

Article

Impact of Phenological and Lighting Conditions on Early Detection of Grapevine Inflorescences and Bunches Using Deep Learning

Rubén Íñiguez ^{1,2} , Carlos Poblete-Echeverría ³, Ignacio Barrio ^{1,2}, Inés Hernández ^{1,2} , Salvador Gutiérrez ⁴, Eduardo Martínez-Cámara ⁵  and Javier Tardáguila ^{1,2,*} 

¹ Televitis Research Group, University of La Rioja, 26006 Logroño, Spain; ruben.iniguez@unirioja.es (R.Í.); ignacio.barrio@unirioja.es (I.B.); ines.hernandez@unirioja.es (I.H.)

² Institute of Grapevine and Wine Sciences, University of La Rioja, Consejo Superior de Investigaciones Científicas, Gobierno de La Rioja, 26007 Logroño, Spain

³ South African Grape and Wine Research Institute (SAGWRI), Stellenbosch University, Private Bag X1, Matieland 7602, South Africa; cpe@sun.ac.za

⁴ Department of Computer Science and Artificial Intelligence (DECSAI), Andalusian Research Institute in Data Science and Computational Intelligence (DaSCI), University of Granada (UGR), 18014 Granada, Spain; salvaguti@decsai.ugr.es

⁵ Department of Mechanical Engineering, University of La Rioja, 26004 Logroño, Spain; eduardo.martinezc@unirioja.es

* Correspondence: javier.tardaguila@unirioja.es

Abstract

Reliable early-stage yield forecasts are essential in precision viticulture, enabling timely interventions such as harvest planning, canopy management, and crop load regulation. Since grape yield is directly related to the number and size of bunches, the early detection of inflorescences and bunches, carried out even before flowering, provides a valuable foundation for estimating potential yield far in advance of veraison. Traditional yield prediction methods are labor-intensive, subjective, and often restricted to advanced phenological stages. This study presents a deep learning-based approach for detecting grapevine inflorescences and bunches during early development, assessing how phenological stage and illumination conditions influence detection performance using the YOLOv11 architecture under commercial field conditions. A total of 436 RGB images were collected across two phenological stages (pre-bloom and fruit-set), two lighting conditions (daylight and artificial night-time illumination), and six grapevine cultivars. All images were manually annotated following a consistent protocol, and models were trained using data augmentation to improve generalization. Five models were developed: four specific to each condition and one combining all scenarios. The results show that the fruit-set stage under daylight provided the best performance ($F1 = 0.77$, $R^2 = 0.97$), while for inflorescences, night-time imaging yielded the most accurate results ($F1 = 0.71$, $R^2 = 0.76$), confirming the benefits of artificial lighting in early stages. These findings define optimal scenarios for early-stage organ detection and support the integration of automated detection models into vineyard management systems. Future work will address scalability and robustness under diverse conditions.

Keywords: yield prediction; inflorescence; grape bunch; precision viticulture; deep learning; object detection; yolov11



Academic Editor: Dengpan Xiao

Received: 14 May 2025

Revised: 26 June 2025

Accepted: 9 July 2025

Published: 11 July 2025

Citation: Íñiguez, R.; Poblete-Echeverría, C.; Barrio, I.; Hernández, I.; Gutiérrez, S.; Martínez-Cámara, E.; Tardáguila, J. Impact of Phenological and Lighting Conditions on Early Detection of Grapevine Inflorescences and Bunches Using Deep Learning. *Agriculture* **2025**, *15*, 1495. <https://doi.org/10.3390/agriculture15141495>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Accurate and timely yield estimation plays a crucial role in vineyard management, guiding decisions ranging from irrigation and fertilization to harvest logistics and winery operations. Early insights into grape yield potential enable winemakers and vineyard managers to optimize agricultural practices and streamline winemaking processes [1]. Traditionally, yield forecasting has relied on manual bunch counting, destructive sampling, or empirical models based on historical data [2,3]. While these methods offer a baseline, they are labor-intensive, subjective, and often fail to capture the spatial and temporal variability characteristic of vineyards, particularly in heterogeneous terrains or when managing large estates [4,5]. Moreover, their limited temporal resolution impedes early-season interventions, reducing their usefulness for proactive crop management. Even non-invasive approaches that rely on proximal sensing or manual visual assessments still face scalability issues and depend on well-developed bunches for reliable measurements [6,7].

Conventional methods generally require significant bunch development, limiting their applicability to late phenological stages such as veraison or ripening. This temporal delay constrains their utility for management decisions such as green thinning, canopy adjustment, or irrigation scheduling. To overcome these limitations, there is growing interest in early-stage yield estimation, ideally during flowering or fruit set, when the grapevine reproductive potential is first revealed [8,9].

Detecting inflorescences at these early stages allows for yield prediction months in advance, supporting timely interventions such as crop thinning, nutrient adjustments, and pest control [4,10]. This is especially valuable in cool climates or for cultivars with erratic flowering, where fruit set can be strongly influenced by environmental stressors [9]. Previous studies have shown that pre-veraison phenotyping enables more flexible and targeted management, aligning with sustainability goals in viticulture [11].

Advancements in remote sensing, computer vision, and artificial intelligence have opened new avenues for automating yield estimation in viticulture. Precision agriculture technologies such as UAV imaging, multispectral sensors, and LiDAR have been explored in various crops to assess biomass, count fruits, or estimate productivity [12,13]. In orchards, vision-based systems are now widely used for flower and fruit detection in apples, cherries, and citrus, showcasing high accuracy in structured plantation systems [14]. In vineyards, however, the application of such systems presents unique challenges. Grapevine canopies are highly variable and often contain complex arrangements of leaves, shoots, tendrils, and fruiting structures. Occlusions caused by foliage and trellis wires, irregular lighting, and varying object sizes complicate the segmentation and detection of grape bunches or inflorescences, particularly in field conditions [15,16]. These constraints are even more critical when attempting to detect yield-forming organs before veraison, when visual contrast is low and object morphology is less distinct [17,18].

These challenges are further compounded by environmental factors. Variations in illumination conditions have a critical influence on image quality and, consequently, on the performance of automated detection systems in agricultural environments. While natural daylight is readily available, it introduces challenges such as shadows, reflections, and inconsistent ambient light, all of which can hinder the visibility of small structures like grapevine inflorescences [19]. In contrast, the use of controlled artificial illumination, such as LED lighting during night-time imaging, allows for more uniform scene lighting, improved contrast of regions of interest, and reduced visual noise [20]. Field-based studies have shown that computer vision systems operating under artificial lighting achieve higher detection consistency, particularly for floral structures and fruits in early phenological stages [21]. In addition to lighting conditions, cultivar-specific characteristics such as canopy architecture, flowering density, and bunch morphology introduce considerable

visual variability. These traits play a significant role in model performance, especially during early growth stages when reproductive structures may be occluded or poorly exposed [22]. For instance, cultivars with denser foliage or compact inflorescences pose higher occlusion levels, which can reduce detection accuracy [23]. Incorporating multiple grapevine cultivars during model development contributes to improving robustness and ensures applicability across diverse vineyard scenarios [24].

To mitigate these limitations, computer vision methods increasingly integrate auxiliary sensors and unsupervised segmentation algorithms, as well as geometry-based filters and temporal tracking [18,25]. Geometry-based filters leverage prior knowledge about the expected shapes, sizes, and spatial arrangements of target structures, such as the compact, roughly conical form of grape bunches or the linear patterns of inflorescences, to improve detection accuracy in cluttered environments. These enhancements contribute to robustness in real-world environments, but challenges such as occlusion, background clutter, and phenotypic variability remain active areas of research [7,26]. Computer vision approaches based on deep learning have become particularly prominent in this context. Convolutional Neural Networks (CNNs) have demonstrated remarkable capabilities in feature extraction and object localization under complex visual contexts [27]. In viticulture, specific CNN architectures such as SegNet with a VGG19 encoder have been successfully applied for tasks like grapevine flower and inflorescence detection, achieving F1 scores of 0.93 and 0.73, respectively, and a determination coefficient (R^2) of 0.91 with respect to manual counts [28]. CNN-based systems have also been employed to segment berries and detect grape bunches [29], and recent works have explored hybrid CNN–Transformer architectures for adaptive learning in heterogeneous vineyard conditions [25]. Moreover, CNNs have been used for disease detection in vineyards, including downy mildew through binary segmentation approaches [30], black rot using super-resolution enhancement and deep learning classification [31], and other foliar symptoms via real-time CNN detectors [32]. These methods often rely on semantic segmentation or classification pipelines and have demonstrated promising accuracy levels for early detection and disease management.

Among deep learning architectures, the YOLO (You Only Look Once) family stands out for its balance between speed and accuracy in object detection [33]. Earlier versions have been applied in viticulture for bunch detection under variable lighting and occlusion conditions [34,35]. Recent models incorporate architectural improvements like transformer blocks, spatial attention modules, and enhanced feature fusion, which boost performance for small or occluded targets [36–38]. Despite these advances, most applications focus on late stages such as harvest, when bunches are large and visible [39,40], although early-stage detection during flowering is gaining attention for its potential to reduce occlusion and improve prediction accuracy [10,41].

In this context, the present work proposes a novel methodology for determining the number of grapevine inflorescences and bunches under real, open-field conditions. This study focuses on early developmental stages and considers two key factors that may influence detection performance: the phenological stage (pre-bloom vs. fruit-set) and the illumination conditions (daylight vs. night-time with artificial lighting). To this end, we implement the YOLOv11 architecture, a state-of-the-art deep learning model, and apply both classification and regression-based evaluation metrics. The dataset comprises RGB images acquired across six cultivars under varied vineyard conditions, enabling a comprehensive assessment of model performance and its potential application in early-season yield forecasting and precision viticulture.

2. Materials and Methods

2.1. Experimental Sites

The experiment was conducted during the 2018 season in a commercial vineyard located in Vergalijo, Navarra, Spain (latitude 42°27'46.0" N, longitude 1°48'13.1" W). Images were acquired at two key phenological stages according to the Baillod and Baggiolini scale [42]: pre-bloom (stage H), corresponding to separated floral buttons during flowering (Figure 1); and fruit-set (stage K), when green berries reach approximately 7 mm in diameter (Figure 2). Images were collected under two distinct lighting conditions: during the day, using uncontrolled natural sunlight; and at night, using an artificial LED spotlight.

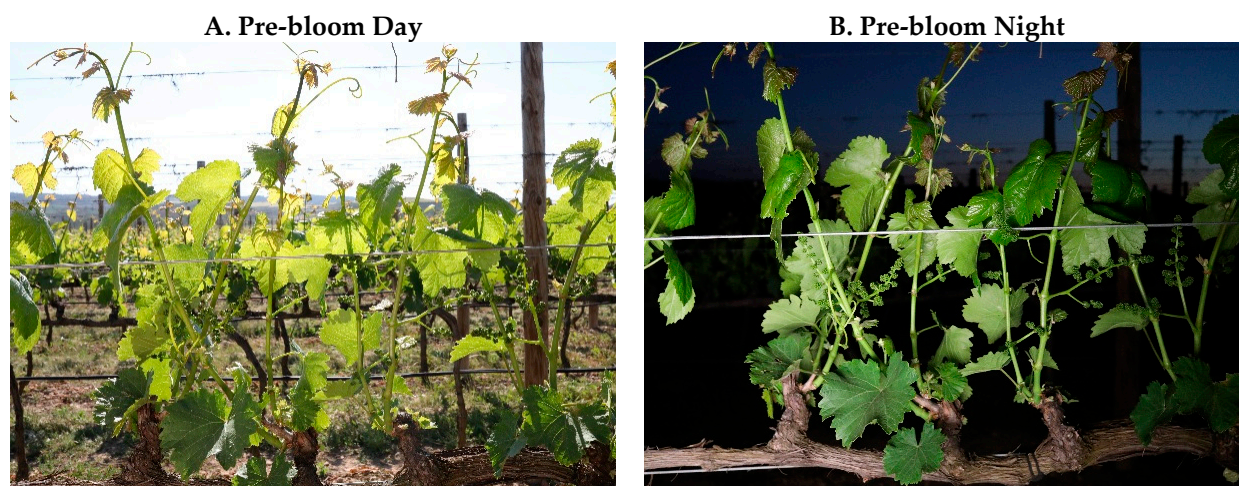


Figure 1. Representative RGB images of a single vine at the pre-bloom stage captured under daylight (A) and night-time artificial lighting (B).

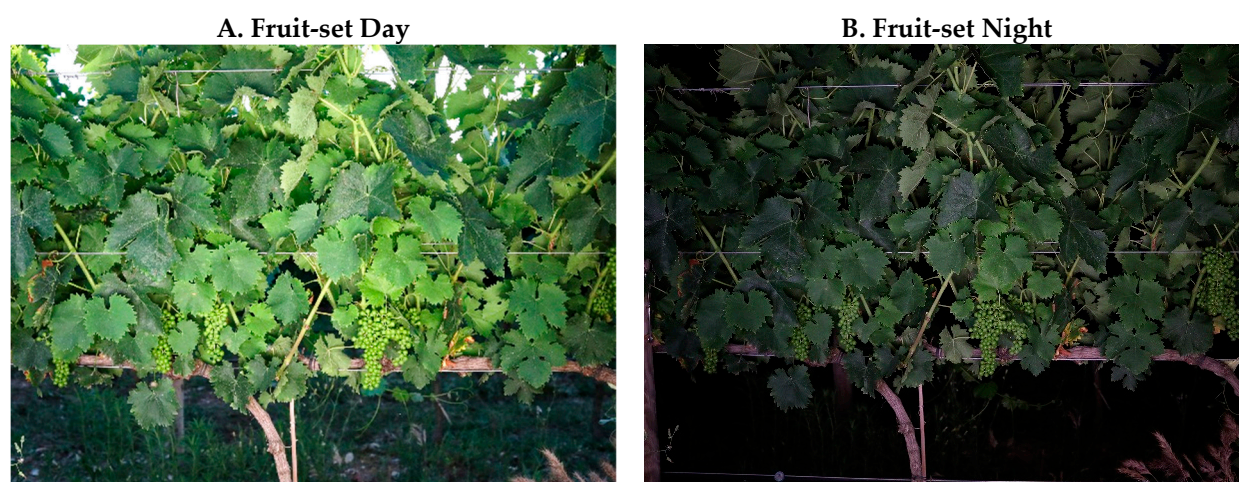


Figure 2. Representative RGB images of a single vine at the fruit-set stage captured under daylight (A) and night-time artificial lighting (B).

A total of six grapevine cultivars were evaluated (Figure 3): three red varieties (Tempranillo, Syrah, and Cabernet Sauvignon) and three white varieties (Malvasia, Moscatel, and Verdejo). The vines were trained on a Vertical Shoot Positioning (VSP) system with 2 m row spacing and 1 m vine spacing. All vines were managed following standard viticultural practices throughout the growing season. In some cases, leaf removal was performed around the fruiting zone, following common viticultural practices in the region, particularly at the berry development stage.

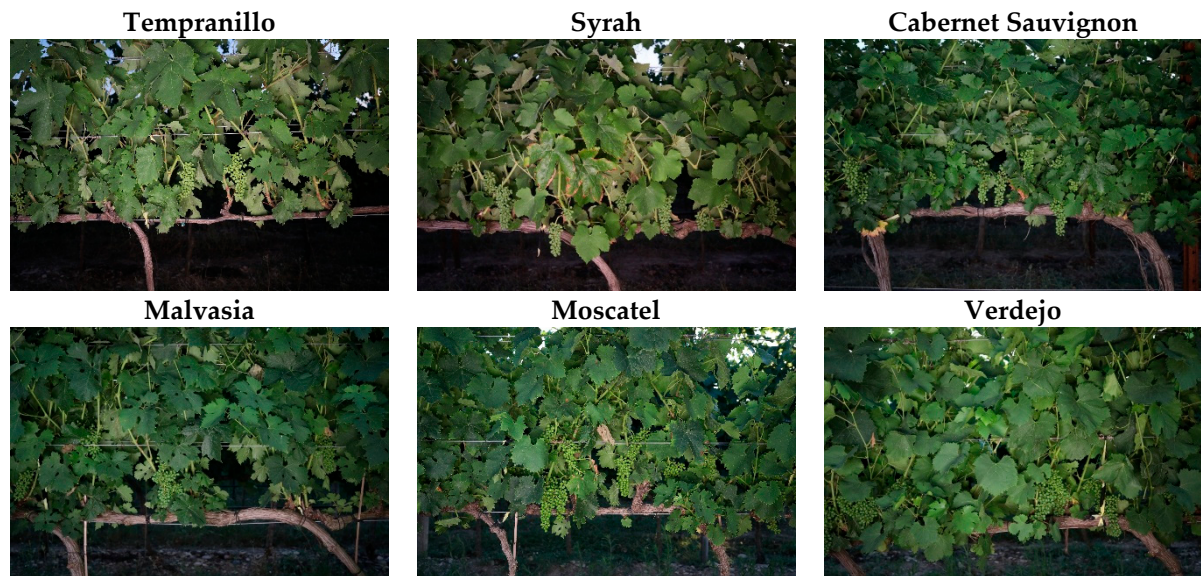


Figure 3. Representative canopy images of the six grapevine cultivars evaluated in this study. The top row shows the red varieties (Tempranillo, Syrah, and Cabernet Sauvignon), while the bottom row displays the white varieties (Malvasia, Moscatel, and Verdejo).

2.2. Image Acquisition

RGB canopy images were acquired using a mobile sensing platform developed at the University of La Rioja. The platform consisted of a modified all-terrain vehicle (ATV) (Trail Boss 330, Polaris Industries, Medina, MN, USA) equipped with a custom aluminium structure integrating a Canon EOS 5D Mark IV RGB camera (Canon Inc., Tokyo, Japan) configured for the native resolution of the sensor at 6720×4480 pixels (30.1 MP) and an artificial LED illumination system for night-time operation. The camera was mounted at a height of 1.0 m from the ground and at a distance of 0.80 to 1.80 m perpendicular to the canopy, depending on the campaign. Image acquisition was performed while the vehicle moved along the vineyard rows at a speed of 5 km/h.

Images were taken at two different phenological stages: pre-bloom (corresponding to Baillod and Baggiolini stage H), 109 days before harvest, and fruit-set (stage K, pea-size berries), 66 days before harvest. For both stages, image acquisition was carried out on the go, under both natural daylight and night-time conditions, using artificial LED illumination. This setup enabled the evaluation of model performance across varying lighting environments. Each image typically captured multiple vines along the row, although in some cases, the framing corresponded to a single plant. Since the objective was to detect all visible bunches within each frame, the number of vines per image was not considered a limiting factor. The artificial lighting provided homogeneous illumination and reduced interference from adjacent rows, thereby facilitating image segmentation and object annotation tasks. All vines included in the image acquisition underwent a partial defoliation prior to flowering, consisting of the manual removal of three basal leaves per shoot in the fruiting zone. This operation, performed uniformly across cultivars, aimed to enhance cluster visibility and reduce occlusions during both pre-bloom and fruit-set imaging stages.

Camera settings were manually fixed at the beginning of each acquisition session and remained constant throughout. Typical parameters included an exposure time of $1/1600$ s in order to avoid blur caused by the motion and vibrations of the vehicle, ISO between 2000 and 5000, and an f-number between $f/5.6$ and $f/8$, depending on ambient lighting. Images were acquired using a Canon EF 35 mm $f/2$ IS USM lens.

This acquisition strategy ensured consistent, high-resolution images for model training, aligned with previous protocols described in [28,29].

2.3. Dataset

A new dataset was compiled to support the training, validation, and evaluation of object detection models for identifying grapevine inflorescences and bunches at early developmental stages. The dataset consisted of 436 RGB images acquired under real field conditions using the mobile sensing platform described in Section 2.2. Images were captured at two distinct phenological stages, namely, pre-bloom and fruit-set, under both day and night illumination conditions. This variability ensured robustness in different light environments and canopy configurations.

Table 1 summarizes the distribution of images according to cultivar, phenological stage, and illumination condition. A total of six grapevine cultivars were included: Cabernet Sauvignon, Malvasia, Moscatel, Syrah, Tempranillo, and Verdejo.

Table 1. Distribution of RGB canopy images by grapevine cultivar, phenological stage (pre-bloom and fruit-set), and illumination condition (daylight and night-time with artificial lighting). The table summarizes the number of annotated images used for model training, validation, and testing across the six cultivars and four experimental conditions.

	Pre-Bloom	Pre-Bloom	Fruit-Set	Fruit-Set
Cultivar	Day	Night	Day	Night
Cabernet Sauvignon	16	30	10	10
Malvasia	18	32	10	10
Moscatel	6	32	10	10
Syrah	20	32	10	10
Tempranillo	19	38	10	10
Verdejo	35	38	10	10
Total	114	202	60	60

The full dataset was subsequently divided with a stratified 70/20/10% split into subsets for training, validation, and testing, ensuring representation across phenological stages, illumination conditions, and cultivars for each one of the splits. Table 2 reports the exact number of images allocated to each subset for every experimental condition, as well as the total per partition “Combined”. The training subset was used to fit the models by optimizing their parameters based on annotated data, while the validation subset was employed to monitor performance during training and guide hyperparameter tuning. The testing subset was reserved exclusively for evaluating model performance on previously unseen data, providing a robust framework for model comparison under heterogeneous conditions.

Table 2. Distribution of RGB canopy images across the training, validation, and testing subsets. Values are given for each experimental condition (Pre-bloom Day, Fruit-set Day, Pre-bloom Night, Fruit-set Night); the “Combined” column indicates the total number of images per partition.

	Pre-Bloom	Pre-Bloom	Fruit-Set	Fruit-Set	
Partition	Day	Night	Day	Night	Combined
Training (70%)	78	142	42	42	304
Validation (20%)	21	40	12	12	85
Testing (10%)	15	20	6	6	47
Total	114	202	60	60	436

Although the dataset was designed to ensure balanced representation across cultivars, phenological stages, and lighting conditions, an overrepresentation of pre-bloom images was observed, particularly under night-time illumination. This was primarily due to

variability in image framing and spacing during acquisition: pre-bloom images were captured at closer range, resulting in a higher number of individual images per vine. In contrast, fruit-set images were often taken from greater distances, covering multiple vines per frame and thus generating fewer usable samples. Additionally, some images, especially under challenging lighting, were discarded due to motion blur or a lack of focus. Despite these factors, the final dataset ensured adequate representation across conditions, supporting robust and realistic model training and evaluation.

2.4. Labeling

All images were manually annotated following a consistent protocol to ensure high-quality ground truth data for supervised training (Figure 4). The annotation focused on identifying visible grapevine organs, either inflorescences or bunches, depending on the phenological stage, and delineating them with bounding boxes. “Visible” elements were defined as those that could be directly observed and clearly identified in the image without severe occlusion or obstruction by leaves or neighboring structures.

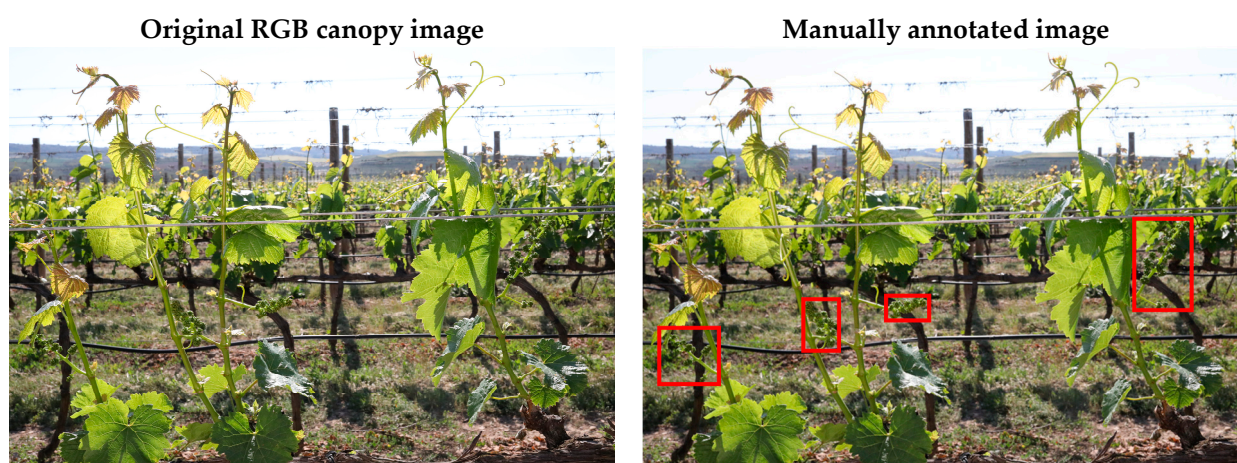


Figure 4. Example of manual annotation process used for model training. The left panel shows the original RGB canopy image, while the right panel displays the corresponding manually annotated version. Red bounding boxes indicate visible grapevine inflorescences or bunches.

The annotation process was carried out using the open-source software LabelImg [43], which enabled efficient and precise labelling. All annotations were carefully reviewed to maintain consistency across the dataset and ensure reliability during the model training and evaluation phases.

2.5. Computational Setup

The training of deep learning models was performed on a high-performance computing server equipped with an AMD Ryzen Threadripper 3970X 32-core CPU (Santa Clara, CA, USA), 256 GB of ECC DDR4 RAM, and multiple NVIDIA GeForce RTX 4090 GPUs (24 GB of VRAM each). For this experiment, a virtual machine was configured with 16 CPU cores, 100 GB of RAM, and access to a single NVIDIA RTX 4090 GPU (Santa Clara, CA, USA). This hardware configuration ensured efficient processing and training of the YOLOv11 object detection models without computational bottlenecks. Model development was conducted using Python 3.10 and the PyTorch 2.2.2 deep learning framework. The YOLOv11 implementation was provided by Ultralytics, and training was accelerated using CUDA-enabled GPU support.

2.6. Model Architecture and Training Strategy

The object detection architecture selected for this study was YOLOv11, a state-of-the-art deep learning model recognized for its balance between accuracy and computational efficiency [37]. YOLOv11 incorporates advanced modules for feature extraction and spatial reasoning, enabling it to detect small and partially occluded inflorescences under highly variable vineyard conditions [33].

The model was initialized with pretrained weights available from the official Ultralytics repository and fine-tuned on the annotated grapevine dataset using transfer learning. All layers of the network were retrained to adapt the model specifically to inflorescence detection in RGB images taken under field conditions [44]. Transfer learning has proven effective in agricultural image analysis by allowing models to leverage previously learned features and adapt them to specific tasks with limited domain data [45].

Model training was optimized using the AdamW optimizer, with an initial learning rate of 0.0001 dynamically adjusted according to validation performance [46]. The key training parameters were

- Batch size: 32;
- Image resolution: 800×800 pixels;
- Maximum epochs: 200;
- Loss function: Objectness + Classification + Bounding Box regression (YOLOv11 default).

This configuration was selected to maximize the model's ability to learn complex spatial patterns and object boundaries while maintaining training efficiency [47].

To evaluate detection performance under different phenological and lighting conditions, five independent YOLOv11 models were fine-tuned, each corresponding to a specific experimental scenario. Four of them were trained separately using images from the Pre-bloom Day, Pre-bloom Night, Fruit-set Day, and Fruit-set Night subsets. A fifth model, referred to as the Combined model, was trained on a pooled dataset including all conditions. This strategy enabled a comparative analysis of condition-specific specialization versus generalization capacity across heterogeneous scenarios.

Each model used the same hyper-parameters listed above, with weights re-initialized from the base checkpoint provided by Ultralytics, which is pretrained on the COCO 2017 dataset. Training was conducted independently for each case, and performance was evaluated using the corresponding held-out test subset to ensure an unbiased assessment.

2.7. Data Augmentation

To enhance model generalization and improve robustness under diverse field conditions, data augmentation techniques were applied during training. The YOLOv11 training pipeline includes a series of online augmentations that are automatically applied to each image batch in real time. These transformations help prevent overfitting and allow the model to learn from a wider range of scenarios without the need to expand the original dataset [48,49].

The main augmentation methods included the following:

- Mosaic augmentation, which combines four images into one during training to improve detection of small and densely clustered objects [50].
- Random horizontal flipping, to introduce variability in canopy orientation.
- Color space augmentations (hue, saturation, and brightness), to simulate differences in natural lighting.
- Affine transformations, such as scaling, translation, and rotation, to account for camera perspective and positioning variability in mobile acquisition.

These augmentations were performed dynamically at each training iteration using the default augmentation settings provided in the Ultralytics YOLOv11 repository. This ensured that the model was exposed to a constantly changing set of visual patterns, thereby enhancing its capacity to detect inflorescences or bunches across a wide range of vineyard conditions.

2.8. Statistical Tools for Model Evaluation

The statistical methods used for evaluation are described to support reproducibility. All analyses were conducted using Python, combining the scikit-learn (v1.4.2), NumPy (v1.26.4), and Seaborn (v0.13.2) libraries. Standard object detection metrics, consisting of mean Average Precision (mAP), mean Average Precision at IoU = 0.50 (mAP@0.50), precision, recall, and F1-score were computed at the image level. Specifically, mAP represents the average of the Average Precision values calculated across multiple Intersection over Union (IoU) thresholds, ranging from 0.50 to 0.95 in increments of 0.05, following the COCO evaluation protocol. The mAP@0.50 metric corresponds to the Average Precision computed at a single IoU threshold of 0.50, providing a more permissive measure of detection performance. Precision quantifies the proportion of true positive detections among all positive predictions, recall measures the proportion of true positives relative to all ground truth instances, and the F1-score is the harmonic mean of precision and recall, summarizing their balance in a single value. All metrics were calculated independently for the validation and test splits to ensure an unbiased assessment of model generalization. These metrics were based on Intersection over Union (IoU) thresholds consistent with common benchmarking practices.

In addition to detection metrics, further regression-based analyses were conducted to provide a deeper understanding of model performance. These included scatter plots comparing predicted versus annotated object counts, and the calculation of regression metrics such as coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE) by comparing the number of predicted and annotated object counts per image. This combination of classification and regression metrics enabled a comprehensive evaluation of model accuracy, consistency, and error distribution across varying phenological stages and lighting conditions.

Importantly, object detection models such as YOLO rely on two key threshold parameters during inference: the confidence threshold and the Intersection over Union (IoU) threshold.

- The confidence threshold defines the minimum probability required for the model to consider a detected object as valid. Lower thresholds may increase sensitivity but also introduce more false positives.
- The IoU threshold determines the minimum overlap between a predicted bounding box and a ground truth annotation for the prediction to be considered correct. A higher IoU enforces stricter spatial agreement but may penalize slightly off-center detections.

Optimizing these thresholds is essential to balance detection accuracy and counting reliability. In this analysis, we performed a grid search across five values for each parameter (0.10, 0.25, 0.30, 0.50, and 0.75), resulting in 25 threshold combinations per model. For each combination, we computed three regression metrics to quantify counting performance:

- Coefficient of determination (R^2): reflects the consistency of predictions relative to ground truth counts.
- Root mean square error (RMSE): quantifies the average magnitude of prediction errors.
- Mean absolute error (MAE): captures the average absolute deviation from ground truth.

3. Results and Discussion

3.1. Sample Distribution Analysis

To better understand the object distribution across conditions, a descriptive statistical analysis was performed on the number of annotated grapevine organs (inflorescences or bunches) per image. Table 3 summarizes the key distribution metrics for each subset corresponding to the four experimental conditions.

Table 3. Descriptive statistics of the number of inflorescences or bunches per image across experimental conditions.

Subset	Mean	Median	SD	Min	Max	CV (%)	Skew	Kurt
Pre-bloom Day	7.46	7	3.74	1	18	50.15	0.61	−0.19
Pre-bloom Night	5.29	5	2.7	0	15	51.07	0.79	0.43
Fruit-set Day	8.35	8	3.52	1	16	42.12	0.16	−0.93
Fruit-set Night	7.53	7	3.7	1	17	49.15	0.81	0.16

SD: standard deviation; Min: minimum; Max: maximum; CV: coefficient of variation; Skew: skewness; Kurt: kurtosis.

The results reveal that fruit-set images tend to contain a higher number of annotated objects per image, with mean values of 8.35 (day) and 7.53 (night). In contrast, pre-bloom subsets, especially under night-time conditions, show lower mean object counts and greater rightward skewness, indicating a larger proportion of images with few detectable inflorescences. The coefficient of variation (CV) is also highest in Pre-bloom Night (51.07%), reflecting greater relative variability. Skewness and kurtosis values suggest that most subsets exhibit light-tailed, moderately asymmetric distributions, with Fruit-set Day being closest to a normal distribution (kurtosis = −0.93, skewness = 0.16).

These patterns are further illustrated in the grouped histogram (Figure 5), which reveals distinct trends in the frequency of object counts across subsets grouped into five classes: 0–2, 3–5, 6–8, 9–11, and 12+. The distribution shows that most images from the Fruit-set Day condition fall into the intermediate to high object count classes (6–8 and 9–11), while Pre-bloom Night images are more concentrated in the 3–5 range.

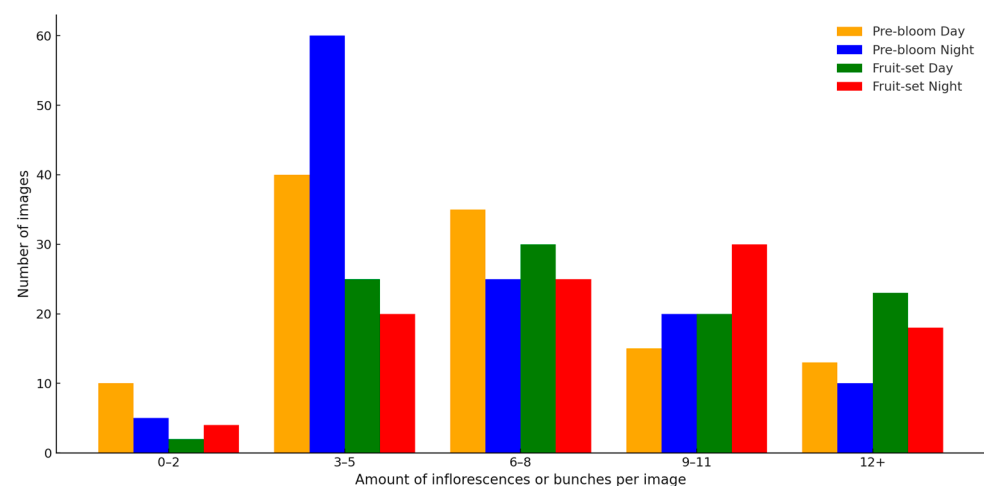


Figure 5. Grouped histogram showing the frequency of images grouped by number of annotated inflorescences or bunches per image across the four experimental conditions. Each group of bars corresponds to a range of object counts (0–2, 3–5, 6–8, 9–11, and 12+), with individual bars representing pre-bloom and fruit-set stages under day and night conditions.

In contrast, the Pre-bloom Night subset shows a relatively high proportion of images with fewer than six objects, confirming the lower mean and higher skewness previously

observed. Conversely, fruit-set images, particularly under daylight, show more images with high object counts, reflecting the increased visibility and compact structure of bunches at this developmental stage.

Despite these differences, all conditions display a similar overall range, typically between 1 and 17 objects per image, confirming that the dataset includes both sparse and dense canopy configurations. These distributional characteristics reinforce the need for training models on diverse image subsets, as the varying object densities and visibility levels observed across conditions (e.g., lower object counts and higher skewness in Pre-Bloom Night) may influence detection consistency and model generalization.

3.2. Performance of YOLOv11 Models Across Conditions

To evaluate detection performance under different phenological and lighting conditions, five object detection models were trained using the YOLOv11 architecture. Four models were trained independently using images from each specific subset: Pre-bloom Day, Pre-bloom Night, Fruit-set Day, and Fruit-set Night. In addition, a fifth model was trained on the Combined dataset (hereafter referred to as the “Combined” model), which integrated all images regardless of condition. This allowed us to compare both condition-specific specialization and generalization capacity across subsets.

The results of the validation phase are presented in Table 4, which shows the detection metrics for each model.

Table 4. Model performance on the validation dataset. Detection metrics were calculated for each model trained on a specific subset, as well as the combined model.

Metric	Pre-Bloom Day	Fruit-Set Day	Pre-Bloom Night	Fruit-Set Night	Combined
mAP50	0.50	0.81	0.73	0.83	0.71
mAP	0.20	0.52	0.37	0.55	0.38
Precision	0.64	0.82	0.75	0.85	0.82
Recall	0.53	0.78	0.69	0.78	0.63
F1 Score	0.58	0.80	0.72	0.81	0.72

The Fruit-set Night model achieved the highest validation scores overall, with an mAP50 of 0.83, F1 Score of 0.81, and strong balance between precision and recall. Similarly, the Fruit-set Day model performed very well (mAP50 = 0.81, F1 = 0.80), confirming that bunches at this stage are more easily detected, likely due to their greater size, compactness, and visibility.

The pre-bloom models showed lower performance, particularly the Pre-bloom Day model, which had the lowest values across all metrics (mAP = 0.20, F1 = 0.58). This is consistent with the visual complexity and high occlusion in inflorescences under daylight conditions. The Combined model achieved intermediate performance, showing strong precision (0.82) but lower recall (0.63), indicating it was more conservative but less sensitive across diverse conditions.

The performance on the test dataset (evaluating each model on new, unseen images) is summarized in Table 5.

The test results confirm the trends observed during validation. The fruit-set models remained the best-performing across all metrics, showing strong generalization ability. Notably, the Fruit-set Day model achieved the highest precision (0.89), while the Fruit-set Night model maintained high recall (0.69) and F1 score (0.76).

Table 5. Model performance on the test dataset. Detection metrics were obtained on the unseen test set for each trained model.

Metric	Pre-Bloom Day	Fruit-Set Day	Pre-Bloom Night	Fruit-Set Night	Combined
mAP50	0.50	0.81	0.68	0.82	0.66
mAP	0.24	0.50	0.35	0.54	0.36
Precision	0.66	0.89	0.71	0.84	0.75
Recall	0.52	0.68	0.72	0.69	0.63
F1 Score	0.58	0.77	0.71	0.76	0.68

The Pre-bloom Night model performed reasonably well in testing, with a similar F1 score (0.71) compared to validation, suggesting robustness despite lower image quality. In contrast, the Pre-bloom Day model yielded the lowest F1 score (0.58) and mAP (0.24), which may reflect the complexity of detecting small and partially occluded inflorescences under heterogeneous natural illumination.

The Combined model showed a more balanced performance across conditions, but with slightly lower recall and F1 compared to the best individual models. This suggests that while generalist models are useful in heterogeneous environments, condition-specific training may yield better performance in more focused applications.

The overall performance of the YOLOv11 models highlights their suitability for detecting grapevine organs across heterogeneous field conditions. While some performance drop was observed between validation and test sets, especially in the Pre-bloom Day condition, several models maintained high F1 scores and precision values, confirming their robustness under previously unseen scenarios. These results are in line with similar YOLO-based applications in viticulture, where generalization is often challenged by canopy variability and fluctuating illumination [34].

Moreover, training times remained relatively short across all models. The combined model required approximately 23.8 min to complete training, while the condition-specific models ranged from 9 to 12 min. This efficiency is particularly advantageous in precision agriculture scenarios that demand rapid model updates for in-season decision making. As noted in recent studies [38], YOLO-based models offer an effective compromise between detection accuracy and computational cost, making them practical tools for dynamic vineyard monitoring.

These results emphasize the relevance of selecting both the phenological stage and the lighting conditions when implementing automated yield estimation systems in vineyards. The superior performance of models trained on fruit-set images under daylight highlights this stage as the most favorable window for accurate bunch detection, which can be leveraged to support critical vineyard tasks such as early yield forecasting or selective thinning. Conversely, the challenges observed in detecting inflorescences during pre-bloom daylight conditions suggest that early-season monitoring may benefit from night-time imaging or targeted lighting control.

Beyond standard detection metrics, we also aimed to evaluate the quantitative accuracy of each model in estimating the number of grapevine organs (inflorescences or bunches) per image. In practical applications, such as early yield estimation or developmental monitoring, accurate object counts are often more critical than precise localization. To this end, we conducted a regression-based analysis comparing predicted object counts to manually annotated ground truth values.

These metrics were calculated for both the validation and test datasets to evaluate model performance and generalization. The best-performing threshold pair for each model was selected based on validation results, and the final outcomes for both datasets are summarized in Table 6.

Table 6. Optimal performance metrics (R^2 , RMSE, and MAE) for each model on validation and testing datasets, based on the best combination of confidence and IoU thresholds.

Model	IoU	Confidence	R^2		RMSE		MAE	
			Validation	Testing	Validation	Testing	Validation	Testing
Pre-bloom Day	0.10	0.10	0.77	0.72	1.76	2.02	1.36	1.61
Pre-bloom Night	0.30	0.50	0.88	0.76	0.96	1.09	0.69	0.82
Fruit-set Day	0.10	0.25	0.83	0.97	1.09	0.83	0.96	0.78
Fruit-set Night	0.50	0.50	0.80	0.79	1.68	1.97	1.18	1.77
Combined	0.25	0.30	0.81	0.71	1.50	2.08	1.11	1.39

The best validation performance was achieved by the Pre-bloom Night model, with an R^2 of 0.88 and the lowest error metrics (RMSE = 0.96, MAE = 0.69). The Fruit-set Day model also performed exceptionally well (R^2 = 0.83, RMSE = 1.09), confirming that bunch detection is more consistent and accurate at this developmental stage. Interestingly, even the Combined model achieved a solid R^2 of 0.81 with moderate errors, indicating that joint training across conditions did not drastically compromise counting performance.

The Pre-bloom Day model showed comparatively lower performance (R^2 = 0.77), which is consistent with its weaker object detection results and the challenges associated with detecting inflorescences under daylight. The Fruit-set Night model also presented slightly higher RMSE and MAE values despite a respectable R^2 , suggesting more variability in detection consistency.

The test results confirm the strong generalization capacity of the Fruit-set Day model, which achieved a remarkably high R^2 of 0.97, along with the lowest RMSE and MAE values of all models (0.83 and 0.78, respectively). This indicates that the model was able to count bunches with high accuracy even on previously unseen images.

The Pre-bloom Night model showed stable generalization, with an R^2 of 0.76 and low error values, consistent with its validation performance. Conversely, the Pre-bloom Day and Fruit-set Night models experienced higher errors in the test set, possibly due to greater variability in lighting or occlusion effects that were not fully captured during training.

The Combined model again delivered intermediate results (R^2 = 0.71), suggesting that although it does not outperform specialized models in any condition, it maintains acceptable accuracy across all environments, which may be advantageous in operational settings with unpredictable field conditions.

The regression-based evaluation provides critical insight into the reliability of automated yield estimation during early phenological stages. The outstanding performance of the Fruit-set Day model confirms that this stage, characterized by more compact and distinguishable bunches, is optimal for obtaining accurate object counts, which can be directly translated into early-season yield forecasts. Conversely, the reduced accuracy observed under daylight pre-bloom conditions highlights the challenges posed by inflorescence morphology and natural illumination variability. Notably, the consistent performance of the Pre-bloom Night model suggests that controlled lighting may serve as a practical solution to enhance early detection.

To better understand the quantitative reliability of the models beyond aggregated performance metrics, a regression-based scatter analysis was performed. This analysis compares the number of predicted objects per image (i.e., the number of detected inflorescences or bunches) against the manually annotated ground truth, offering a visual and statistical assessment of estimation accuracy. These scatter plots include a linear regression trendline, confidence intervals, and a 1:1 reference line representing perfect agreement.

This type of evaluation is particularly valuable in vineyard environments, where dense canopy structures, partial occlusions, and lighting variability can lead to systematic over-

or underestimation that may be masked by global metrics such as RMSE or F1 score. As noted in [51], scatter-based regression analysis provides critical insights for agricultural object detection tasks, especially when ground truth counts carry practical significance in early yield estimation.

Figure 6 presents the prediction results under daylight conditions for both validation and testing datasets. The left-hand panels show results for the pre-bloom models, while the right-hand panels correspond to the fruit-set models.

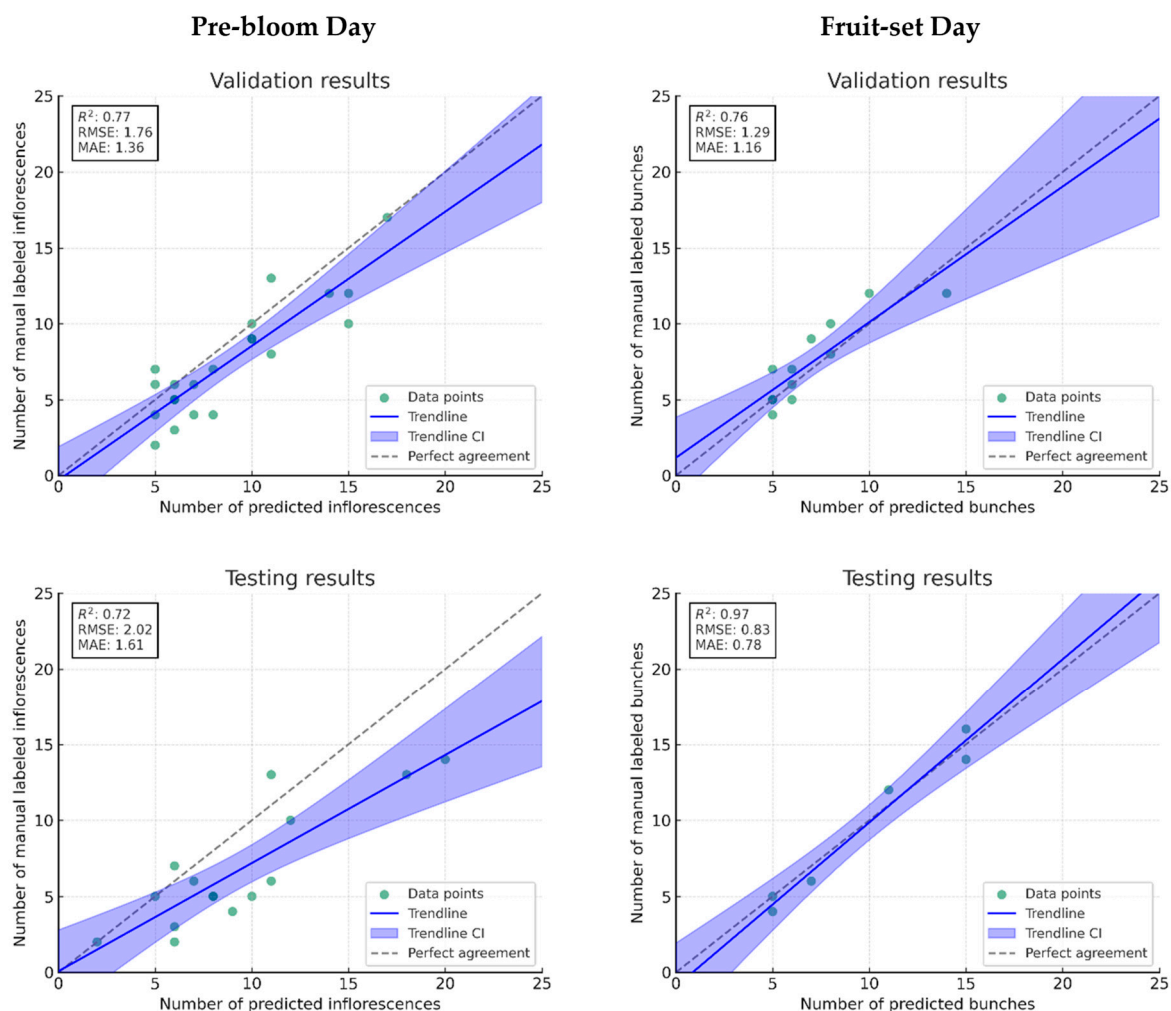


Figure 6. Regression-based scatter plots comparing predicted versus manually annotated inflorescences (**left**) and bunches (**right**) under daylight conditions. Results are shown separately for validation (**top**) and testing (**bottom**) datasets. The shaded area represents the confidence interval of the linear trendline.

During validation, the inflorescence model under daylight achieved an R^2 of 0.77, with an RMSE of 1.76 and MAE of 1.36. These results indicate moderate accuracy, although there is noticeable dispersion among individual data points. In contrast, the bunch detection model for fruit-set under daylight performed more consistently ($R^2 = 0.76$, RMSE = 1.29, MAE = 1.16), with tighter clustering around the regression line.

In the testing dataset, performance trends were preserved. The Fruit-set Day model showed excellent generalization, with an R^2 of 0.97, RMSE = 0.83, and MAE = 0.78, demonstrating high reliability across images with different bunch densities. The Pre-bloom Day model showed a lower generalization capacity ($R^2 = 0.72$), reflecting the known difficulty of inflorescence detection in natural lighting and variable canopy conditions.

Figure 7 shows the same regression analysis under night-time conditions, where artificial LED illumination was used to normalize lighting and minimize interference from adjacent rows.

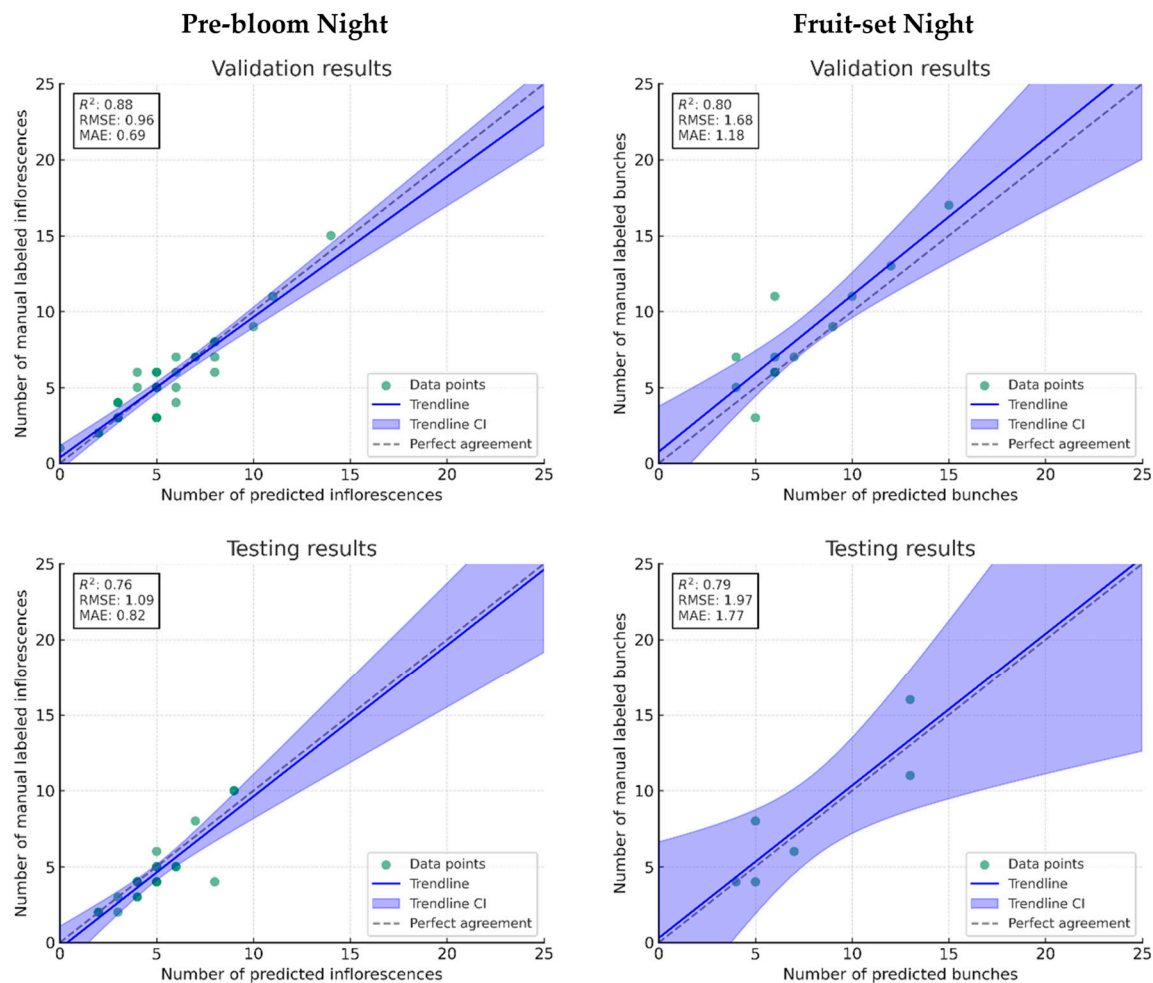


Figure 7. Regression-based scatter plots comparing predicted versus manually annotated inflorescences (**left**) and bunches (**right**) under night-time conditions. Results are shown separately for validation (**top**) and testing (**bottom**) datasets.

Both night-time models demonstrated robust validation performance. The Pre-bloom Night model achieved an R^2 of 0.88 and the lowest overall RMSE (0.96), highlighting the benefit of controlled illumination in early-stage detection. Similarly, the Fruit-set Night model achieved an R^2 of 0.80, with reasonable error margins (RMSE = 1.68, MAE = 1.18).

On the test dataset, the Pre-bloom Night model continued to perform well (R^2 = 0.76, RMSE = 1.09, MAE = 0.82), while the Fruit-set Night model experienced more variability (R^2 = 0.79, RMSE = 1.97), suggesting that bunch visibility under artificial lighting may vary depending on orientation or occlusion.

Overall, the scatter plots confirm the trends observed in the quantitative evaluation (Table 6). Most predictions align closely with the 1:1 reference line, especially in the mid-range of object counts. Greater dispersion is observed at the extremes, typically at high bunch densities, an expected phenomenon in agricultural detection systems with limited training examples for extreme cases [52]. However, the narrow confidence bands in most plots reinforce the consistency and reliability of the models under both day and night conditions.

These findings support the practicality of YOLOv11-based detection systems for early-stage yield estimation in viticulture. As emphasized in [53], regression-based indicators such as RMSE and MAE are often more informative in agricultural contexts than strict per-object matching, especially when the ultimate goal is to estimate counts at the canopy or plot level.

These results demonstrate the potential of deep learning models to provide accurate and scalable estimations of reproductive structures under field conditions. The high agreement between predicted and annotated counts, particularly under night-time illumination, reinforces the value of standardized imaging protocols for early-season vineyard monitoring.

3.3. Visual Assessment of Predictions

In addition to the quantitative metrics discussed above, a visual evaluation was conducted to qualitatively assess the detection performance of the YOLOv11 models. Representative test images were selected for each of the four experimental conditions (Pre-bloom Day, Pre-bloom Night, Fruit-set Day, and Fruit-set Night) to illustrate the model's capacity to localize grapevine organs in real-world vineyard scenarios. Figure 8 displays representative visual results for all four experimental conditions. Ground truth annotations are marked with red bounding boxes, while model predictions appear in blue.

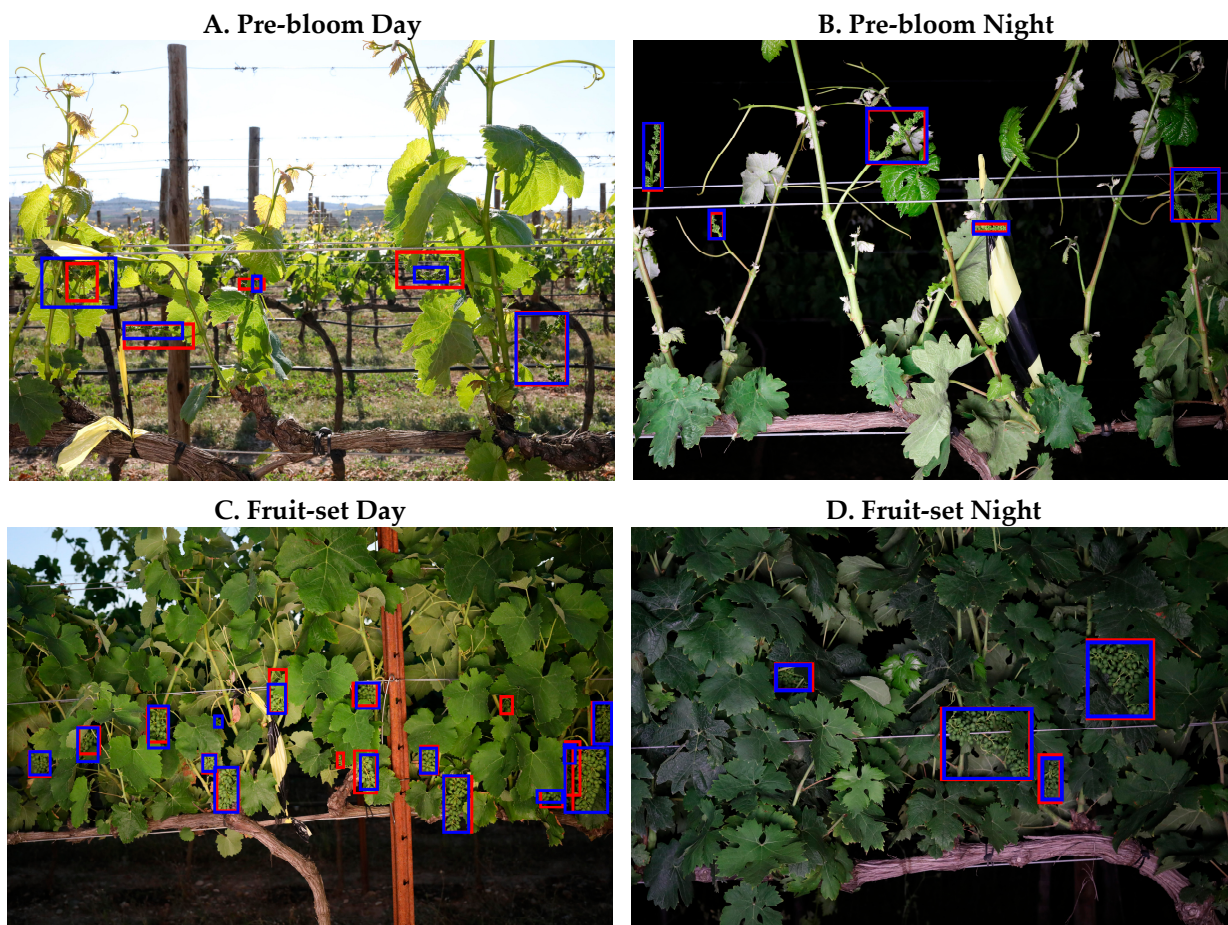


Figure 8. Visual examples of YOLOv11 detection results under all four experimental conditions. Each panel shows one test sample representative of the (A) Pre-bloom Day, (B) Pre-bloom Night, (C) Fruit-set Day, and (D) Fruit-set Night subsets. Red bounding boxes indicate manual annotations (ground truth), while blue boxes represent model predictions.

Visual inspection of the samples in Figure 8 confirms that the model performs robustly across all conditions, with particularly strong agreement observed in fruit-set images under

daylight. In these cases, the predicted bounding boxes show excellent spatial overlap with the manual annotations, reflecting the high precision and recall previously reported in both detection and regression analyses.

These observations are consistent with known challenges in agricultural computer vision. As highlighted in [54], natural vineyard environments often feature complex canopy structures and non-uniform backgrounds, which complicate object localization. In grapevines, these challenges are magnified during early phenological stages due to the small size and sparse distribution of floral structures. Similar findings were reported by [15], who noted a marked drop in detection reliability under such conditions, particularly when occlusion from leaves or trellis wires was present.

Several strategies have been proposed to mitigate these limitations. For instance, [51] advocates the use of 3D reconstruction and multi-angle data acquisition to improve object visibility in complex canopies. Similarly, [55] demonstrated that integrating RGB-D sensing and multi-object tracking significantly improves flower detection under occlusion-prone conditions. Although our current study focuses solely on 2D RGB images, the consistency of model predictions, especially under night-time illumination, suggests that controlled lighting can partially compensate for these challenges.

Dataset diversity also plays a key role in model robustness. The authors of [56] found that exposure to a wide range of canopy complexities and visibility conditions during training significantly improved generalization. This is in line with our findings: while the combined model did not reach the top performance scores of condition-specific models, it still achieved consistent results across all subsets, indicating resilience in heterogeneous environments.

Overall, this visual analysis reinforces the conclusions drawn from quantitative evaluations. YOLOv11 demonstrates a strong capacity to detect grapevine inflorescences and bunches under varying field conditions, though performance is notably influenced by lighting and phenological stage. Future work should explore depth-enhanced imaging, occlusion-aware labeling, and mobile multi-angle acquisition platforms to further improve detection robustness and deployment potential in precision viticulture.

The visual evaluation of detection outcomes highlights the practical relevance of deploying object detection models in diverse vineyard conditions. The ability to correctly identify grapevine inflorescences and bunches, even in complex canopy environments, is essential for tasks such as early yield estimation, selective thinning, and spatial variability analysis. In particular, the effectiveness observed under artificial night-time illumination offers promising avenues for routine scouting during low-light hours, minimizing operational disruption. Furthermore, the variable performance across phenological stages reinforces the need to align phenotyping strategies with crop development cycles, tailoring technological solutions to the specific challenges of each stage.

3.4. Optimal Detection Conditions

After evaluating multiple models under diverse experimental conditions, this section aims to synthesize all detection and regression results to determine the most reliable scenarios for detecting grapevine inflorescences and bunches. The five models developed, each trained under specific lighting and phenological conditions, plus a combined model, have been assessed using both object detection metrics (mAP, precision, recall, and F1 score) and regression-based analysis (R^2 , RMSE, and MAE), providing a comprehensive framework for identifying optimal detection strategies in vineyard environments.

3.4.1. Influence of Phenological Stage

Across all models and conditions, bunch detection at the fruit-set stage proved more robust and accurate than inflorescence detection during pre-bloom. This was reflected in

consistently higher F1 scores, lower RMSE and MAE values, and tighter clustering around the ideal regression line. For instance, the best fruit-set model (Fruit-set Day) achieved an RMSE of 0.83 and MAE of 0.78, whereas the best pre-bloom model (Pre-bloom Night) had higher errors (RMSE = 1.09, MAE = 0.82).

Among the models targeting pre-bloom detection, the Pre-bloom Night model consistently outperformed its daytime counterpart. It achieved the highest R^2 during validation (0.88) and maintained strong generalization in testing ($R^2 = 0.76$, RMSE = 1.09, MAE = 0.82), alongside solid detection metrics (F1 = 0.72 in validation; 0.71 in testing). The improved performance under night-time conditions can be attributed to the uniform artificial illumination, which reduces visual noise and enhances the visibility of small, low-contrast floral structures. These findings align with previous studies in vineyard environments, where night-time imaging with artificial lighting has consistently improved detection accuracy of early-stage reproductive structures due to enhanced contrast and reduced background interference [4,28,57].

In contrast, the Pre-bloom Day model yielded the lowest overall performance across nearly all indicators ($R^2 = 0.72$ in test, F1 = 0.58, RMSE = 2.02, MAE = 1.61). Several factors may explain this outcome, including increased occlusions, variable natural lighting, and the inherent morphological complexity of inflorescences at this stage. Additionally, images captured under daylight were often taken at closer range, occasionally including elements from adjacent rows or trellis structures within the frame. This may have introduced background clutter and reduced the visual separation between the target structures and their surroundings, subtly impacting the model's detection capabilities.

3.4.2. Influence of Illumination Conditions

In terms of lighting, the effect varied depending on the phenological stage. For inflorescences, night-time imaging clearly improved detection outcomes. As discussed above, the Pre-bloom Night model outperformed the Pre-bloom Day model across all metrics. The uniformity of artificial lighting and reduced visual noise appear to be key factors enabling better visibility and localization of small floral structures under night conditions.

In contrast, for bunch detection during the fruit-set stage, the best-performing model was Fruit-set Day. It achieved exceptional test performance ($R^2 = 0.97$, F1 = 0.77, precision = 0.89, RMSE = 0.83, MAE = 0.78), outperforming all other configurations. The validation metrics were also high ($R^2 = 0.83$, RMSE = 1.09, MAE = 0.96), indicating excellent model calibration and generalization. Interestingly, this result contrasts with the expectation that night-time imaging would yield better quality due to controlled illumination. However, it is plausible that the greater structural definition and size of developing bunches, combined with favorable natural daylight, enabled more effective detection during the day. This outcome is consistent with findings from previous work suggesting that the benefits of artificial illumination become less pronounced as target structures grow in size and contrast, allowing natural daylight to offer comparable or even superior detection conditions under certain phenological stages [40,51].

The Fruit-set Night model, while still effective, showed slightly lower regression and detection scores ($R^2 = 0.79$ in test, F1 = 0.76, RMSE = 1.97, MAE = 1.77) and more variability in high-density regions. These results suggest that while night-time lighting enhances performance in challenging early phenological stages, daylight may be sufficient or even advantageous in later stages when bunches are larger and more visible.

The model trained on the full dataset, combining all conditions, showed intermediate performance across all metrics. It did not outperform any of the condition-specific models, but still delivered acceptable accuracy in both validation ($R^2 = 0.80$, F1 = 0.72, RMSE = 1.5, MAE = 1.11) and testing ($R^2 = 0.71$, F1 = 0.68, RMSE = 2.08, MAE = 1.39). This model may

be suitable in operational scenarios where environmental variability is high and phenological uniformity cannot be ensured. Nevertheless, when conditions can be standardized, specialized models clearly provide superior results for both detection and counting tasks.

Therefore, the most reliable scenario for detecting grapevine reproductive structures varies depending on the developmental stage. For inflorescence detection during pre-bloom, the use of night-time artificial illumination provides the highest accuracy, likely due to improved contrast and reduced background interference. In contrast, for bunch detection at the fruit-set stage, daylight imaging yields superior results, benefiting from the larger and more defined morphology of bunches under favorable natural light. These findings suggest that Pre-bloom Night and Fruit-set Day represent the optimal acquisition conditions for early-stage monitoring, balancing detection precision and operational feasibility under field conditions.

Future research should aim to increase robustness and adaptability across diverse field scenarios. Key directions include maintaining a fixed camera-to-canopy distance, avoiding the inclusion of adjacent rows to reduce visual clutter, and refining optical settings to improve depth separation in daylight images. Additionally, it will be important to assess detection performance without prior defoliation, expand data acquisition to multiple cultivars and phenological windows, and evaluate model transferability to other vineyard architectures and growing seasons. While the dataset was curated to ensure representation across cultivars, phenological stages, and illumination conditions, its size and balance remain constrained by the practical challenges of field acquisition. These aspects should be considered when interpreting the results and motivate future efforts toward expanding and further homogenizing the dataset.

Future work should focus on expanding the dataset to enhance model generalization and balance across conditions, particularly in underrepresented scenarios such as night-time fruit-set. Increasing the number of annotated images per condition and including a wider diversity of cultivars, vineyard layouts, and lighting environments will improve the robustness and applicability of the models. In addition, the proposed approach could be extended to predict harvest yield by integrating morphological descriptors with empirical models that relate these traits to final weight. This would allow early-stage estimations of yield potential based on visual organ detection. From an applied perspective, further steps are needed to integrate this method into vineyard management systems. These may include coupling the detection algorithm with georeferenced imaging platforms (e.g., UAVs or tractor-mounted systems), incorporating outputs into spatial decision support tools, and validating performance under operational field conditions.

4. Conclusions

This study presents a robust deep learning-based approach for detecting grapevine inflorescences and bunches at early developmental stages under real field conditions. The methodology relies on YOLOv11 models trained on RGB images acquired across two phenological stages (pre-bloom and fruit-set) and under contrasting illumination scenarios (natural daylight and night-time artificial lighting).

The results demonstrate that the phenological stage is a critical factor, with bunch detection at fruit-set consistently outperforming inflorescence detection at pre-bloom. The greater size and structural definition of bunches during fruit-set enable more reliable identification, supporting the use of this stage for early yield forecasting. Illumination conditions also played a key role. While daylight imaging was sufficient for fruit-set detection, artificial night-time lighting substantially improved accuracy in pre-bloom images by providing homogeneous illumination and reducing visual interference. These

findings suggest that the combination of fruit-set stage and daylight is optimal for bunch detection, while pre-bloom detection benefits from night-time acquisition.

Beyond detection metrics, this work highlights the practical potential of integrating early-stage organ detection into vineyard management workflows. The ability to estimate yield months in advance enables timely decisions regarding crop thinning, irrigation scheduling, and harvest logistics. These early interventions contribute to more sustainable, efficient, and data-driven viticulture practices.

Author Contributions: Conceptualization, R.Í. and J.T.; methodology, R.Í. and J.T.; software, I.B. and I.H.; validation, R.Í., C.P.-E., and S.G.; formal analysis, R.Í.; investigation, R.Í.; resources, E.M.-C.; data curation, R.Í.; writing—original draft preparation, R.Í.; writing—review and editing, R.Í., C.P.-E., S.G., and J.T.; visualization, R.Í.; supervision, S.G. and J.T.; project administration, J.T.; funding acquisition, J.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the FPI Grant 591/2021 from the Universidad de La Rioja and the Gobierno de La Rioja.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The data presented in this study are available upon request from the corresponding author.

Acknowledgments: The authors also would like to acknowledge South African WINE for their collaboration and technical support through the project “Establishment of the Technical and Scientific Bases for AI Applications in Wine Production: A Case Study on Viticulture, Yield, and Phenology AI Models”.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Laurent, C.; Oger, B.; Taylor, J.A.; Scholasch, T.; Metay, A.; Tisseyre, B. A review of the issues, methods and perspectives for yield estimation, prediction and forecasting in viticulture. *Eur. J. Agron.* **2021**, *130*, 126339. [\[CrossRef\]](#)
2. Martin, S.; Dunstone, R.; Dunn, G. *How to Forecast Wine Grape Deliveries Using Grape Forecaster Excel Workbook Version 7*; Department of Primary Industries: Victoria, Australia, 2003.
3. Santos, T.T.; De Souza, L.L.; dos Santos, A.A.; Avila, S. Grape detection, segmentation, and tracking using deep neural networks and three-dimensional association. *Comput. Electron. Agric.* **2020**, *170*, 105247. [\[CrossRef\]](#)
4. Mohimont, L.; Alin, F.; Rondeau, M.; Gaveau, N.; Steffanel, L.A. Computer vision and deep learning for precision viticulture. *Agronomy* **2022**, *12*, 2463. [\[CrossRef\]](#)
5. Dunn, G.M.; Martin, S.R. The current status of crop forecasting in the Australian wine industry. In Proceedings of the ASVO Seminar Series: Grapegrowing at the Edge, Barossa Valley, South Australia, 10 July 2003.
6. Wohlfahrt, Y.; Collins, C.; Stoll, M. Grapevine bud fertility under conditions of elevated carbon dioxide: This article is published in cooperation with the 21th GIESCO International Meeting, June 23–28 2019, Thessaloniki, Greece. Guests editors: Stefanos Koundouras and Laurent Torregrosa. *Oeno One* **2019**, *53*, 279–288. [\[CrossRef\]](#)
7. Nuske, S.; Wilshusen, K.; Achar, S.; Yoder, L.; Narasimhan, S.; Singh, S. Automated visual yield estimation in vineyards. *J. Field Robot.* **2014**, *31*, 837–860. [\[CrossRef\]](#)
8. Khokher, M.R.; Liao, Q.; Smith, A.L.; Sun, C.; Mackenzie, D.; Thomas, M.R.; Wang, D.; Edwards, E.J. Early yield estimation in viticulture based on grapevine inflorescence detection and counting in videos. *IEEE Access* **2023**, *11*, 37790–37808. [\[CrossRef\]](#)
9. Palacios, F.; Diago, M.P.; Melo-Pinto, P.; Tardaguila, J. Early yield prediction in different grapevine varieties using computer vision and machine learning. *Precis. Agric.* **2023**, *24*, 407–435. [\[CrossRef\]](#)
10. Moreira, G.; dos Santos, F.N.; Cunha, M. Grapevine inflorescence segmentation and flower estimation based on Computer Vision techniques for early yield assessment. *Smart Agric. Technol.* **2025**, *10*, 100690. [\[CrossRef\]](#)
11. Aquino, A.; Millan, B.; Diago, M.-P.; Tardaguila, J. Automated early yield prediction in vineyards from on-the-go image acquisition. *Comput. Electron. Agric.* **2018**, *144*, 26–36. [\[CrossRef\]](#)
12. Anderson, N.T.; Walsh, K.B.; Wulfsohn, D. Technologies for forecasting tree fruit load and harvest timing—From ground, sky and time. *Agronomy* **2021**, *11*, 1409. [\[CrossRef\]](#)
13. Tsouros, D.C.; Bibi, S.; Sarigiannidis, P.G. A review on UAV-based applications for precision agriculture. *Information* **2019**, *10*, 349. [\[CrossRef\]](#)

14. Cubero, S.; Aleixos, N.; Moltó, E.; Gómez-Sanchis, J.; Blasco, J. Advances in machine vision applications for automatic inspection and quality evaluation of fruits and vegetables. *Food Bioprocess Technol.* **2011**, *4*, 487–504. [\[CrossRef\]](#)
15. Íñiguez, R.; Palacios, F.; Barrio, I.; Hernández, I.; Gutiérrez, S.; Tardaguila, J. Impact of leaf occlusions on yield assessment by computer vision in commercial vineyards. *Agronomy* **2021**, *11*, 1003. [\[CrossRef\]](#)
16. Liu, S.; Cossell, S.; Tang, J.; Dunn, G.; Whitty, M. A computer vision system for early stage grape yield estimation based on shoot detection. *Comput. Electron. Agric.* **2017**, *137*, 88–101. [\[CrossRef\]](#)
17. Parr, B.; Legg, M.; Alam, F. Analysis of depth cameras for proximal sensing of grapes. *Sensors* **2022**, *22*, 4179. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Tardif, M. Proximal Sensing and Neural Network Processes to Assist in Diagnosis of Multi-Symptom Grapevine Diseases. Ph.D. Thesis, Université de Bordeaux, Pessac, France, 2023.
19. Hu, C.; Sapkota, B.B.; Thomasson, J.A.; Bagavathiannan, M.V. Influence of image quality and light consistency on the performance of convolutional neural networks for weed mapping. *Remote Sens.* **2021**, *13*, 2140. [\[CrossRef\]](#)
20. Arad, B.; Kurtser, P.; Barnea, E.; Harel, B.; Edan, Y.; Ben-Shahar, O. Controlled lighting and illumination-independent target detection for real-time cost-efficient applications. The case study of sweet pepper robotic harvesting. *Sensors* **2019**, *19*, 1390.
21. Fu, H.; Zhao, X.; Tan, H.; Zheng, S.; Zhai, C.; Chen, L. Effective methods for mitigate the impact of light occlusion on the accuracy of online cabbage recognition in open fields. *Artif. Intell. Agric.* **2025**, *15*, 449–458. [\[CrossRef\]](#)
22. Silwal, A.; Parhar, T.; Yandun, F.; Baweja, H.; Kantor, G. A robust illumination-invariant camera system for agricultural applications. In Proceedings of the 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Prague, Czech Republic, 27 September–1 October 2021; IEEE: Piscataway, NJ, USA.
23. Kurtulmus, F.; Lee, W.S.; Vardar, A. Immature peach detection in colour images acquired in natural illumination conditions using statistical classifiers and neural network. *Precis. Agric.* **2014**, *15*, 57–79. [\[CrossRef\]](#)
24. Dhanush, G.; Khatri, N.; Kumar, S.; Shukla, P.K. A comprehensive review of machine vision systems and artificial intelligence algorithms for the detection and harvesting of agricultural produce. *Sci. Afr.* **2023**, *21*, e01798. [\[CrossRef\]](#)
25. Bruni, V.; Dominijanni, G.; Vitulano, D.; Ramella, G. A perception-guided CNN for grape bunch detection. *Math. Comput. Simul.* **2025**, *230*, 111–130. [\[CrossRef\]](#)
26. Mohimont, L.; Hollard, L.; Steffanel, L.A. Smart-Viticulture and Deep Learning: Challenges and Recent Developments on Yield Prediction. In *Smart Life and Smart Life Engineering: Current State and Future Vision*; Springer: Cham, Switzerland, 2025; pp. 187–207.
27. Kamilaris, A.; Prenafeta-Boldu, F. Deep learning in agri-culture: A survey. *Comput. Electron. Agric.* **2018**, *147*, 70–90. [\[CrossRef\]](#)
28. Palacios, F.; Bueno, G.; Salido, J.; Diago, M.P.; Hernández, I.; Tardaguila, J. Automated grapevine flower detection and quantification method based on computer vision and deep learning from on-the-go imaging using a mobile sensing platform under field conditions. *Comput. Electron. Agric.* **2020**, *178*, 105796. [\[CrossRef\]](#)
29. Palacios, F.; Melo-Pinto, P.; Diago, M.P.; Tardaguila, J. Deep learning and computer vision for assessing the number of actual berries in commercial vineyards. *Biosyst. Eng.* **2022**, *218*, 175–188. [\[CrossRef\]](#)
30. Hernández, I.; Gutiérrez, S.; Barrio, I.; Íñiguez, R.; Tardaguila, J. In-field disease symptom detection and localisation using explainable deep learning: Use case for downy mildew in grapevine. *Comput. Electron. Agric.* **2024**, *226*, 109478. [\[CrossRef\]](#)
31. Zhu, J.; Cheng, M.; Wang, Q.; Yuan, H.; Cai, Z. Grape leaf black rot detection based on super-resolution image enhancement and deep learning. *Front. Plant Sci.* **2021**, *12*, 695749. [\[CrossRef\]](#)
32. Xie, X.; Ma, Y.; Liu, B.; He, J.; Li, S.; Wang, H. A deep-learning-based real-time detector for grape leaf diseases using improved convolutional neural networks. *Front. Plant Sci.* **2020**, *11*, 751. [\[CrossRef\]](#)
33. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
34. Sozzi, M.; Cantalamessa, S.; Cogato, A.; Kayad, A.; Marinello, F. Automatic bunch detection in white grape varieties using YOLOv3, YOLOv4, and YOLOv5 deep learning algorithms. *Agronomy* **2022**, *12*, 319. [\[CrossRef\]](#)
35. Guo, C.; Zheng, S.; Cheng, G.; Zhang, Y.; Ding, J. An improved YOLO v4 used for grape detection in unstructured environment. *Front. Plant Sci.* **2023**, *14*, 1209910. [\[CrossRef\]](#)
36. Zhang, T.-Y.; Li, J.; Chai, J.; Zhao, Z.-Q.; Tian, W.-D. Improved yolov5 network with attention and context for small object detection. In *Intelligent Computing Methodologies*; Springer: Cham, Switzerland, 2022.
37. Khanam, R.; Hussain, M. Yolov11: An overview of the key architectural enhancements. *arXiv* **2014**, arXiv:2410.17725.
38. Badeka, E.; Karapatzak, E.; Karampatea, A.; Bouloumpasi, E.; Kalathas, I.; Lytridis, C.; Tziolas, E.; Tsakalidou, V.N.; Kaburlasos, V.G. A deep learning approach for precision viticulture, assessing grape maturity via YOLOv7. *Sensors* **2023**, *23*, 8126. [\[CrossRef\]](#) [\[PubMed\]](#)
39. Aguiar, A.S.; Magalhães, S.A.; Dos Santos, F.N.; Castro, L.; Pinho, T.; Valente, J.; Martins, R.; Boaventura-Cunha, J. Grape bunch detection at different growth stages using deep learning quantized models. *Agronomy* **2021**, *11*, 1890. [\[CrossRef\]](#)
40. Barriguinha, A.; de Castro Neto, M.; Gil, A. Vineyard yield estimation, prediction, and forecasting: A systematic literature review. *Agronomy* **2021**, *11*, 1789. [\[CrossRef\]](#)

41. Schieck, M.; Krajsic, P.; Loos, F.; Hussein, A.; Franczyk, B.; Kozierkiewicz, A.; Pietranik, M. Comparison of deep learning methods for grapevine growth stage recognition. *Comput. Electron. Agric.* **2023**, *211*, 107944. [CrossRef]
42. Baillod, M.; Baggiolini, M. Les stades repères de la vigne. In *Revue Suisse de Viticulture, Arboriculture, Horticulture*; National Library of Australia: Canberra, Australia, 1993; Volume 25.
43. Tzutalin. *LabelImg: Label Image Bounding Boxes for Object Detection*. GitHub repository, 2015. Available online: <https://github.com/tzutalin/labelImg> (accessed on 4 April 2025).
44. Ultralytics. YOLOv11 by Ultralytics. 2024. Available online: <https://github.com/ultralytics/> (accessed on 4 April 2025).
45. Howard, J.; Gugger, S. *Deep Learning for Coders with Fastai and PyTorch*; O'Reilly Media: Newton, MA, USA, 2020.
46. Loshchilov, I.; Hutter, F. Decoupled weight decay regularization. *arXiv* **2017**, arXiv:1711.05101.
47. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
48. Jocher, G.; Chaurasia, A.; Qiu, J. YOLO by Ultralytics. GitHub. 2023. Available online: <https://github.com/ultralytics/ultralytics/blob/main/CITATION.cff> (accessed on 4 April 2025).
49. Shorten, C.; Khoshgoftaar, T.M. A survey on image data augmentation for deep learning. *J. Big Data* **2019**, *6*, 60. [CrossRef]
50. Bochkovskiy, A.; Wang, C.-Y.; Liao, H.-Y.M. YOLOv4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
51. Ferro, M.V.; Catania, P. Technologies and innovative methods for precision viticulture: A comprehensive review. *Horticulturae* **2023**, *9*, 399. [CrossRef]
52. Íñiguez, R.; Gutiérrez, S.; Poblete-Echeverría, C.; Hernández, I.; Barrio, I.; Tardáguila, J. Deep learning modelling for non-invasive grape bunch detection under diverse occlusion conditions. *Comput. Electron. Agric.* **2024**, *226*, 109421. [CrossRef]
53. Ferrara, G.; Marcotuli, V.; Didonna, A.; Stellacci, A.M.; Palasciano, M.; Mazzeo, A. Ripeness prediction in table grape cultivars by using a portable NIR device. *Horticulturae* **2022**, *8*, 613. [CrossRef]
54. Yu, F.; Zhang, Q.; Xiao, J.; Ma, Y.; Wang, M.; Luan, R.; Liu, X.; Ping, Y.; Nie, Y.; Tao, Z. Progress in the application of cnn-based image classification and recognition in whole crop growth cycles. *Remote Sens.* **2023**, *15*, 2988. [CrossRef]
55. Tan, C.; Sun, J.; Paterson, A.H.; Song, H.; Li, C. Three-view cotton flower counting through multi-object tracking and RGB-D imagery. *Biosyst. Eng.* **2024**, *246*, 233–247. [CrossRef]
56. Miranda, J.C. Open Source Software and Benchmarking of Computer Vision Algorithms for Apple Fruit Detection, Fruit Sizing and Yield Prediction Using RGB-D Cameras. Ph.D. Thesis, University of Lleida, Lleida, Spain, 2024.
57. Gatou, P.; Tsiara, X.; Spitalas, A.; Sioutas, S.; Vonitsanos, G. Artificial Intelligence Techniques in Grapevine Research: A Comparative Study with an Extensive Review of Datasets, Diseases, and Techniques Evaluation. *Sensors* **2024**, *24*, 6211. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.