

DeepPhenoTree – Apple Edition: a Multi-site apple phenology RGB annotated dataset with deep learning baseline models

Herearii Metuarea

Université d'Angers

Abdoul Oussen hamza

INRAE, Univ Angers, Institut Agro, IRHS, SFR QUASAV

Walter Guerra

Research Centre Laimburg

Francesca Zuffa

Research Centre Laimburg

Francesco Panzeri

Research Centre Laimburg

Andrea Patocchi

Agroscope

Lidia Lozano

IRTA, Fruit Production

Shauny Hoye

Better3fruit N.V

François Laurens

INRAE, Univ Angers, Institut Agro, IRHS, SFR QUASAV

Jeremy Labrosse

Hiphen

Pejman Rasti

Université d'Angers

David Rousseau

`david.rousseau@univ-angers.fr`

Université d'Angers

database

Keywords: Apple phenology, multi-site dataset, deep learning, RGB imaging, growth stages, dataset paper

Posted Date: February 27th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-8977752/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

DeepPhenoTree – Apple Edition:
a Multi-site apple phenology RGB annotated
dataset with deep learning baseline models

Herearii Metuarea^{1,2}, Abdoul djalil Ousseni hamza²,
Walter Guerra³, Francesca Zuffa³, Francesco Panzeri³,
Andrea Patocchi⁵, Lidia Lozano⁴, Shauny Van Hoya⁶,
François Laurens², Jeremy Labrosse⁷, Pejman Rasti^{1,2},
David Rousseau^{1,2*}

¹Université d'Angers, Angers, 49000, France.

²INRAE, Univ Angers, Institut Agro, IRHS, SFR QUASAV , F-49000,
Angers, France.

³Research Centre Laimburg, Auer, 39040, Italy.

⁴IRTA, Fruit Production, Fruitcentre, Lleida, Catalonia, 25003, Spain.

⁵Agroscope, Wädenswil, 8820, Switzerland.

⁶Better3fruit N.V., Rillaar, 3202, Belgium.

⁷Hiphen, Avignon, 84140, France.

*Corresponding author(s). E-mail(s): david.rousseau@univ-angers.fr;

Contributing authors: herearii.metuarea@univ-angers.fr;

abdoul-djalil.ousseini-hamza@inrae.fr; Walter.Guerra@laimburg.it;

Francesca.Zuffa@laimburg.it; Francesco.Panzeri@laimburg.it;

andrea.patocchi@agroscope.admin.ch; lidia.lozano@irta.cat;

shauny@better3fruit.com; francois.laurens@inrae.fr;

jlabrosse@hiphen-plant.com; pejman.rasti@univ-angers.fr;

Abstract

In machine learning-driven plant phenotyping, well-annotated image datasets are essential for developing robust models capable of capturing phenological variability across environments. Here, we introduce *DeepPhenoTree – Apple Edition*, a multi-site, multi-variety RGB image dataset dedicated to the detection of key phenological stages in apple trees. The dataset comprises 48,320 time-stamped RGB images acquired across four European orchards of the Apple REFPOP consortium under contrasting climatic conditions. From this large corpus, a carefully curated subset of 808 representative images was manually annotated. It includes 241,600 expert annotations covering developmental stages from dormant bud to fruit maturity.

Images were acquired using a standardized tractor-mounted phenotyping platform equipped with active flash illumination, ensuring consistent lighting conditions across sites and acquisition dates. Phenological structures were annotated following the BBCH scale, with bounding boxes adapted to organ visibility and developmental stage. In addition to the dataset, we provide baseline deep learning experiments to illustrate detection performance and assess model generalization across locations.

Keywords: Apple phenology, multi-site dataset, deep learning, RGB imaging, growth stages, dataset paper

1 Background

Apple (*Malus domestica*) is a perennial crop that characterized by a succession of phenological stages, strongly influenced by climatic variability. Accurate monitoring of these stages is essential for orchard management, pest and disease forecasting, irrigation scheduling, and yield prediction.

Phenological monitoring in orchards is traditionally performed through manual field scoring based on standardized scales such as the extended Biologische Bundesanstalt, Bundessortenamt et Chemische Industrie (BBCH) scale (Meier et al., 1994), Correspondence with the historical Fleckinger scale (Fleckinger, 1948), or North American USDA descriptors (Ferree and Warrington, 2003). While these approaches are well established, they are time-consuming and provide limited temporal resolution. As a result, they are poorly suited to large-scale or high-frequency phenotyping.

In recent years, advances in machine learning and computer vision have enabled automated phenotyping approaches, particularly in annual crops. However, publicly available image datasets for perennial crop phenotyping—especially in orchard environments—remain limited. The scarcity is even more pronounced for datasets that simultaneously encompass multi-site acquisitions, real-world environmental variability, fine-grained organ-level annotations, and dense temporal monitoring.

To address these limitations, we introduce *DeepPhenoTree – Apple Edition*, a multi-site RGB image dataset dedicated to apple tree phenology. The dataset was acquired under open field conditions and annotated by experts following BBCH scale (Meier et al., 1994). It is designed to support the development and evaluation of computer vision models for phenological stage detection, tree-level phenotyping, and orchard-scale analysis.

2 Construction and content

2.1 Orchard site and material

The dataset consists of RGB images acquired at multiple time points during the 2025 different phenological phases across four European apple orchards, using the same tractor-mounted phenotyping platform.

Apple REFPOP orchard (Jung et al., 2020) is an European network, which spans six locations in total across the European continent and represents diverse climatic regions. Each site hosts the same Apple REFPOP core collection, comprising 534 apple genotypes, with two trees per genotype arranged in a vertically trained trellis system. The network was established to support genomic and phenotypic research on apple trees, with particular emphasis on genotype–environment interactions, climate response, genomic prediction, and cultivar evaluation. Within each orchard, two management regimes are implemented to assess genotypic responses to contrasting agricultural practices. One regime follows conventional regional management, while

139 the second applies alternative practices, such as low-input systems. At the time of
140 data acquisition in 2025, trees were nine years old.

142 DeepPhenoTree was collected at Apple REFPOP orchards : Institute of Research
143 and Technology in Agro-Food in Gimenells, Spain; the Research Centre Laimburg
144 (RCL) in Auer, Italy; Agroscope in Wädenswil, Switzerland; and Better3fruit N.V. in
145 Aarschot, Belgium.



166 **Fig. 1** PhenoMobile mounted on tractor at Auer, Italy, in December 2024. The platform integrates
167 two RGB cameras, two GNSS antennas, two flash units, and one LiDAR sensor.

169 Each site was equipped with a mobile tractor-based phenotyping platform, referred
170 to as the PhenoMobile (Fig. 1). The PhenoMobile consists of a sensor head mounted on
171 an adjustable pole and integrates four types of sensors: two high-resolution RGB cam-
172 eras arranged in a stereovision configuration, two xenon flash units, a two-dimensional
173 Light Detection and Ranging (LiDAR) sensor, and a dual-antenna Global Navigation
174 Satellite System (GNSS) receiver. A detailed technical specification of each sensor is
175 provided in table 1.

181 Data acquisition was controlled through scripted sensor activation based on a
182 georeferenced orchard map. This map is composed of microplots covering both sides
183
184

of the orchard rows. Each microplot corresponds to a spatial unit containing two adjacent trees, which define the boundaries of the microplot. For each microplot, a LiDAR point cloud was acquired, and a color image was captured at the center of the microplot.

Color image acquisition was synchronized with flash activation within 200 μ s, ensuring controlled illumination and complete visibility of the two trees within the field of view. Flash units were configured to emit 100 J per trigger, effectively suppressing ambient sunlight and standardizing lighting conditions across all images. As a result, image brightness was homogenized across microplots and acquisition dates.

Due to flash recharge constraints, the tractor speed was limited around 2 km h⁻¹. Figure 2 illustrates the georeferenced microplot map at Italy, with the PhenoMobile positioned at the center of each microplot during data acquisition.

Inter-row spacing varied across sites, and camera height was therefore adapted to maximize canopy visibility, particularly in the middle canopy where apple reproductive structures predominantly occur. Inter-row distances are 4.0 m at Spain, 2.5 m at Italy, and 3.5 m at both Switzerland and Belgium, while inter-tree spacing ranges from 0.7 to 1.3 m depending on the location.. In Spain, Belgium, and Switzerland, inter-row spacing was sufficiently large to capture both the trunk and upper canopy, whereas in Italy, narrower inter-row spacing limited visibility primarily to the middle and upper canopy.

Camera field of view (FOV) and mounting height were adjusted at each site according to the distance between the cameras and the tractor center, as well as to ensure optimal visibility of the two target trees. Although the FOV differed among sites, image resolution varied accordingly, with dimensions of 3088 $\times$ 5120 pixels in Switzerland, 3560 $\times$ 5120 pixels in Belgium, 3656 $\times$ 5120 pixels in Spain, and 5120 $\times$ 5120 pixels in Italy.

Table 1 List of the main sensors integrated into the PhenoMobile.

Sensor	Model	Number	Details
Colour camera	HIKRobot MV-CH250-90GC	2	5120 (H) x 5120 (W) px = 25Mpx Genicam interface Electrical triggering of image capture Global shutter trigger Pixel size: 2.5 μm Focal length: 8.5 mm (FOV 77° x 77°)
LIDAR	Sick LMS 4000 LiDAR	1	Scanning frequency: 600 Hz Operating area: 0.5 m ... 5.5 m Opening angle: 70°
Flash	Phoxene SxIP type Xenon	2	Lighting field: 80° x 50° Adjustable energy per trigger: 20–120 J

The acquisition protocol was designed to collect data for monitoring apple tree phenology as well as fruit growth, fruit color, quality, and health status. Based on these objectives, data collection was organized into three main phenological periods: flowering, fruitlet development, and fruit maturity. Together, these periods encompass five principal BBCH growth stages: bud development (BBCH 00), inflorescence emergence (BBCH 51-59), flowering (BBCH 61-69), fruit development (BBCH 71-77), and fruit and seed maturity (BBCH 81-89).

Data were collected under open field orchard conditions. As a result, the main sources of variability included fluctuations in ambient light due to changing weather conditions (e.g., cloud cover) and variations in tree visibility caused by overlapping canopies and branch movement.

The acquisition schedule was adapted to the site-specific timing of phenological development, as climatic differences among locations resulted in asynchronous phenological responses across orchards. Phenological data acquisition depended on seasonal timing, weather conditions, and field staff availability. Data were typically acquired on a weekly basis, although acquisition frequency varied depending on local constraints.

The acquisition strategy was designed to ensure sufficient coverage of three key phenological datasets: Flowering related to BBCH 00-69, Fruitlet related to BBCH 71-77, and Fruit related to BBCH 81-89. During each acquisition, the PhenoMobile captured images from both sides of the tree rows under both management regimes. Each acquisition covered the entire orchard and generated more than 2,000 images associated with individual microplots.

During data acquisition, the platform control interface of PhenoMobile indicated sensor activation but did not provide real-time feedback on data quality. A summary of the BBCH stages included in the dataset and the number of acquisition rides conducted at each location is provided in Table 2.

Table 2 Summary of orchard acquisition rides per phenological period during 2025 season.

Period	Location	Ride	Period
Flowering	Switzerland	5	14-04 to 30-04
	Italy	7	02-04 to 22-04
	Belgium	4	08-04 to 22-04
	Spain	8	24-03 to 22-04
Fruitlet	Switzerland	2	25-06 to 09-07
	Italy	1	11-06
	Belgium	1	14-05
	Spain	4	27-05 to 02-07
Fruit	Switzerland	3	02-08 to 26-08
	Italy	10	07-08 to 01-10
	Belgium	4	26-08 to 18-09
	Spain	4	31-07 to 01-10

Illumination variability was mitigated through synchronized activation of the RGB cameras and flash units. However, effective flash triggering was not guaranteed at higher vehicle speeds. Speeds exceeding 2–4 km h⁻¹ resulted in underexposed images due to incomplete flash activation. When the flash was properly triggered, image

brightness was standardized across microplots, and background visibility beyond the target rows was largely suppressed.

After each acquisition ride, data were downloaded and securely stored. Sensor data associated with each microplot were compressed and stored in HDF5 format. A single acquisition ride consisting solely of RGB images generated approximately 1 TB of data. Due to the large data volume, data transfer from the PhenoMobile to local servers was performed manually using SD cards, which was time-consuming.

Once stored on local servers, data were uploaded via command-line tools to a secure virtual server hosted by the European Grid Infrastructure (EGI). Within the virtual server, data were organized hierarchically by acquisition year, phenological period (flowering, fruitlet, and fruit), and site. For each site, individual subdirectories corresponded to acquisition rides and contained the associated HDF5 data files.

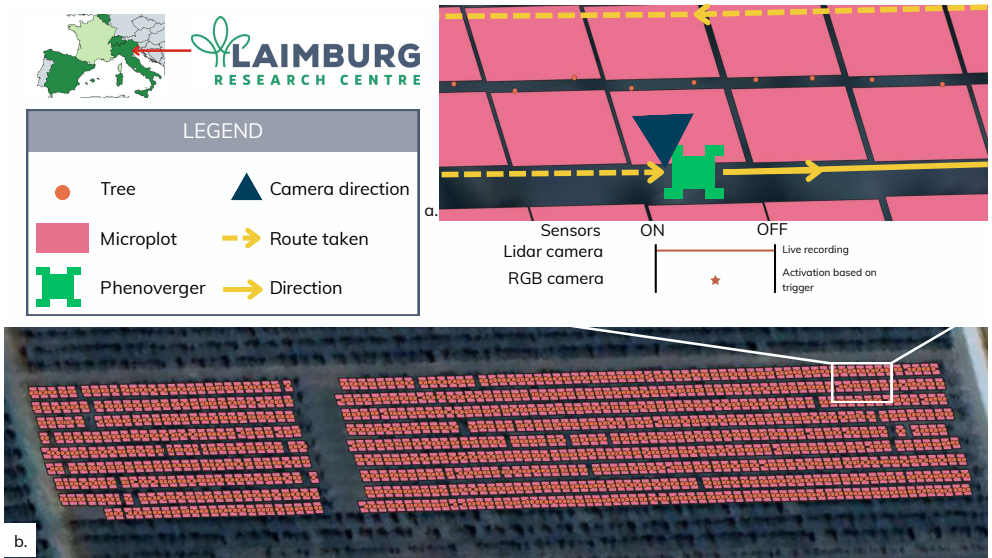


Fig. 2 Orchard map of the Apple REFPOP plot at Laimburg Research Centre. Figure a. shows data acquisition by the PhenoMobile at the microplot level. Figure b. shows the microplots covering the orchard, each corresponding to a pair of trees. The map illustrates the division of the orchard into two management regimes, with Management 1 located on the right and Management 2 on the left.

2.2 Data description

Although the acquisition campaign generated 48,320 images, manual annotation of the entire dataset was not feasible due to time and resource constraints. Therefore, a curated subset of 808 images was selected to ensure the coverage of phenological variability, genotypes and locations representation.

The dataset combines RGB images acquired at four European orchard sites. Figure 3 illustrates representative images collected in Spain, spanning phenological stages from flowering to harvest. The initial raw dataset was imbalanced both in the number of images per site and in terms of color variability.

To reduce inter-site and intra-site color variability, a manual white-balance normalization was applied to all images. Gray reference objects visible in the scenes were used as calibration targets, including poles in Italy and Spain, small poles in Switzerland, and protective nets in Belgium. For each acquisition ride, a reference target was manually selected, and a rectangular region of interest (ROI) was defined on the target using ImageJ 1.54p (Schneider et al., 2012).

The RGB distribution within each ROI was extracted, and mean R , G , and B values were used to compute per-channel correction factors (f_R , f_G , and f_B), where $f_R = R/G$, $f_G = 1$, and $f_B = B/G$. These factors were applied to the full image to match predefined reference RGB values. Because flash power was fixed during acquisition, lighting conditions remained stable over time, enabling consistent per-ride color normalization.

In parallel, a preprocessing step was implemented to balance the number of images across sites, resulting in a curated dataset that is homogeneous with respect to location. Images were selected based on genotype visibility under flash illumination and with the objective of maximizing genotype coverage. The curated selection includes images covering at least 15 genotypes per acquisition ride, with the aim of representing all genotypes at least once in the final dataset.

To limit annotation time and cost, the number of images retained per curated dataset was deliberately restricted between 15 and 150 images per site and phenological period, while ensuring that all sites and genotypes were represented.

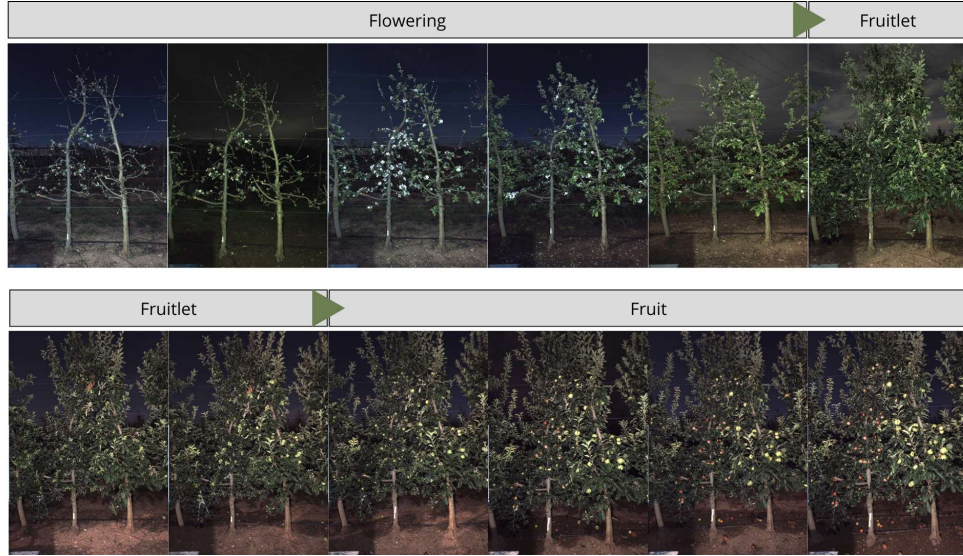


Fig. 3 Representative images covering the three acquisition periods, captured with the PhenoMobile in Spain.

Annotation was performed by Human-in-the-loop, a private European annotation company, following detailed annotation guidelines developed by Université d’Angers. These guidelines were derived from established manual phenotyping protocols used within the Apple REFPOP framework. Figure 4 shows representatives samples of each location with ground truth annotations.

Annotations consist of bounding boxes drawn around phenological structures corresponding to BBCH stages, as summarized in Table 3. Figure 5 illustrates representative BBCH stages as observed in the dataset.

In the Flowering dataset, bounding boxes individually enclose flowering from BBCH 00 to BBCH 59. From BBCH stages 61 to 69, leaf emergence and flower opening resulted in substantial object overlap. In these cases, annotators enclosed



Fig. 4 Sample images with ground-truth bounding boxes in green.

flower clusters rather than individual organs, provided that the clusters were clearly distinguishable from the background.

In the Fruitlet dataset, bounding boxes enclose fruit clusters for stages BBCH 71 and 72, and subsequently enclose individual fruits for stages BBCH 74 to 77. In the Fruit dataset, fruits are annotated individually, including in cases of partial overlap. For BBCH stages 81 to 89, fruit instances are annotated only when at least two characteristic apple features—a rounded shape and characteristic coloration—are clearly visible, either against the foliage or on the ground.

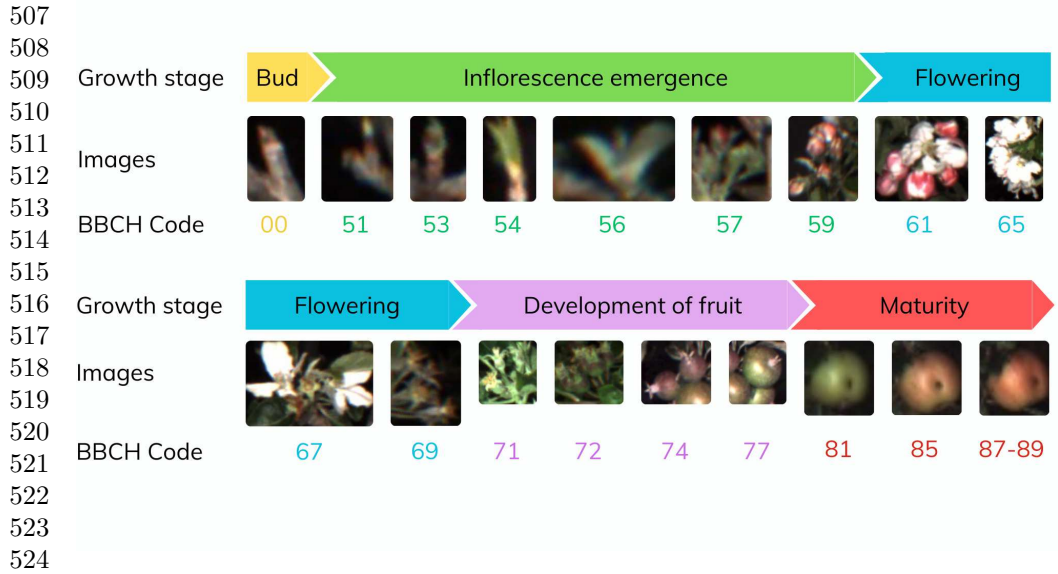


Fig. 5 Phenological growth stages and corresponding BBCH codes of pome fruit observed in images acquired in Switzerland and Spain.

The label schema and the definition of the annotation unit for each dataset are summarized in Table 4.

Quality control was performed in two stages. During the initial annotation phase, 30% of the images from each location were manually reviewed to assess annotation consistency. Feedback and corrections were provided to ensure that the guidelines were correctly understood and applied. After validation of the annotation procedure, the remaining 70% of the dataset was annotated and subsequently manually re-checked.

Final annotations consist of bounding boxes associated with each detected object. Labels are stored in YOLO format as `class_id x_center y_center width height`, with coordinates normalized by image width and height. A summary of the dataset composition, including the number of images and annotated bounding boxes per location, is provided in Table 5. The overall experimental and technical workflow is illustrated in Figure 7.

The dataset described in Table 5 is publicly available through the DATA INRAE repository [Metuarea et al. \(2026\)](#) (will be open after acceptance). The corresponding

Table 3 BBCH stages (Meier et al., 1994) included in each dataset.

Dataset	Principal growth stage		BBCH	
	Code	Description	Code	Description
Flowering	0	Bud development	00	Dormancy
	5	Inflorescence emergence	51	Inflorescence buds swelling
			53	Bud burst
			54	Mouse-ear stage
			56	Green bud stage
			57	Pink bud stage
			59	Most flowers with petals forming a hollow ball
	6	Flowering	61	About 10% of flowers open
			65	Full flowering : at least 50% of flowers open
			67	Flowers fading
			69	All petals fallen
			71	Fruit size up to 10mm
			72	Fruit size up to 20mm
			74	Fruit size up to 40mm
Fruitlet	7	Development of fruit	77	Fruit about 70% of final size
			81	Beginning of ripening
			85	Advanced ripening
			87-89	Fruit ripe
Fruit	8	Maturity of fruit and seed		

Table 4 Label schema and annotation unit definition. All annotations use a single YOLO class.

Class ID	Class name	Subset	Annotated instance definition
0	item	Flowering	Bounding box around a visible phenological structure. From BBCH 00–59, items are annotated individually when separable; from BBCH 61–69, annotators enclose a cluster of flowers when individual organs overlap.
0	item	Fruitlet	BBCH 71–72: bounding boxes enclose fruit clusters; BBCH 74–77: fruits are annotated individually.
0	item	Fruit	Individual mature fruits are annotated (regardless of size) when the rounded fruit contour is clearly visible.

Table 5 Synthesis of dataset per location during 2025 season.

Dataset	Location	Raw data		Curated data	
		Images	Bounding box	Images	Bounding box
Flowering	Switzerland	98	13383	94	12985
	Italy	135	21181	94	14815
	Belgium	94	17824	94	17824
	Spain	186	44733	94	24832
Fruitlet	Switzerland	90	7620	56	5038
	Italy	56	6499	56	6499
	Belgium	63	10230	56	9387
	Spain	248	42296	56	8988
Fruit	Switzerland	52	2538	52	2538
	Italy	120	6596	52	4284
	Belgium	107	14327	52	9618
	Spain	189	37883	52	20996

Digital Object Identifier (DOI) is <https://doi.org/10.57745/NORPF1>. The overall file structure of the repository is summarized in Figure 6.

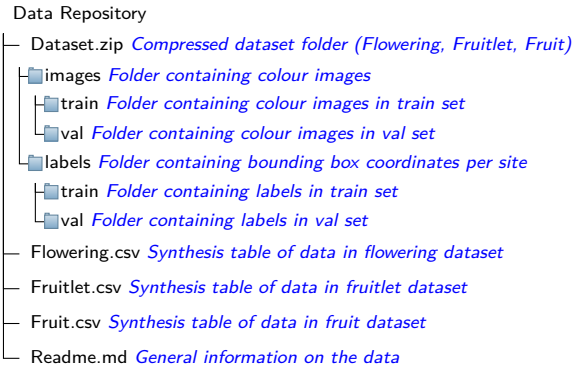


Fig. 6 Data repository structure in the DATA INRAE repository, DOI: <https://doi.org/10.57745/NORPF1> (the data set will be open upon acceptance) .

This dataset is proposed to be positioned within the context of existing literature on orchard imaging datasets. In order to position our dataset relative to existing

datasets, we comparatively assess previously published datasets and highlight their main limitations.

Most publicly available datasets related to apple orchards focus primarily on the fruit ripening stage (Miranda, 2024; Gené-Mola et al., 2020; Chu et al., 2023; Johanson et al., 2024; Häni et al., 2020; Gené-Mola et al., 2019; Bargoti and Underwood, 2017; Sun et al., 2021; James et al., 2024; Yuan, 2024; Koirala et al., 2019; Santos et al., 2020; Gremes et al., 2023; Yuan, 2024). These datasets are mainly designed to address fruit detection, counting, or tracking tasks in orchard environments. In particular, they emphasize robustness to occlusion of fruits (Chu et al., 2023), detection of small fruits under low-contrast conditions (Johanson et al., 2024), low contrast imagery (Gremes et al., 2023), or variability in illumination and complex backgrounds (Häni et al., 2020).

In most cases, the images capture only portions of the tree rather than the entire canopy. Moreover, these datasets are not explicitly intended for detailed biological or phenological analysis, as they typically rely on low-resolution imagery and uncontrolled lighting conditions. More recently, von Hirschhausen et al. (2025) introduced a dataset that includes temporal BBCH tracking and apple counting under real orchard conditions, with a slightly broader BBCH coverage (stages 0 to 9). However, the lack of controlled illumination in this dataset limits fine-scale visual analysis at the organ level and restricts phenological assessment primarily to coarse tree-level stages.

Overall, the widespread use of low-resolution images and uncontrolled lighting conditions in existing orchard datasets substantially limits their suitability for fine-grained phenological phenotyping. A summary of the key characteristics of publicly available fruit datasets is provided in Table 6.

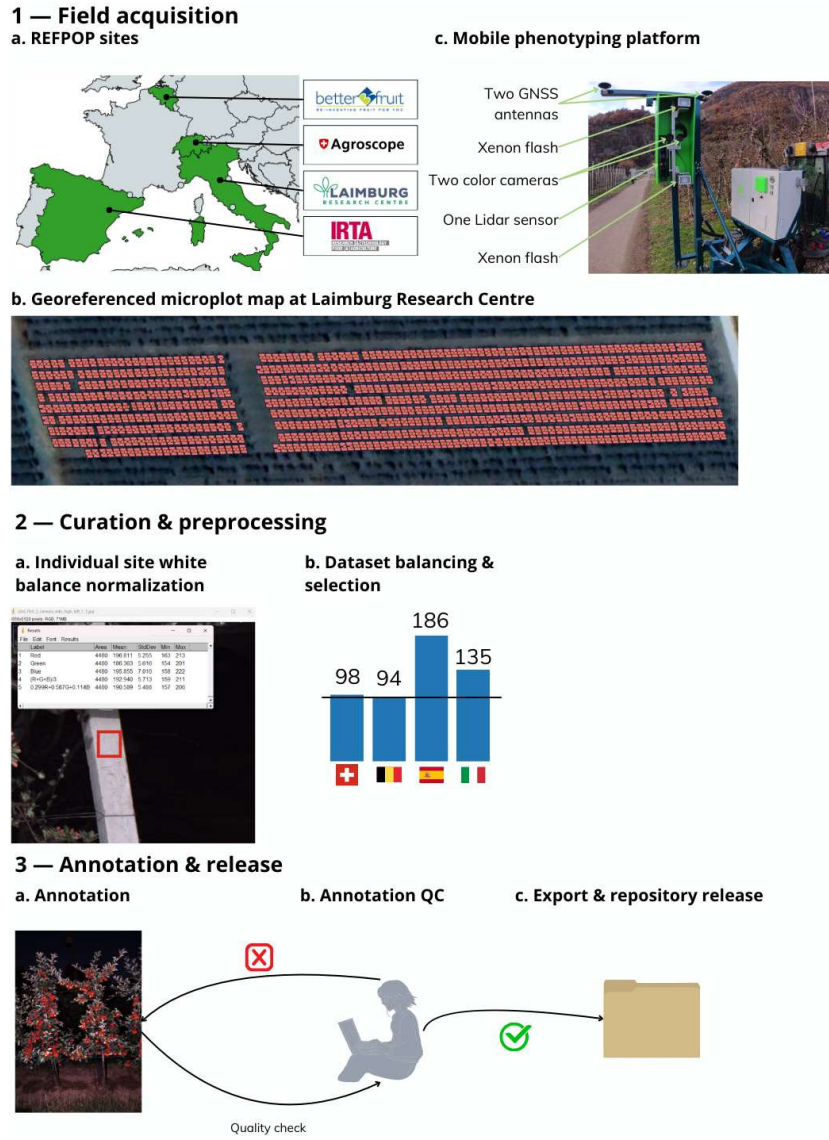


Fig. 7 Experimental and technical workflow of the DeepPhenoTree–Apple Edition dataset.

In contrast, the proposed dataset is characterized by high-resolution image acquisition under controlled lighting conditions, covering the main developmental stages of apple trees as summarized in Table 3. This design enables detailed observation

of phenological transitions, including early developmental stages, and supports the extraction of precise morphological and visual traits.

In addition, the multi-site acquisition protocol allows the assessment of model robustness across locations and environmental conditions. Each site includes a high level of genetic diversity, comprising more than 534 genotypes distributed across multiple locations throughout the European continent. Together, these features provide a robust framework for longitudinal and comparative phenotyping studies.

The multi-site design further enables the analysis of phenological traits across diverse genotypes under four distinct soil and climatic conditions.

3 Utility and discussion

This section evaluates the practical utility of the DeepPhenoTree–Apple Edition dataset for automated phenological analysis. We first establish a baseline performance for phenological object detection to quantify its immediate applicability. We then explore broader research opportunities enabled by the dataset, particularly in the context of fine-grained phenotyping and cross-site generalization.

3.1 Baseline: phenological object detection

We provide a coarse baseline in which BBCH sub-stages are merged into broader phenological categories; however, the released annotations retain sufficient detail to support future fine-grained BBCH sub-stage modeling. Each dataset enables the development of automated detection of phenological stages within an image, without explicitly distinguishing individual BBCH sub-stages at inference time.

Real-Time Detection Transformer (RT-DETR) architecture proposed by [Zhao et al. \(2023\)](#), which has been shown to outperform You Only Look Once (YOLO)-based detectors in terms of average precision. Prior to training, images from the combined training and validation sets were divided into smaller tiles of 640×640

Table 6 Fruit dataset comparison considering image resolution, lighting conditions, BBCH stage coverage, varietal diversity, and country coverage. The symbol “—” indicates that the corresponding information is not reported in the referenced study.

Reference	Fruit	High Resolution	CL	BBCH	Multi-Varty	Multi-Ctry
DeepPhenoTree - Apple Edition (Our)	Apple	Yes	Yes	00-89	Yes	Yes
von Hirschhausen et al. (2025)	Apple	Yes	No	00-91	—	No
Miranda (2024)	Apple	No	No	87-89	Yes	No
Gené-Mola et al. (2020)	Apple	No	No	87-89	Single	No
Chu et al. (2023)	Apple	No	No	87-89	—	No
Johanson et al. (2024)	Apple	No	No	87-89	—	No
Häni et al. (2020)	Apple	No	No	87-89	—	No
Gené-Mola et al. (2019)	Apple	No	No	87-89	Single	No
Bargoti and Underwood (2017)	Apple	No	No	87-89	Yes	No
	Mango					No
	Almond					No
Sun et al. (2021)	Apple	No	No	65	—	No
	Peach					No
James et al. (2024)	Citrus	No	No	87-89	—	No
Yuan (2024)	Peach	No	Yes	81-89	—	No
Koirala et al. (2019)	Mango	No	No	81-89	—	No
Santos et al. (2020)	Grape	No	No	81-89	Yes	No
Gremes et al. (2023)	Orange	No	No	81-89	—	No
Yuan (2024)	Apple	No	No	51-69	—	No

CL: Controlled Light; Multi-Varty: Multi-variety; Multi-Ctry: Multi-Country.

pixels with 30% overlapping, strides equal to 448 pixels and filtering of tiles without annotation. Given the high spatial resolution of the original images, each tile preserves sufficient visual detail to capture phenological structures.

Visual inspection of the training data revealed substantial variability in object appearance, including differences in color, illumination, spatial position, and degrees of occlusion. To increase robustness to this variability, data augmentation techniques were applied during training, including color transformations in HSV space, mosaic augmentation, horizontal and vertical flipping, translation, scaling, and rotation.

The RT-DETR [Zhao et al. \(2023\)](#) is an end-to-end real-time object detector based on a Transformer architecture. It combines a convolutional neural network backbone with a hybrid encoder to extract multi-scale features while incorporating global contextual information. A Transformer decoder then processes a fixed set of object queries, each corresponding to a potential object instance. During training, predicted objects are matched to ground-truth annotations using Hungarian assignment.

In this study, RT-DETR is used as a baseline detection framework. For each object query, the model predicts a bounding box (x, y, w, h) , an object class, and a confidence score. Training relies on global bipartite matching using the Hungarian algorithm, with a loss composed of classification, L_1 bounding box regression, and IoU-based localization terms, following the original implementation.

The model was trained for a maximum of 100 epochs on a subset of the dataset, using a batch size of 8. Data loading was accelerated using eight worker threads. Optimization was performed using the Adam optimizer with an initial learning rate of 1×10^{-4} , combined with a final learning rate factor of 0.01 to progressively reduce the learning rate and stabilize convergence. A weight decay of 0.05 was applied to regularize the model and mitigate overfitting.

Early stopping was implemented with a patience of 10 epochs, such that training was automatically terminated if no improvement was observed for 10 consecutive epochs following the best-performing model. For evaluation, the trained model was combined with the Slicing Aided Hyper Inference (SAHI) framework ([Akyon et al., 2022](#)) to perform object detection on full-resolution images, enabling performance assessment at the image level. We used SAHI with a 640×640 sliding window and a 0.4×0.4 overlap, then merged the detections using GreedyNMS.

Detection performance on the test set was quantified using Average Precision (AP), which jointly accounts for precision (the proportion of correct predictions among all

875 predicted bounding boxes) and recall (the proportion of ground-truth objects cor-
876 rectly detected). We report AP at an Intersection over Union (IoU) threshold of 0.5
877 (AP@0.5), as well as mean AP computed across IoU thresholds from 0.5 to 0.95
878 in increments of 0.05 (AP@[0.5:0.95]). Experiments were conducted with Python
880 3.10.18, PyTorch 2.7.1, SAHI 0.11.36, and Ultralytics 8.3.214 on 4× **NVIDIA**
883 **H100 GPUs**. For practical usability, the trained models were embedded into a napari
884 plugin DeepPhenoTree (<https://napari-hub.org/plugins/deepphenotree>), providing
885 an end-to-end framework for model inference. In addition, we provide representative
886 test images for flowering, fruitlet, and fruit stages from each acquisition site.
887
888

890 This experiment does not aim to propose a new detection architecture, but to
891 establish a reference benchmark for object detection on the DeepPhenoTree–Apple
892 Edition dataset. In particular, a reference benchmark for the detection of all BBCH 00-
893 69 merged into Flowering; all BBCH 71-77 into Fruitlet and all BBCH 81-89 into Fruit.
894
895 As a result, each bounding box corresponds to a phenological object instance (flower,
896 fruitlet, or fruit), without distinguishing its exact BBCH sub-stage. This simplified
897 setup allows the benchmark to assess object detectability and visual consistency across
898 phenological phases, while leaving fine-grained BBCH classification to future studies.
899

900 The results reported in Table 7 primarily reflect intrinsic differences in visual com-
901 plexity among phenological stages rather than limitations of the detection model itself.
902 For fruit detection, performance is most stable at Belgium, with low variability across
903 all evaluation metrics and the highest average mAP@0.5. Italy achieves the strongest
904 localization performance, yielding the highest mean mAP@0.5:0.95. In contrast, Spain
905 exhibits lower recall and reduced mAP values, indicating more challenging detection
906 conditions. Switzerland shows intermediate but consistent performance across folds,
907 with limited variability and balanced precision–recall behavior.
908
909

910 Overall, fruitlet detection demonstrates robust performance and good general-
911 ization across locations. Precision and recall are consistently high across sites, with
912
913
914
915
916
917
918
919
920

Table 7 Average performance of trained RT-DETR on test set for each dataset. Standard deviation is applied on 5-fold.

Dataset	Location	Precision	Recall	mAP@.5	mAP@.5:.95
Flowering	Overall (4 sites)	0.69 ± 0.01	0.58 ± 0.02	0.65 ± 0.02	0.37 ± 0.02
	Switzerland	0.73 ± 0.02	0.60 ± 0.04	0.68 ± 0.03	0.40 ± 0.04
	Belgium	0.72 ± 0.02	0.63 ± 0.03	0.69 ± 0.03	0.40 ± 0.03
	Spain	0.66 ± 0.01	0.53 ± 0.05	0.60 ± 0.03	0.30 ± 0.02
	Italy	0.69 ± 0.04	0.61 ± 0.03	0.67 ± 0.04	0.40 ± 0.04
Fruitlet	Overall (4 sites)	0.85 ± 0.02	0.73 ± 0.02	0.82 ± 0.02	0.53 ± 0.01
	Switzerland	0.86 ± 0.04	0.78 ± 0.04	0.84 ± 0.06	0.56 ± 0.04
	Belgium	0.83 ± 0.03	0.65 ± 0.04	0.77 ± 0.04	0.52 ± 0.14
	Spain	0.86 ± 0.02	0.72 ± 0.03	0.81 ± 0.03	0.52 ± 0.03
	Italy	0.88 ± 0.01	0.80 ± 0.01	0.88 ± 0.01	0.61 ± 0.01
Fruit	Overall (4 sites)	0.87 ± 0.01	0.79 ± 0.01	0.86 ± 0.01	0.57 ± 0.01
	Switzerland	0.86 ± 0.03	0.80 ± 0.02	0.87 ± 0.02	0.59 ± 0.01
	Belgium	0.90 ± 0.01	0.84 ± 0.01	0.90 ± 0.01	0.63 ± 0.02
	Spain	0.86 ± 0.02	0.75 ± 0.02	0.84 ± 0.02	0.51 ± 0.03
	Italy	0.88 ± 0.02	0.84 ± 0.03	0.90 ± 0.02	0.66 ± 0.02

Overall (4 sites) denotes the aggregated evaluation across the four experimental sites (Switzerland, Belgium, Spain, and Italy).

mAP@0.5 values ranging from 0.77 to 0.88 depending on the location. This stability can be attributed to the relatively homogeneous morphology and improved visibility of fruitlets, which is further reflected in sustained performance at stricter localization thresholds, as indicated by high mAP@0.5:0.95 values.

In contrast, flowering stage detection is more challenging, as evidenced by a pronounced drop between mAP@0.5 and mAP@0.5:0.95, indicating difficulties in accurately localizing detected objects. This degradation is particularly pronounced at certain sites, notably Spain with mAP@0.5 equal to 0.60 and mAP@0.5:0.95 equal to 0.30 in means. Flowering structures are highly context-dependent: flowers often appear partially occluded, clustered, and visually intertwined with surrounding foliage, branches, and background elements. Their appearance also varies substantially with respect to color, illumination, and phenological progression.

967 By comparison, fruitlet and fruit detection are more object-centric tasks. These
968 structures exhibit more compact and homogeneous morphologies, clearer boundaries,
969 and improved visual separability from the background, facilitating more consistent
970 localization across acquisition conditions. Collectively, these observations indicate that
971 the dataset is well suited for fruitlet and fruit detection using standard object detection
972 approaches, whereas flowering stage detection likely requires more advanced strategies
973 to better capture visual and phenological variability, such as multi-scale learning as
974 introduced in [Liu et al. \(2025\)](#); [Aldubaikhi \(2025\)](#), domain adaptation as explored by
975 [Mohamadi \(2024\)](#); [Yao et al. \(2025\)](#), or segmentation-based methods as proposed by
976 [Li et al. \(2023\)](#).

983 Furthermore, figure 8 presents qualitative examples comparing ground-truth anno-
984 tations and RT-DETR predictions across the three dataset (Flowering, Fruitlet, and
985 Fruit). For the Flowering dataset, flowers are accurately detected with good spatial
986 alignment to the ground truth. A similar behavior is observed for the Fruitlet sub-
987 set, where fruitlets are consistently and correctly localized. In the Fruit subset, fruits
988 are reliably detected both on the trees and on the ground, demonstrating the model’s
989 ability to generalize across varying spatial configurations and object contexts.
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012

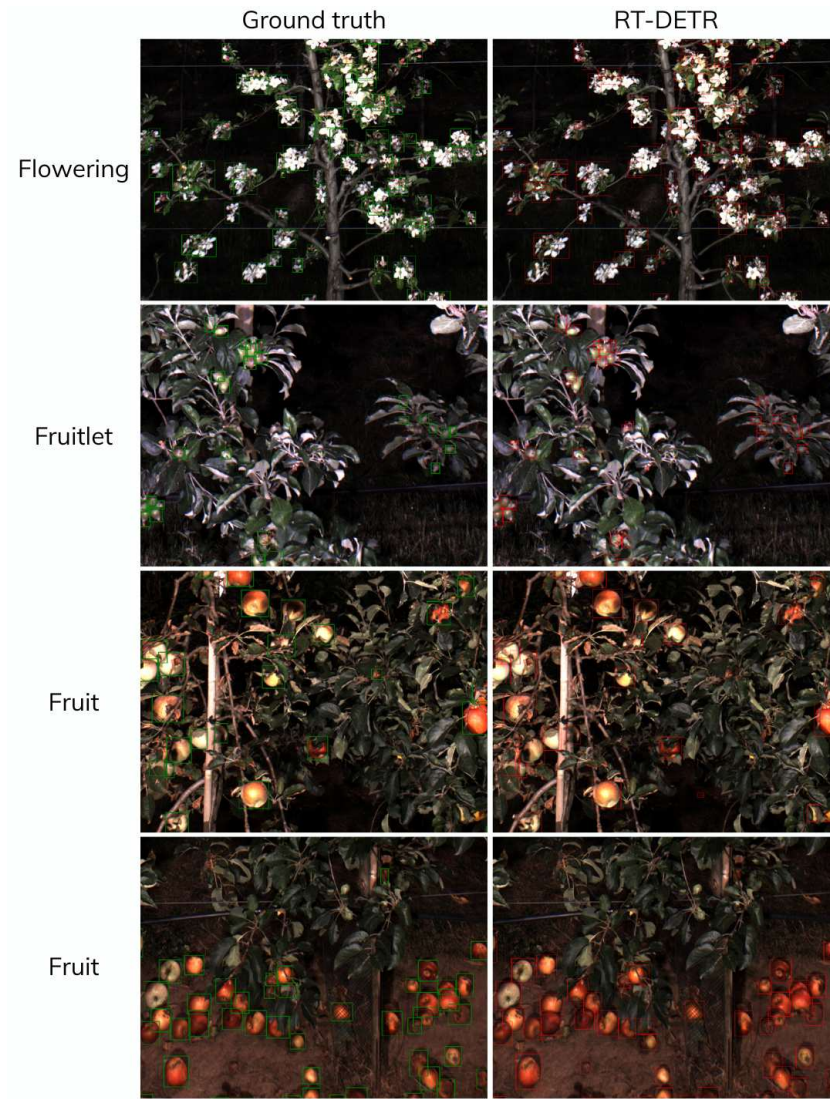


Fig. 8 Illustration of ground truth annotations and predicted bounding boxes in each dataset. Bounding box in red (resp. green) means predicted (resp. ground truth) bounding box

Figure 9,10 and 11 provide a qualitative analysis of representative detection outcomes across the three dataset (Flowering, Fruitlet, Fruit), highlighting intrinsic ambiguities in the annotations and challenging visual patterns present in the dataset. In particular, figure 11-b shows the presence of fruitlets within the Fruit subset. The

acquisition dates were determined by agronomic experts to capture periods where the majority of genotypes reached a similar phenological stage. However, due to intrinsic differences in developmental dynamics between genotypes, slight phenological asynchronies may occur within the same acquisition session. Consequently, images acquired at a given time point may contain trees exhibiting minor stage shifts relative to the majority. This variability reflects realistic field conditions and enhances the ecological validity of the dataset.



Fig. 9 Qualitative analysis in flowering dataset showing several sources of discrepancy : (a) fluorescence emergence during flowering leads to systematic over-detection, as these visual cues are consistently detected by the model but not annotated in the ground truth; (b) some annotated buds remain undetected due to their small size and low visual contrast; and (c) buds are correctly detected by the model but not seen by manual annotator. Color : ground truth in green, prediction in red.

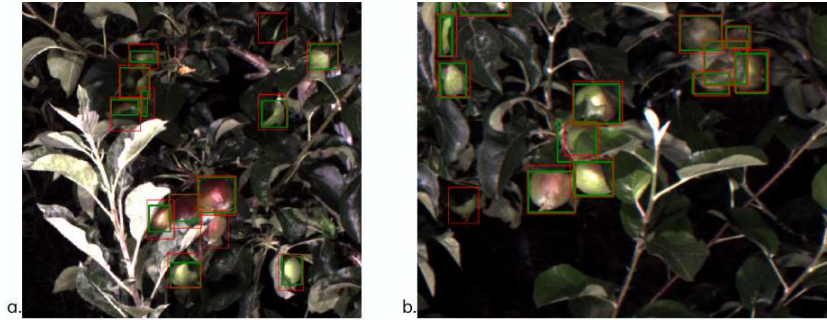


Fig. 10 Qualitative analysis in fruitlet dataset showing challenges for fruit detection : over-detections arise from unannotated fruitlets (a), while strong spatial overlap between foliage and fruitlets, as well as fruitlet–fruitlet interactions (b). Color : ground truth in green, prediction in red.



Fig. 11 Qualitative analysis in fruit dataset : severe fruit–foliage overlap (a); the presence of fruitlets within the Fruit subset (b) indicating phenological asynchrony between genotypes. We notice some detections of background fruits not included in the ground-truth annotations (c). Color : ground truth in green, prediction in red.

1151 3.2 Exploring Research Trends

1152

1153 Beyond BBCH stage recognition, the *DeepPhenoTree–Apple Edition* dataset supports
1154 a wide range of emerging research directions in computer vision-based orchard phe-
1155 notyping. Its combination of high-resolution RGB imagery, controlled illumination,
1156 dense temporal sampling, and multi-site acquisition makes it particularly well suited
1157 for investigating model robustness, generalization, and biological relevance.

1161 First, the dataset enables **cross-site and cross-environment generalization**
1162 **studies**, allowing researchers to quantify how phenological models trained under one
1163 set of climatic and agronomic conditions perform when applied to different orchard
1164 contexts. This remains a key challenge for operational phenotyping systems, as orchard
1165 environments vary widely in climate, training systems, canopy architecture, and
1166 background complexity. In this respect, the Apple REFPOP network design further
1167 supports genotype-by-environment interaction analyses based on vision-derived traits.

1173 Second, the dataset provides a foundation for **temporal phenology modeling**,
1174 in which sequential observations of the same genotypes across BBCH stages can be
1175 exploited to model phenological dynamics. Such approaches may aim to predict pheno-
1176 logical transitions, estimate developmental rates, or align phenological timelines across
1177 sites, for example by combining object detection outputs with time-series models or
1178 biologically informed constraints.

1183 Third, *DeepPhenoTree–Apple Edition* is well suited for **multi-task and rep-**
1184 **resentation learning** approaches, where a single model jointly addresses multiple
1185 phenotyping tasks such as phenological stage detection, organ counting, size estima-
1186 tion, or visual trait extraction (e.g., color and shape descriptors). The availability
1187 of fine-grained annotations across both early and late developmental stages is par-
1188 ticularly valuable for learning transferable visual representations that extend beyond
1189 fruit-centric datasets.

1194

1195

1196

Finally, the dataset can serve as a reference benchmark for **annotation-efficient learning strategies**, including semi-supervised learning, weak supervision, and active learning. Given the high cost and expertise required for phenological annotation, such strategies are essential for scaling phenotyping pipelines to larger spatial and temporal domains. By providing a curated, diverse, and openly accessible dataset, this work aims to foster methodological advances at the interface between plant science and computer vision.

4 Future extensions and perspectives

The current release of *DeepPhenoTree–Apple Edition* has been designed as a modular phenotyping resource that can be incrementally enriched in subsequent releases. A first axis of extension concerns **phenological annotation refinement**. While the dataset already supports phenology-oriented analyses, future work may provide additional expert-derived BBCH labels at increased granularity (e.g., finer macro-/micro-stage delineations) and harmonization guidelines across sites and seasons. Such refinements would strengthen biological interpretability, enable more consistent longitudinal comparisons, and facilitate methodological benchmarking. Comparable phenology-focused resources and stage-recognition pipelines have recently been developed in other fruit systems, including grapevine growth-stage recognition at BBCH micro-levels and curated image datasets for viticulture [Schieck et al. \(2023\)](#), as well as phenology-labelled detection/segmentation benchmarks in berry crops such as raspberry [Jafary et al. \(2024\)](#).

A second major direction is the integration of **3D sensing and multimodal RGB–3D learning**. The availability of complementary LiDAR acquisitions offers the possibility to release co-registered point clouds (or derived structural descriptors) alongside RGB imagery, thereby enabling the extraction of architectural traits (e.g., crown volume, branch-related descriptors) and more robust inference under occlusion.

Recent work in orchard phenotyping illustrates the value of fusing LiDAR geometry with colour information to quantify canopy and floral traits at scale in pear and apple orchards [Xia et al. \(2025\)](#). In addition, LiDAR-derived canopy descriptors have been shown to correlate with key biophysical variables such as leaf area across pome-fruit orchards and vineyards, supporting the biological relevance of 3D structural measurements [Sanz et al. \(2018\)](#). Making such 3D information available would broaden the scope of *DeepPhenoTree–Apple Edition* beyond 2D phenological recognition toward integrated structural–functional phenotyping.

Finally, extending the dataset over **multiple growing seasons** and potentially incorporating auxiliary data layers (when available) would support genotype-by-environment analyses, quantify inter-annual phenological plasticity, and improve model robustness under climatic variability. This perspective is aligned with ongoing developments in other perennial fruit systems, including multi-view fruit detection/tracking for yield estimation in commercial citrus orchards [Santos et al. \(2023\)](#), and UAV-driven phenology/health/yield modelling frameworks developed for mango orchards [Afsar et al. \(2025\)](#). Altogether, these envisioned extensions position *DeepPhenoTree–Apple Edition* as an evolving benchmark that can progressively integrate richer phenological, structural, and temporal information, fostering advances at the interface between agronomy, plant science, and computer vision.

5 Conclusion

This data paper introduces *DeepPhenoTree – Apple Edition*, a large-scale, multi-site, and expertly annotated RGB dataset dedicated to apple tree phenology. By covering key BBCH stages from dormancy to fruit maturity under controlled illumination and real orchard conditions, the dataset fills an important gap in publicly available resources for perennial crop phenotyping.

The dataset is distinguished by its high spatial resolution, standardized acquisition protocol, and genetic and environmental diversity across four European countries. The inclusion of baseline deep learning experiments demonstrates its immediate usability for phenological stage recognition while highlighting remaining challenges related to occlusion, inter-site variability, and class imbalance.

Beyond the baseline results, *DeepPhenoTree – Apple Edition* is intended as a long-term reference resource for developing robust, biologically meaningful, and transferable computer vision models in orchards. It supports a wide range of applications, from phenology monitoring and yield-related trait extraction to genotype evaluation and climate adaptation studies.

By making this dataset openly available, we aim to foster reproducibility, cross-study comparisons, and interdisciplinary collaboration between plant scientists, agronomists, and computer vision researchers. Future extensions may include additional years, modalities (e.g., LiDAR or multispectral data), and phenological traits, further strengthening the role of vision-based phenotyping in sustainable orchard management.

6 Declarations

Ethics approval and consent to participate: Not applicable. Consent for publication: Not applicable. Availability of data and materials: The datasets generated and/or analyzed during the current study are available in the DATA INRAE repository, DOI: <https://doi.org/10.57745/NORPF1> (will be open upon acceptance). Competing interests: Author Jeremy Labrosse was employed by the company Hiphen; Shauny Van Hoyer was employed by the company Better3fruit N.V. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. Authors' contributions: HM, AO, WG, AP, LL, SVH, FL, DR conceived and designed this work. HM, WG,

1335 FZ, FP, AP, LL, SVH carried out the acquisitions. HM managed the data storage.
 1336
 1337 HM managed image annotations. AO performed the deep learning experiment. HM,
 1338 AO, DR interpreted all the data. HM, DR wrote the manuscript. WG, FP, FZ, AP,
 1339 LL, SVH, FL, JL, PR revised the manuscript. HM, WG, AP, LL, SVH, FL, JL, PR,
 1340
 1341 DR supervised the work. All authors read and approved the final manuscript.
 1342

1343

1344

1345 **7 Acknowledgments**

1346

1347 The authors wish to thank the heads of each Apple REFPOP site for their collab-
 1348 oration and support, especially to the technical staff of the experimental units for
 1349 their valuable assistance: technical staff at Gimènells (IRTA, Spain), at Aarschot
 1350 (Better3fruit N.V, Belgium), at Wädenswil (Agroscope, Switzerland) and at Auer
 1351 (Research Centre Laimburg, Italy). The authors also gratefully acknowledge Oumaima
 1352 Karia and the interns for their contribution to data labeling : Lou Decastro, Jade
 1353 Marhadour, Lorène Masson and Marie Kourkouvelis-Rodostamos.
 1354

1355

1356

1360 **8 Funding**

1361

1362

1363 This work was funded by the European Union’s Horizon Europe research inno-
 1364 vation programme under grant agreement No 101094587 and the PEPR program
 1365 Agroécologie et Numérique AGROECOPHEN led by INRAe within the french infras-
 1366 tructure on plant phenotyping PHENOME-EMPHASIS and in partnership with
 1367 INRIA. This work was granted access to the HPC resources of IDRIS under the
 1368 allocation 2024-AD010115553 made by GENCI.
 1369

1370

1371

1372

1373

1374

1375

1376

1377

1378

1379

1380

References

Meier U, Graf H, Hack H, Hess M, Kennel W, Klose R, et al. Phenological growth
 stages of pome fruits (*Malus domestica* Borkh. and *Pyrus communis* L.), stone fruits
 (*Prunus* species), currants (*Ribes* species) and strawberry (*Fragaria × ananassa*)

Duch.). Nachrichtenblatt des Deutschen Pflanzenschutzdienstes. 1994;46(7):141–	1381
153.	1382
	1383
	1384
Fleckinger J. Stades repères du pommier. Annales de l'École Nationale d'Agriculture	1385
de Montpellier. 1948;21:1–35.	1386
	1387
	1388
Ferree DC, Warrington IJ. Apples: Botany, Production and Uses. Wallingford, UK:	1389
CABI Publishing; 2003.	1390
	1391
	1392
Jung M, Roth M, Aranzana MJ, Auwerkerken A, Bink M, Denancé C, et al. The	1393
apple REFPOP—a reference population for genomics-assisted breeding in apple.	1394
Horticulture research. 2020;7.	1395
	1396
	1397
	1398
Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image	1399
analysis. Nature methods. 2012;9(7):671–675.	1400
	1401
	1402
	1403
Metuarea H, Ousseni Hamza AD, Decastro L, Marhadour J, Karia O, Masson L, et al.	1404
DeepPhenoTree - Apple Edition; 2026. Dataset, Version V5.	1405
	1406
	1407
Miranda JC. Open source software and benchmarking of computer vision algorithms	1408
for apple fruit detection, fruit sizing and yield prediction using RGB-D cameras	1409
[PhD thesis]. Universitat de Lleida. Lleida, Spain; 2024.	1410
	1411
	1412
	1413
Gené-Mola J, Sanz-Cortiella R, Rosell-Polo JR, Morros JR, Ruiz-Hidalgo J, Vilaplana	1414
V, et al. Fuji-SfM dataset: A collection of annotated images and point clouds for Fuji	1415
apple detection and location using structure-from-motion photogrammetry. Data	1416
in brief. 2020;30:105591.	1417
	1418
	1419
	1420
Chu P, Li Z, Zhang K, Chen D, Lammers K, Lu R. O2RNet: Occluder-occludee	1421
relational network for robust apple detection in clustered orchard environments.	1422
Smart Agricultural Technology. 2023;5:100284.	1423
	1424
	1425
	1426

1427 Johanson R, Wilms C, Johannsen O, Frintrap S. S3ad: Semi-supervised small apple
 1428 detection in orchard environments. In: Proceedings of the IEEE/CVF Winter
 1429 Conference on Applications of Computer Vision; 2024. p. 7076–7085.
 1430
 1431
 1432 Häni N, Roy P, Isler V. MinneApple: a benchmark dataset for apple detection and
 1433 segmentation. IEEE Robotics and Automation Letters. 2020;5(2):852–858.
 1434
 1435
 1436 Gené-Mola J, Vilaplana V, Rosell-Polo JR, Morros JR, Ruiz-Hidalgo J, Gregorio E.
 1437 KFuji RGB-DS database: Fuji apple multi-modal images for fruit detection with
 1438 color, depth and range-corrected IR data. Data in brief. 2019;25:104289.
 1439
 1440
 1441 Bargoti S, Underwood J. Deep fruit detection in orchards. In: 2017 IEEE international
 1442 conference on robotics and automation (ICRA). IEEE; 2017. p. 3626–3633.
 1443
 1444
 1445 Sun K, Wang X, Liu S, Liu C. Apple, peach, and pear flower detection using semantic
 1446 segmentation network and shape constraint level set. Computers and Electronics
 1447 in Agriculture. 2021;185:106150.
 1448
 1449
 1450 James JA, Manching HK, Mattia MR, Bowman KD, Hulse-Kemp AM, Beksi WJ. Cit-
 1451 det: A benchmark dataset for citrus fruit detection. IEEE Robotics and Automation
 1452 Letters. 2024;.
 1453
 1454 Yuan H. Peach Maturity Dataset. Figshare. 2024;Dataset. [https://doi.org/10.6084/
 1455 m9.figshare.27125373.v2](https://doi.org/10.6084/m9.figshare.27125373.v2).
 1456
 1457
 1458 Koirala A, Walsh KB, Wang Z, McCarthy C. Deep learning–Method overview and
 1459 review of use for fruit detection and yield estimation. Computers and electronics
 1460 in agriculture. 2019;162:219–234.
 1461
 1462 Santos TT, De Souza LL, dos Santos AA, Avila S. Grape detection, segmentation, and
 1463 tracking using deep neural networks and three-dimensional association. Computers
 1464
 1465
 1466
 1467
 1468
 1469
 1470
 1471
 1472

and Electronics in Agriculture. 2020;170:105247.	1473
	1474
Gremes MF, Fermo IR, Krummenauer R, Flores FC, Andrade CMG, Lima OCdM.	1475
System of counting green oranges directly from trees using artificial intelligence.	1476
AgriEngineering. 2023;5(4):1813–1831.	1477
	1478
	1479
	1480
Yuan W. AriAplBud: An aerial multi-growth stage apple flower bud dataset for	1481
agricultural object detection benchmarking. Data. 2024;9(2):36.	1482
	1483
	1484
von Hirschhausen LS, Magnusson JS, Kovalenko M, Boye F, Rawat T, Eisert P,	1485
et al. AppleGrowthVision: A large-scale stereo dataset for phenological analysis,	1486
fruit detection, and 3D reconstruction in apple orchards. In: Proceedings of the	1487
Computer Vision and Pattern Recognition Conference (CVPR) Workshops; 2025.	1488
	1489
	1490
p. 5443–5450.	1491
	1492
	1493
	1494
Zhao Y, Lv W, Xu S, Wei J, Wang G, Dang Q, et al. DETRs Beat YOLOs on	1495
Real-time Object Detection. arXiv preprint arXiv:230408069. 2023;.	1496
	1497
	1498
Akyon FC, Altinuc SO, Temizel A. Slicing aided hyper inference and fine-tuning for	1499
small object detection. In: 2022 IEEE international conference on image processing	1500
(ICIP). IEEE; 2022. p. 966–970.	1501
	1502
	1503
	1504
Liu F, Zheng Q, Tian X, Shu F, Jiang W, Wang M, et al. Rethinking the multi-scale	1505
feature hierarchy in object detection transformer (DETR). Applied Soft Computing.	1506
2025;175:113081. https://doi.org/10.1016/j.asoc.2025.113081 .	1507
	1508
	1509
	1510
Aldubaikhi A. Advancements in Small-Object Detection (2023–2025). Applied	1511
Sciences. 2025;15(22):11882.	1512
	1513
	1514
Mohamadi H. Feature-based methods in domain adaptation for object detection: a	1515
review. arXiv preprint. 2024;.	1516
	1517
	1518

1519 Yao H, Zhao S, Lu S, Chen H, Li Y, Liu G, et al. Source-Free Object Detection with
1520 Detection Transformer. arXiv preprint. 2025;.

1522

1523 Li F, Zhang H, Xu H, Liu S, Zhang L, Ni LM, et al. Mask DINO: Towards a Unified
1524 Transformer-based Framework for Object Detection and Segmentation. In: Proceed-
1525 ings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition
1526 (CVPR); 2023. p. 2925–2934.

1528

1529

1530 Schieck M, Krajsic P, Loos F, Hussein A, Franczyk B, Koziarkiewicz A, et al. Compar-
1531 ison of deep learning methods for grapevine growth stage recognition. Computers
1532 and Electronics in Agriculture. 2023;211:107944. [https://doi.org/10.1016/j.compag.](https://doi.org/10.1016/j.compag.2023.107944)
1533 [2023.107944](https://doi.org/10.1016/j.compag.2023.107944).

1535

1536

1537

1538 Jafary P, Bazangeya A, Pham M, Campbell LG, Saeedi S, Zareinia K, et al. Raspberry
1539 PhenoSet: A Phenology-based Dataset for Automated Growth Detection and Yield
1540 Estimation. arXiv preprint arXiv:241100967. 2024;.

1542

1543

1544 Xia Y, Li H, Zhang F, Sun G, Qi K, Jackson R, et al. OrchardQuant-3D: combining
1545 drone and LiDAR to perform scalable 3D phenotyping for characterising key canopy
1546 and floral traits in fruit orchards. Plant Biotechnology Journal. 2025;23(11):4910–
1547 4929. <https://doi.org/10.1111/pbi.70229>.

1549

1550

1551 Sanz R, Llorens J, Escolà A, Arnó J, Planas S, Román C, et al. LIDAR and
1552 non-LIDAR-based canopy parameters to estimate the leaf area in fruit trees and
1553 vineyard. Agricultural and Forest Meteorology. 2018;260-261:229–239. [https:](https://doi.org/10.1016/j.agrformet.2018.06.017)
1554 [//doi.org/10.1016/j.agrformet.2018.06.017](https://doi.org/10.1016/j.agrformet.2018.06.017).

1556

1557

1558

1559 Santos TT, de Souza KXS, Camargo Neto J, Koenigkan LV, Moreira AS, Ternes S.
1560 A pipeline for multiple orange detection and tracking with 3-D fruit relocalization
1561 and neural-net based yield regression in commercial citrus orchards. arXiv preprint
1562 and neural-net based yield regression in commercial citrus orchards. arXiv preprint
1563
1564

arXiv:231216724. 2023;.	1565
	1566
Afsar MM, Iqbal MS, Bakhshi AD, Hussain E, Iqbal J. MangiSpectra: A Multivariate	1567
Phenological Analysis Framework Leveraging UAV Imagery and LSTM for Tree	1568
Health and Yield Estimation in Mango Orchards. Remote Sensing. 2025;17(4):703.	1569
https://doi.org/10.3390/rs17040703 .	1570
	1571
	1572
	1573
	1574
	1575
	1576
	1577
	1578
	1579
	1580
	1581
	1582
	1583
	1584
	1585
	1586
	1587
	1588
	1589
	1590
	1591
	1592
	1593
	1594
	1595
	1596
	1597
	1598
	1599
	1600
	1601
	1602
	1603
	1604
	1605
	1606
	1607
	1608
	1609
	1610