Data in Brief 43 (2022) 108466

Contents lists available at ScienceDirect

Data in Brief

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

Data Article

wGrapeUNIPD-DL: An open dataset for white grape bunch detection ☆

Marco Sozzi [a,*], Silvia Cantalamessa [b], Alessia Cogato [c], Ahmed Kayad [d], Francesco Marinello [a]

[a] Department of Land Environment Agriculture and Forestry, University of Padova, Legnaro 35020, Italy
[b] Department of Agronomy, Food, Natural Resources, Animals, and Environment, University of Padova, Legnaro 35020, Italy
[c] Department of Agricultural and Environmental Sciences, University of Udine, Udine 33100, Italy
[d] Department of Botany and Plant Sciences, University of California, Riverside, CA 92521, United States

ARTICLE INFO

Article history:
Received 20 May 2022
Revised 26 June 2022
Accepted 7 July 2022
Available online 13 July 2022

Keywords:
Object detection
Early yield estimation
Cluster detection
Digital agriculture
Crop load mapping
Grape yield
Precision viticulture

ABSTRACT

National and international Vitis variety catalogues can be used as image datasets for computer vision in viticulture. These databases archive ampelographic features and phenology of several grape varieties and plant structures images (e.g. leaf, bunch, shoots). Although these archives represent a potential database for computer vision in viticulture, plant structure images are acquired singularly and mostly not directly in the vineyard. Localization computer vision models would take advantage of multiple objects in the same image, allowing more efficient training. The present images and labels dataset was designed to overcome such limitations and provide suitable images for multiple cluster identification in white grape varieties. A group of 373 images were acquired from later view in vertical shoot position vineyards in six different Italian locations at different phenological stages. Images were then labelled in YOLO labelling format. The dataset was made available both in terms of images and labels. The real number of bunches counted in the field, and the number of bunches visible in the image (not covered by other

☆ Francesco Marinello has a role as Editorial Board Member of this journal but had no involvement in the peer-review of this article and has no access to information regarding its peer-review.

* Corresponding author.
E-mail address: marco.sozzi@unipd.it (M. Sozzi).
Social media: @3MSozzi (M. Sozzi), @AGKayad (A. Kayad), @marinello_unipd (F. Marinello)

https://doi.org/10.1016/j.dib.2022.108466

vine structures) was recorded for a group of images in this dataset.

Specifications Table

Subject	Agriculture engineering
Specific subject area	Application of computer vision to agriculture for grape variety classification and yield estimation
Type of data	Image Label (bounding boxes)
How the data were acquired	Image were acquired by using a Nikon D300 (Nikon Corporation, Shinjuku, Japan) camera equipped with Sigma 18-200 optics (Sigma Corporation, Kawasaki, Japan), and different smartphones. Image were manually labelled using Yolo_label V2 project
Data format	Classified
Description of data collection	Sideview images of vineyard canopy were acquired by the authors in 6 experimental vineyards located in different Italian regions during the 2020 growing season. Most of the images were calibrated for size and color using Macbeth color references. All images were acquired with a distance from the side canopy from 1.5 up to 3 meters.
Data source location	- Institution: Department of Land Environment Agriculture and Forestry, University of Padova; - City: Legnaro; - Country: Italy;
Data accessibility	Repository name: Zenodo Data identification number: 10.5281/zenodo.4066730 Direct URL to data: https://zenodo.org/record/4066730#.YofMr9hBxPY Instructions for accessing these data: data are Open Access in Creative Commons Attribution 4.0 International
Related research article	Sozzi, M., Cantalamessa, S., Cogato, A., Kayad, A., & Marinello, F. (2022). Automatic Bunch Detection in White Grape Varieties Using YOLOv3, YOLOv4, and YOLOv5 Deep Learning Algorithms. Agronomy, 12(2), 319. 10.3390/agronomy12020319
Reference of dataset	Marco Sozzi, Silvia Cantalamessa, Alessia Cogato, Ahmed Kayad, & Francesco Marinello. (2022). wGrapeUNIPD-DL: an open dataset for white grape bunch detection [Data set]. Zenodo 10.5281/zenodo.4066729

Value of the Data

- This dataset can be used to train classification and object detection algorithm of cluster on white grape varieties.
- Researchers, professional, and data scientist can benefit of this dataset to train models for phenological stage recognition, cluster counting, and variety classification.
- This dataset can be used to train deep learning crop load estimation. In addition, it can be reused for white grape varieties classification.
- This dataset can be combined with national and international Vitis variety catalogues for computer vision application in viticulture.
- This dataset overcomes the limitation of national and international Vitis variety catalogues where plant structures images are acquired singularly and mostly not directly in the vineyard since multiple objects are present allowing a more efficient training.

1. Data Description

The database is divided into three levels. The primary dataset folder contains two sub-folders, named Calibrated_Images (271 images) and Uncalibrated_Images (102 images), respectively. These two folders represent the first level. The Calibrated_Imgaes folder contains two sub-folders (which represent the second level), named with_Counting (24 images) and without_Counting (247 images). Uncalibrated_imgaes folder, at the second level, contains only one sub-folder, named without_Bunches, where all images are included. The with_Counting folder comprises a text file (_counting.txt) with the real number of bunches counted in the field and the number of bunches visible in the image (not covered by other vine structures). The third level is represented by folders contained in the second level sub-folder, which names correspond to the varieties, the phenology and the acquisition date of the included images (e.g. Chardonnay_BBCH75_20_06_20). All images contained in each folder are matched with the associated label, which has the same name. The dataset structure is visible in Fig. 1.

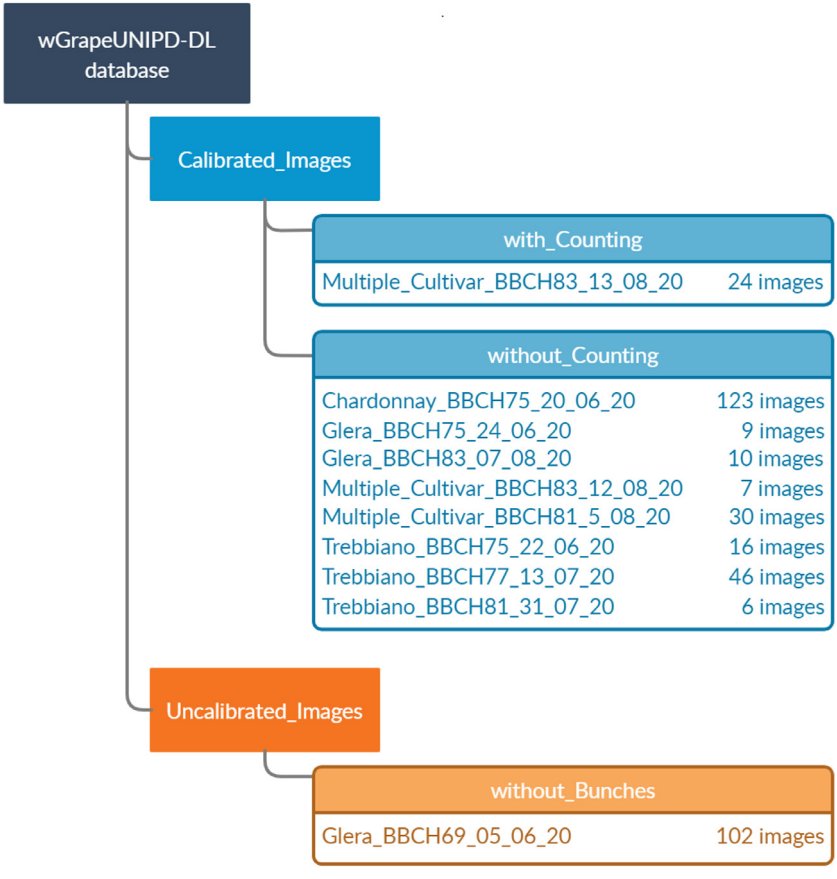


Fig. 1. wGrapeUNIPD-DL folder names and structure.

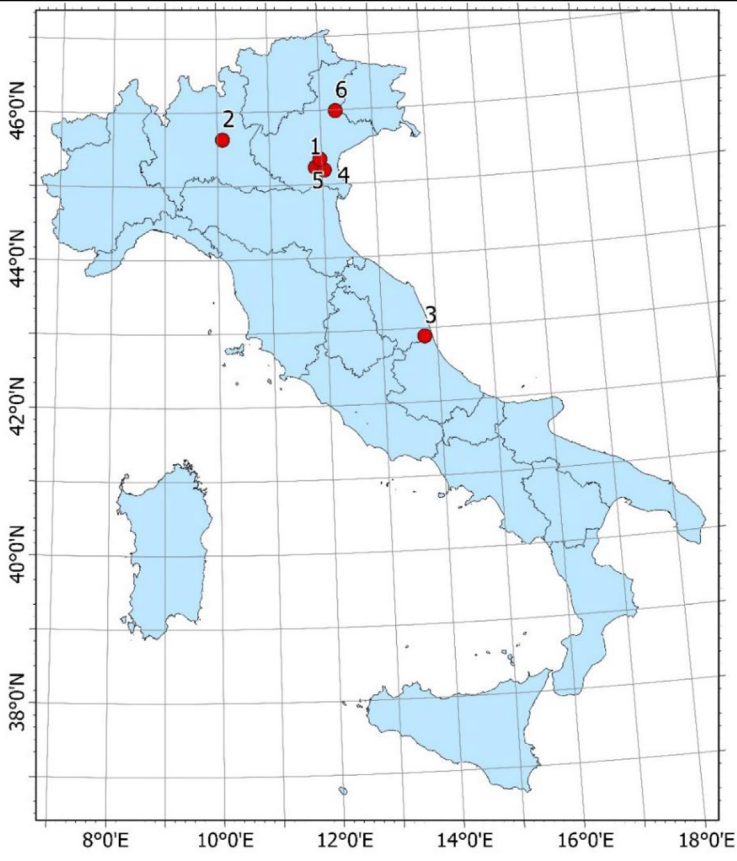
2. Experimental Design, Materials and Methods

A group of 373 images were acquired at field conditions [1]. Sideview images of vineyard canopy were acquired during 2020 growing season in 6 different vineyards located in North and Central Italy, from BBCH 69 (late blooming) up to BBCH 83 (veraison) phenological stages (Fig. 2). Legend of BBCH scale is showed in Table 1. Most of the images (271) were calibrated for size and colour using Macbeth colour references (Fig. 3) [2]. The availability of colour reference makes colour-based classification ML possible for potential users. One set of data (Uncalibrated_Images, 102 images) was acquired without size and colour reference as images were acquired at BBCH69, and bunches were not visible. Most of the images (203) were acquired with Nikon D300 (Nikon Corporation, Shinjuku, Japan) equipped with Sigma 18–200 optics (Sigma Corporation, Kawasaki, Japan), while the remaining images were acquired with different smartphones. All images were acquired with a distance from the canopy wall from 1.5 up to 3 meters, allowing the application of retrieved classification models on all ground vehicle and agricultural robots [3]. Vines where images were acquired where selected in order to avoid abiotic and biotic stress (e.g. water stresses) [4]. Sensors and optical features have been included in the image metadata. Dataset is mainly composed of images acquired on Chardonnay (123 images) and Glera (121 images) varieties, while 68 images were acquired on Trebbiano varieties; 61 images were acquired in a vineyard from the University of Padova, characterized by several varieties, which were not identified. Example of vines at different phenological stage is visible in Fig. 4.

Table 1
Description of phenological stages.

BBCH-scale	Description
69	End of flowering
75	Berries pea-sized, bunches hang
77	Berries beginning to touch
81	Beginning of ripening: berries begin to develop variety-specific colour
83	Berries developing colour

Data annotation (labelling) was manually performed by the authors, drawing bounding boxes on each bunch in the image using Yolo_label V2 project [5]. Yolo_label allows to create annotation (label) for object detection algorithm using the Yolo label format, which consists of five columns for each object (object-class, x, y, width, and height). As only one class (bunches) was used to label the dataset of this study, all label text files start with 0, which is the identification of the first index in Python.



ID	Cultivar	Phenology	Number of Images	Place and Coordinates	Date
1	Glera	BBCH 69	102	Conselve (PD) -ITA 45.316180 11.914806	05/06/2020
2	Chardonnay	BBCH 75	123	Monticelli Brusati (BS) - ITA 45.632234 10.099519	20/06/2020
3	Trebbiano	BBCH 75 BBCH 77 BBCH 81	16 46 6	Spinetoli (AP) - ITA 42.889327 13.773421	22/06/2020 13/07/2020 31/07/2020
4	Glera	BBCH 75	9	Cona (VE) - ITA 45.192860 12.039130	24/06/2020
5	Multiple Cultivar	BBCH 81 BBCH 83 BBCH 83	30 7 24	Legnaro (PD) - ITA 45.343678 11.962747	05/08/2020 12/08/2020 13/08/2020
6	Glera	BBCH 81	10	Vittorio Veneto (TV) - ITA 45.989834 12.296151	07/08/2020

Fig. 2. Location, phenology and dataset dimension for each data acquisition campaign.

Fig. 3. Example of Macbeth color reference (source Wikipedia CC BY-SA 4.0).

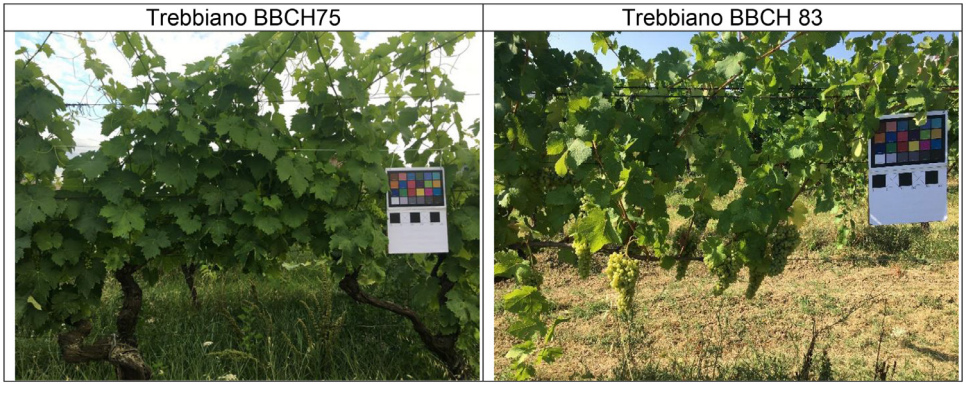


Fig. 4. Example of acquired images in different phonological stages.

Ethics Statements

Dataset do not include human subjects, animal in experiments or data collected from social media platforms.

Declaration of Competing Interest

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

Data Availability

wGrapeUNIPD-DL: an open dataset for white grape bunch detection (Original data) (Zenodo).

CRediT Author Statement

Marco Sozzi: Conceptualization, Methodology, Data curation, Writing – original draft; **Silvia Cantalamessa:** Conceptualization, Methodology, Data curation, Writing – original draft; **Alessia Cogato:** Data curation, Writing – original draft; **Ahmed Kayad:** Writing – review & editing; **Francesco Marinello:** Supervision.

Acknowledgments

The authors would like to acknowledge Dr. Diego Tomasi and Dr. Davide Boscaro (Council for Agricultural Research and Economics-Research Centre for Viticulture and Enology -Conegliano, Italy), Dr. Franco Meggio (DAFNAE dept. University of Padova), Giuliano Preghenella, Tenuta Civrana S.R.L., Società Agricola Castelveder, Società Agricola Grappolo D'oro, and Azienda Agricola Saladini Pilastri for their technical support.

This research was financially supported by the Land Environment Resources and Health (L.E.R.H.) doctoral course.

References

[1] M. Sozzi, S. Cantalamessa, A. Cogato, A. Kayad, F. Marinello, Automatic bunch detection in white grape varieties using YOLOv3, YOLOv4, and YOLOv5 deep learning algorithms, *Agronomy*, *12*(2), 319. MDPI AG. http://dx.doi.org/10.3390/agronomy12020319 (2022).

[2] K.P. Seng, L.M. Ang, L.M. Schmidtke, S.Y. Rogiers, Computer vision and machine learning for viticulture technology, IEEE Access 6 (2018) 67494–67510, doi:10.1109/ACCESS.2018.2875862.

[3] M. Vieri, R. Lisci, M. Rimediotti, D. Sarri, The RHEA-project robot for tree crops pesticide application, J. Agric. Eng. 44 (2013) 359–362, doi:10.4081/JAE.2013.(S1):E71.

[4] A. Cogato, L. Wu, S.Y.Y. Jewan, F. Meggio, F. Marinello, M. Sozzi, V. Pagay, Evaluating the spectral and physiological responses of grapevines (Vitis vinifera L.) to heat and water stresses under different vineyard cooling and irrigation strategies, Agronomy 11 (2021) 1940, doi:10.3390/AGRONOMY11101940.

[5] Y. Kwon, D. Marrable, R. Abdulatipov, J. Loïck, Yolo_label, (2020). https://github.com/developer0hye/Yolo_Label. Accessed September 1, 2020.