DOI: 10.2507/35th.daaam.proceedings.xxx

LEARNING WITH SMALL DATA: A NOVEL FRAMEWORK FOR BLUEBERRY ROOT COLLAR DETECTION

A G M Zaman[1], Nesma Mahmoud[2], Indrek Virro[1], Olga Liivapuu[1], Tormi Lillerand[1], Kallol Roy[2] & Jüri Olt[1]

[1]Estonian University of Life Sciences and [2]University of Tartu, Tartu, Estonia

This Publication has to be referred as: Zaman, A [G M]; Mahmoud, N[esma]; Virro, I[ndrek]; Liivapuu, O[lga]; Lillerand, T[ormi]; Roy, K[allol]; Olt, J[üri] (2024). Learning with Small Data: A Novel Framework for Blueberry Root Collar Detection, Proceedings of the 35th DAAAM International Symposium, pp.xxxx-xxxx, B. Katalinic (Ed.), Published by DAAAM International, ISBN 978-3-902734-xx-x, ISSN 1726-9679, Vienna, Austria
DOI: 10.2507/35th.daaam.proceedings.xxx

Abstract

The integration of artificial intelligence and data science in agriculture is significantly transforming traditional human knowledge-based agricultural practices into a data-driven decision-making domain, known as precision agriculture. This study focuses on enhancing precision farming by using computer vision models to detect root collar of blueberry plants. Accurate detection of the root collar is crucial for optimizing agricultural practices such as fertilizer application, pesticide usage, and water management. Integrating computer vision model with agrobotic promises to maximize food production, reduce operational costs, minimize environmental impact, and ensure food quality and sustainability. We collected 118 images of blueberry plants from a plantation in Vehendi village, Tartu, Estonia, using an Insta360 One X2 camera mounted 1.6 meters above the plant rows. We developed and trained a machine learning model to detect blueberry root collar in a diverse environment. A pre-trained YOLOv5 model is trained and tested for blueberry root collar detection. The model's performance was evaluated, achieving 74.7% precision, 78.3% recall, and a 76.2% mAP50. Our results demonstrate the model's generalization capability with limited numbers of training samples. Future work will focus on expanding the dataset and integrating the model end-to-end with agrobotic systems to further advance precision farming practices.

Keywords: precision farming; YOLO; agrobotic; computer vision; machine learning.

1. Introduction

The world's population reached 8 billion in November 2022, raises its implications on lives, rights, health, and future generations [1]. Modern technologies have enabled human society to meet the increasing demand for food production. However, numerous challenges persist in ensuring food security, including plant diseases, climate change, extreme weather events, conflict and instability, water scarcity, land degradation, land grabbing, and many more. To address the

crucial need for food safety and also meeting the present challenges in agricultural sustainability, integration of advanced technologies such as machine learning becomes central to smart farming [2] and precision agriculture [3]. Precision farming practices optimize productivity, reduce environmental impact, ensures food security, and promote sustainability. These practices have the potential to significantly increase global food production while maintaining availability, accessibility, and nutritional quality worldwide. Moreover, sustainable farming practices can safeguard natural ecosystems, benefiting the future generations [4], [5].

The combination of machine vision, big data, IOT, and robotics has emerged as a promising set of tools for smart farming and precision agriculture. In precision farming, object detection plays a crucial role in different automated applications, such as, root collar detection, leaf detection, fruit counting, weed identification, phenotyping, and many more, by applying the computer vision and machine learning techniques.

The collaboration of these technologies is a key enabler for precision farming, with the potential to transform the entire agricultural lifecycle from planting to harvesting. To achieve such advanced automation, a high-accuracy object detection model is a primary requirement. In this study, we applied the transfer learning technique using the pre-trained YOLOv5 [6], small variant model to train and test the detection of the root collar of low bush blueberry (*Vaccinium angustifolium*) plants. Accurate detection of the plant's root collar can enhance fertilization use, optimize irrigation, and enable targeted pesticides application, ultimately promoting sustainable agriculture and increase crop yield, and lowering production costs.

This study makes several key contributions to precision agriculture. Firstly, a specialized platform was designed and developed for collecting high-quality images of blueberry plants, which enables precise data acquisition. Secondly, a dataset of high-quality blueberry images was assembled, which is crucial for training and evaluating detection models. Finally, the effectiveness of a pre-trained YOLOv5 model for detecting root collars was demonstrated, offering valuable insights into its applicability in precision farming.

2. Literature review

Machine vision is pivotal in precision agriculture, enhancing efficiency through automated analysis of plant health, fruit count, weed detection, crop monitoring, disease identification, and yield estimation. This technological advancement parallels the evolutionary significance of vision, as zoologist Andrew Parker links to the Cambrian explosion [7], [8]. Precision agriculture, initially conceptualized in the 1980s by Dr. Pierre C. Robert, is often credited as the cornerstone of modern precision agriculture. It is defined as an information revolution driven by new technologies, leading to a more precise farm management system. Later, it evolved to be defined as "the application of modern information technologies to provide, process and analyze multisource data of high spatial and temporal resolution for decision making and operations in the management of crop production" [9]. Precision agricultural applications and experiments can be classified into four distinct groups, such as, crop management, soil management, water management and livestock management [10]. Its main goal is to promote sustainable, efficient, and resilient farming practices that address the challenges of growing global population, environmental concerns, and changing climate conditions, while optimizing resource use and ensuring food security. Recent advancements in artificial intelligence and computer vision have significantly heightened research interest in this field.

In the field of object detection, numerous experiments have been conducted such as, plant diseased detection, yield estimation, root collar detection, autonomous robotic harvesting guidance, vegetation index calculation and many more. Plant disease detection from images is a highly explored area and can be categorized into three main categories, i.e., diseases detection, measurement of disease severity, and classification [11]. Among different image processing techniques convolutional neural network (CNN) has achieved impressive results in the field of image classification and detection. CNN has shown remarkable performance in image classification, successfully identifies 13 distinct types of plant diseases with accuracy ranging from 91% to 98% in various individual class tests [12]. In another study, disease detection from cucumber leaves employed a classical method involving three key stages i.e., segmenting diseased leaf images using the K-means clustering algorithm, extracting shape and color features from data and classifying diseased leaf images through sparse representation. This approach effectively recognized seven major cucumber diseases, with an overall recognition rate of 85.7%, surpassing the performance of alternative methods [13].

A novel neural network model integrated with semantic graphics has been proposed for detecting the crop lines and weeds in paddy fields, where the detected line serves as a guiding path for autonomous weeding robot aimed to enhance agricultural efficiency [14]. Another study employed a deep learning-based two-step method to identify weeds. This method used a CenterNet model trained to detect vegetables, achieving impressive performance with a precision 95.6%, recall 95.0%, and an F1 score of 0.953. Any remaining green objects in the image were subsequently classified as weeds [15]. Additionally, the study [16] shows that an automated system for targeted spraying in orchards using RGB imaging demonstrated a 23% reduction in pesticide use without compromising spray coverage by utilizing machine vision to adjust the pesticide spray based on the detected shape of the apple tree canopy.

A real-time computer vision approach was proposed for locating crop stems, with applications in mechanical hoeing within precision farming. This method integrates an object-detecting neural network to identify stems in RGB images, followed by the utilization of an aggregation algorithm that enhances detection by considering temporal data from consecutive frames. Testing was conducted on maize and bean crops in two configurations demonstrated promising

results, achieving F1-score of 94.74% and 93.82%, respectively, for detecting maize and bean stems, with a location accuracy of 0.7 cm and 0.5 cm [17].

In this study, we focus on blueberry root collar detection, an increasingly important crop in Europe and North America [18]. Due to high production costs and a labour shortage, there is a rising need for the successful deployment of agrobotic platforms [19]. Accurate root collar detection could enable the precise application of fertilizers, pesticides, and watering to plants in an informed manner leading to efficient farming practices.

Additionally, developing detection models for blueberry root collar holds the potential for generalization to other bush-structured fruits such as raspberries and blackberries, thereby broadening the application of autonomous robots in precision farming.

3. Materials and methods

In this study, we employ advanced object detection techniques that have evolved considerably over the past two decades. Initially dependent on handcrafted feature detection methods, the field underwent a significant transformation in 2012 with the introduction of convolutional neural networks (CNNs) [20]. Modern object detectors are categorized into two-stage and one-stage approaches [21]. For our study, we utilize YOLO (You Only Look Once) [22], a single-stage object detector. YOLO simultaneously performs feature extraction and region of interest pooling by framing detection as a regression problem. It predicts bounding boxes and class probabilities directly from a grid, which provides enhanced processing speed and competitive accuracy, making it particularly suited for detecting blueberry root collars. Building on these advantages, our future goal is to deploy this machine vision model within an agricultural robot to automate precision farming tasks. This integration will also address constraints such as computational resources, power consumption, and environmental uncertainties, enabling real-time detection of blueberry root collars in a field environment. The agrobotic system, which is currently under development is shown in figure 1 below.

Fig. 1. Agricultural robot for precision farming applications

The figure 2 below shows the full process of the blueberry root collar detection experiment. This process has four fundamental stages: data acquisition, data pre-processing and labelling, YOLOv5 architecture and training, and results and discussion.

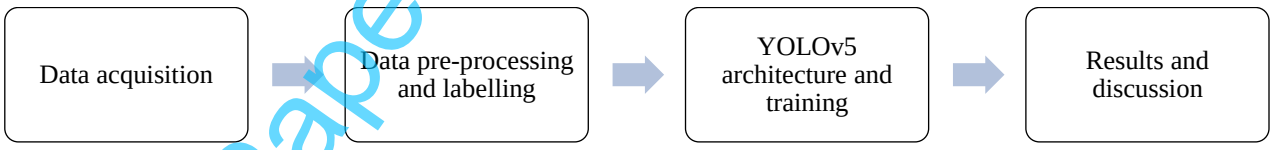
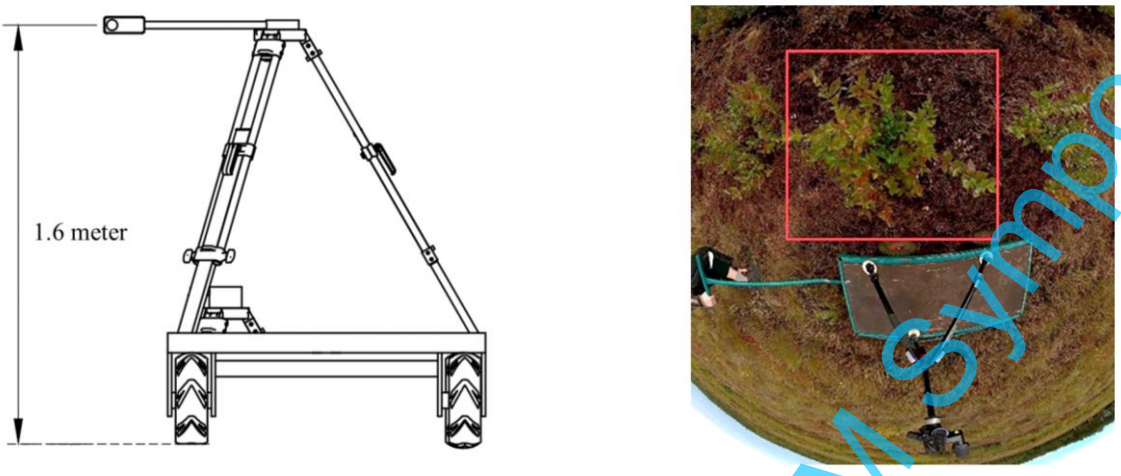


Fig. 2. End-to-end process for YOLOv5 based blueberry root collar detection

3.1. Data acquisition

Agricultural data is often not readily available in open sources due to high costs, the need for specialized equipment, extensive manual effort, and the significant time required for collection. In this initial phase, we focused on capturing images of individual low bush blueberry plants from above, with an average height of 1.5 meters. The data was collected from a plantation field in Vehendi village, Elva Municipality, Tartu County (GPS 58.20, 26.13), Estonia. The Insta360 One X2 camera was used for image acquisition, mounted on a mobile platform that navigates through the blueberry field rows and capture continuous frames of the plants. The design of the data acquisition platform, shown in figure 3(a), was specifically developed for this task. It features an adjustable arm reaching a height of 1.6 meters and ensures stability and mobility in the field. Figure 3(b) illustrates the on-field data collection process, highlighting the camera's positioning and

the captured image frame in a red bounding box. This setup allowed for efficient and consistent data collection across the field. The collected blueberry dataset was relatively small, consisting of 118 blueberry images in total.

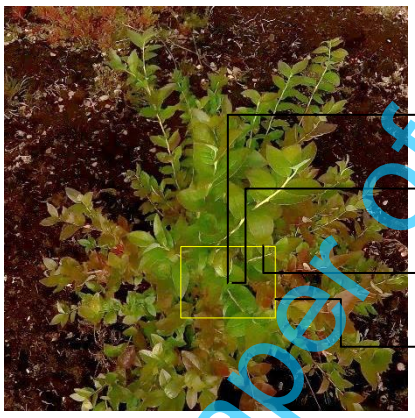


a) Data acquisition platform b) On-field data collection process

Fig. 3. Blueberry data collection platform and process

3.2. Data pre-processing and labelling

Specific frames of blueberry images were extracted using the Insta360 Studio software. The original images have a resolution of 1440 by 1440 pixels and a density of 24 dpi. Accurate annotation of bounding boxes and assignment of class labels are essential for effective supervised learning. To facilitate this process, we used Roboflow software [23] which allows for direct annotation within its interface. Each bounding box, i.e., the yellow rectangle outlines a distinct spatial region within the image and is associated with a class label - in this study, "blueberry root collar" as shown in figure 4(a). Figure 4(b) depicts the annotation process, detailing how the class label, the coordinates (x, y), and dimensions (width, height) of the bounding box are specified in normalized units relative to the image dimensions. This structured data format enables the model to learn to detect and classify objects with precision during both training and inference stages.



Root collar coordinates details:

0	Blueberry root collar (class label)
0.49	Normalized centre X-coordinate with respect to image width
0.69	Normalized centre Y-coordinate with respect to image height
0.24	Normalized width of bounding box with respect to image width
0.18	Normalized height of bounding box with respect to image height

a) Bounding box (yellow) for root collar b) YOLO-specific bounding box coordinates

Fig. 4. Bounding box and annotations for blueberry root collar

3.3. YOLOv5 architecture and training

In this study we present a baseline method for blueberry root collar detection utilizing the YOLOv5 [6] model. This model has consistently demonstrated high detection accuracy and rapid inference speed across variety of applications [24], [25], [26]. The YOLOv5 model is available in different configurations: nano (n), small (s), medium (m), large(l), and extra-large (x). Though larger models with a greater number of parameters generally exhibit increased accuracy, the trade-off between accuracy and inference speed makes lighter models more suitable for real-time, edge-based applications due to their lower computational complexity and reduced resource requirements.

The architecture of YOLOv5 consists primarily of three main components: the backbone, neck, and head. The figure 5 below illustrates how these components work together to process input images and predict bounding boxes [6].

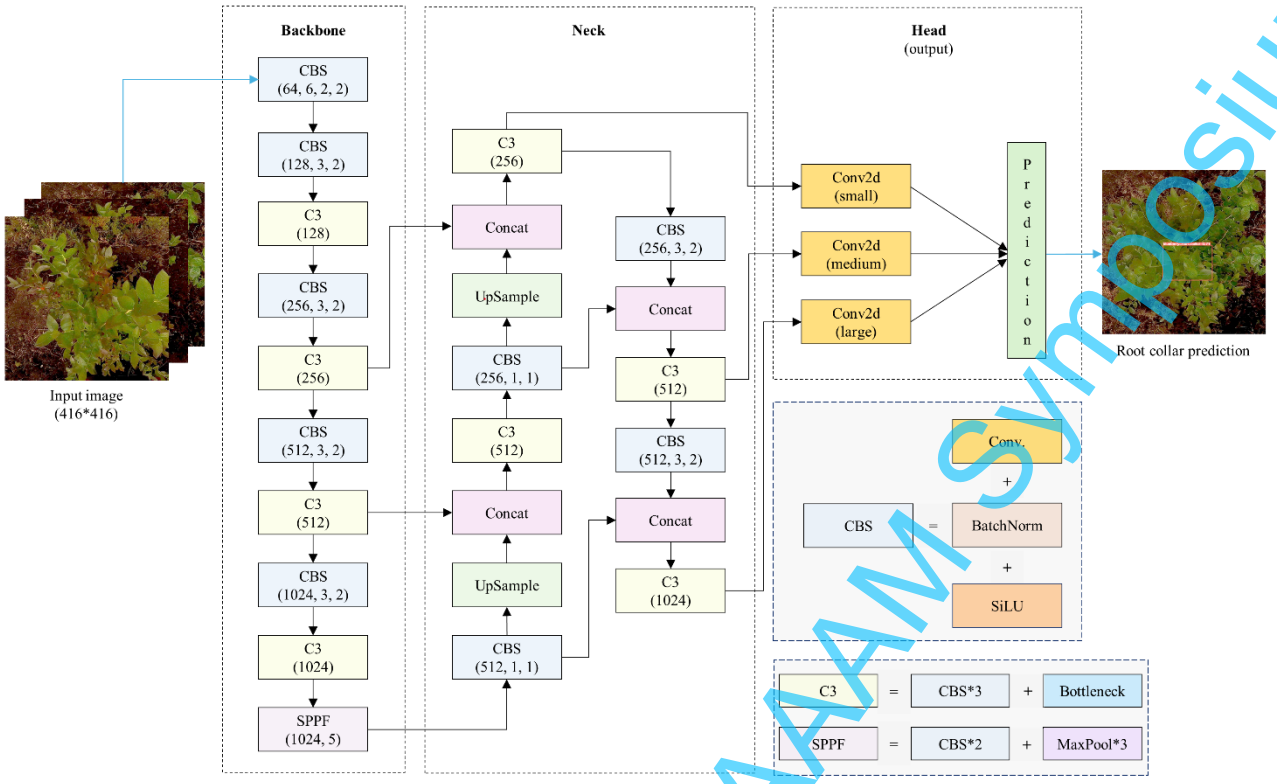


Fig. 5. Simplified YOLOv5 network architecture

The model processes the input image through a comprehensive end-to-end pipeline, where the model extracts relevant features and regresses the coordinates of bounding boxes, alongside calculating the confidence scores for each class. This approach allows YOLOv5 to efficiently detect objects by simultaneously considering spatial and class information, resulting in a fast and accurate detection process suitable for real-time applications. Throughout the network, the bounding boxes are refined, ensuring precise localization and classification of objects in diverse and complex scenes.

The backbone of YOLOv5 employs CSPDarknet53 [27], which enhances feature extraction while minimizing computational complexity. The backbone consists of multiple CBS (Conv. + BatchNorm + SiLU) modules, C3 modules, and an SPPF (Spatial Pyramid Pooling-Fast) module. The CBS layers initiate feature extraction with convolution operations, batch normalization, and the SiLU (Sigmoid Linear Unit) activation function, refining the learning process. For instance, an input blueberry image resized to 416x416 pixels passes through the first CBS layer configured with 64 filters, a kernel size of 6, stride of 2, and padding of 2, resulting in a 208x208x64 feature map. This step not only extracts essential features but also normalizes the outputs through batch normalization, ensuring a more stable and consistent range of activations. The SiLU activation function introduces non-linearity, enabling the network to capture complex patterns. These combined processes refine the feature representation, enhancing the model's capability to distinguish between subtle details and variations in the data, which is crucial for accurate detection in challenging conditions.

The backbone also includes C3 layers, composed of three consecutive CBS layers followed by a bottleneck layer, which capture spatial hierarchies and contextual information. The final feature maps undergo processing in the SPPF layer, which differs from the original SPPNet implementation [28]. In YOLOv5, the SPPF layer does not output a fixed-size vector but instead maintains spatial dimensions that depend on the input size, resulting in a fixed number of channels. The SPPF layers aggregate context at multiple scales, enhancing the receptive field and improve detection of objects at various levels of abstraction.

The neck component receives various features from the backbone ranging from low to high level, and applies fusion techniques, that combine principles from FPN (Feature Pyramid Network [29]) and PANet (Path Aggregation Networks) [30]. This architecture utilizes UpSample layers and Concat operations to construct a multi-scale feature pyramid, crucial for detecting small objects and preserving fine details. The UpSample layers enhance the spatial resolution, while the Concat layers merge these up-sampled features with lower-level maps, enabling a comprehensive representation across object sizes. The integration of PANet principles further refines and aggregates features through CBS and C3 modules, enhancing the model's localization and detection accuracy.

The head section is responsible for the final predictions, which includes object classes, objectness scores, and bounding box coordinates. It utilizes anchor boxes, which are pre-defined shapes used to assist in predicting object locations and sizes, facilitating detection across different scales - small, medium, and large. The head processes inputs from three distinct feature maps obtained from the neck, each corresponding to a different level of down-sampling. This approach enables multi-scale predictions, allowing the model to effectively identify objects of various sizes within an

image, from small details to larger objects, thereby enhancing overall accuracy and robustness of the detection system. By integrating feature maps of different resolutions, the head section refines and synthesizes both spatial and semantic information, ensuring that fine-grained details and broader contextual context are captured. This detailed, layered prediction strategy is crucial for accurate detection across complex and varied images.

In our experiments for blueberry root collar detection, the pre-trained YOLOv5 ‘s’ variant was utilized which was initially pre-trained on the COCO dataset [31]. The input image size for the model was set to 416x416 pixels. Various augmentation techniques were applied to the training images to enhance the diversity and robustness of the dataset. These common augmentations include random scaling, cropping, flipping, colour adjustments, and affine transformations. Such techniques help the model generalize better by simulating different variations and distortions that might occur in real-world scenarios. This process enhances model performance on unseen data and prevents the model from overfitting to the training data.

We have trained the model using stochastic gradient descent (SGD) optimizer, with a learning rate of 0.01, a batch size of 16, a weight decay of 0.0005, and a momentum value set at 0.937 to ensure efficient convergence. The loss function is defined as the sum of three losses (i) bounding box regression loss (ii) classification loss, and (iii) objectness loss, as shown in equation (1). The overall loss is a weighted sum of these individual losses, reflecting their relative contributions to the training objective.

$$loss = loss_{box} + loss_{cls} + loss_{obj} \tag{1}$$

The bounding box regression loss function measures how well the predicted bounding box matches the ground truth bounding box. This loss is defined in equation (2) below. Where λ_{coord} is a hyperparameter that regulates the relative contribution of bounding box regression loss compared to other loss factors, s^2 represents the number of grid cells into which the image is divided. The indicator $I_{obj}^{(i,j)}$ is set to 1 if the j^{th} bounding box in the i^{th} grid cell is responsible for the prediction. The term $b_j(2 - w_i h_i)$ acts as a scaling factor, emphasizing the significance of different bounding box sizes, smaller bounding boxes are given higher influence, while larger bounding boxes receive relatively lower influence. The relative coordinates and dimensions of the ground truth bounding box are denoted by x_i, y_i, w_i, h_i where x_i and y_i represent the relative coordinates of the centre of the root collar bounding box, and w_i and h_i represents its relative width and height, respectively. Similarly, $\hat{x}_i, \hat{y}_i, \hat{w}_i, \hat{h}_i$ denote the coordinates and dimensions of the predicted bounding box, where $\hat{x}_i$ and $\hat{y}_i$ are the predicted relative coordinates of the centre, and $\hat{w}_i$ and $\hat{h}_i$ are the predicted relative width and height.

$$loss_{box} = \lambda_{coord} \sum_{i=0}^{s^2} \sum_{j=0}^{B} I_{obj}^{(i,j)} b_j (2 - w_i h_i) \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 + (w_i - \hat{w}_i)^2 + \left(h_i - \hat{h}_i\right)^2 \right] \tag{2}$$

The classification loss function, as shown in equation (3), quantifies the accuracy of predicted class probabilities against the ground truth class. Here, λ_{cls} is a loss weighting factor for the classification loss. The terms, $p_i(c)$ and $\hat{p}_i(c)$ denote the ground truth and predicted probabilities for class c in the i^{th} grid cell, respectively. This formulation ensures that the model’s prediction aligns closely with the actual class distribution, thereby enhancing classification accuracy.

$$loss_{cls} = \lambda_{cls} \sum_{i=0}^{s^2} \sum_{j=0}^{B} I_{obj}^{(i,j)} \sum_{c \in classes} p_i(c) \log\left(\hat{p}_i(c)\right) \tag{3}$$

The objectness loss function, as represented in equation (4), evaluates the accuracy of the predicted bounding box’s confidence score in relation to the ground truth. The parameters, λ_{no_obj}, and λ_{obj} act as weighting factors for the loss in scenarios where no object is present and where an object is present, respectively. The terms c_i and $\hat{c}_i$ are the ground truth and predicted confidence score for the i^{th} grid cell, respectively. This loss function ensures that the model effectively distinguishes between object and non-object regions, thereby enhancing its detection performance.

$$loss_{obj} = \lambda_{no\_obj} \sum_{i=0}^{s^2} \sum_{j=0}^{B} I_{no\_obj}^{(i,j)} (c_i - \hat{c}_i)^2 + \lambda_{obj} \sum_{i=0}^{s^2} \sum_{j=0}^{B} I_{obj}^{(i,j)} (c_i - \hat{c}_i)^2 \tag{4}$$

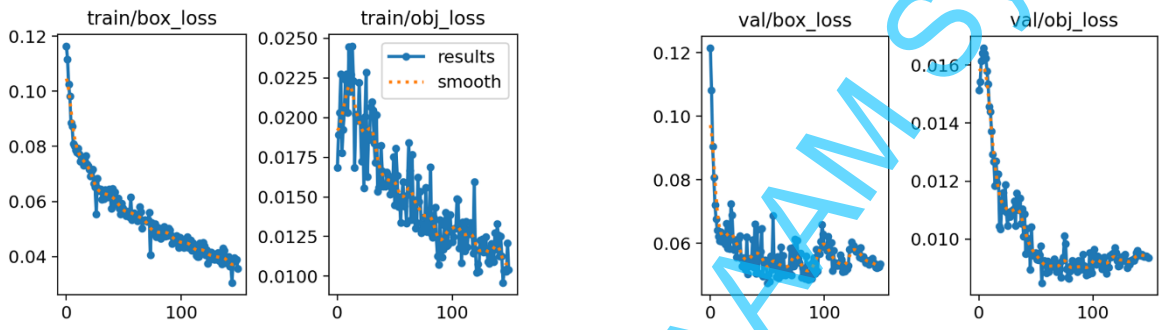
The YOLOv5 model significantly enhances object detection accuracy through its advanced use of anchor boxes and Intersection over Union (IoU). It automatically generates anchor boxes by applying K-means clustering to the bounding box dimensions in the training dataset, and then refines these boxes using genetic algorithm to optimize their shapes and sizes for better matching of the objects being detected. By assigning these dynamically generated anchors based on the highest IoU and utilizing multiple anchors per grid cell effectively identifies and localizes small or closely spaced objects. Its multi-scale prediction approach and non-maximum suppression (NMS) further refine detection results, ensuring that

overlapping detections are minimized and only the most confident predictions are retained. In our blueberry root collars detection it shows high accuracy, even in smaller image sizes such as 118 pixels.

4. Results and discussion

The model training was conducted on a system running Windows 10 Pro with Python 3.12. The hardware configuration included an Intel® Core™ i5-7400 CPU @ 3.00 GHz with 4 cores, 16 GB of RAM, and the training was performed exclusively on the CPU.

The model was trained and validated over 150 epochs using the datasets. The training curves are represented by the box loss and object loss curves for both training and validation dataset, as shown in figure 6. The curves for box loss and object loss reflect the model's convergence and performance improvement throughout the training process. Specifically, the reduction in validation loss values indicates effective adaptation and learning, demonstrating the model's ability to generalize and optimize performance over successive epochs.



a) Model training loss b) Model validation loss

Fig. 6. Convergence of model training and validation loss

The performance of the model is assessed using precision (P), recall (R), and average precision (AP) metrics, which are defined in equations (5), (6), and (7), respectively.

$$P = \frac{TP}{(TP + FP)} \tag{5}$$

$$R = \frac{TP}{(TP + FN)} \tag{6}$$

$$AP = \int_0^1 P(R)dR \tag{7}$$

Precision (P) measures the proportion of true positive (TP) detections among all positive detections, which includes both true positives (TP) and false positives (FP). Recall (R), on the other hand, measures the proportion of true positive (TP) detections among all actual positive examples, which includes both true positives (TP) and false negatives (FN). Given that our evaluation involves a single class, the average precision is synonymous with the mean average precision (mAP) computed at a 50% confidence threshold. This metric represents the area under the precision-recall curve, which is derived by integrating the precision with respect to recall.

Figure 7 presents the test results of the model's performance metrics including precision, recall, mAP. The precision

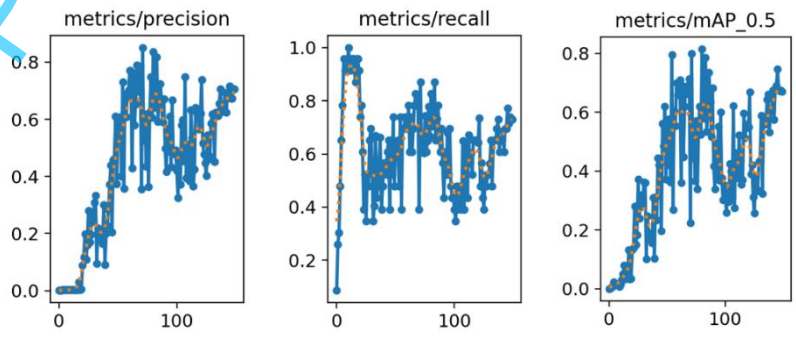


Fig. 7. Evaluation metrics: precision, recall, and mAP@50 results

of the model is 74.7%, indicating that 74.7% of the predicted positive detections are true positives. The recall is 78.3%, reflecting that 78.3% of actual positive examples are correctly detected. Additionally, the mAP@50 is 76.2%, shows the model's effectiveness in detecting root collars with a high level of accuracy at a 50% IoU threshold. These results collectively show the model's strong performance in object detection.

Figure 8 shows three predictions of blueberry root collar detection. The first two images show correct predictions, with the red bounding boxes representing the model's predictions with confidence score, and the yellow boxes denoting the ground truth. In contrast the third image (right most) shows a failed detection, where the predicted bounding box does not align with the actual root collar, resulting in a low IoU score. These examples show the challenges of the detection task, highlighting the difficulty due to the high similarity of features within the image context.

Fig. 8. Blueberry root collar detection: correct and misaligned predictions

Our results demonstrate the potential of automated machine vision in scenarios with scarce data, particularly in agricultural applications where large datasets are often unavailable, making model training challenging. Despite the limited size of our dataset, the model exhibited significant convergence and robust performance, effectively detecting blueberry root collars. This highlights the model's potential for practical applications in precision farming, suggesting it could be a valuable tool for enhancing agricultural practices and decision-making.

5. Conclusion

In this study, advanced machine learning techniques were utilized to detect blueberry root collar positions using a pre-trained YOLOv5s model. Initially, a data acquisition platform was designed and developed to gather images from blueberry plantations in Tartu, Estonia. The dataset underwent pre-processing for supervised learning, where the root collar bounding box coordinates were prepared in a specified format. The model was systematically trained, validated, and tested on this dataset. Our results indicate that the model effectively learned from the dataset despite its small size and the complex learning environment, which included similar features from other plants such as grass and weeds. The model achieved a precision of 74.7%, a recall of 78.3%, and a mAP@50 of 76.2%, marking a significant advancement in the field of precision farming.

Future work will focus on expanding the dataset and integrating the model into an agrobotic platform for real-time field operations. This integration aims to optimize agricultural practices by accurately identifying the root collar positions to facilitate precise applications of fertilizers, pesticides, and watering.

6. Acknowledgments

This research was supported by development fund PM210001TIBT from the Estonian University of Life Sciences and proof-of-concept grant EAG304 from the Estonian Research Council.

7. References

[1] The UNFPA State of World Population 2023 report. (2023). Population bomb, bust - or boon? New UNFPA report debunks 8 myths about a world of 8 billion, The UNFPA State of World Population 2023 Report, Available from: https://www.unfpa.org/news/population-bomb-bust-or-boon-new-unfpa-report-debunks-8-myths-about-world-8-billion Accessed: 2024-09-06

[2] Tyagi, A. C. (2016). Towards a Second Green Revolution, Irrigation and Drainage, Vol. 65, No. 4, pp. 388-389, ISSN 1531-0361, DOI: 10.1002/IRD.2076

[3] Zhang, N.; Wang, M. & Wang, N. (2002). Precision agriculture a worldwide overview, Computers and Electronics in Agriculture, Vol. 36, No. 2-3, pp. 113-132, ISSN 1872-7107, DOI: 10.1016/S0168-1699(02)00096-0
[4] Gebbers, R. & Adamchuk, V. I. (2010). Precision agriculture and food security, Science, Vol. 327, No. 5967, pp. 828-831, ISSN 0036-8075, DOI: 10.1126/science.1183899
[5] Vekic, A.; Borocki, J.; Stankovski, S. & Ostojic, G. (2017). Development of innovation in field of precision agriculture, Proceedings of the 28th DAAAM International Symposium, pp.0787-0795, B. Katalinic (Ed.), Published by DAAAM International, ISBN 978-3-902734-11-2, ISSN 1726-9679, Vienna, Austria, DOI: 10.2507/28th.daaam.proceedings.111
[6] Jocher, G. (2020). Ultralytics YOLOv5, DOI: 10.5281/zenodo.3908559, Available from: https://github.com/ultralytics/yolov5 Accessed: 2024-09-06
[7] Douglas, E. & Valentine, J. W. (2013). The Cambrian explosion: the construction of animal biodiversity, Roberts and Company Publishers Inc., ISBN 1936221039, Greenwood Village, CO, USA
[8] Parker, A. (2003). In the blink of an eye: how vision sparked the big bang of evolution, Basic Books, ISBN 0465054382, New York, USA
[9] National Research Council & others (1997). Precision Agriculture in the 21st Century: Geospatial and Information Technologies in CROP Management, National Academy Press, ISBN 978-0-309-05893-3, DOI: 10.17226/5491, Washington, D.C., USA
[10] Liakos, K.; Busato, P.; Moshou, D.; Pearson, S. & Bochtis, D. (2018). Machine Learning in Agriculture: A Review, Sensors, Vol. 18, No. 8, pp. 2674, ISSN 1424-8220, DOI: 10.3390/s18082674
[11] Garcia, J. & Barbedo, A. (2013). Digital image processing techniques for detecting, quantifying and classifying plant diseases, SpringerPlus, Vol. 2, No. 1, pp. 660-672
[12] Sladojevic, S.; Arsenovic, M.; Anderla, A.; Culibrk, D. & Stefanovic, D. (2016). Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification, Computational Intelligence and Neuroscience, Vol. 2016, No. 1, ISSN 16875273, DOI: 10.1155/2016/3289801
[13] Zhang, S.; Wu, X.; You, Z. & Zhang, L. (2017). Leaf image based cucumber disease recognition using sparse representation classification, Computers and Electronics in Agriculture, Vol. 134, pp. 135-141, ISSN 0168-1699, DOI: 10.1016/j.compag.2017.01.014
[14] Adhikari, S. P.; Yang, H. & Kim, H. (2019). Learning semantic graphics using convolutional encoder–decoder network for autonomous weeding in paddy, Frontiers in Plant Science, Vol. 10, pp. 1404, ISSN 1664-462X, DOI: 10.3389/fpls.2019.01404
[15] Jin, X.; Che, J. & Chen, Y. (2021). Weed identification using deep learning and image processing in vegetable plantation, IEEE Access, Vol. 9, pp. 10940-10950, ISSN 2169-3536, DOI: 10.1109/ACCESS.2021.3050296
[16] Stajnko, D.; Vindis, P. & Mursec, B. (2011). Automated system for targeted spraying in orchards by using RGB imaging, Chapter 23 in DAAAM International Scientific Book 2011, pp. 283-300, B. Katalinic (Ed.), Published by DAAAM International, ISBN 978-3-901509-84-1, ISSN 1726-9687, Vienna, Austria, DOI: 11.2507/daaam.scibook.2011.23
[17] Lac, L.; Da Costa, J.-P.; Donias, M.; Keresztes, B. & Bardet, A. (2022). Crop stem detection and tracking for precision hoeing using deep learning, Computers and Electronics in Agriculture, Vol. 192, pp. 106606, ISSN 01681699, DOI: 10.1016/j.compag.2021.106606
[18] Medeiros, J. G. S.; Biasi, L. A.; Bona, C. M. D. & Cuquel, F. L. (2018). Phenology, production and quality of blueberry produced in humid subtropical climate, Revista Brasileira de Fruticultura, Vol. 40, No. 3, pp. e- -520, ISSN 1806-9967, DOI: 10.1590/0100-29452018520
[19] Stefanović, D.; Antić, A.; Otlokan, M.; Ivošević, B.; Marko, O.; Crnojević, V. & Panić, M. (2022). Blueberry row detection based on uav images for inferring the allowed ugv path in the field, Proceedings of the fifth Iberian Robotics Conference, pp. 401-411, Published by Springer Cham, ISBN 978-3-031-21062-4, Zaragoza, Spain, DOI: 10.1007/978-3-031-21062-4_33
[20] Krizhevsky, A.; Sutskever, I. & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks, Proceedings of twenty fifth Advances in Neural Information Processing Systems, ISBN 9781627480031, pp. 1097-1105
[21] Zou, Z.; Chen, K.; Shi, Z.; Guo, Y. & Ye, J. (2023). Object detection in 20 years: A survey, Proceedings of the IEEE, Vol. 111, No. 3, pp. 257-276, ISSN 0018-9219, DOI: 10.1109/JPROC.2023.3238524
[22] Redmon, J.; Divvala, S.; Girshick, R. & Farhadi, A. (2016). You only look once: Unified, real-time object detection, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), ISBN 978-1-4673-8851-1, pp. 779-788, IEEE, DOI: 10.1109/CVPR.2016.91
[23] Dwyer, B.; Nelson, J.; Hansen, T. & et al. (2024). Roboflow (Version 1.0), Available from: https://roboflow.com Accessed: 2024-09-06
[24] Ge, Z.; Liu, S.; Wang, F.; Li, Z. & Sun, J. (2021). Yolox: Exceeding yolo series in 2021, arXiv preprint arXiv:2107.08430
[25] Qi, D.; Tan, W.; Yao, Q. & Liu, J. (2022). YOLO5Face: Why reinventing a face detector, Proceedings of European Conference on Computer Vision, ISBN 978-3-031-25072-9, pp. 228-244, Tel Aviv, Israel, DOI: 10.1007/978-3-031-25072-9_15

[26] Wu, W.; Chang, H.; Zheng, Y.; Li, Z.; Chen, Z. & Zhang, Z. (2022). Contrastive learning-based robust object detection under smoky conditions, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, ISBN 978-1-6654-8739-9, pp. 4295-4301, New Orleans, LA, USA, DOI: 10.1109/CVPRW56347.2022.00475

[27] Wang, C.-Y.; Liao, H.-Y. M.; Wu, Y.-H.; Chen, P.-Y.; Hsieh, J.-W. & Yeh, I.-H. (2020). CSPNet: A new backbone that can enhance learning capability of CNN, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, ISBN 978-1-7281-9360-1, pp. 390-391, Seattle, WA, USA, DOI: 10.1109/CVPRW50498.2020.00203

[28] He, K.; Zhang, X.; Ren, S & Sun, J. (2015). Spatial pyramid pooling in deep convolutional networks for visual recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 37, No. 9, pp. 1904-1916, ISSN 0162-8828, DOI: 10.1109/TPAMI.2015.2389824

[29] Lin, T.-Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B. & Belongie, S. (2017). Feature pyramid networks for object detection, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, ISBN 978-1-5386-0457-1, pp. 936-944, IEEE, Honolulu, HI, USA, DOI: 10.1109/CVPR.2017.106

[30] Liu, S.; Qi, L.; Qin, H.; Shi, J. & Jia, J. (2018). Path aggregation network for instance segmentation, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, ISBN 978-1-5386-6420-9, pp. 8759-8768, Salt Lake City, UT, USA, DOI: 10.1109/CVPR.2018.00913

[31] Lin, TY. & et al. (2014). Microsoft COCO: Common Objects in Context, Proceedings of Computer Vision - ECCV 2014, ISBN 978-3-319-10602-1, pp. 740-755, Springer, Zurich, Switzerland, DOI: 10.1007/978-3-319-10602-1_48