ANALYSIS OF IMAGES OF VINEYARD WITH COMPUTER VISION MODELS FOR DATA UNDERSTANDING INTERPRETATION

Francesco Vicino, Francesco Paciolla, Tommaso Quartarella, Simone Pascuzzi
University of Bari Aldo Moro, Italy
francesco.paciolla@uniba.it

Abstract: The wine-growing sector represents one of the most important sectors of the agri-food system in the Italian region of Apulia. Suffice it to say that wine-growing alone has about 11 thousand farms and 600 wine cellars. The regional area allocated to this sector is about 89,000 hectares, which represents about 10% of the national figure. In addition, the data indicate that in recent years the production of wine from Puglia has followed an increasing trend, covering about 20% of the national total. However, it is the production of table grapes in this region that boasts 60% of all Italian production. The latest ISTAT data available show that, with over 24,000 hectares, Puglia is the first region for area destined to the cultivation of table grapes. To support this large agri-food chain there are investments related to the purchase of tangible and intangible assets aimed at improving product quality, greater organizational efficiency to increase market demand, and increased competitiveness. To achieve these objectives, agricultural practices related to agriculture 4.0 and precision farming are increasingly used. In this study we want to highlight how, starting from RGB images that can be captured by different devices (drone, robot or camera), it is possible to create a dataset to train computer vision models and obtain potentially usable data at an agronomic level, as a decision support. In particular, the study proposes a computer vision model trained with images of table grapes captured in a vineyard in Apulia (Italy). The objective is to have a greater understanding of the data available in the field from different devices, obtaining a more complete frame of the state of the art of the vineyard for more efficient agronomic management.

Keywords: analysis of images, computer vision, vineyards, agriculture 4.0.

1. Introduction

The detection and tracking of fruits and vegetables is a fundamental aspect of precision agriculture, as detailed in several key studies [1-3]. Testing in low variability environments, characterized by lower fruit position uncertainty, reduced target occlusion and uniform light conditions, facilitates the implementation of learning-based detection systems [4; 5]. Environmental variability significantly affects the performance of these systems, with models trained in controlled light conditions achieving an accuracy of more than 95% compared to those operating in natural light [6; 7]. Furthermore, controlled environments facilitate model training, as in the case of Kim J-SG [8], where a non-destructive system was developed to predict the fresh weight of head lettuce in an industrial plant, using images collected by an automatic image acquisition system. Although artificial vision systems are more effective in controlled environments, open field use is an evolving research challenge, with promising results already highlighted in recent [9; 10]. This study proposes a computer vision system based on smartphone images to develop a model for counting table grape bunches in the open field, with the aim of estimating yields. The objective is to ensure flexibility and adaptability of the model to different cultivars and the typical form of rearing in tent widespread in the region of Puglia, Italy. Previous research has used image analysis to estimate cluster compactness as an alternative to traditional visual methods, but few studies have faced the challenge of working with noisy image datasets [11; 12]. This is crucial for cross-sectional applications, including the monitoring of ripening, the detection of visual damage and the prediction of harvest to optimise response to market demand. The developed model has provided results consistent with the set objectives and is implementable on smartphones, and therefore within reach of farmers who will be able to benefit from a low-cost technology for yield estimation. Future prospects include the integration of the system with an autonomous agricultural robot, able to move between rows and provide real-time data for a more accurate yield estimate.

2. Materials and methods

The images were acquired in several vineyards of table grapes located in the agricultural area of Noicattaro, Puglia, Italy, with the initial objective of assessing the state of ripening of grapevines and any damage to support agronomic analyses and not with the intention of training a computer vision model. Only later was the image dataset used to develop a computer vision model, aimed at providing both a visual overview of plant and cluster health, to allow the counting of bunches for the estimation

of the yield in the different local cultivars. Fig. 1 shows some representative examples of the images used.

Fig. 1. **Images taken in the field with smartphones for the evaluation of the progress of ripening and agronomic aspects**

It is important to note that many of the images acquired were discarded due to the quality being too low for the purpose of the work. In fact, image processing is a fundamental step for the efficient functioning of intelligent models [13; 14]. After cleaning the images, we then proceeded to create the dataset that was used for model training, explained in section 2.1.

2.1. Overview of the Roboflow platform

The platform used for model training is Roboflow. The work focuses on developing an object detection model that identifies the presence of grape clusters in images and then classifies these objects into relevant classes, as has been done in other recent works [15; 16]. The individual training steps are explained in the following sections.

2.2. Labelling of images

At this stage each image has been tagged with only one category: Brunches. Therefore, on the whole dataset containing 150 images of different table grape cultivars, each cluster has been associated with a label, thus adding an additional information to the image. A bounding box has been drawn around the target to have a higher level of detail, understood as the addition of information related to the shape of the cluster and the category of belonging (Ground truth). Fig. 2 shows an example of this.

Fig. 2. **Bounding box of the cluster to which a label has been assigned; the model interprets the label as Ground truth**

2.3. Test split

To train a computer vision model, the image dataset must be divided into three parts.

1. Training split (or Train): composed of those images that will be used to train the model.

2. Validation set (or Val): represented by those images useful to monitor the performance of the model during training and prevent overfitting.
3. Test set (or Test): are all those images that are displayed by the model after training, simulating real data never seen before. The Test serves to separate a part of the dataset that will be used only at the end of the training process to have an objective evaluation of the model.

In this work, the dataset was divided into 86% of images from Train, 9% of images from Validation and 5% of images from Test. The total number of images in the dataset is implemented during the preprocessing and image augmentation phase explained in section 2.4.

2.4. Preprocessing and imagine augmentation

In this step of normalization and standardization, the dataset has been scaled to a fixed size of 640 by 640, avoiding problems related to the variation of sizes between images. An auto-orientation has also been applied to correct any misalignment in the vertical arrangement. To increase the robustness of the model with respect to orientation variations, the dataset has been artificially increased by generating images rotated 90° clockwise and counterclockwise. Before the training, several copies of the original images were created, bringing the dataset to a total of 320 images (maximum dataset size). This process has resulted in a well-structured dataset, improving both the image quality (despite smartphone capture) and the number of samples, with the aim of optimizing model metrics.

3. Results and discussion

The results of training can be explained by Precision and recall values, but especially model metrics. The Precision was calculated as the ratio between *true positive/(true positive – false positive)* and measures how often the model prediction is correct. Its value for this model is 94.9%. The Recall was calculated as the ratio between *true positive/(true positive – false negative)* and measures the percentage of relevant labels correctly identified. The Recall value for this model is 75.5%.

3.1. Training Graphs

Mean Average Precision (mAP) was used to evaluate model performance, being a standard metric for computer vision models. The mAP is calculated as the weighted average of the accuracies at different thresholds, with the previous threshold being used as a weight, and is useful for comparing both different models and different versions of the same model. The value obtained for this model is 86.4%. During training, the behaviour of mAP is illustrated in Fig. 3.

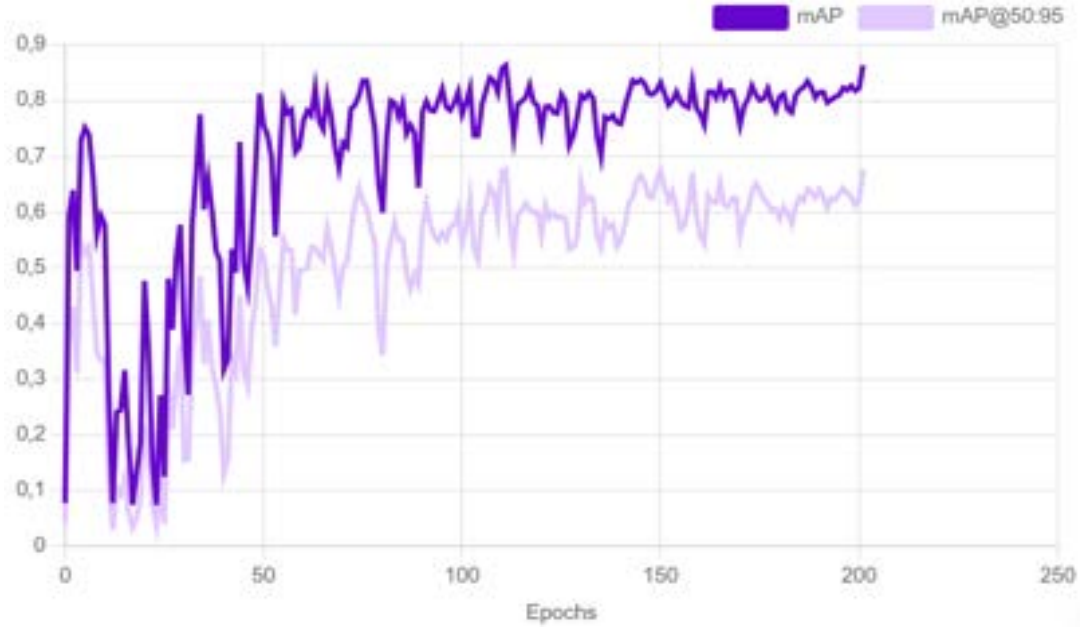

Fig. 3. **Performance of mAP during model training.** Dark purple line (mAP@50) shows the average accuracy considering IoU (Intersection over Union) 50%. Light purple line (mAP@50:95) considers the average accuracy over multiple IoU thresholds, from 50% to 95% in 5% increments

The dark purple line (mAP@50), which considers IoU (Intersection over Union) ≥ 50%, shows a rapid growth in the first epochs, stabilising between 0.75 and 0.8, indicating good learning in cluster detection with a certain margin of error in the bounding box. The light purple line (mAP@50:95), which evaluates accuracy on IoU thresholds between 50% and 95% in increments of 5%, represents a more rigorous metric, requiring greater accuracy in localization. As expected, the values are lower (0.55-0.60), reflecting the higher severity of the criterion. Both curves show an increase with the number of epochs, signaling an improvement in performance. However, after about 100-150 epochs, the pattern ceases to improve significantly, suggesting convergence has been reached and training is potentially sufficient.

3.2. Model vision

In Fig. 4, the image of the first labelling phase is shown on the left (Ground Truth) and on the right, what the model returns to us after training (Model Predictions). From the test images, the model gives us a forecast of the number of clusters present where they can be derived, establishing the confidence threshold parameter, information on agronomic parameters of grape clusters, as well as information on trust and category.

Fig. 4. **On the left the Ground Truth assigned during the image labelling phase; on the right the model predictions that give us the total number of clusters present in the image**

Conclusions

The yield of a table grape vineyard can be estimated by counting the number of clusters from images using computer vision models. In this work a dataset of 150 images taken with a smartphone in different fields with different table grape cultivars was created. The starting images were improved in quality due to the high initial noise, and later a label was assigned to the grape bunches. The dataset was implemented to a total of 320 images and split into Train, Validation and Test images before being subjected to training. Precision and recall values reached 94.9% and 75.5%, respectively. The Mean Average Precision (mAP) metric used to measure the performance of the computer vision model is 86.4%. In the future, it is planned to test the model directly in the field to validate the predictive results of the computer vision system with those measured by an operator in the field during the grape harvesting phase.

Author contributions

All authors contributed equally to the work.

References

[1] Shi X., Wang S., Zhang B., Ding X., Qi P., Qu H., Li N., Wu J., Yang H. Advances in Object detection and Localization Techniques for Fruit Harvesting Robots. Agronomy 2025, 15, 145. DOI: 10.3390/agronomy15010145

[2] Sandeep K., Chiranjoy C., Suman K. Enhancing fruit and vegetable detection in unconstrained environment with a novel dataset. Scientia Horticulturae Volume 338 ,1 dicembre 2024, 113580. DOI: 10.1016/j.scienta.2024.113580
[3] Nuske S., Wilshusen K., Achar S., Yoder L., Narasimhan S., Singh S. Automated Visual Yield Estimation in Vineyards. J. Campo Robotica, 31 (5) (2014), pp. 837-860, DOI: 10.1002/rob.21541
[4] Lee K.H. Samsuzzaman., Reza MN., Islam S., Ahmed, S., Cho YJ., Noh DH., Chung S.-O. Valutazione di modelli di apprendimento automatico per la classificazione dei sintomi di stress delle piantine di cetriolo coltivate in un ambiente controllato. Agronomy 2025 , 15 , 90. DOI: 10.3390/agronomy15010090
[5] Naito H., Shimomoto K., Fukatsu T., Hosoi F., Ota T. Interoperability Analysis of Tomato Fruit Detection Models for Images Taken at Different Facilities, Cultivation Methods, and Times of the Day. AgriEngineering 2024, 6, 1827-1846. DOI: 10.3390/agriengineering6020106
[6] Thomas A.C., Ionut M.M., Leonardo S., Mulham F., Alberto S., Daniele N. Weakly and semi-supervised detection, segmentation and tracking of table grapes with limited and noisy data. Computers and Electronics in Agriculture, Volume 205, February 2023, 107624. DOI: 10.1016/j.compag.2023.107624
[7] Mojtaba D., Yousef A-G., Tarahom M-G., Sajad S., Juan I.A. A stereoscopic video computer vision system for weed discrimination in rice field under both natural and controlled light conditions by machine learning. Measurement, volume 237, 30 September 2024, 115072. DOI: 10.1016/j.measurement.2024.115072
[8] Kim J-SG., Moon S., Park J., Kim T., Chung S. (2024). Development of a machine vision-based weight prediction system of butterhead lettuce (Lactuca sativa L.) using deep learning models for industrial plant factory. Front. Plant Sci. 15:1365266. doi: 10.3389/fpls.2024.1365266
[9] Ameer T.K., Signe M.J., Abdul R.K. Advancing precision agriculture: A comparative analysis of YOLOv8 for multi-class weed detection in cotton cultivation. Artificial Intelligence in Agriculture, volume 15, Issue 2, June 2025, Pages 182-191. DOI: 10.1016/j.aiia.2025.01.013
[10] Akshay D., Satish C. Deep learning based weed classification in corn using improved attention mechanism empowered by Explainable AI techniques. Crop Protection, volume 190, April 2025, 107058. DOI: 10.1016/j.cropro.2024.107058
[11] Cubero S., Diago M.P., Blasco J., Tardaguila J., Prats-Montalbán J.M., Ibáñez J., Tello J., Aleixos N. A new method for assessment of bunch compactness using automated image analysis. Wiley Online Library, 20 January 2015. DOI: 10.1111/ajgw.12118
[12] Diana-Carmen R-L., Diana-Margarita C-E., José M. Á-A., Juan T., Julio-Alejandro R-G., Juvenal Rodríguez-Reséndiz. Trends in Machine and Deep Learning Techniques for Plant Disease Identification: A Systematic Review. Agriculture 2024, 14, 2188. DOI: 10.3390/agriculture14122188
[13] Udeogu C.U., Nwakanma C.I., Ayoade I.A., Amadi C.S., Eze U.F. Agro-vision IoT-enabled Crop Pest Recognition System based on VGG-16. In Proceedings of the 2023 2nd International Conference on Multidisciplinary Engineering and Applied Science (ICMEAS), Abuja, Nigeria, 1–3 November 2023; IEEE: Piscataway Township, NJ, USA, 2023; pp. 1–5.
[14] Shiyu L., Yiannis A., Congliang Z., Won S.L. AI-driven time series analysis for predicting strawberry weekly yields integrating fruit monitoring and weather data for optimized harvest planning. Computers and Electronics in Agriculture, volume 233, June 2025, 110212. DOI: 10.1016/j.compag.2025.110212
[15] Joseph R., Santosh D., Ross G., Ali F. You only look once: Unified, real-time object detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2016), pp. 779-788.
[16] He Z., Karkee M., Zhang Q. Detecting and localizing strawberry centers for robotic harvesting in field environment. IFAC-PapersOnLine, 55 (32) (2022), pp. 30-35