edge computing devices for grape bunches and

trunks detection using accelerated object detection single shot multibox deep learning models. *Engineering Applications of Artificial Intelligence*, 117:105604, 2023. 1, 2

[27] Alessandro Matese and Salvatore Filippo Di Gennaro. Technology in precision viticulture: A state of the art review. *International journal of wine research*, pages 69–81, 2015. 2

[28] José Gilberto Sousa Medeiros, Luiz Antonio Biasi, Claudine Maria de Bona, and Francine Lorena Cuquel. Phenology, production and quality of blueberry produced in humid subtropical climate. *Revista Brasileira de Fruticultura*, 40, 2018. 1

[29] Rafael Padilla, Wesley L Passos, Thadeu LB Dias, Sergio L Netto, and Eduardo AB Da Silva. A comparative analysis of object detection metrics with a companion open-source toolkit. *Electronics*, 10(3):279, 2021. 7

[30] Delong Qi, Weijun Tan, Qi Yao, and Jingfeng Liu. Yolo5face: Why reinventing a face detector. In *Computer Vision–ECCV 2022 Workshops: Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part V*, pages 228–244. Springer, 2023. 6

[31] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 779–788, 2016. 7

[32] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018. 6

[33] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28, 2015. 2

[34] Bryan C Russell, Antonio Torralba, Kevin P Murphy, and William T Freeman. Labelme: a database and web-based tool for image. *Int. J. of Computer Vision*, 77(1), 2005. 5

[35] Luís Santos, André Aguiar, and Filipe Santos. VineSet - Vine Trunk Image/Annotation Dataset. https://zenodo.org/record/5362354, 2021. [Online; accessed 17-March-2023]. 2

[36] Dimitrije Stefanović, Aleksandar Antić, Marko Otlokan, Bojana Ivošević, Oskar Marko, Vladimir Crnojević, and Marko Panić. Blueberry row detection based on uav images for inferring the allowed ugv path in the field. In *ROBOT2022: Fifth Iberian Robotics Conference: Advances in Robotics, Volume 2*, pages 401–411. Springer, 2022. 2, 3

[37] Fei Su, Yanping Zhao, Yanxia Shi, Dong Zhao, Guanghui Wang, Yinfa Yan, Linlu Zu, and Siyuan Chang. Tree trunk and obstacle detection in apple orchard based on improved yolov5s model. *Agronomy*, 12(10):2427, 2022. 2

[38] William Swenson, David Sloan Wilson, and Roberta Elias. Artificial ecosystem selection. *Proceedings of the National Academy of Sciences*, 97(16):9110–9114, 2000. 1

[39] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016. 2

[40] Mingxing Tan, Ruoming Pang, and Quoc V Le. Efficientdet: Scalable and efficient object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10781–10790, 2020. 2

[41] Ultralytics. YOLOv5 Github Repository. https://github.com/ultralytics/yolov5. [Online; accessed 17-March-2023]. 2, 6

[42] Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv preprint arXiv:2207.02696*, 2022. 2

[43] Wei Wu, Hao Chang, Yonghua Zheng, Zhu Li, Zhiwen Chen, and Ziheng Zhang. Contrastive learning-based robust object detection under smoky conditions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4295–4302, 2022. 6

[44] Syed Sahil Abbas Zaidi, Mohammad Samar Ansari, Asra Aslam, Nadia Kanwal, Mamoona Asghar, and Brian Lee. A survey of modern deep learning based object detection models. *Digital Signal Processing*, page 103514, 2022. 1, 3

[45] Qian Zhang, Yeqi Liu, Chuanyang Gong, Yingyi Chen, and Huihui Yu. Applications of deep learning for dense scenes analysis in agriculture: A review. *Sensors*, 20(5):1520, 2020. 1

[46] Jianjun Zhou, Siyuan Geng, Quan Qiu, Yang Shao, and Man Zhang. A deep-learning extraction method for orchard visual navigation lines. *Agriculture*, 12(10):1650, 2022. 1, 2