



**HAL**  
open science

# Ripening dynamics revisited: an automated method to track the development of asynchronous berries on time-lapse images

Benoit Daviet, Christian Fournier, Llorenç Cabrera-Bosquet, Thierry Simonneau, Maxence Cafier, Charles Romieu

## ► To cite this version:

Benoit Daviet, Christian Fournier, Llorenç Cabrera-Bosquet, Thierry Simonneau, Maxence Cafier, et al.. Ripening dynamics revisited: an automated method to track the development of asynchronous berries on time-lapse images. *Plant Methods*, 2023, 19, pp.146. 10.1186/s13007-023-01125-8. hal-04515162

**HAL Id: hal-04515162**

**<https://hal.inrae.fr/hal-04515162>**

Submitted on 21 Mar 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

METHODOLOGY

Open Access



# Ripening dynamics revisited: an automated method to track the development of asynchronous berries on time-lapse images

Benoit Daviet<sup>1</sup>, Christian Fournier<sup>1\*</sup>, Llorenç Cabrera-Bosquet<sup>1</sup>, Thierry Simonneau<sup>1</sup>, Maxence Cafier<sup>2</sup> and Charles Romieu<sup>2\*</sup>

## Abstract

**Background** Grapevine berries undergo asynchronous growth and ripening dynamics within the same bunch. Due to the lack of efficient methods to perform sequential non-destructive measurements on a representative number of individual berries, the genetic and environmental origins of this heterogeneity, remain nearly unknown. To address these limitations, we propose a method to track the growth and coloration kinetics of individual berries on time-lapse images of grapevine bunches.

**Results** First, a deep-learning approach is used to detect berries with at least  $50 \pm 10\%$  of visible contours, and infer the shape they would have in the absence of occlusions. Second, a tracking algorithm was developed to assign a common label to shapes representing the same berry along the time-series. Training and validation of the methods were performed on challenging image datasets acquired in a robotised high-throughput phenotyping platform. Berries were detected on various genotypes with a F1-score of 91.8%, and segmented with a mean absolute error of 4.1% on their area. Tracking allowed to label and retrieve the temporal identity of more than half of the segmented berries, with an accuracy of 98.1%. This method was used to extract individual growth and colour kinetics of various berries from the same bunch, allowing us to propose the first statistically relevant analysis of berry ripening kinetics, with a time resolution lower than one day.

**Conclusions** We successfully developed a fully-automated open-source method to detect, segment and track overlapping berries in time-series of grapevine bunch images acquired in laboratory conditions. This makes it possible to quantify fine aspects of individual berry development, and to characterise the asynchrony within the bunch. The interest of such analysis was illustrated here for one cultivar, but the method has the potential to be applied in a high throughput phenotyping context. This opens the way for revisiting the genetic and environmental variations of the ripening dynamics. Such variations could be considered both from the point of view of fruit development and the phenological structure of the population, which would constitute a paradigm shift.

**Keywords** High-throughput phenotyping, Computer vision, Grapevine berry, Fruit detection, Fruit segmentation, Tracking

\*Correspondence:

Christian Fournier  
christian.fournier@inrae.fr  
Charles Romieu  
charles.romieu@inrae.fr

Full list of author information is available at the end of the article



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Background

Unlike climacteric fruits (e.g. bananas, apples or mangoes) which accumulate sufficient starch reserve to achieve post-harvest ripening, the grape berry necessarily ripens on the vine, at the rate of translocation of water and sucrose through the phloem. Ripening involves the sudden activation of the apoplastic pathway of phloem unloading [52], which leads to the second growth phase during which each berry accumulates about 1 M hexoses and becomes coloured [5, 22, 36, 42]. Then, the definitive stop of phloem unloading triggers a more or less pronounced shrivelling period known as overripening [30, 41]. It is widely accepted that these dynamic processes are under strong developmental and transcriptomic control [12, 39], and may vary according to the genotype and its interaction with environmental conditions ( $G \times E$ ), particularly light, temperature, and water availability [45]. A considerable research effort is devoted to understanding the phenological, physiological and molecular origins of such processes to better anticipate the effects of global change on grapevine yield and quality [37].

However, due to the lack of efficient non-destructive phenotyping methods to study berries individually, the body of knowledge is mainly based on measurements of the average evolution in periodic samples of randomly selected berries. This approach overlooks the heterogeneity of the fruit population representative of the future harvest. Moreover, scarce studies on single berries recently revealed how chimerical these population-relevant samples are, on both the phenology and metabolic point of view. For example, the fact that the developmental lag between two berries can be almost as long as their growth duration leads to a two fold overestimation of the duration of the second growth period, when considering the average evolution among several berries [5, 42]. Furthermore, mixing growing and shrivelling berries leads to the averaging bias that a constant volume is maintained during late ripening, and that excess water from the phloem mass flow must be released into the xylem back-flow [25]. There is thus a clear need to develop methods for temporal and non-destructive monitoring of cohorts of individual fruits.

Non-destructive spectrometric methods such as NIR, fluorescence and hyperspectral imaging have received considerable attention for harvest date anticipation based on berry ripeness assessment (e.g. [13, 24, 33]). The major interest of these methods is that they eliminate the need for solute extraction and physico-chemical tests, and make it possible to objectivise the heterogeneity of maturities at plot level. However, such data acquisition may be practically as tedious as harvesting representative samples. It also misses the kinetics of volume growth, which is critically needed to predict yield and distinguish

the sugar accumulation phase from its final concentration. Alternatively, time-lapse RGB imaging of a grapevine bunch could be used to monitor the evolution of the external aspect of individual berries over time, such as berry volumes or berry colour. Indeed, the dynamics of such external features are closely linked to internal physico-chemical changes occurring during berry ripeness, and thus could be used as proxies to study ripening dynamics. While the efficiency of this non-destructive approach was demonstrated with manual annotation of the images [28], only the automation of such tasks would allow a large enough sampling to get a representative view of the ripening process and its variability.

The first task to be automated is the detection and segmentation of individual berries. This task is challenging, due to the natural variability of the aspect of berries (e.g. shape, size, colour, degree of light exposure) and to the fact that they frequently overlap with other berries and plant parts. Deep-learning has shown to be an effective solution to this problem for a number of fruits such as oranges [16], blueberries [18, 34], apples [17, 23], strawberries [38] and grapevine berries [43]. In all these studies, an instance segmentation model (e.g. Mask R-CNN [20]) was trained on manual annotations of visible fruit parts to retrieve the apparent contour of each fruit. This strategy is suitable for measuring their colour [43], counting them to estimate yield [51], or locating them for automatic fruit picking [47]. However, it misses the occluded parts of berries that are partially covered by neighbouring fruits, which frequently occurs in ordinary bunches, thus preventing the deduction of statistics related to their real shape such as volume. To cope with this, [31] used ellipse fitting as a post-processing of the segmented contours to infer a plausible intrinsic contour of individual berries. Alternatively, deep-learning models can be trained on annotations guessing the shape each fruit would have in the absence of occlusions, so that predictions of the segmentation model directly infer complete fruit shapes, including their hidden parts [1, 9, 27]. The annotation protocol and the extent to which the hidden parts can be deduced from the visible ones are crucial in such cases, as annotation errors will be learned by the models and will directly alter predictions. Higher level of occlusions can be addressed by training the model with synthetic images for which various levels of occlusion can be generated, by artificially superposing images of isolated fruit and other plant elements [21] or by rendering plant models in a 3D graphics software [2].

The second task to be automated is the tracking of segmented berries over successive time steps, to deduce individual volume and colour kinetics. The majority of fruit tracking algorithms addressed the issue of matching segmented instances between different viewpoints, or

over time on short videos (seconds to minutes) of several frames per second [27, 50, 53]. Hondo et al. [21] managed to track apples over periods of several weeks, but for a very limited number of instances (two well separated apples), which is far from the issue of tracking dozens to hundreds of overlapping instances over a long period of time, as needed for following berry ripening.

In this paper, we introduce a fully automated method to measure and track the size and colour of individual berries on time-lapse images of grapevine bunches. The method starts with a detection model to recognize berries that are sufficiently visible to reasonably infer their size. Second, a segmentation model was trained to infer both the visible and hidden contours of individual berries, using a training dataset derived from a fast and original annotation method. Ellipses are further fitted on the segmented contours to compute position and shape parameters for each berry. Finally, we adapted a tracking algorithm to assign time-consistent labels to the detected berries while handling global deformations of the bunch. This method was tested on image time-series acquired at the PhenoArch platform [8], to assess the quality and the limits of the method at quantifying individual berry growth kinetics. We finally showed how this unprecedented data analysis can provide new insights on the ripening dynamics of grape berries.

**Materials and methods**

**Plant material, image acquisition and dataset composition**

The complete pipeline (segmentation and tracking) was tested on an image dataset from two independent experiments conducted in 2020 and 2021, spanning 51 and 32 days respectively, each containing 9 grapevine (*Vitis vinifera* L.) plants. For each plant, RGB images (2048×2448 px) of a selected grapevine bunch were taken every 8 h. An additional dataset including bunch images of 78 grapevine genotypes from a diversity panel maximising genetic diversity [35] was used to robustify the training and evaluation of the berry segmentation pipeline (without tracking). All experiments were conducted in the PhenoArch phenotyping platform ([https://www6.montpellier.inrae.fr/lepse\\_eng/Phenotyping-platforms-M3P/Montpellier-Plant-Phenotyping-Platforms-M3P/PhenoArch](https://www6.montpellier.inrae.fr/lepse_eng/Phenotyping-platforms-M3P/Montpellier-Plant-Phenotyping-Platforms-M3P/PhenoArch)), hosted at the M3P (Montpellier Plant Phenotyping Platforms) [8].

Images were shot using an imaging cabin of PhenoArch, which involves an RGB camera (Grasshopper3, Point Grey Research, Richmond, BC, mounted in a robotized XYZ arm and LED illumination (5050–6500 K colour temperature). For each plant, the bunch position was manually recorded at the beginning of the experiment, and a robotic arm (see [7] for details) was then used to automatically position the camera to a fixed

time-consistent position along the experiment allowing to get a detailed shot of the bunch.

**Detection, segmentation and features extraction of individual berries**

The objective of this step is to (i) detect berries suitable for shape inference, defined as berries with more than half of their contours visible in a grapevine bunch image, (ii) infer their complete shape, and (iii) extract features that allow quantifying their size and colour. The first two sub-steps rely on deep-learning models which have to be trained on annotations of complete berry contours inferring their hidden part.

- a) Construction of the annotation dataset

The annotated dataset contains 159 images, sampled from the three experiments (Table 1). The sampling was done to best cover all stages of growth, and maximise the visual diversity of the berries in the dataset in terms of size, shape, colour, texture, blurring and shading. It also includes various levels of occlusions between berries or with other plant organs and objects.

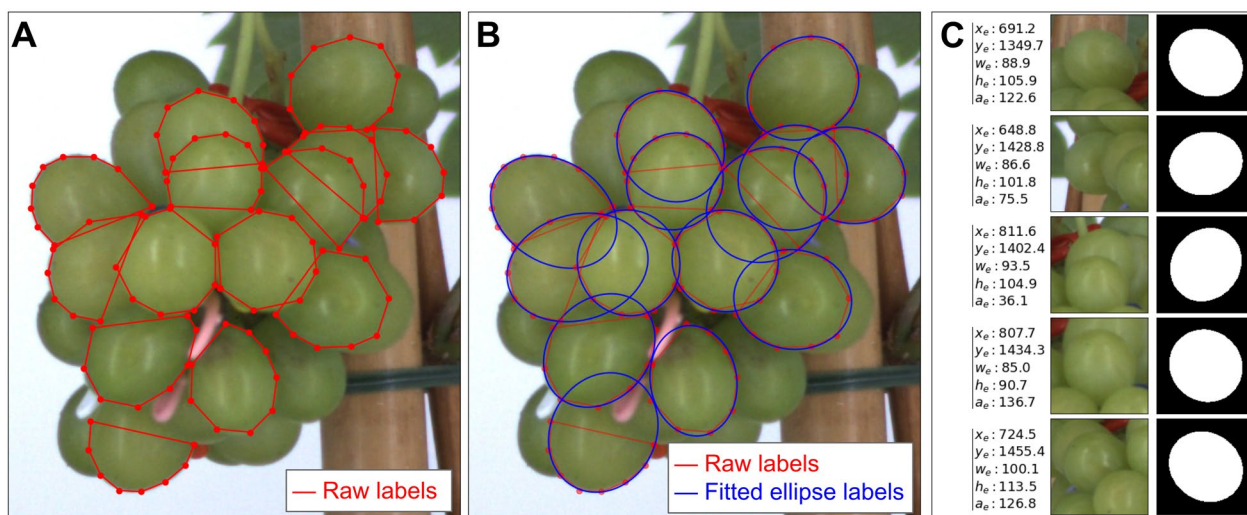
A total of 6134 berries were manually annotated as polygons using Labelme [49]. Similarly to [31], only berries with at least half of their contours visible were annotated. Berries that did not reach pea size stage were rejected based on the assessment of their morphological characteristics, as they are not relevant for studying ripening. For each berry, an average of only 8 points (at least 5) were placed along the uncovered parts of its contours (Fig. 1A). Then, least-square ellipse fitting [14] was used to fit 5 ellipse parameters ( $x_e, y_e, w_e, h_e, a_e$ ) to the set of points (Fig. 1B; blue lines), with  $(x_e, y_e)$  the centre coordinates of the ellipse,  $w_e$  and  $h_e$  the respective length of minor and major ellipse axis, and  $a_e$  the ellipse rotation.  $w_b$  and  $h_b$  were further deduced as the width and height of the smallest box enclosing the ellipse.

This dataset (Table 1) was then split into training (129 images; 4447 labels), validation (10 images; 814 labels) and test (20 images; 873 labels) subsets. Each subset includes different plants, to better assess the generalisability of the detection and segmentation models. We

**Table 1** Annotated dataset

Experiment year	n. of images	n. of plants	n. of genotypes	n. of annotated berries
2020	8	8	1	536
2021	33	22	3	1670
2022	118	109	78	3928





**Fig. 1** Berry annotation procedure. **A** Raw labels, consisting of simple polygons (5 to 10 points) drawn manually along the edges of berries with at least 50% of their contour visible. **B** Gussed actual contour of berries, obtained by an automatic ellipse fitting (blue) on the annotated points. **C** Instances generated from the annotations dataset, used to train the segmentation model. Each instance corresponds to one berry, for which we show the fitted ellipse parameters, the image input and the targeted binary segmentation mask

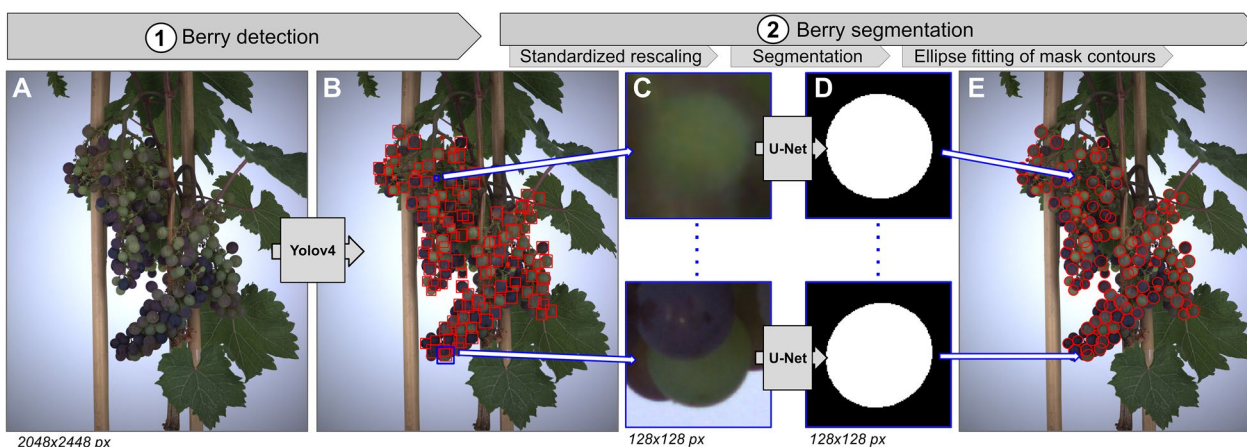
also restricted the test subset to genotypes not present in the training subset, to ensure that our method generalises to a wide range of genetic diversity. The test subset was therefore only based on images from the 2022 experiment, whose batches contained fewer berries on average.

b) Detection of measurable berries

To detect measurable berries on an image of a grapevine bunch, a Yolov4 deep-learning object detection

model [6] was trained to find bounding boxes around berries with at least 50% visible contour in 416×416 px sub-parts of the image (Fig. 2A, B).

20,000 training instances were generated by cropping 416×416 px sub-parts of the training images, each being labelled by the list of parameters of the boxes entirely included in it. This dataset was further augmented with random adjustments of vignettes hue, saturation and brightness, and random flips of image-label pairs. It was then used to train the model, using the yolov4-tiny



**Fig. 2** Berry detection and segmentation pipeline. **A** RGB image of a grapevine bunch acquired in the PhenoArch platform [8]. **B** Bounding boxes (red rectangles) detected by a Yolov4 deep-learning model trained to identify berries with at least 50% visible contour. **C** Vignettes cropped around the centre coordinates of detected boxes, and resized to 128×128 px. The resizing ensures that berries occupy a similar space in the vignette regardless of their size. **D** Binary segmentation masks predicted by a U-Net deep-learning model on berry vignettes. The model was trained to infer the shape of berries in the absence of occlusions. **E** Ellipse fitting of the contour points extracted from a segmentation mask, and projection of the ellipse (red) on the original image

architecture and default hyperparameters [6]. Model weights were stored every 500 iterations for a total of 65,000 iterations. The weights leading to the highest Average Precision (AP) on the validation dataset were saved.

For predictions, the 2048×2448 px source image is split into image sub-parts cropped over the entire pixel range with a maximum spacing of 270 px, which are then fed to the detection model, resulting in a set of predicted parameters of the box dimensions ( $\widehat{w}_b, \widehat{h}_b$ ) and centre coordinates ( $\widehat{x}_b, \widehat{y}_b$ ). Because of sub-image overlaps, the same berry can be detected more than once. To remove these redundancies, non-maximum suppression is used to avoid having box pairs with an intersection over union above 70%. Berries detected with a confidence score below a threshold  $s = 0.89$  are filtered out. This value of  $s$  was chosen to maximise the F1-score on the validation subset.

c) Segmentation of berries

For each detected berry, a square vignette of size  $s = \max(\widehat{w}_b, \widehat{h}_b) / z$  is cropped around its box centre coordinate ( $\widehat{x}_b, \widehat{y}_b$ ). A constant value  $z = 0.75$  is used to ensure that all berries are entirely contained within their respective vignette, and occupy a similar space regardless of their size (Fig. 2C). Each vignette is then resized to 128×128 px by bilinear interpolation, and fed to a U-Net [40] deep-learning model with a VGG16 [44] backbone. The model was trained to output a binary mask representing the shape of the berry as if it were not occluded by any other element present in the image (Fig. 2D).

To train the segmentation model, 40,000 vignettes were extracted from the annotation dataset using the cropping method described above. Elliptic mask labels were directly generated using the annotated ellipse parameters (Fig. 1C). Random noise was applied to the centre coordinate and value of  $z$  during cropping, to help the model handle detection inaccuracies. This was supplemented by the augmentation scheme explained earlier in the detection section. Therefore, all the masks generated had ellipse shapes of similar sizes, in order to restrict the learning domain of the model. These vignettes and mask labels were then used as inputs and output to train the model, using categorical cross-entropy loss, Adam optimizer, and a learning-rate of 0.0001. The number of iterations was automatically chosen with early stopping, and the model weights leading to the minimal validation loss were saved.

d) Extraction of berry morphology and colour features

Assuming that the resulting mask has an elliptic shape, its contour points are extracted as in [46], to fit  $(\widehat{x}_e, \widehat{y}_e, \widehat{w}_e, \widehat{h}_e, \widehat{a}_e)$  ellipse parameters [14]. These parameters are then rescaled to the original image coordinate space (Fig. 2E). For each berry the following features are computed:

Colour: the raw hue  $h_{raw}$  of a berry is computed as the circular mean of the hue angle of the pixels contained inside the ellipse, after removing the pixels that are less than  $dp = \max(3, w_e/4)$  px away from the ellipse's edges, and removing the pixels shared by other ellipses. Given  $h_{50} = 100^\circ$  the mean value of  $h_{raw}$  for grape berries that are halfway through their colour change from green to black in our dataset, the centred berry hue  $H$  is defined as:

$$H = (180 - h_{raw} - h_{50})\%180 \tag{1}$$

Volume: Berry volume  $V$  is estimated as the volume of the sphere that has the same projection area  $A$  as the ellipse fitting the individual berry shape, as in [11]:

$$A = \frac{\widehat{w}_e}{2} \times \frac{\widehat{h}_e}{2} \times \pi \tag{2}$$

$$V = \frac{4\pi}{3} \times \sqrt[3]{\frac{A}{\pi}} \tag{3}$$

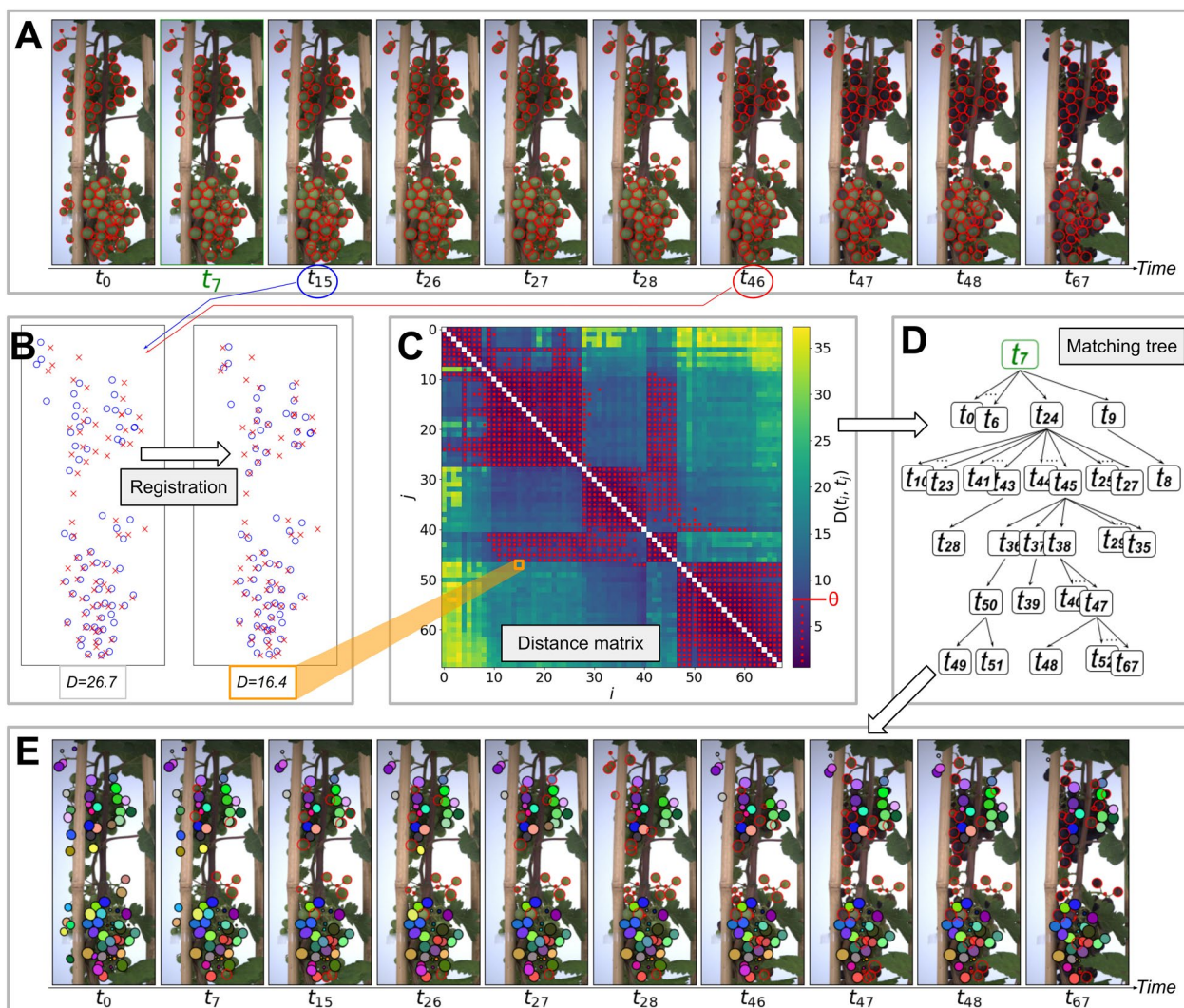
It should be noticed that  $V$  is only a geometric transform of the measured projected area that we found convenient for comparing our result with other studies. We do not further investigate the accuracy of such an estimate as our study focuses on relative variations in  $V$ .

The camera height was individually adjusted to the height of each bunch. Assuming that the distance between a bunch and the camera was uniform across plants, and relatively large compared to the differences in distance to the camera across berries, a constant calibration factor of  $3.94 \times 10^{-6} \text{ mLpx}^{-3}$  was used to express  $V$  in mL. This ratio was calculated by capturing an image of a chessboard with known dimensions and comparing its pixel representation with its actual size.

**Time-lapse tracking of individual berries**

This step aims to track individual berries over successive segmented images of a grapevine bunch, that is to associate a unique label to each berry over time (Fig. 3A, E). To that end, three independent methods were combined (i.e.





**Fig. 3** Berry time-lapse tracking pipeline. **A** 10 segmented RGB images sampled from a 68 images time series representing the evolution of one bunch over time. Raw images were captured with a median interval of 8 h. **B** Scatter plots of the coordinates of the berry ellipses centres detected at two time steps ( $t_{15}$ ; blue circles,  $t_{46}$ ; red crosses), before (left) and after (right) registration. The distance metrics  $D$  between the two point sets is given below each plot. **C** Heat map of the distance matrix, storing the distance between all pairs of time-points after registration. Red points correspond to matrix values below the threshold  $\theta = 8\text{px}$ . **D** Matching tree, determining the order in which labels are propagated during tracking. Each rectangle represents a time-step. The highest one corresponds to  $t_{root}$ , used to initialise tracking labels. **E** Labelled segmented images after tracking. Each colour corresponds to one tracking label. Segmented berries without label (no match found with  $t_{root}$ ) are drawn as red empty ellipses

Baseline, Registration and Matching Tree). First, an original algorithm (Matching tree) was used to both find the best starting point  $t_{root}$  to initialise the labels, and optimally reorder the way these labels are propagated to other time-steps (Fig. 3D). This algorithm is based on the construction of a distance matrix that quantifies the dissimilarity between all possible pairs of time steps (Fig. 3C). The tracking itself is based on an iterative matching of the central coordinates of the ellipses between two time steps (Baseline), and includes a pre-processing step to

better manage the global movements of the bunch (Registration, Fig. 3B).

- a) Baseline: matchings of berry centre coordinates between two time steps

For any time-point  $t$ , the segmentation provides a point set  $S_t = \{c_k\}$  containing the ellipse centre-point coordinates  $c_k = (\hat{x}_e, \hat{y}_e)_k$  of each berry detected. Assuming all berries remain in the image frame with

approximately constant relative positions, the tracking of berries from time steps  $i$  to  $j$  is treated as a bipartite matching of point sets  $S_i$  and  $S_j$ . The correspondence between two point sets is done by associating to each centre  $c_i$  in  $S_i$  its nearest neighbour  $c_j$  in  $S_j$  in euclidean distance.

Each point can only be paired once, the closest pairs are matched first, and pairs with a distance above a threshold  $\delta = 16px$  are discarded. The value of  $\delta$  was chosen as the quarter of the median value of  $w_e$  in our annotation dataset, with the idea that such a low value strongly limits mismatches, even in dense areas of the bunch. This algorithm can be applied successively to pairs of sets  $(S_t, S_{t+1})$  along a time-series of  $N$  images, to propagate the correspondence of the initial set of labels.

- b) Registration: estimating global bunch deformations prior to matching

Even if the berries in a bunch maintain the same relative arrangement, their absolute positions may change between two time steps  $i$  and  $j$  due to relative movements of the bunch and the camera, or due to internal deformations and movements of the bunch. Assuming that the resulting deformation of the point cloud in the image coordinate system is affine, the Coherent-Point Drift algorithm [32] was used to realign the two sets prior to matching, by finding the affine transformation  $\tilde{S}_j$  of  $S_j$  that minimises the distance to  $S_i$  (Fig. 3B).

- c) Matching tree: processing the time steps in an optimal order

The matching algorithm can be applied to any pair of sets  $(S_i, S_j)$  from the time-series. Propagating the matching to successive pairs  $(S_t, S_{t+1})$  in chronological order is a common choice in multiple object tracking [29]. However, this option may not always be optimal, as in our case where the camera may for example move unexpectedly at a time step and then return to its original position (see video in Additional file 1 for examples). Here we propose to match the most similar pairs of sets in priority, to avoid errors that could occur and propagate from a pair of sets that are too dissimilar. To do so, we define a metric  $d$  quantifying the distance from  $S_i$  to  $S_j$ , based on the euclidean distance function  $e$ :

$$d(S_i, S_j) = \text{median}_{c \in S_i} (\min_{c_k \in S_j} e(c, c_k)) \tag{4}$$

Since  $d$  is not commutative (e.g. one point in a set could be the closest neighbour of many points in the other set), we define the distance  $D$  between two sets  $S_i$  and

$S_j$  as the average of  $d(S_i, S_j)$  and  $d(S_j, S_i)$ . Unlike during the matching (*Baseline*), the computation of  $D$  does not involve a bipartite pairing of points, which is time consuming because of the iterative nature of the algorithm. It therefore allows for a fast quantification of the dissimilarity  $m_{ij} = D(S_i, \tilde{S}_j)$  between all pairs of time-steps, which are stored in a distance matrix  $M = (m_{ij})$  (Fig. 3C). This matrix is used to arrange the order of the successive matchings through a layered tree (Fig. 3D). Unlike a linear ordering of the successive time-steps, arranging them through a tree structure reduces the average number of intermediate steps between  $t_{root}$  and other time-steps, thus limiting the risk of propagation of matching errors.

The matching tree contains a root node  $S_{root}$ , and each node has a depth  $k$  equal to the length of its path to the root. To select the nodes of depth  $k$ , we iteratively connect the candidate set  $S_i$  to another set  $S_j$  of depth less than  $k$ , such that  $d_{min} = \min(M_{ij}, M_{ji})$  is minimised. This is repeated as long as  $d_{min} < \theta$ , with  $\theta = 8px$  a threshold controlling the ratio between the width and depth of the tree. If no candidate set meets this criterion for depth  $k$ , a single *long-distance* edge is built between layers  $k - 1$  and  $k$  with the minimum possible distance. This process is iterated for successive depths until the tree contains all sets of  $S$ .  $t_{root}$  is selected exhaustively as the value allowing to place the most nodes in the tree before reaching a *long-distance* edge, and secondarily by maximising the number of points in  $S_{root}$ .

### Evaluation of the method

To evaluate the berry detection on a given image, the predicted ellipses whose Intersection over Union (IoU) are greater than 0.5 with a labelled ellipse are classified as True Positives (TP), indicating correct identifications. Predicted ellipses falling below this IoU threshold are False Positives (FP), representing incorrect identifications, while labelled ellipses without a corresponding predicted ellipse above this threshold are False Negatives (FN), signifying missed detections.

Precision, Recall and F1-score metrics are then deduced as follow:

$$\text{Precision} = TP / (TP + FP) \tag{5}$$

$$\text{Recall} = TP / (TP + FN) \tag{6}$$

$$\text{F1-score} = (2 \times \text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}) \tag{7}$$

For segmentation evaluation, the area of the segmented ellipses was compared with ground-truth observations



using the following metrics: bias, root mean-square error (RMSE), mean absolute percentage error (MAPE) and coefficient of determination ( $R^2$ ).

Berry tracking was evaluated by two metrics, namely the coverage  $T_c$  and the precision  $T_p$ .  $T_c$  is defined as the percentage, over the full time-series, of segmented berries that could be matched by the tracking algorithm to a berry segmented at  $t_{root}$  (coloured ellipses in Fig. 3E).  $T_p$  is the percentage of labels that point to the same berry over time, and was estimated by manually checking random samples of 10 time-steps per bunch in each time-series.

One bunch of the 2020 experiment was further analysed to assess the potential of the method at capturing and quantifying berry development and its asynchrony (demonstration dataset). 81 berries were measured by applying the full image analysis pipeline on 3 time-series of 138 images from 3 different camera views (120° difference) of the same grapevine bunch. The use of image time-series with different views was facilitated by the PhenoArch platform’s capacity to rotate a plant’s pot while images are being taken. (see [7] for details). We further selected berries tracked over at least 90% of the experiment duration. For each berry, an 8-days moving median was used to smooth the raw volume measurements over time (Fig. 7A and Additional file 2A; red curves), and a MAPE value was computed between the raw and smoothed volume data. The 10% berries with the highest MAPE were excluded from the analysis to reduce the noise, resulting in a final dataset of 73 berries. For each variable  $X$  observed (either  $V$  or  $H$ ), a relative value  $X_r$  and a scaled value  $X_s$  were computed as:

$$X_r = (X - X_0)/X_0 \tag{8}$$

$$X_s = (X - X_0)/(X_{max} - X_0) \tag{9}$$

where  $X_0$  is the median of  $X$  over the first 8 days.  $X_{max}$  is the maximum of the smoothed  $X$  values over the last 8 days for  $V$ , and the median of  $X$  over the last 8 days for  $H$ .

With these statistics, we compute the following descriptors of the ripening dynamics:

*Ripening duration (RD)* was estimated from the time interval ( $\Delta t$ ) between  $V_s=0.15$  and  $V_s=0.85$ :

$$RD = \frac{\Delta t}{0.85 - 0.15} \tag{10}$$

*Ripening relative speed (RS)* was defined as the variation of the relative volume ( $\Delta V_r$ ) during  $\Delta t$ :

$$RS = \frac{\Delta V_r}{\Delta t} \tag{11}$$

Finally, growth resumption time and coloration start time were defined as the time when  $V_s= 0.15$  and  $H_s = 0.15$  respectively. All these statistics were also computed for the “mean berry”, using daily average of individual berries volume or colour as input.

All the method conception and the data analysis were performed in Python.

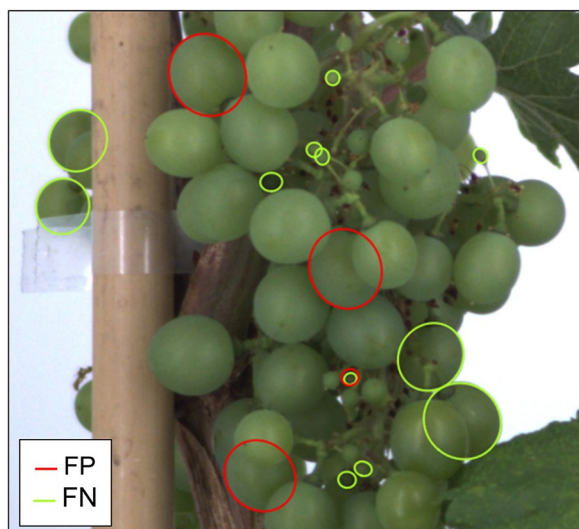
## Results

### Deep-learning segmentation allows accurate and robust shape inference of partly hidden berries

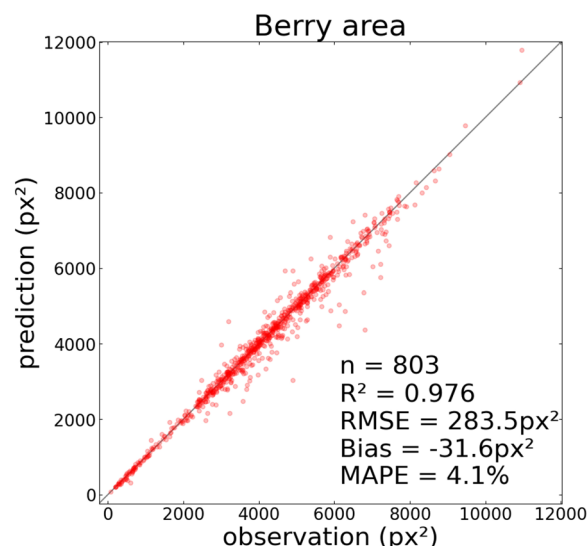
Berry segmentation was performed on the 2020 and 2021 datasets (21,744 images), resulting in an average detection of 64 berries per image. Figure 4 provides some examples of detection on the test subset, showing that



**Fig. 4** Examples of segmented grapevine bunch images. Output of the berry detection and segmentation pipeline on bunch images from 12 grapevine genotypes. Images come from the test subset, and none of these genotypes were used to train the model. Only a 500 × 500 px subpart of each image is shown



**Fig. 5** Example of mismatches between the berry detections and annotations. False positives (FP, red) and false negatives (FN, green) found when comparing berries detected by the pipeline to manually annotated berries, on a grapevine bunch image from the test subset. Only a subpart of the full image is shown



**Fig. 6** Accuracy of berry area measurement. Comparison of the area of manually annotated berries (observation) with those from the detection and segmentation pipeline (prediction). *n* number of points, *RMSE* root-mean square error, *MAPE* mean absolute percentage error

the model was able to infer the full contour of overlapping berries from different genotypes varying in size, shape and aspect, even when these contours were not fully visible. Predictions on the full test subset were compared with ground-truth annotations to quantify both detection and segmentation accuracies.

The detection of measurable berries had a Precision of 92.3% and a Recall of 87.5% on the test subset, resulting in a F1-score of 89.8%. The remaining errors (64 FPs, 109 FNs) were further investigated (e.g. Fig. 5) through a manual classification (Additional file 3A). This revealed that most errors (59% of FPs and 52% of FNs) correspond either to berries with a visible contour fraction within a  $50 \pm 10\%$  range, or to small underdeveloped berries (around pea size stage). Both situations are close to the selection criteria used when annotating berries, and the assessment of whether or not these criteria have been crossed may be ambiguous for both the annotator and the model. For FPs (i.e. detected but not annotated berries), errors were evenly distributed across berry sizes (Additional File 3B). 56% of them correspond to berries within the  $50 \pm 10\%$  visible contours range, sometimes due to an error by the annotator detected a posteriori. Considering that berries within this 10% error range are still good candidates to shape inference, the precision of the method at detecting measurable (even if not annotated) berries can thus be re-estimated to 96.0% (F1-score=91.8%). Concerning FNs (i.e. missed detections), pea sized berries alone account for 27% of the cases, which result in a

slight under-representation of this class in the histogram of the size of berries detected. (Additional File 3B).

The area of the ellipses segmented by the model closely matched those of the manual annotations on the test subset (Fig. 6; MAPE=4.1%,  $R^2=0.976$ ), with a low bias of  $-32 \text{ px}^2$ . This demonstrates that the segmentation model was able to accurately infer the size of berries with up to 50% of their contours hidden. A similar MAPE around 4% was obtained on genotypes either present ( $n=440$ ) or absent ( $n=363$ ) from the training subset, suggesting that the segmentation generalised well to the genetic diversity in our dataset.

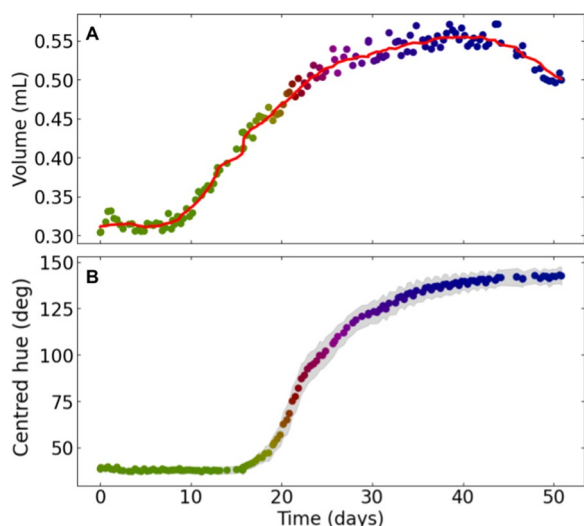
**An almost error-free tracking of 50 to 80% of the segmented berries**

Berry tracking was performed on 9 grapevine bunches from different plants in each of the 2020 and 2021 experiments, observed over an average of 65 and 136 time-steps respectively, with the same median interval of 8 h between images. The video from Additional file 1 shows examples of tracking outputs for three different bunches. It was observed that image time-series exhibit periods during which the relative position of the camera and the plant remain stable, resulting in a fixed positioning of the bunch in the image, but also include irregular movement of the plant and of the camera despite the robotisation of the image acquisition, combined with irregular movements of both the bunch and the leaves.

**Table 2** Coverage index ( $T_c$ ) of the tracking method (M4) for two experiments

Tracking method	Mean coverage ( $T_c$ )	
	2020 (n=9)	2021 (n=9)
M1: Baseline	5.1%	40.8%
M2: M1 + Registration	32.3%	62.8%
M3: M1 + Matching tree	35.6%	63.2%
M4: M1 + Matching tree + Registration	53.4%	74.2%
S1: M4 on an increased time-step (8 to 80 h)	54.1%	70.4%
S2: M4 (and M1) restricted to k stable periods (k: mean value of k)	77.1% (24.1%) (k = 3.4)	82.0% (52.2%) (k = 1.8)

Lines correspond to different combinations of the tracking algorithm elements (M1 to M3) and two sub-sampling strategies of the data (S1, S2)



**Fig. 7** Growth and coloration kinetics of an individual grapevine berry. Volume (A) and Centred hue (B) measured over time on an individual berry of the demonstration dataset. All points are coloured using the corresponding average hue values. In A, the red curve corresponds to an 8-days moving median smoothing. In B, the grey area corresponds to the standard deviation of the centred hue value observed within the berry segmentation mask

The coverage  $T_c$  of the berry tracking method was assessed for each of the 18 bunches (M4, Table 2). The individual effects of the method components (*Registration*, *Matching tree*) were evaluated by re-running the tracking without them (M1 to M3, Table 2). Two sub-sampling scenarios were also used to assess the effect of increasing the time step from 8 to 80 h (S1, Table 2) and restricting the time-series to periods of stable shooting conditions (S2, Table 2). These periods were manually identified by careful examination of the stability of the image acquisition over time.

$T_c$  had an average value of 53.4% and 74.2% for 2020 and 2021 experiments respectively. The precision was very high in both 2020 ( $T_p=96.7%$ , 623 labels) and 2021 ( $T_p=99.2%$ , 793 labels) experiments. This indicates that the tracking method is more accurate than exhaustive, which is appropriate for studying berry growth kinetics, since an accurate monitoring of a representative sub-sample of berries is sufficient to reflect the whole bunch dynamics. Such a high precision might be ensured by the low value chosen for the distance threshold which determines whether two segmented berries can be matched. Using point-set registration (Fig. 3B) and a matching tree (Fig. 3D) during tracking both contribute to maintain sufficiently high coverage, as it increases  $T_c$  by a factor of 10.4 and 1.8 for 2020 and 2021 experiments respectively, compared to a regular succession of point-sets matchings (Table 2; M2-4 vs M1).

However, a significant amount of segmented berries (2020: 46.6%, 2021: 24.8%) remained unmatched to  $t_{root}$ . The time interval between images was not likely to explain these losses, since a tenfold decrease in the image frequency did not significantly modify  $T_c$  for the same duration (Table 2; S1). Instead, further examination of the distance matrices computed during tracking highlighted periods with a strong temporal consistency (i.e. low distance between point sets of ellipse centres), separated by abrupt transitions which were often associated with a drop in  $T_c$  (Additional file 4). 30 transitions were empirically annotated using these matrices, to identify their cause on their corresponding images (Additional file 4A; red lines). Most transitions coincided with a bunch rotation (70%), a strong shift in camera position causing berry apparitions or disappearances (13%), or a deformation within the bunch (10%). These situations correspond to the actual limitations of our registration method, but most of them could have been avoided by a better management of the experimental conditions. Performing tracking independently in each time consistent period increased  $T_c$  to 77.1% and 82.0% for 2020 and 2021 experiments respectively (Table 2; S2). These metrics probably reflect the performance of our method under experimental conditions where the instability of the image acquisition is better managed.

**Robust measurement of single berry dynamics, differing from the usual “mean berry” approach**

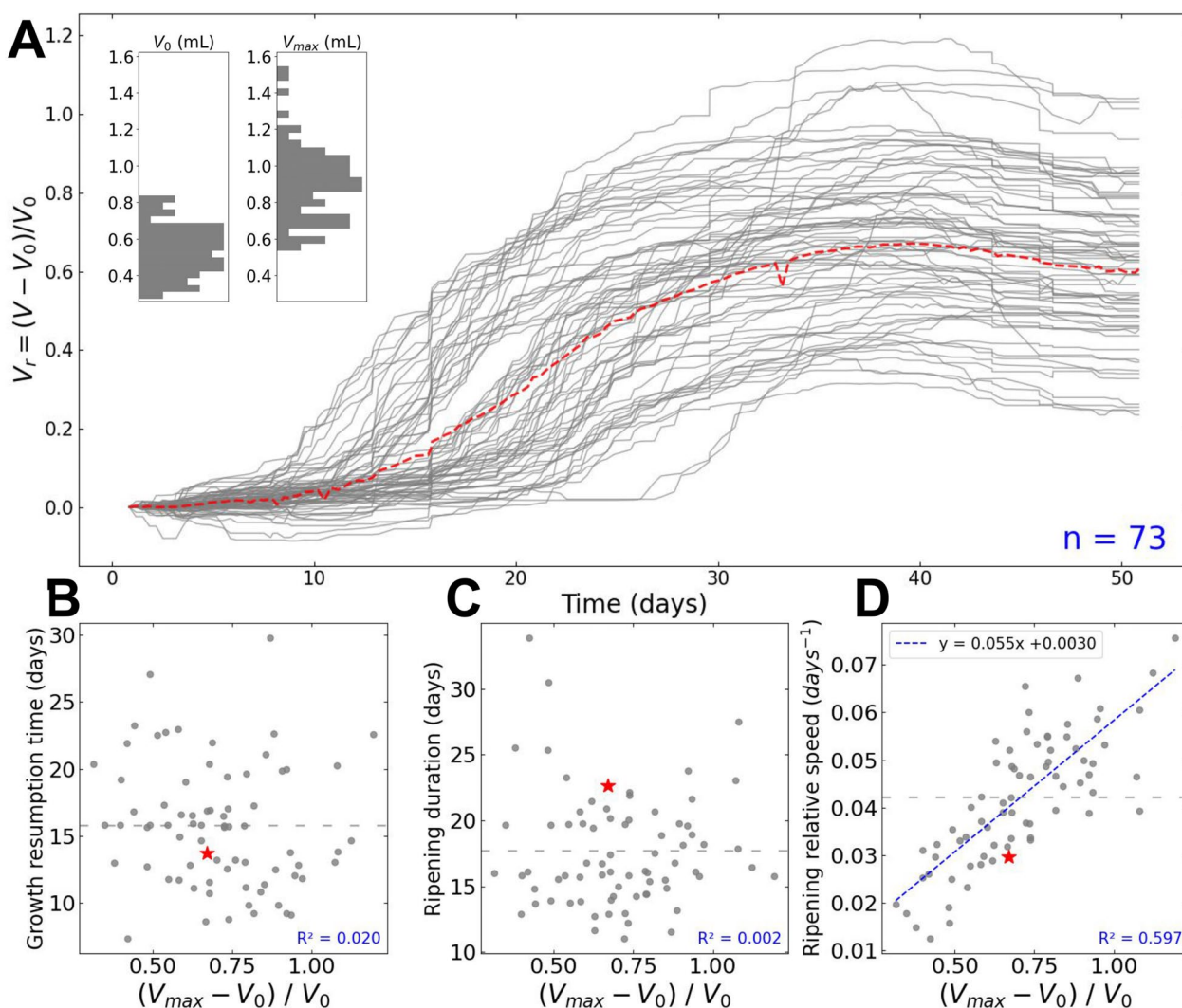
Combining the tracking labels with the features extracted from the segmented berries allowed to monitor the growth of a single berry over time with high accuracy and temporal resolution, both in terms of volume and colour (Fig. 7). While volume measurements can be noisier due to variations of just a few pixels in the image, colour measurements are more reliable because they are derived



from averaging a larger number of pixels. These kinetics exhibit smooth patterns over time, using high frequency measurements of a large number of berries in several bunches (Additional file 2), which supports the suitability of this method to high-throughput phenotyping conditions.

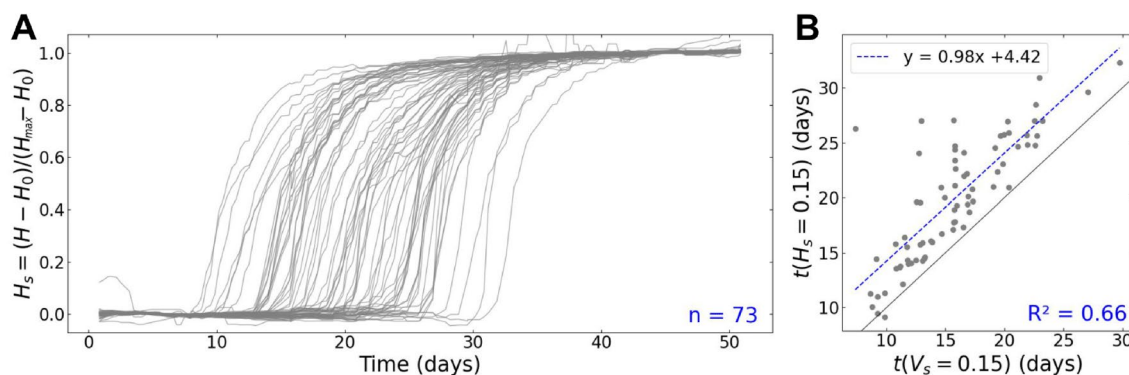
The potential of the method to reveal new characteristics of berry ripening and bunch population structure was further assessed on 73 individual berries tracked on a bunch from the 2020 experiment, on the *Vitis vinifera* cv Alexandrouli, a black hermaphrodite cultivar of Georgian origin used for wine making (<https://www.vivc.de/?r=passport%2Fview&id=>

263). The bunch was observed for a period of 50 days, corresponding to the second growth period of berries, which goes from the end of the green phase to over-ripening (Figs. 8, 9). To our knowledge, this is the first report on the growth and colouring dynamics of a statistically relevant number of individual fruits. We confirm here that, at similar developmental stages, the individual berry volume exhibits 2–threefold variations with considerable deviations from Gaussian distribution inside a single bunch (Fig. 8A). On average in this bunch, an individual berry increased its volume by 60% (Fig. 8A) in 18 days before reaching its maximal volume (Fig. 8C; grey dotted line), and underwent



**Fig. 8** Individual growth kinetics and ripening statistics of berries within a grapevine bunch. **A** Smoothed relative volume ( $V_r$ ) as a function of time for the  $n = 73$  berries of the demonstration dataset (grey lines), and for the daily averaged ‘mean berry’ (red dotted line). Inset: histogram of initial ( $V_0$ ) and maximum ( $V_{max}$ ) volumes of individual berries. **B**, **C** and **D** respectively show the growth resumption time, ripening duration and ripening relative speed, as a function of the maximum relative volume  $(V_{max} - V_0) / V_0$ . The grey dotted horizontal line represents the mean value for the considered statistics. The red stars indicate the values for the daily averaged “mean berry”. In **D**, the blue dotted line corresponds to the linear regression between x and y axis





**Fig. 9** Individual coloration kinetics of berries within a grapevine bunch. Centred hue ( $H$ ) kinetics were automatically computed for the  $n=73$  berries of the demonstration dataset. **A** Scaled coloration kinetics ( $H_s$ ) of each measured berry (grey lines), computed using their initial ( $H_0$ ) and final ( $H_{max}$ ) centred hue. **B** Relation between growth resumption time  $t(V_s=0.15)$  and coloration start time  $t(H_s=0.15)$ . The grey line is the  $x=y$  diagonal, and the blue dotted line shows the linear regression between growth resumption time and coloration start time

a more or less intense shrivelling period once phloem unloading in berries definitively stopped [41]. Such a relative expansion rate is in line with the approximate doubling of berry volume during the ripening of most *V. vinifera* cultivars [4, 5, 22], which took three weeks to complete on individual fruits of Meunier, Syrah, Zinfandel or ML1 [42], Cabernet Sauvignon and Pinot [15]. Further studies are needed to establish if the slightly shorter growth duration and expansion of Alexandrouli’s berry is truly of genetic origin, or is the result from tests on fairly young own rooted potted plants in greenhouse conditions. In any case, we confirm here that the duration of ripening of an individual berry, when measured directly, is at least 30–50% shorter than the consensual duration of ripening reported in textbooks [12, 26, 45]. To us, such a discrepancy occurs because ripening duration is routinely inferred upon calculating the average weight and composition on hundreds of asynchronous berries representing fruit diversity at the plot scale, before checking its time evolution. Indeed, present data even shows that for a single bunch, which undoubtedly underestimates asynchrony at plot scale, the global growth curve recalculated for all detected berries noticeably overestimates the average duration of the second growth period (Fig. 8C; red star) and underestimates the maximum growth rate (Fig. 8D; red star). These statistical biases clearly result from adding the asynchrony to the real, but previously unknown, duration of the second growth period in average representative samples. Moreover, the fact that asynchrony and growth duration last approximately as long in a single bunch (Similar ranges for the y-axis of Fig. 8B, C) means that conventional

random samples combine berries of very different phenological stages, which is a major drawback for tackling fruit development biology. Real time monitoring of berry growth allows to constitute synchronised berry samples, and more conveniently than marking each berry fecundation or softening dates. In this respect, coloration has been proposed as a proxy for the induction of ripening [48]. Our data suggested that growth resumption is a more pertinent indicator of the onset of ripening, as it precedes coloration by more than four days on average, and the delay can vary from one day up to two weeks (Fig. 9). Thus, this variability clearly limits the use of coloration alone in building an effective sampling strategy. Finally, present original data allow us to test hypotheses and give first insights on the drivers of the dynamic structure of berry cohorts within a bunch. First, our data do not suggest that the acceleration of the ripening program in the late berries [19] is accompanied by an acceleration of the berry growth rate, as no correlation was found between berry relative expansion and the growth resumption time (Fig. 8B;  $R^2=0.02$ ). Our observations does not support either the idea that ripening berries compete for water or photoassimilates, as their relative growth rate does not vary consistently with the number of berries growing simultaneously. Nevertheless, individual berries largely differed in their maximum relative expansion, which was clearly related to their maximal growth rate (Fig. 8D;  $R^2=0.59$ ), not by growth duration (Fig. 8C;  $R^2=0.002$ ). This first approach on the dynamic structure of berry population based on the discretisation of single berry dynamics clearly constitutes a paradigm shift from modelling the future crop as an average ideal fruit [54].

## Discussion

### A method to quantify the variability of ripening kinetics within an asynchronous cohort of growing berries

The automated tracking of the asynchronous ripening of individual berries within one grapevine bunch allowed us to revisit the basic growth rates of ripening berries, and to propose, for the first time, an analysis of their growth and colour kinetics for a statistically significant number of observations, with an unprecedented time resolution (less than one day). In line with preliminary reports on other cultivars [5, 42], we found that the ripening duration of individual berry does not exceed three weeks, which differs from the 32 to 56 days duration of sugar loading reported in a panel of 36 international cultivars, following random sampling and averaging 50 berries over time [45]. This confirms that by neglecting berry asynchrony as a confounding variable, the ripening duration of daily averaged ‘mean berry’ could be up to two fold overestimated compared to individual fruit kinetics. Such a gap could have strong impacts on deconvolving the effects of annual variations in light, temperature and rainfalls on growth and sugar loading intensities. It should be noted that the intra-bunch variability documented here far exceeds the phenological and compositional drifts observed over the last half-century, as consequences of climate change [3]. It is therefore likely that minor phenological changes affecting the population age pyramid may have been previously misinterpreted as kinetic or metabolic changes intrinsic to the ripening process. Hopefully, the method presented here will help in future investigations to better document which part of the G×E interaction is mediated by the temporal structure of the population and the fundamentally different part linked to metabolic variation during berry ripening.

### A generic method to infer the shape of partially hidden fruits on an image

While numerous computer vision approaches have been developed to identify and measure fruits using deep-learning [16–18, 23, 34, 38, 43], they mostly aim at inferring their occlusion boundaries (i.e. visible edges), which differs from the true contours of the object of interest in the case of overlapping fruits, thus preventing to access their actual size. Instead, our segmentation method was designed to infer the non-visible part of overlapping fruit shapes, directly in the deep-learning process. The original and fast annotation strategy we introduced allowed to implicitly constrain the model during training to produce elliptical masks,

without the need to pass by a long annotation of image edges, or to make this constrain explicit in the model architecture, as in Ellipse R-CNN [10]. The counterpart of this strategy is to restrict the inference to sufficiently visible berries, which we managed to do by training a model at detecting only berries with more than 50% visible contours. Using such a binary criterion can lead to ambiguities during both annotation and prediction, but our results suggest that it does not degrade the biological outputs. Still, including this filter in the detection step does not allow to detect all visible berries, which would be a limit for counting. An alternative would be to first use a more exhaustive fruit detector, and delegate the task of filtering measurable berries to a classifier. This would allow for classical counting strategies (e.g. [51]) to be combined with our physiological measurements in a single pipeline. Although this study focuses on grapevines, we think that our method could be applied to any other fruit that can be approximated to have an ellipsoidal shape.

### The performance of berry tracking relies on the stability of the image acquisition setup

This work was carried out on images from a high-throughput phenotyping platform, where controlled conditions and standardised image acquisition facilitated the temporal tracking of individual berries. While we adapted the tracking algorithm to better tolerate slight movements of the bunch and camera, our results still showed that the tracking performance can be improved by almost 50% by stabilising the image acquisition. In particular, tracking performance may greatly improve by avoiding non-linear relative movements (e.g. rotations) of the bunch and the camera, preventing deformations inside the bunch, and keeping the entire bunch in the camera’s field of view. While more advanced image analysis methods may address these issues in the future, we argue that it is more effective to avoid such situations during image acquisition. Still, our method allows to visualize the aforementioned discontinuities in the image time-series via the computation of a distance matrix (Fig. 3C). This could allow to improve the tracking performance in a semi-automatic way, by re-running the automatic tracking for all periods delineated by discontinuities, and then manually mapping berry labels at each discontinuity. Lastly, choosing the right image acquisition timings is essential for subsequent analysis. Indeed, quantifying rapid dynamics such as berry colour changes needs a sufficiently high frequency of image capture, and the standardisation of ripening dynamics requires including both their initial and final plateaus during the observation period.

### Can the method be adapted outside the controlled conditions of an indoor phenotyping platform?

While our method was only evaluated in controlled conditions within the PhenoArch platform, it could probably be adapted to other environments such as field conditions, with a few adaptations.

For the berry detection and segmentation, the performance of the models used might decrease for images where the appearance of the berries differs from those in the training dataset. For instance, preliminary tests on a field image (Additional file 5) showed that berries illuminated by the sun were rarely detected, while performance was better for berries in the shade, whose appearance resembled the berries observed on a platform. This drop in performance might be corrected by re-training the models, which would be facilitated by our open-source implementation that allows re-use, and by our fast and robust annotation strategy (about 100–150 berries per hour), compared to traditional approaches that rely on annotating visible edges.

For berry tracking, we showed in the previous section that the stabilisation of the image acquisition setup is essential for good performance. This might be harder to achieve outside the controlled conditions of the greenhouse, for example in windy conditions, and might require adaptations specific to the experimental setup used.

Finally, sufficient berries need to be detected in the images to correctly quantify their ripening heterogeneity. This could be a limit in experimental conditions closer to real conditions, for example with more compact bunches and leaves hiding the berries. Especially as it is more complicated to rotate the plant as in our platform to capture more berries. Image acquisition should therefore be carefully considered so as to prioritise the largest bunches, remove any leaves obscuring the bunches, or capture several bunches with the same image.

### Conclusion

We introduce a fully-automatic open-source method to detect, segment and track overlapping berries in a time-series of grapevine bunch images in laboratory conditions. This non-destructive method gives direct access to the growth and colour kinetics of individual berries within a bunch. Coupled with high frequency image capture, this makes it possible to quantify undocumented aspects of individual fruit development, and to characterise their asynchrony at the population level. Using this method in real time during future experiments could allow the design of new sampling strategies that will consider the bunch as a population of unsynchronized berries, rather than an ideal, average

berry, and lead to a complete revisitation of the ripening dynamics. In particular, the GxE effects could be more clearly attributed not only to physiological changes in the ripening process, but also to changes in the age structure of the whole population of berries. The complete automation of our method is also fully compatible with high-throughput phenotyping, providing the opportunity to study these detailed GxE interactions on physiology and asynchrony of berry ripening for large plant panels.

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13007-023-01125-8>.

**Additional file 1:** Video (.mp4) of berry segmentation, detection and tracking outputs for 3 grapevine bunches. For each plant, the video displays a time-series of 62 to 66 labelled segmented RGB images, obtained after running the full berry segmentation, detection and tracking pipeline. Raw images were captured with a median interval of 8h. Each colour corresponds to one tracking label. Segmented berries without labels are drawn as white empty ellipses. *t* indicates the order of each image in the time-series

**Additional file 2:** Growth and coloration kinetics of several individual grapevine berries. Repetition of the results shown in Fig. 7 for more berries. Each subplot displays the Volume (mL) (**A**) or Centred hue (deg) (**B**) measured over time (days) on an individual berry, after running the full image analysis pipeline on a time-series of 138 images, from 3 different camera views (120° difference) of the same grapevine bunch. All points are coloured using the corresponding average hue values. In **A**, the red curve corresponds to a 8-days moving median smoothing. In **B**, the grey area corresponds to the standard deviation of the centred hue value observed within the berry segmentation mask.

**Additional file 3:** Analysis of berry detection errors in the test subset. Analysis of the False Positive (FP) and False Negative (FN) errors found when comparing berries detected by the pipeline to manually annotated berries, on the grapevine bunch images from the test subset. **A** Manual classification of detection errors as pea-sized berries, non-small (i.e. not pea-sized) berries, and non-berry objects. Non-small berries are further classified according to their percentage of visible contours (ct). **B** Distribution of detected berry sizes after segmentation, for all berries (top subplot), FP (middle subplot) and FN (bottom subplot). *n*: number of detected berries.

**Additional file 4:** Analysis of abrupt transitions in time-series of grapevine bunch images. **A** Heat map of the distance matrices obtained after tracking berries in time-series of 138 grapevine bunch images, for 6 different plants. Vertical red lines correspond to the empiric annotation of time-steps exhibiting abrupt transitions in these matrices. **B** Tracking coverage ( $T_c$ ) over time obtained for these time-series. The dashed blue vertical line represents the time step  $t_{root}$  used to initialise the tracking.

**Additional file 5:** Detection and segmentation of berries in field conditions. Output of the berry detection and segmentation pipeline on an image of grapevine bunches taken in the field. This is a preliminary result.

### Acknowledgements

We are grateful to all members at the M3P platforms for providing technical support, conducting the experiments and collecting data.

### Author contributions

LCB supervised the experiment and acquired the data. BD and MC tested and compared different image analysis pipelines. MC provided the first proof of concept on grapevine in natural conditions. BD and CF designed the final pipeline and analysed the resulting data. LCB and CR provided advice on the

conception of the pipeline. BD and CR wrote the manuscript and CF, LCB, CR, TS reviewed and edited it. All the authors have approved the manuscript and have made all required statements and declarations.

#### Funding

This work was supported by the Agence Nationale de la Recherche (G2WAS project, ANR-19-CE20-002) and by the EU project STARGATE H2020 952339 (<https://stargate-hub.eu/>).

#### Availability of data and materials

The source code for both training and prediction, notebook examples and trained model are available on Github ([github.com/openalea/deepberry](https://github.com/openalea/deepberry)) under an Open Source licence (Cecill-C).

#### Declarations

##### Ethical approval and consent to participate

Not applicable.

##### Consent for publication

Not applicable.

##### Competing interests

The authors declare that they have no competing interests.

#### Author details

<sup>1</sup>LÉPSE, Univ Montpellier, INRAE, Institut Agro, Montpellier, France. <sup>2</sup>AGAP Institut, Univ Montpellier, CIRAD, INRAE, Institut Agro, Montpellier, France.

Received: 17 July 2023 Accepted: 8 December 2023

Published online: 14 December 2023

#### References

- Bargoti S, Underwood J. Deep fruit detection in orchards. In: 2017 IEEE International Conference on Robotics and Automation (ICRA). 2017. p. 3626–33.
- Barth R, IJsselmuiden J, Hemming J, Henten EJV. Data synthesis methods for semantic segmentation in agriculture: a *Capsicum annum* dataset. *Comput Electron Agric*. 2018;144:284–96.
- Bécart V, Lacroix R, Puech C, de Cortázar-Atauri IG. Assessment of changes in Grenache grapevine maturity in a Mediterranean context over the last half-century. *OENO One*. 2022;56(1):53–72.
- Bigard A, Romieu C, Sire Y, Veyret M, Ojeda H, Torregrosa L. The kinetics of grape ripening revisited through berry density sorting. *Oeno One*. 2019;53(4):1–16.
- Bigard A, Romieu C, Ojeda H, Torregrosa L. The sugarless grape trait characterized by single berry phenotyping. *bioRxiv*. 2022. <https://doi.org/10.1101/2022.03.29.486323v1>.
- Bochkovskiy A, Wang CY, Liao HY. Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv: 2004.10934*. 2020.
- Briche N, Fournier C, Turc O, Strauss O, Artzet S, Pradal C, Welcker C, Tardieu F, Cabrera-Bosquet L. A robot-assisted imaging pipeline for tracking the growths of maize ear and silks in a high-throughput phenotyping platform. *Plant Methods*. 2017;13(1):1–2.
- Cabrera-Bosquet L, Fournier C, Briche N, Welcker C, Suard B, Tardieu F. High-throughput estimation of incident light, light interception and radiation-use efficiency of thousands of plants in a phenotyping platform. *New Phytol*. 2016;212(1):269–81.
- Dolata P, Wróblewski P, Mrzygłód M, Reiner J. Instance segmentation of root crops and simulation-based learning to estimate their physical dimensions for on-line machine vision yield monitoring. *Comput Electron Agric*. 2021;190:106451.
- Dong W, Roy P, Peng C, Isler V. Ellipse R-CNN: learning to infer elliptical object from clustering and occlusion. *IEEE Trans Image Process*. 2021;30:2193–206.
- Dubois C, Irisson JO, Debreuve E. Correcting estimations of copepod volume from two-dimensional images. *Limnol Oceanogr Methods*. 2022;20(6):361–71.
- Fasoli M, Richter CL, Zenoni S, Bertini E, Vitolo N, Dal Santo S, et al. Timing and order of the molecular events marking the onset of berry ripening in grapevine. *Plant Physiol*. 2018;178(3):1187–206.
- Fernández-Navales J, Garde-Cerdán T, Tardáguila J, Gutiérrez-Gamboa G, Pérez-Álvarez EP, Diago MP. Assessment of amino acids and total soluble solids in intact grape berries using contactless Vis and NIR spectroscopy during ripening. *Talanta*. 2019;199:244–53.
- Fitzgibbon A, Pilu M, Fisher RB. Direct least square fitting of ellipses. *IEEE Trans Pattern Anal Mach Intell*. 1999;21(5):476–80.
- Friend AP, Trought MCT, Creasy GL. The influence of seed weight on the development and growth of berries and live green ovaries in *Vitis vinifera* L. cvs. Pinot Noir and Cabernet Sauvignon. *Aust J Grape Wine Res*. 2009;15(2):166–74.
- Ganesh P, Volle K, Burks TF, Mehta SS. Deep orange: mask R-CNN based orange detection and segmentation. *IFAC-PapersOnLine*. 2019;52(30):70–5.
- Gené-Mola J, Sanz-Cortiella R, Rosell-Polo JR, Morros JR, Ruiz-Hidalgo J, Vilaplana V, et al. Fruit detection and 3D location using instance segmentation neural networks and structure-from-motion photogrammetry. *Comput Electron Agric*. 2020;169:105165.
- Gonzalez S, Arellano C, Tapia JE. Deepblueberry: quantification of blueberries in the wild using instance segmentation. *IEEE Access*. 2019;7:105776–88.
- Gouthu S, O'Neil ST, Di Y, Ansarolia M, Megraw M, Deluc LG. A comparative study of ripening among berries of the grape cluster reveals an altered transcriptional programme and enhanced ripening rate in delayed berries. *J Exp Bot*. 2014;65(20):5889–902.
- He K, Gkioxari G, Dollár P, Girshick R. Mask r-CNN. In: *Proceedings of the IEEE international conference on computer vision*, 2017. p. 2961–69.
- Hondo T, Kobayashi K, Aoyagi Y. Real-time prediction of growth characteristics for individual fruits using deep learning. *Sensors*. 2022;22(17):6473.
- Houel C, Martin-Magniette ML, Nicolas SD, Lacombe T, Le Cunff L, Franck D, et al. Genetic variability of berry size in the grapevine (*Vitis vinifera* L.). *Aust J Grape Wine Res*. 2013;19(2):208–20.
- Jia W, Tian Y, Luo R, Zhang Z, Lian J, Zheng Y. Detection and segmentation of overlapped fruits based on optimized mask R-CNN application in apple harvesting robot. *Comput Electron Agric*. 2020;172:105380.
- Kalopesa E, Karyotis K, Tziolas N, Tsakiridis N, Samarinas N, Zalidis G. Estimation of sugar content in wine grapes via in situ VNIR–SWIR point spectroscopy using explainable artificial intelligence techniques. *Sensors*. 2023;23(3):1065.
- Keller M, Zhang Y, Shrestha PM, Biondi M, Bondada BR. Sugar demand of ripening grape berries leads to recycling of surplus phloem water via the xylem. *Plant, Cell Environ*. 2015;38(6):1048–59.
- Krasnow MN, Shackel KA, Matthews MA. Modelling water and sugar flux to developing berries suggests early cessation of sugar accumulation and substantial xylem backflow *Actes du XVIIIth International GIESCO meeting*, 8–11 July 2013, Oporto, Portugal.
- Liu X, Chen SW, Aditya S, Sivakumar N, Dcunha S, Qu C, et al. Robust fruit counting: combining deep learning, tracking, and structure from motion. In: *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 2018. p. 1045–52.
- Lou Y, Miao Y, Wang Z, Wang L, Li J, Zhang C, et al. Establishment of the soil water potential threshold to trigger irrigation of Kyoho grapevines based on berry expansion, photosynthetic rate and photosynthetic product allocation. *Aust J Grape Wine Res*. 2016;22(2):316–23.
- Luo W, Xing J, Milan A, Zhang X, Liu W, Kim TK. Multiple object tracking: a literature review. *Artif Intell*. 2021;1(293):103448.
- McCarthy MG. Weight loss from ripening berries of Shiraz grapevines (*Vitis vinifera* L. cv Shiraz). *Aust J Grape Wine Res*. 1999;5:10–6. <https://doi.org/10.1111/j.1755-0238.1999.tb00145.x>.
- Miao Y, Huang L, Zhang S. A two-step phenotypic parameter measurement strategy for overlapped grapes under different light conditions. *Sensors*. 2021;21(13):4532.
- Myronenko A, Song X. Point set registration: coherent point drift. *IEEE Trans Pattern Anal Mach Intell*. 2010;32(12):2262–75.
- Navrátil M, Buschmann C. Measurements of reflectance and fluorescence spectra for nondestructive characterizing ripeness of grapevine berries. *Photosynthetica*. 2016;54(1):101–9.



34. Ni X, Li C, Jiang H, Takeda F. Three-dimensional photogrammetry with deep learning instance segmentation to extract berry fruit harvestability traits. *ISPRS J Photogramm Remote Sens.* 2021;171:297–309.
35. Nicolas SD, Péros JP, Lacombe T, Launay A, Le Paslier MC, Bérard A, et al. Genetic diversity, linkage disequilibrium and power of a large grapevine (*Vitis vinifera* L) diversity panel newly designed for association studies. *BMC Plant Biol.* 2016;16(1):74.
36. Ojeda H, Deloire A, Carbonneau A, Ageorges A, Romieu C. Berry development of grapevines : relations between the growth of berries and their DNA content indicate cell multiplication and enlargement. *Vitis.* 1999;38(4):145.
37. Pastore C, Frioni T, Diago MP. Editorial: resilience of grapevine to climate change: from plant physiology to adaptation strategies. *Front Plant Sci.* 2022;9(13):994267.
38. Perez-Borrero I, Marin-Santos D, Vasallo-Vazquez MJ, Gegundez-Arias ME. A new deep-learning strawberry instance segmentation methodology based on a fully convolutional neural network. *Neural Comput Applic.* 2021;33(22):15059–71.
39. Rienth M, Torregrosa L, Sarah G, Ardisson M, Brillouet JM, Romieu C. Temperature desynchronizes sugar and organic acid metabolism in ripening grapevine fruits and remodels their transcriptome. *BMC Plant Biol.* 2016;16:164.
40. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF, editors. International conference on medical image computing and computer-assisted intervention. Berlin: Springer; 2015. p. 234–41.
41. Savoi S, Torregrosa L, Romieu C. Transcripts switched off at the stop of phloem unloading highlight the energy efficiency of sugar import in the ripening *V. vinifera* fruit. *Hortic Res.* 2021;8:193.
42. Shahood R, Torregrosa L, Savoi S, Romieu C. First quantitative assessment of growth, sugar accumulation and malate breakdown in a single ripening berry. *Oeno One.* 2020;54(4):1077–92.
43. Shen L, Chen S, Mi Z, Su J, Huang R, Song Y, et al. Identifying veraison process of colored wine grapes in field conditions combining deep learning and image analysis. *Comput Electron Agric.* 2022;200:107268.
44. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. In: International Conference on Learning Representations (ICLR). 2015.
45. Suter B, Destrac Irvine A, Gowdy M, Dai Z, van Leeuwen C. Adapting wine grape ripening to global change requires a multi-trait approach. *Front Plant Sci.* 2021. <https://doi.org/10.3389/fpls.2021.624867>.
46. Suzuki S, Be K. Topological structural analysis of digitized binary images by border following. *Comput Vis Graph Image Process.* 1985;30(1):32–46.
47. Tang Y, Chen M, Wang C, Luo L, Li J, Lian G, et al. Recognition and localization methods for vision-based fruit picking robots: a review. *Front Plant Sci.* 2020. <https://doi.org/10.3389/fpls.2020.00510>.
48. Vondras AM, Gouthu S, Schmidt JA, Petersen AR, Deluc LG. The contribution of flowering time and seed content to uneven ripening initiation among fruits within *Vitis vinifera* L. cv. Pinot noir clusters. *Planta.* 2016;243(5):191–202.
49. Wada K. labelme: image polygonal annotation with python. 2018. <https://github.com/wkentaro/labelme>. Accessed 12 Dec 2023.
50. Wang Z, Walsh K, Koirala A. Mango fruit load estimation using a video based MangoYOLO-Kalman Filter-Hungarian algorithm method. *Sensors.* 2019;19(12):2742.
51. Zabawa L, Kicherer A, Klingbeil L, Töpfer R, Kuhlmann H, Roscher R. Counting of grapevine berries in images via semantic segmentation using convolutional neural networks. *ISPRS J Photogramm Remote Sens.* 2020;164:73–83.
52. Zhang XY, Wang XL, Wang XF, Xia GH, Pan QH, Fan RC, et al. A shift of phloem unloading from symplasmic to apoplasmic pathway is involved in developmental onset of ripening in grape berry. *Plant Physiol.* 2006;142(1):220–32.
53. Zhang W, Wang J, Liu Y, Chen K, Li H, Duan Y, et al. Deep-learning-based in-field citrus fruit detection and tracking. *Hortic Res.* 2022;9:uhac003.
54. Zhu J, Génard M, Poni S, Gambetta GA, Vivin P, Vercambre G, et al. Modelling grape growth in relation to whole-plant carbon and water fluxes. *J Exp Bot.* 2019;70(9):2505–21.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

