ORIGINAL RESEARCH **Open access**

Deep Learning-Based Grapevine Inflorescence Segmentation and Flower Detection for Early Yield Prediction in Precision Viticulture

Krzysztof Wiśniewski[1*], Małgorzata Nowak[1], Tomasz Adamczyk[2], Piotr Kowalski[3]

Abstract

Forecasting crop production is exceptionally valuable in modern grape growing for refining harvest planning and quality assurance. The count of inflorescences and individual blossoms per plant is a key determinant, serving as an early indicator that explains 85%-90% of seasonal yield fluctuations. This paper presents an advanced pipeline that integrates benchmarking across multiple state-of-the-art deep learning and traditional image processing techniques to automate inflorescence segmentation and solitary flower identification, thereby delivering accurate, early, and contact-free yield estimates in viticulture. The YOLOv8n architecture achieved the highest scores in inflorescence localization (F1-ScoreBox = 95.9%) and flower pinpointing (F1-Score = 91.4%), while YOLOv5n performed best at the segmentation task (F1-ScoreMask = 98.6%). Both architectures displayed tight alignment (R2 > 90.0%) between algorithm-detected and human-visible flower tallies within inflorescences. Statistical testing corroborated the system's dependability, with the YOLOv8 variant once again excelling by showing no meaningful discrepancy in error rates across distinct vine architectures and cultivars, affirming its suitability for broad deployment. The evidence confirms that these networks can greatly enhance the reliability of early yield forecasts, providing a non-intrusive, scalable tool for Precision Viticulture. The discoveries spotlight the promise of Computer Vision tools to enhance vineyard oversight, enabling smarter resource stewardship and higher-quality fruit.

Keywords Deep learning, Object detection, Digital phenotyping, Precision viticulture, Yield components

*Correspondence:
Krzysztof Wiśniewski
krzysztof.wisniewski@sggw.edu.pl
[1] Department of AI for Winter Wheat and Rye Growth Stage Classification, Faculty of Agriculture, Warsaw University of Life Sciences, Warsaw, Poland
[2] Department of Soil Moisture Prediction using Neural Networks, Faculty of Agriculture, Jagiellonian University, Krakow, Poland
[3] Department of Smart Disease Forecasting, Faculty of Agriculture, University of Agriculture Krakow, Krakow, Poland

Introduction

The international wine business is a trillion-dollar sector that encompasses a vast array of commercial undertakings and is an essential pillar of global economic momentum [1]. Grape producers encounter acute pressures—climatic (e.g., shifting climate norms, unpredictable meteorological patterns, compressed growing windows, aridity, heat spikes) and financial (e.g., workforce scarcity and escalating input expenditures) [2]—and to stay viable across domestic and overseas arenas, they must trim operating and production outlays while fine-tuning cultivation cycles to lift both output and grade [3].

A linchpin for securing peak results in grape cultivation is yield forecasting—projecting the amount and standard of fruit a plot will deliver each season. This predictive capacity not only guides choices regarding logistics and asset deployment

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC BY-NC-SA 4.0), which permits non-commercial use, distribution, reproduction, and adaptation in any medium or format, provided that appropriate credit is given to the original author(s) and source, a link to the license is provided, and any changes made are indicated. Adapted material must be distributed under the same license.

© 2023 The Author(s). | Published by Agro Intelligence Research Press

but also shapes vineyards' overall fiscal health and the quality of the bottled product [4]. Conventionally, estimation hinges on three principal yield constituents: bunch number per vine, berry number per bunch, and individual berry weight, each contributing partially to the year-on-year spatial inconsistency in harvest [5]. One of the earliest windows for prediction opens during spring flush, since the initiation of inflorescence initials (floral primordia) sets the upper limit for cluster numbers, while the blossoms formed on each inflorescence establish the potential berry count on that cluster [6]. Yet, because these jobs are performed by hand and appraised by naked-eye scrutiny, they become costly, sluggish, and susceptible to inaccuracies—repetitive and exacting, they breed operator weariness and depend excessively on personal training and dexterity.

In contemporary times, the confluence of viticulture with emergent innovations has catalysed transformative leaps, propelling more applied and forward-looking approaches that reconfigure the sector's contours [7]. The power to autonomously spot and tally grapevine yield components afield not only relieves the manual burden of traditional counting but also opens the door to instantaneous, evidence-led administration. Pinpointing and enumerating inflorescences and/or blossoms is a comparatively nascent theme, with its inaugural work surfacing in 2014 [8]. The prevailing and most widely used technique in this space is Computer Vision (CV), employed to distil salient cues from physical entities in still imagery or moving footage [9]. Since that seminal moment, varied strategies and tools have been deployed, as recapped in Section 2.

Nonetheless, the route toward a fully honed agricultural CV solution remains unfinished. Despite scholarly work suggesting the merits of migrating from conventional CV routines to those built on Deep Learning (DL) frameworks, the extant literature still harbours conspicuous gaps that merit scrutiny [10]. Farming's inherently tangled, unstructured settings erect formidable hurdles that can degrade the effectiveness of these tools. The within-class variability of the target yield constituents runs high, shifting markedly across distinct phenological phases, camera angles, and visual signatures [11]. Concerning inflorescences specifically, their hue often merges with the backdrop and other vegetative structures—canes, foliage, and shoots—which likewise provoke blockage and shadow difficulties. Their minuscule dimensions and cultivar-specific shape variations further reduce conspicuousness, and given their clustered growth habit, the task becomes more complex when the aim is to resolve individual blossoms or berries. Beyond these domain-intrinsic hurdles, hardware-linked constraints surface, particularly regarding the ability of edge appliances—handheld gadgets, terrestrial rovers, or aerial drones—to host robust compute modules. High-performing DL architectures demand substantial processing muscle, obligating premium hardware. This bottleneck principally lengthens inference runtimes since the full workflow—spanning input transformation, forward propagation, and output refinement—hinges on such hardware [11].

Furthermore, many inquiries in this niche are hampered by data scarcity and poor data quality. They frequently rest on meagre image collections that fail to mirror the breadth of conditions and circumstances met in genuine commercial growing. Investigators must strive to construct expansive, contemporaneous data repositories, ideally equipped with field-reference measurements such as blossom numbers per inflorescence or inflorescence counts per vine. More holistic appraisals must be pursued, obstacles must be isolated, and the variables that sway algorithmic performance and early production foresight in viticulture must be laid bare. Current detection paradigms and analytical pipelines may be suboptimal for the idiosyncratic demands of agrarian settings and must evolve to be tailored to the nuances of this field.

Thus, in view of the enumerated hurdles, the present inquiry sets out to scrutinize the capabilities of various compact, state-of-the-art CV architectures via a pipeline that permits delineation- and outline-based isolation of grapevine inflorescences, alongside the assessment of individual flower populations. The chief novel inputs delivered are the following:

• A wide-ranging survey of scholarly articles centred on grapevine inflorescence and single blossom identification and quantification;

© 2023 The Author(s). | Published by Agro Intelligence Research Press

• The assembly and public dissemination of a current dataset numbering 539 high-definition frames (8000x6000 px), purpose-built for inflorescence and flower spotting, amassed under authentic field circumstances across three vine growth intervals, annotated in the YOLO (You Only Look Once) schema (https://doi.org/10.5281/zenodo.8332171), a realm presently underserved in the scale and fidelity of labelled data encapsulating real-world vineyard conditions;

• A comparative evaluation framework expressly devised for benchmarking lightweight CV networks within precision grape-growing contexts, foregrounding real-time field deployability. Whereas antecedent studies generally benchmark models without factoring in agri-hardware ceilings, our framework gauges lightweight YOLO nano variants' aptitude across grapevine inflorescence spotting and mask extraction tasks;

• A side-by-side test of the most recent YOLO nano iterations against a classical Watershed-driven routine, furnishing a balanced contrast. This juxtaposition not only validates the potency of modern DL routes but also exposes scenarios where legacy methods may retain an edge, supplying a layered outlook to steer future model picking and field implementation for grape yield estimation;

• Granular understanding of how vine cultivar and inflorescence architecture modulate the fidelity of flower tally approximation, as such traits can inject pronounced within-class divergence. A statistical evaluation was undertaken to gauge the influence of grape variety and floral cluster geometry on blossom counts and to isolate specific attributes that can either enhance or degrade counting accuracy.

Cumulatively, these contributions broaden the analytical repertoire available to the research community and have tangible implications for strengthening early-season productivity projections in viticulture. By laying a foundation of robust data and penetrating analysis, this study represents a decisive step toward bringing this technology into practice, potentially arming growers with real-world tools for sharper, data-centric vineyard governance in the near future.

Related Works

The earliest technical attempts to forecast grapevine yield components in advance drew on more traditional image manipulation and analytical routines, with the primary goal of counting blossoms on each inflorescence. Standard practice at the time involved snapping pictures within controlled settings against artificial backgrounds, using inflorescences already clipped from the vine. A key signal used for blossom identification is the pattern formed when light glances off the surface of individual buds. Diago *et al.* [8] turned to the CIELAB colour system paired with an algorithm known as "H-maxima transform" to isolate local brightness peaks in the imagery—these being the spots where light rebounded from the buds—and managed to secure an R^2 of 76% when correlating their automatic flower counts against actual hand-counted figures. That same algorithmic core was subsequently enhanced by Millan *et al.* [12], who trialled it together with both a nonlinear and a linear model, each designed to work across multiple cultivars. The nonlinear variant outperformed its counterpart, posting an R^2 of 91%. Operating from the same conceptual base, Aquino *et al.* [13] developed the vitisFlower® app for Android, and their experimental results showed that more than 84% of the blossoms were collected. Aquino *et al.* [14] then improved the earlier method proposed by Diago *et al.* [8] by integrating mathematical morphology with pyramidal decomposition, resulting in a Precision score of 83%. Still, features tied to light reflection lack robustness and can vary widely across grape varieties, a weakness noted by Liu *et al.* [15]. These authors propose a setup in which vine stems and inflorescences are segmented using an active contour propagation routine, after which blossoms are extracted using the generic SURF (Speeded-Up Robust Features) keypoint detector. When compared with manual counts, the system achieved 84.3% accuracy. In a more recent contribution, Tello *et al.* [16] resorted to particle analysis driven by colour and shape cues for blossom detection, notching an R^2 of 91%. That said, a large share of these efforts side-step the challenge of segmenting against a genuine field background, making an artificial substitute a necessity [8, 12, 13, 15, 16]. Classical pipelines are thus principally constrained by the need to hand-pick suitable operators for tasks such as feature distillation, shape delineation, and class assignment, and they routinely require at least partial control over the environment [10]. Command of that sort is never total, however, since one cannot govern phenotypic traits of the plants themselves—for instance, the span and architecture an inflorescence can assume, or the diversity among the

© 2023 The Author(s). | Published by Agro Intelligence Research Press

flowers it bears. What is more, given how readily these older algorithms are thrown off by image resolution, target object pixel footprint, and shooting distance, such pathways lose their edge the moment they are tested under mismatched conditions.

Over the last several years, Deep Learning (DL) architectures have emerged as formidable tools, radically reshaping the trajectory of CV algorithm design thanks to their knack for making sense of and coping with complex visual environments [17]. Convolutional neural networks (CNNs) are widely adopted in agriculture [18], serving as the go-to DL paradigm and the current gold standard for object classification, spatial localization, and recognition. Zooming in on viticulture, two towering obstacles for CV-powered platforms are the ease with which yield components can be visually accessed and their conspicuousness. Inflorescences and bunches often endure occlusion levels well above 50% [19]. DL-powered models have opened the door to non-invasive predictive pipelines that function outside controlled environments, moving past simple blossom-per-cluster counting to encompass inflorescences-per-vine tallies as well, since these learners prove heartier and cope better when objects are hidden or overlap. Grimm *et al.* [20] developed a semantic segmentation DL framework that leveraged SegNet, U-Net, and FCN backbones to detect and count grapevine organs. When detecting isolated flower buds, the pipeline achieved an F1 Score of 87%. However, the pipeline's sheer computational appetite—chiefly on the FCN side—made wrangling large-resolution frames a genuine struggle. To enumerate both whole inflorescences and solitary blooms, Rudolph *et al.* [21] coupled an FCN-based segmenter for cluster-level localization with a Circular Hough Transform routine for pulling out individual flowers, netting an F1-Score of 75%. Even so, the margins of inflorescences routinely triggered false negatives, eroding repeatability. Khokher *et al.* [22] built a video corpus and assembled an inflorescence spotter based on ResNeXt, FPN, and RetinaNet, achieving an F1-Score of 80.5%. Occlusion from foliage, together with motion blur, occasionally degraded fidelity during clip processing. Another construct that has seen frequent deployment, above all for single-blossom counting, is Mask-RCNN. Rahim *et al.* [23] used the image collection assembled by Liu *et al.* [15] to test this model, achieving an F1-Score of 98% for blossom detection. Pahalawatta *et al.* [24], on the other hand, trained it to separate open from closed buds in grape inflorescence images, yielding an R^2 of 74.0%. Telling one floral stage from another called for extremely fine calibration, given how similar open and shut buds can appear.

To side-step the heavy lifting demanded by image processing chains and to inflate success rates, several research groups have opted to gather imagery after dark under lamps, which imposes greater uniformity and lets each vine be isolated—an especially handy move when the target is detecting and tallying inflorescences on a per-plant basis. Palacios *et al.* [25] measured SegNet's behaviour when paired with two distinct backbones (VGG16 and VGG19), with the latter hitting an F1-Score of 93% on the inflorescence segmentation front and 73% when singling out flowers. Keeping accuracy stable under shifting light and movement, however, meant retuning the model, most critically for in-field flower-level capture. Jaramillo *et al.* [26] fielded Faster R-CNN for inflorescence detection and enumeration. When pitched against a suite of object detectors, Faster R-CNN came out ahead with a mean Average Precision (mAP) of 64%, even though limited transferability across vineyard blocks was flagged. Rahim *et al.* [27] harnessed a Mask-RCNN setup to segment each inflorescence occurrence and to pick out solitary flowers, landing F1-Scores of 94% and 90% on the segmentation and single-flower tasks, respectively. Beyond yield projection, DL-driven inflorescence detection has found ancillary uses, including cluster thinning. Buayai *et al.* [28] described a novel end-to-end system for measuring inflorescence length that leans on the DetectoRS model for segmentation, notching an mAP of 99%. Du *et al.* [29] refined YOLOv5s to achieve an Average Precision (AP) of 96.3%, though heavy computation remained a bottleneck for broader, on-the-fly deployment.

Embedding DL within yield estimation workflows has catalysed marked leaps in how automatically and faithfully yield constituents can be captured. Still, obstacles such as outsized computational loads, fragility in the face of weather and lighting fluctuations, and uneven behaviour under operational conditions have constrained the scalability and trustworthiness of these techniques for heterogeneous, real-time vineyard use, underscoring the need for further tuning to enhance their practical value for forecasting harvests **(Table 1)**.

© 2023 The Author(s). | Published by Agro Intelligence Research Press

Table 1. State-of-the-art computer vision techniques for grapevine inflorescence and single flower detection and estimation.

Study focus	Operational conditions	Analytical framework	Detection/Segmentation performance	Estimation/Counting performance	Citation
Automated flower quantification	Controlled indoor environment; plain artificial background; destructive acquisition	Conventional computer vision techniques	Precision reached 92.90%, while recall attained 74.30%	Coefficient of determination (R^2) = 76.00%	Diago *et al.* [8]
Flower number assessment	Indoor imaging conditions with a synthetic uniform backdrop; destructive sampling	Image-processing-based workflow	Precision = 95.98%; Recall = 87.52%	Not reported	Aquino *et al.* [13]
Segmentation of inflorescences together with flower quantification	Outdoor vineyard environment; natural scenery; non-invasive acquisition	Vision-based image analytics	Flower precision achieved 83.38%, with a recall of 85.01%	R^2 value reached 97.78%	Aquino *et al.* [14]
Floral object counting	Indoor conditions using an artificial homogeneous background; destructive setup	Image analysis algorithms	Precision = 84.00%; Recall = 86.00%	R^2 = 96.00%	Millan *et al.* [12]
Flower detection and enumeration	Outdoor setting with a synthetic solid background; both destructive and non-destructive experiments	Classical image analysis	F1-score recorded at 93.00%	Overall accuracy = 84.00%	Liu *et al.* [15]
Flower quantity estimation	Outdoor conditions involving natural and artificial backgrounds; non-destructive measurements	Deep neural network models	F1-score = 86.70%	Not available	Grimm *et al.* [20]

© 2023 The Author(s). | Published by Agro Intelligence Research Press

Inflorescence segmentation coupled with flower counting	Outdoor environment with natural background; non-destructive imaging	Hybrid deep learning and image-analysis approach	mIoU for inflorescences = 87.60%; flower F1-score = 75.20%	Not specified	Rudolph *et al.* [21]
Flower estimation procedure	Indoor acquisition using an artificial solid backdrop; destructive methodology	Digital image processing	Not available	R^2 = 91.00%	Tello *et al.* [16]
Simultaneous inflorescence segmentation and blossom counting	Outdoor nighttime environment with artificial lighting and natural background; non-destructive	Integrated deep learning and image-analysis system	Inflorescence F1-score = 93.00%; flower F1-score = 73.00%	R^2 = 91.00%	Palacios *et al.* [25]
Flower segmentation and numerical estimation	Outdoor setup with an artificial plain background; non-destructive acquisition	Deep learning architecture	Not available	R^2 reached 81.50%	Pahalawatta *et al.* [24]
Inflorescence localization and counting	Outdoor nighttime scene under artificial illumination with natural surroundings; non-destructive	Deep learning strategy	mAP = 63.57%	R^2 = 73.95%	Jaramillo *et al.* [26]
Automated blossom counting	Outdoor experiments with an artificial solid backdrop; destructive and non-destructive configurations	Deep-learning-based method	Not available	R^2 = 96.00%	Rahim *et al.* [23]
Inflorescence region segmentation	Outdoor natural environment; non-destructive data collection	Deep learning framework	mAP achieved 99.00%	Not reported	Buayai *et al.* [28]
Inflorescence segmentation alongside	Outdoor nighttime conditions with artificial	Deep learning approach	F1-score = 94.30% for inflorescences and 90.30% for flowers	Not available	Rahim *et al.* [27]

© 2023 The Author(s). | Published by Agro Intelligence Research Press

flower enumeration	illumination and natural background; non-destructive				
Inflorescence recognition	Outdoor field conditions with natural background; non-destructive operation	A combination of deep learning and image-analysis methods	AP reached 96.30%	Not reported	Du *et al.* [29]
Detection of inflorescences	Outdoor natural-field environment; non-destructive imaging	Deep neural network technique	F1-score = 80.48%	R^2 = 86.00%	Khokher *et al.* [22]

[a]: The best results achieved, regardless of the grapevine variety or equipment surveyed.

Materials and Methods

Data acquisition and processing

A bank of RGB imagery depicting grapevine inflorescences was built up over the course of three separate growth intervals, following the expanded Biologische Bundesanstalt, Bundessortenamt und Chemische Industrie (BBCH) classification framework [30]:

• BBCH Code 53 - Inflorescences readily apparent to the eye

• BBCH Code 55 - Inflorescences beginning to swell, with individual buds packed tightly together

• BBCH Code 57 - Inflorescences at full expansion; blossoms starting to pull apart from one another

Testing the proposed detection and enumeration strategies required image capture within a research vineyard plot situated on the Agrarian Campus of Vairão, operated by the Faculty of Sciences of the University of Porto (41°24’12.2 “N 2°10’26.5 “W), within the boundaries of the Vinho Verde Demarcated Region. Pictures were snapped with the dual-lens setup on a Xiaomi Redmi Note 7, yielding 8000x6000-pixel images. The image library spans a set of indigenous Portuguese varieties alongside internationally cultivated ones, each identified through its Vitis International Variety Catalogue (VIVC) accession number (consult https://www.vivc.de, accessed on 1 August 2023):

• Red-skinned types:

o Touriga Nacional (VIVC-12594);

o Barroca (VIVC-12462);

o Tinta Roriz (VIVC-12350);

o Cabernet Sauvignon (VIVC-1929).

• White-skinned types:

© 2023 The Author(s). | Published by Agro Intelligence Research Press

o Viosinho (VIVC-13109);

o Trajadura (VIVC-12629).

Hue does not yet offer a reliable means of telling cultivars apart at these early growth milestones; nonetheless, both colour classes of grape were deliberately incorporated, chiefly because of the pronounced contrasts in how their inflorescences—and eventually their fruit clusters—are sized and shaped [31]. As a case in point, Touriga Nacional and Cabernet Sauvignon stand out for bearing dense, relatively slim clusters, whereas Trajadura tends toward a looser, broader architecture, signalling a less tightly packed arrangement of both blooms and later berries. Viosinho typically forms cone-like bunches that may produce wing-like appendages. In contrast, Tinta Barroca can yield conical clusters with wings or occasionally adopt a more cylindrical silhouette, a trait that speaks to its versatility across different agronomic conditions. Spanning such a breadth of physical variation renders it indispensable to include multiple cultivars so that the efficacy, precision, and resilience of the CV pipelines under scrutiny can be subjected to a genuinely demanding trial. What is more, the photographs were deliberately taken across a spectrum of natural illumination regimes and camera orientations, routinely capturing instances where inflorescences are either partially concealed or superimposed by assorted obstructions—originating from the vine's own anatomy (i.e., leaf canopy, shoots, trunk, or adjacent inflorescences) or from the support and training apparatus (i.e., cordon wire or foliage catch wires)—thereby weaving in a dense fabric of challenging, heterogeneous visual cues **(Figure 1)**. Altogether, the compiled repository amounts to 539 stills, deposited for unrestricted access on the Zenodo digital archive: https://doi.org/10.5281/zenodo.8332171.



Figure 1. Juxtaposition of two distinct inflorescence profiles, contrasting the compactness and floral spacing seen in the red cultivar Tinta Barroca (left) against those of the white cultivar Trajadura (right).

© 2023 The Author(s). | Published by Agro Intelligence Research Press

The sheer pixel density of these captures means that the CV models must wade through enormous volumes of raw data, a burden that swells computing overhead and stretches out runtime [32]. Pinpointing an acceptable compromise between preserving enough detail and keeping file sizes manageable thus becomes a pivotal concern. In response, the original dimensions were reduced to 1254x1672 pixels, a step that maintained the width-to-height ratio while removing little of the instructive content the algorithms require.

Upon completion of this downsampling operation, the entire collection was carved into three non-overlapping portions: (i) Training (60%); (ii) Validation (20%); and (iii) Test (20%). **Table 2** enumerates, for each slice, the number of images for each combination of cultivar and phenological window. To artificially swell the Training and Validation pools, the Python library Albumentations [33] was brought to bear, a toolkit purpose-built for image augmentation that synthesizes fresh exemplars from pre-existing ones. The distortion recipes were hand-picked to produce nothing but plausible vineyard-like imagery. They can be sorted into two broad categories: pixel-level modifications, which act exclusively on the raw input frame while preserving any associated annotation layers—masks, bounding boxes, or the like—exactly as they were, and spatial-level modifications, which warp the geometry of the input frame and its paired annotations in lockstep. **Table 3** lists the selected operations, and **Figure 2** displays them on a single representative photograph.

Table 2. Per-subset breakdown of image counts according to grapevine variety and growth stage.

Variety	Phenological stage	Total	Data sets		
			Test	Validation	Train
Touriga nacional	BBCH 53	78	2	2	19
	BBCH 55		7	4	17
	BBCH 57		4	3	20
Barroca	BBCH 53	99	2	6	18
	BBCH 55		7	8	21
	BBCH 57		7	10	20
Tinta roriz	BBCH 53	82	2	5	15
	BBCH 55		6	5	16
	BBCH 57		6	8	19
Cabernet sauvignon	BBCH 53	76	3	4	18
	BBCH 55		5	5	17
	BBCH 57		2	5	17
Viosinho	BBCH 53	95	7	5	23
	BBCH 55		5	3	18
	BBCH 57		9	6	19
Trajadura	BBCH 53	109	9	8	15
	BBCH 55		12	8	14
	BBCH 57		12	12	19
Total			107	107	325

© 2023 The Author(s). | Published by Agro Intelligence Research Press

		539			

Table 3. Rundown of the pixel-space and geometric augmentation routines enlisted to expand the training and validation subsets.

Transformation category	Augmentation technique	Reformulated description
Pixel-based augmentation	CLAHE	Performs Contrast Limited Adaptive Histogram Equalization on the original image to improve local contrast.
Pixel-based augmentation	Emboss	Generates an embossed representation of the image and blends it with the original input.
Pixel-based augmentation	Sharpen	Enhances edge details by sharpening the image and merging the output with the source image.
Pixel-based augmentation	ISONoise	Introduces sensor-like camera noise into the image.
Pixel-based augmentation	Random Fog	Artificially creates fog effects within the image scene.
Pixel-based augmentation	Spatter	Mimics visual obstruction caused by rain droplets or mud splashes on a camera lens.
Pixel-based augmentation	Random Brightness Contrast	Randomly adjusts image brightness and contrast.
Pixel-based augmentation	Blur	Applies blurring using kernels of randomly selected sizes.
Pixel-based augmentation	Gaussian Noise	Injects Gaussian-distributed noise into the input image.
Spatial-based augmentation	Horizontal Flip	Mirrors the image horizontally across the y-axis.
Spatial-based augmentation	Shift Scale Rotate	Resizes the image with a scaling factor of 0.7.
Spatial-based augmentation	Shift Scale Rotate	Enlarges the image using a scaling factor of 1.3.
Spatial-based augmentation	Shift Scale Rotate	Rotates the image clockwise by +15º.
Spatial-based augmentation	Shift Scale Rotate	Rotates the image counterclockwise by -15º.

© 2023 The Author(s). | Published by Agro Intelligence Research Press



Figure 2. A visual walkthrough of the augmentation routines administered to one sample image from the dataset.

Each of these routines was not only deployed solo; composite stacks were also assembled, with every geometric transform paired sequentially with every pixel-space transform, yielding 59 distinct augmentation pipelines per source image across the two targeted subsets. By the time this expansion phase concluded, the dataset had grown to 26,027 frames. The training and validation partitions swelled to 19,500 and 6,420 images, respectively, leaving the test partition at its original size of 107 untouched images.

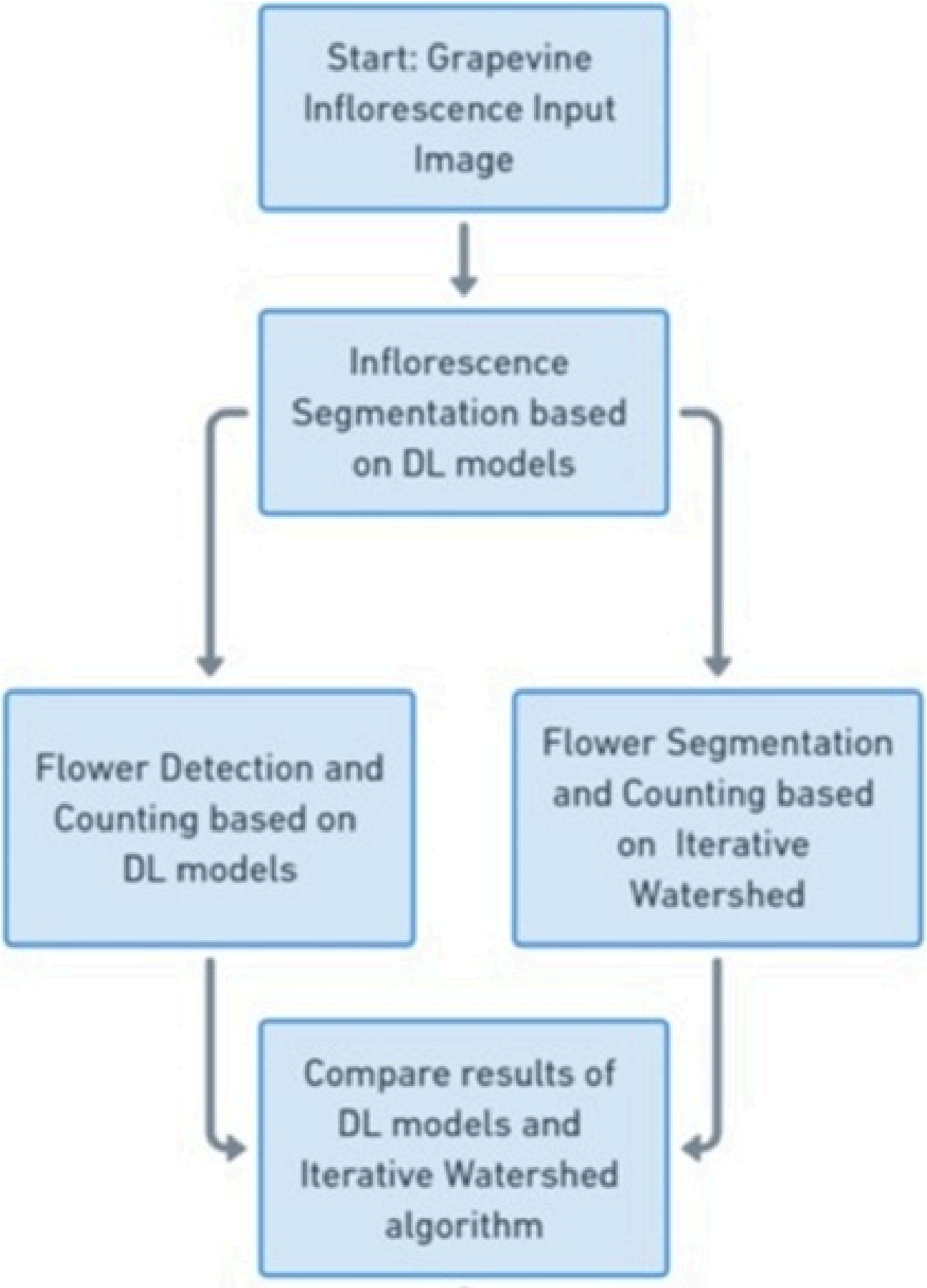
Grapevine inflorescence segmentation and flower counting framework

A hybrid processing chain blending DL architectures with time-tested image analysis procedures was engineered to first locate inflorescences and then gauge how many solitary flowers each one harbours **(Figure 3)**. The DL modules shoulder the responsibility of demarcating and isolating the region of interest (RoI)—specifically, the inflorescence itself—thus enabling a clean cut-away from everything else in the scene, discarding distracting clutter that might otherwise skew the intended outputs. During the follow-on stage, where the goal shifts to singling out and totalling independent blooms nested within that delimited area, the DL-driven candidates were compared with a classical iterative watershed routine from the image processing literature. This bifurcated evaluation strategy not only bolsters confidence in the floral tallies

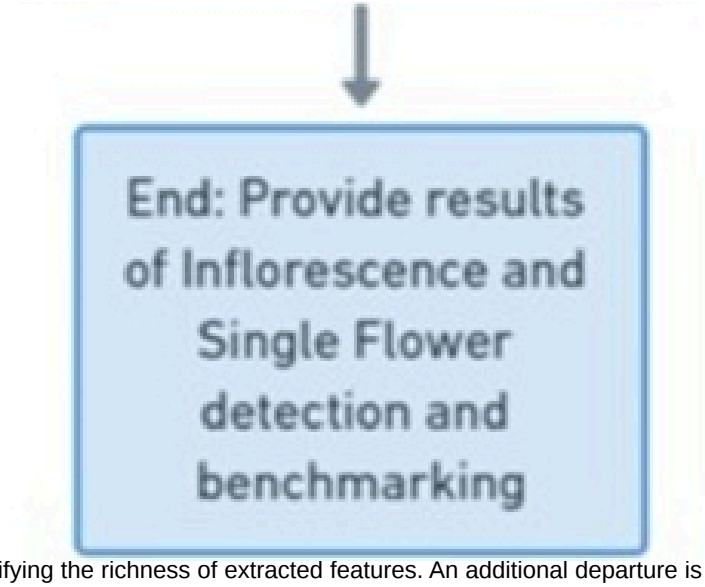
© 2023 The Author(s). | Published by Agro Intelligence Research Press

obtained but also reveals where contemporary CV tooling outshines its more established predecessors—and, just as tellingly, where the reverse may hold. Pitting these alternative philosophies against one another empowers researchers to select the fittest instrument for a given scenario, rebalance where computational or human effort is invested, and conceivably stitch together composite models that fuse the complementary advantages of both paradigms. Such an exhaustive comparative exercise ultimately becomes an engine of methodological progress, helping ensure that whatever techniques prevail are sturdy, capable of operating at scale, and genuinely suited to the messiness of on-farm deployment.

© 2023 The Author(s). | Published by Agro Intelligence Research Press



© 2023 The Author(s). | Published by Agro Intelligence Research Press



away duplication and amplifying the richness of extracted features. An additional departure is the adoption of a decoupled detection head, in which the twin objectives of category assignment and spatial regression are handled by separate branches, thereby sharpening both bounding-box precision and class discrimination [36]. The nano-sized, latency-optimized incarnations of each model were deliberately chosen to keep the computational burden light, with an eye toward eventual embedding within mobile or edge devices.

Initialization of both networks relied on weights pre-trained against Microsoft's COCO (Common Objects in Context) benchmark [37]; transfer learning then steered this generic knowledge toward the specialized job of detecting and masking grapevine inflorescences. Operating under a supervised regime meant the gathered imagery required manual labelling to provide ground truth references for both fitting and assessment. This markup effort was conducted inside the open-source Computer Vision Annotation Tool (CVAT) (see https://cvat.org/, accessed on 5 June 2024). Every annotation bundles a bounding rectangle delineating each object's spatial footprint, placement, and type, alongside a pixel-wise mask that binds every enclosed pixel to a designated class, thereby carving the RoI—in this instance, the floral clusters—apart from background clutter and non-target structures like peduncles **(Figure 4)**.

Figure 4. Hand-crafted groundtruth labels pairing bounding boxes with segmentation masks to cordon off the flower clusters.

© 2023 The Author(s). | Published by Agro Intelligence Research Press



Figure 4. Hand-crafted groundtruth labels pairing bounding boxes with segmentation masks to cordon off the flower clusters.

© 2023 The Author(s). | Published by Agro Intelligence Research Press

Training proceeded from pre-trained YOLO snapshots, advancing through 20 epochs with mini-batches of 16 frames and a learning rate of 0.01. Computation was handled by an NVIDIA GeForce RTX 4060 graphics processing unit (GPU) furnished with 8 gigabytes (GBs) of RAM.

Flower detection and counting based on DL models

Having walled off the inflorescence zones, the subsequent challenge lay in teasing apart the constituent blossoms packed within. The identical model pair deployed during the segmentation phase was enlisted anew, this time joined by two further entrants: YOLOv9t and YOLOv10n. The YOLOv9 formulation introduces two hallmark contributions: the integration of Vision Transformers (ViTs), which strengthen the capture of long-range spatial relationships, and the embedding of Programmable Gradient Information (PGI), a mechanism engineered to arrest the decay of informative signals as they propagate through deeper network strata. A further structural refinement is a Generalized Efficient Layer Aggregation Network (GELAN), which boosts parameter frugality and computational muscle, thereby making the architecture well-suited to detection tasks spanning diverse object scales [38]. YOLOv10 extends this trajectory by engineering an architectural pathway that renders the customary Non-Maximum Suppression (NMS) stage obsolete, thereby eliminating post-processing overhead and shaving inference time. Internally, it employs a twin-stream topology: one stream focuses on the meticulous isolation of individual targets, while the other handles broader detection, thereby raising overall correctness. The design further incorporates partial self-attention (PSA), negotiating a middle ground between globe-spanning feature awareness and leaner computational cost, a balance that underpins its standout real-time capability [39]. Individual blossoms were annotated by hand, each receiving its own bounding box **(Figure 5)**. The transfer learning recipe was reused, training the models on the same GPU with a batch size of 16 and a learning rate of 0.01. However, the training duration was substantially extended to 500 epochs.

© 2023 The Author(s). | Published by Agro Intelligence Research Press

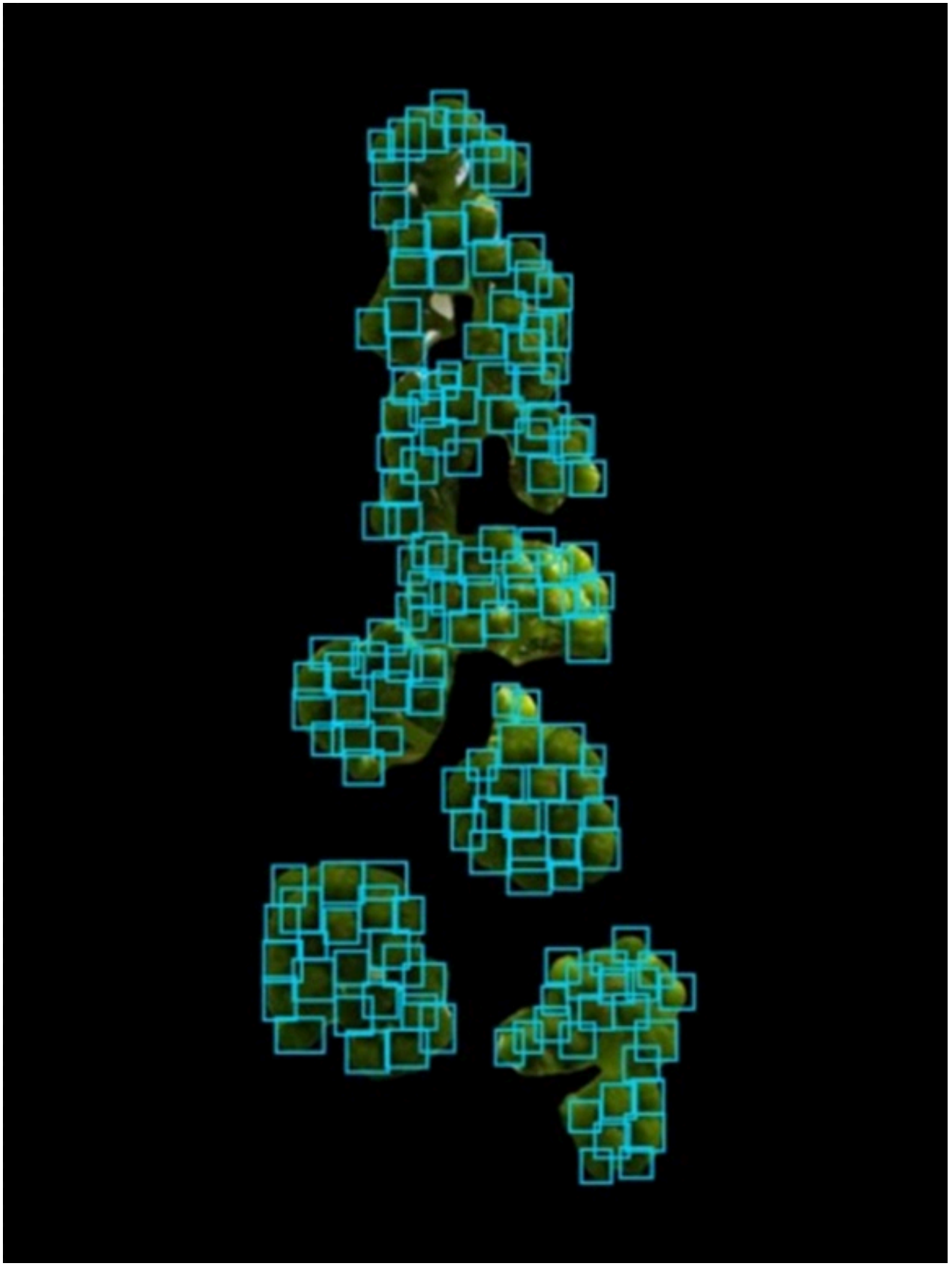


Figure 5. Manual groundtruth labelling of single flowers through bounding boxes.

© 2023 The Author(s). | Published by Agro Intelligence Research Press

Flower segmentation and counting based on iterative watershed

To serve as a reference point against the most sophisticated DL pipelines, blossom tallies were also generated using the "Marker-controlled Watershed" paradigm. This technique interprets an image as a topographic surface and carves it into catchment basins. The procedure first unearths local minima—the deepest pits in this intensity relief—and elevates them to seed markers. Typically, spurious fluctuations and segmentation artefacts are tamed by pre-filtering operations such as the "h-minima" transform, which squashes any local minimum whose depth fails to exceed a supplied threshold, h. Calibrating this parameter carries significant weight, as blossoms can exhibit markedly dissimilar visual signatures within a single frame, and a uniform h may fail to correctly designate the markers. Cognizant of this, the assembled processing chain incorporates an original routine dubbed "Iterative H-minima Based Marker-Controlled Watershed," implemented in MATLAB [40], which probes a succession of h levels to progressively home in on its seed markers. The sterling outcomes this algorithm had previously delivered for segmenting cell nuclei, as showcased by Koyuncu *et al.* [41], coupled with the morphological and spatial affinities between nuclei clusters and floral groupings, motivated its adaptation to the task of grapevine flower enumeration **(Figure 6)**.

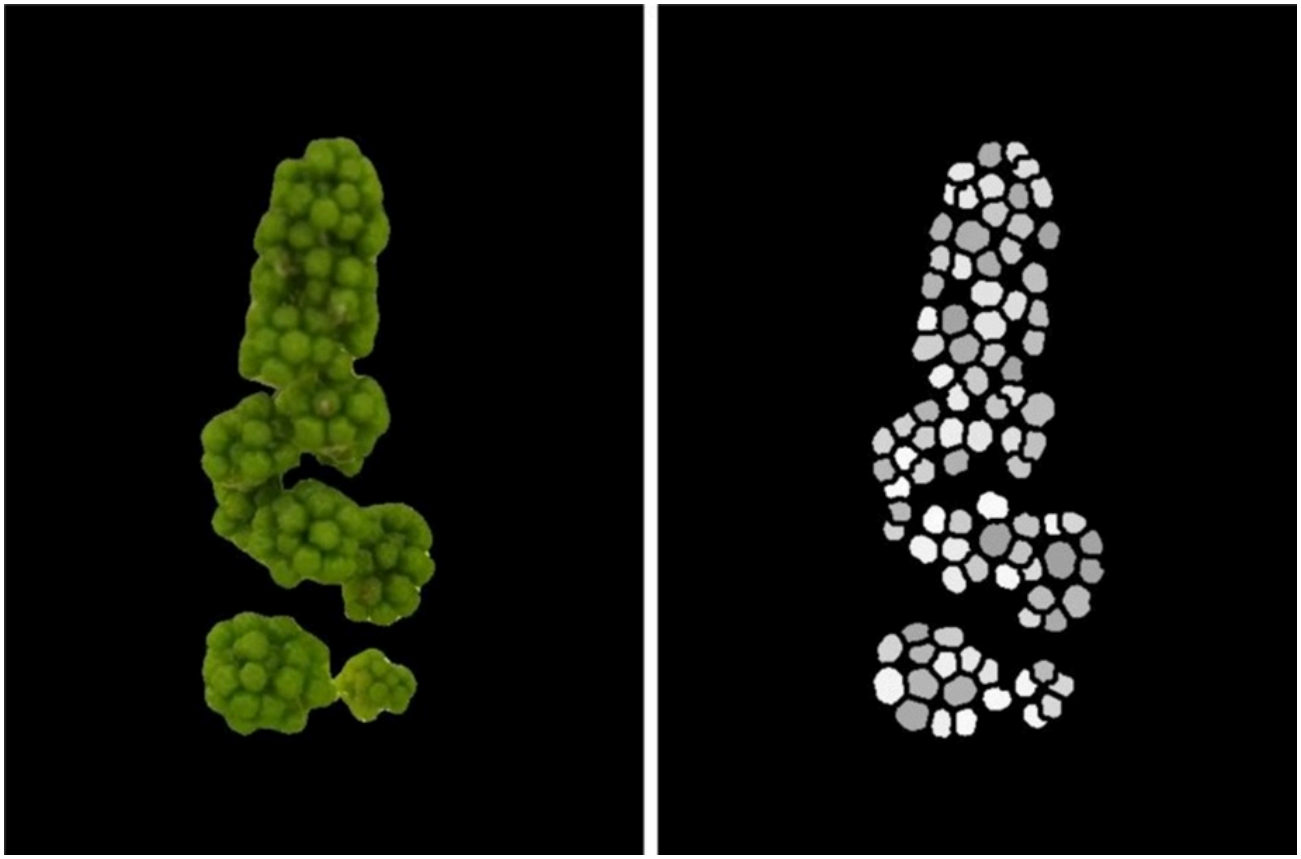


Figure 6. Solitary flower segmentation via the "Iterative H-minima Based Marker-Controlled Watershed" method.

The routine exposes four user-facing parameters:

• Tarea – an area cutoff employed to discard the most minuscule marker fragments during the marker-selection phase;

• Dsize – the radius of the disk-shaped structuring element leveraged to construct the topographic map and pinpoint the markers;

• α – an angular sweep defining the termini of an arc whose constituent pixels supply the halting rule for region growth;

© 2023 The Author(s). | Published by Agro Intelligence Research Press

• p – an upper bound on how many pixels a marker may annex before the stopping condition takes effect.

Groundtruthing was handled by annotating each flower with a keypoint situated at the centroid of the bounding box drawn during the prior labelling step. An early scoping analysis showed that the specific choices for α and p did not affect the final blossom count returned by the algorithm. Accordingly, the search space was narrowed to every combination drawn from: tarea = {5, 10, 20, 30}, dsize = {5, 7, 10, 13}, α = {30}, and p = {2}.

Model inference

A forecast is deemed on-target when the region of overlap between the predicted entity and its groundtruth counterpart is as large as possible, a yardstick conventionally captured by the Intersection over Union (IoU) ratio. The pool of groundtruth instances corresponds to the sum of True Positives and False Negatives (TP + FN), while the pool of model outputs equates to True Positives plus False Positives (TP + FP). Consequently, to benchmark all models along both segmentation and detection axes, three standard figures of merit were derived: Precision (1), which captures how well the model avoids false alarms; Recall (2), which captures how completely it retrieves the objects of interest; and the F1-Score (3), the harmonic mean that fuses Recall and Precision into a single summary statistic.

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

$$F1\text{-}Score = \frac{2TP}{2TP+FP+FN} \quad (3)$$

A critical lever when operationalizing DL-based predictors is the fine-tuning of the confidence cutoff. A cross-validation scheme was pursued for this purpose. The F1-Score was plotted across all possible confidence values (0% to 100% in 1% increments) and evaluated on the validation fold. The threshold yielding the maximal F1-Score was frozen, after which model performance was evaluated on the sequestered test set, applying an IoU floor of ≥ 80% for inflorescence segmentation and > 50% for single-flower detection. Regarding inference from the iterative Watershed pipeline, the complication arises that its outputs are segmentation masks, whereas the ground-truth blossoms were demarcated using bounding boxes. To bridge this mismatch, a keypoint was synthesized from the centre of each groundtruth bounding box, and the following scoring convention was adopted: any mask harbouring a groundtruth keypoint was classed as a TP, while masks devoid of one were counted as FP; groundtruth keypoints residing outside any mask were tagged as FN, and in those instances where a single mask subsumed several groundtruth keypoints, one was credited as TP and the excess were relegated to FN.

The fidelity with which the DL models could quantify flower numbers was appraised via the coefficient of determination (R^2) (4), the root mean square error (RMSE) (5), and the mean absolute percentage error (MAPE) (6), as specified in the equations that follow, where n denotes the count of test-set images, y the actual hand-verified flower tally, and $\hat{y}$ the model-estimated flower tally.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \quad (4)$$

© 2023 The Author(s). | Published by Agro Intelligence Research Press

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (5)$$

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \quad (6)$$

To supplement this examination, a Univariate Analysis of Variance (ANOVA) was used to quantify the sway exerted by the tested models and several categorical predictors on error distributions in the held-out test portion. Three principal axes of comparison were defined for the models **(Figure 7)**:

• Cultivar colour class:

1. White-fruited types;

2. Red-fruited types.

• Cluster architecture stratified by blossom quantity:

1. [0,140) flowers;

2. [140,175) flowers;

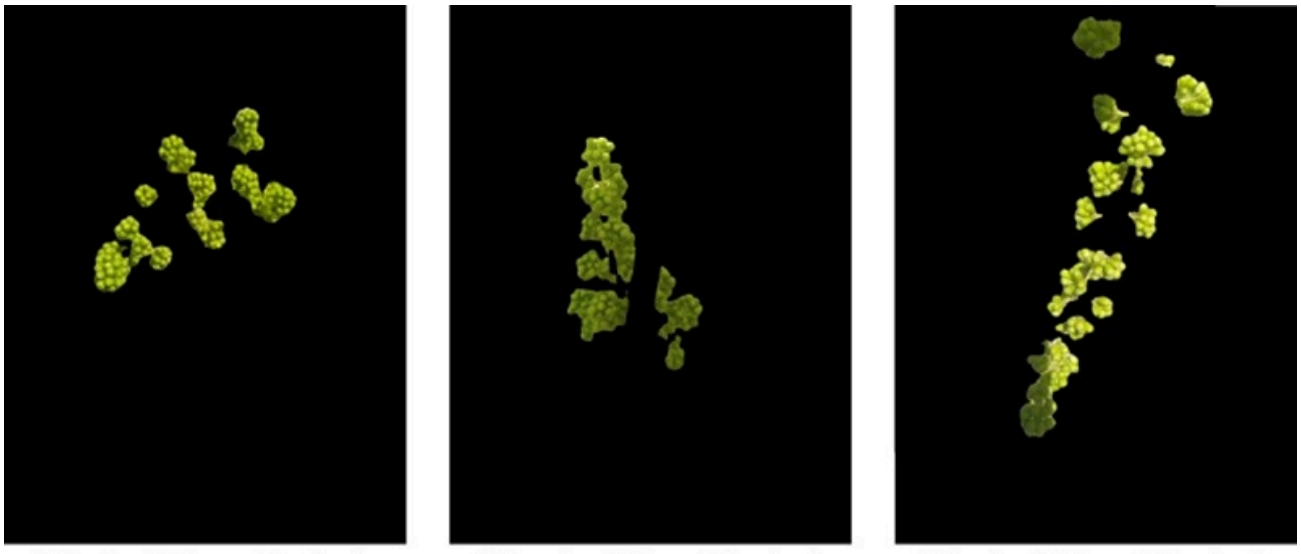
3. [175,∞) flowers.

• Cluster architecture stratified by longitudinal extent (along the bounding box's major axis):

1. [0,650) px;

2. [650,800) px;

3. [800,∞) px.



Red variety, 176 flowers, 629 px length | White variety, 135 flowers, 721 px length | Red variety, 172 flowers, 1307 px length

© 2023 The Author(s). | Published by Agro Intelligence Research Press

Figure 7. Visual contrast between the cultivar groupings and inflorescence properties factored into the statistical evaluation, drawing attention to how blossom abundance and cluster elongation diverge between red and white grape types.

A two-way analysis of variance (ANOVA) was executed, relying on the Fisher test for significance determination and, in instances where effects proved meaningful, the Duncan post hoc procedure ($\alpha = 0.05$) was subsequently applied to disentangle mean discrepancies in floral enumeration errors linked to model choice, cultivar grouping, or inflorescence morphology bin, together with their combined influences. Embracing this statistical framework enabled an exhaustive examination of the extent to which dissimilar inflorescence configurations and grape lineages condition model outputs, surfacing instructive regularities in algorithmic reliability across a broad range of realistic vineyard scenarios.

Results and Discussion

Inflorescence segmentation

Deriving meaningful performance readouts first required identifying the confidence operating points that maximize the F1-Score, before broader metric comparisons. Generally speaking, more stringent confidence thresholds lift Precision—the share of issued detections that hit true targets—yet they can simultaneously depress Recall, which gauges completeness in retrieving every object of interest. **Table 4** gathers the metric tallies. The confidence values shown were cherry-picked to strike a harmonious trade-off between Precision and Recall; both architectures achieved their peak F1 Scores at thresholds above 70%. In particular, YOLOv8n reached its apex F1-Score at a 73.0% threshold.

On the whole, both networks delivered impressive performances, with every single metric surpassing the 90% bar. YOLOv8n demonstrated a particular flair for object placement, posting an F1Box of 95.9%, while YOLOv5n proved superior at producing faithful segmentation masks, yielding an F1Mask of 98.6%. Another dimension that carries weight for time-critical deployment scenarios is runtime latency. Both architectures required less than 0.01 seconds to perform combined detection and masking of grapevine inflorescences, aligning with expectations from their deliberately minimalist model sizes.

Table 4. Detection and segmentation results for grapevine inflorescences on the test set under optimized confidence thresholds. (P = Precision; R = Recall; F1 = F1-Score).

Model	Confidence threshold	Speed (s)	$F1_{Mask}$	R_{Mask}	P_{Mask}	$F1_{Box}$	R_{Box}	P_{Box}
YOLOv5n	0.716	0.005	0.986	0.972	1.00	0.943	0.917	0.971
YOLOv8n	0.730	0.007	0.981	0.982	0.980	0.959	0.963	0.955

Delving deeply into test-set exemplars is critical for assembling a full portrait of model behavior, particularly for cataloguing breakdown patterns and pinpointing avenues for refinement. Even though both architectures post striking numbers, there are sporadic cases where extra tuning could yield dividends. For instance, on some occasions, entire flower clusters escaped detection. At the same time, in other cases, inflorescences that the model correctly flagged were absent from the ground truth because they appeared only partially within the frame boundaries **(Figure 8)**.

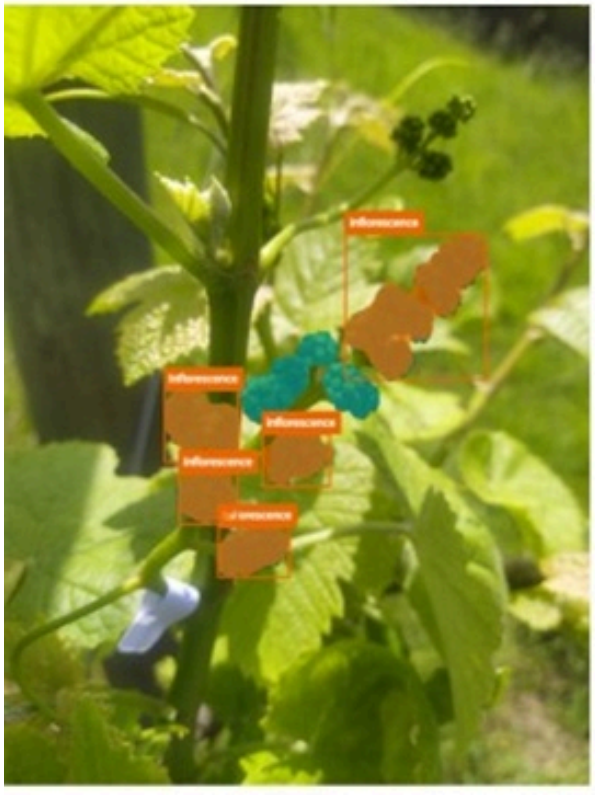
© 2023 The Author(s). | Published by Agro Intelligence Research Press



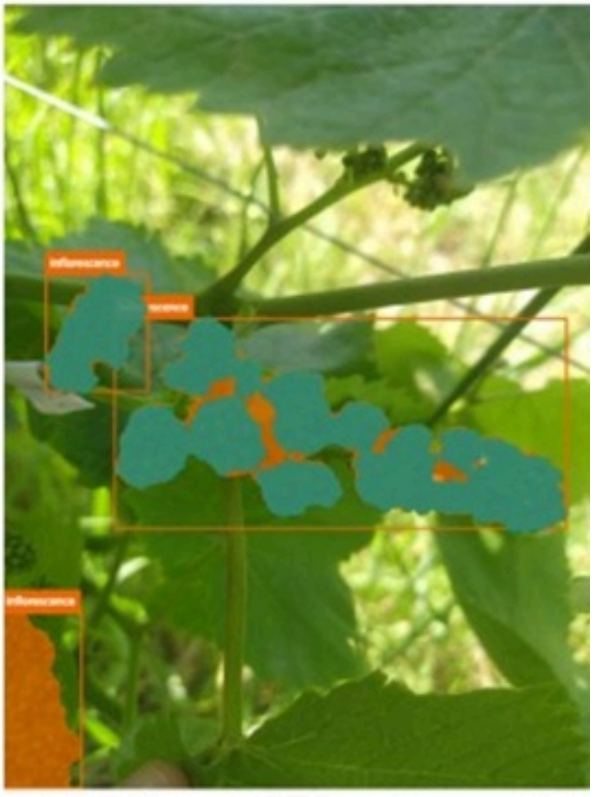
(a)



(b)



(c)



(d)

© 2023 The Author(s). | Published by Agro Intelligence Research Press

Figure 8. Successful isolation of a tightly bunched inflorescence (YOLOv5n) (a) and a stretched-out one (YOLOv8n) (b); omission of a flower cluster from segmentation output (YOLOv5n) (c); detection and masking of inflorescences lacking groundtruth entries (YOLOv8n) (d). Orange-filled masks correspond to model-generated regions; blue-filled masks represent human-annotated reference regions.

As far as the authors can ascertain, the models under evaluation set a new high-water mark relative to previously published inflorescence segmentation results, while operating on a robust dataset collected in unconstrained natural settings. Some research teams have deliberately chosen to gather imagery after dark with supplemental lighting. This strategy imposes visual uniformity and partially side-steps the complications introduced by natural scene clutter. This describes the approaches of Palacios *et al.* [25] and Rahim *et al.* [27], who leveraged SegNet (VGG19) and Mask R-CNN, respectively, achieving F1 scores of 93.0% and 94.3%. It bears underscoring that those captures were made from considerably greater standoff distances, which inherently magnifies the difficulty of the detection and segmentation tasks. Counterbalancing this, the architectures they adopted are vastly more intricate and heavyweight, possessing far greater parameter volumes and imposing significantly higher computational demands, in marked contrast to the streamlined, nimble YOLO nano variants trialled in the present work. Image scarcity remains an additional bottleneck, with a preponderance of published works relying on datasets with fewer than 10,000 samples. Rudolph *et al.* [21], by way of illustration, validated an AlexNet-grounded FCN on a sparse collection of just 10 frames, securing a mean IoU of merely 76.0%.

When considered collectively, the evidence presented here supports an optimistic outlook on the feasibility of automated inflorescence detection and segmentation. That said, lingering impediments—most notably the fragility characterizing many existing datasets and the frequent lack of rigour in specifying evaluation benchmarks—must be squarely tackled for the discipline to make genuine strides toward operationalizing these processes.

Flower detection

Deep learning models

The strong fidelity demonstrated by the models during the inflorescence segmentation stage furnishes a solid foundation of confidence for advancing to the more granular challenge of resolving individual grapevine blossoms. Accordingly, beyond the pair of networks trialled in the preceding step, the freshest entries in the YOLO family tree—version 9 tiny and version 10 nano—were folded into the comparative assessment, given that their most compact object-detection flavours had become available by the time these experiments were conducted. As shown in **Table 5**, the full cohort of models delivered respectable performances, all clearing F1-Scores above 88%, yet YOLOv8n distinguished itself from the pack by achieving an F1-Score of 91%. A noteworthy observation is that the more contemporary models show a downturn in performance, with YOLOv10n suffering the steepest drop. Although its F1-Score remains broadly on par with its rivals and it actually claims the top Precision figure, its ability to surface all relevant objects falls far short, managing a Recall of just 82%. For all that it represents the latest architectural iteration and might therefore be presumed superior in principle, its markedly more elaborate design could prove a liability in the present context, where the objective is to discern a dense multitude of minute targets. The greater architectural sophistication is also reflected in the inference speed measurements: although all models posted exemplary latencies (none exceeding 0.015 seconds), the two newer incarnations exhibit a slight slowdown compared to their predecessors.

Table 5. Single-flower detection outcomes on the test set derived from DL models under optimized confidence thresholds.

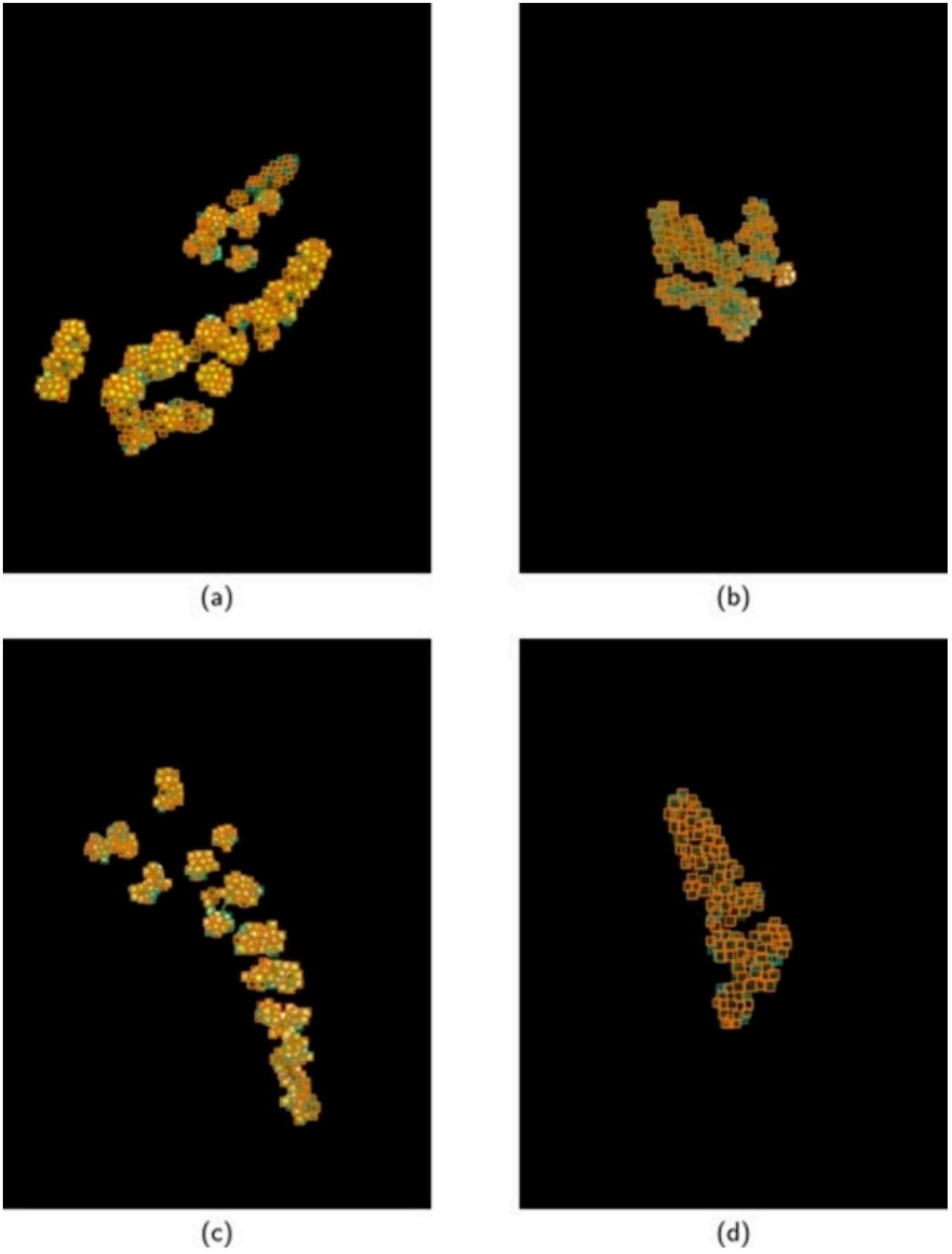
Models	Confidence threshold	Speed (s)	*F1-Score*	*Recall*	*Precision*
YOLOv5n	0.197	0.007	0.900	0.880	0.920
YOLOv8n	0.194	0.007	0.914	0.895	0.933
YOLOv9t	0.179	0.015	0.899	0.879	0.920

© 2023 The Author(s). | Published by Agro Intelligence Research Press

YOLOv10n	0.234	0.011	0.881	0.821	0.950

Looking across the board, the models consistently display sound and dependable detection capabilities, rendering faithful results on imagery featuring single or multiple inflorescences of diverse morphologies, captured from varying standoffs and across a spectrum of illumination and optical sharpness conditions **(Figure 9)**. Nevertheless, it remains essential to map the specific scenarios that give rise to FP and FN and to understand the mechanisms underpinning their occurrence. Such diagnostic clarity is a prerequisite for engineering mitigations that curb their prevalence. The emergence of FP, which exacts a toll on Precision, was predominantly driven by two failure patterns: portions of the peduncle being mistaken for blossoms **(Figure 10a)**, an error mode especially pronounced in inflorescences that are elongated and have wide spacing between buds, given the chromatic affinity between the two tissue types; and redundant double firings targeting an already-counted bloom **(Figure 10b)**, a phenomenon partly attributable to the reality that an individual flower may occasionally present more than a single specular hotspot, and light reflectance patterns constitute one of the signals that these networks internalize and exploit when forming predictions [15]. Meanwhile, Recall, depressed by FN, suffered attrition from the models' difficulties in registering exceptionally small blossoms **(Figure 10c)**, whether due to the intrinsically modest dimensions of the inflorescence or the extended capture distance, and from buds that were partially or wholly obscured by adjacent buds **(Figure 10d)**, a frequent situation within densely packed, tight-clustered inflorescences.

© 2023 The Author(s). | Published by Agro Intelligence Research Press



© 2023 The Author(s). | Published by Agro Intelligence Research Press

Figure 9. Flawlessly executed detection of individual flowers across scenes containing one or multiple inflorescences (YOLOv5n) (a), with contrasting morphologies (YOLOv8n and YOLOv9t) (b and c), and under varied illumination and focus regimes (YOLOv10n) (d). Orange bounding boxes signal model predictions; blue bounding boxes signal ground truth annotations.

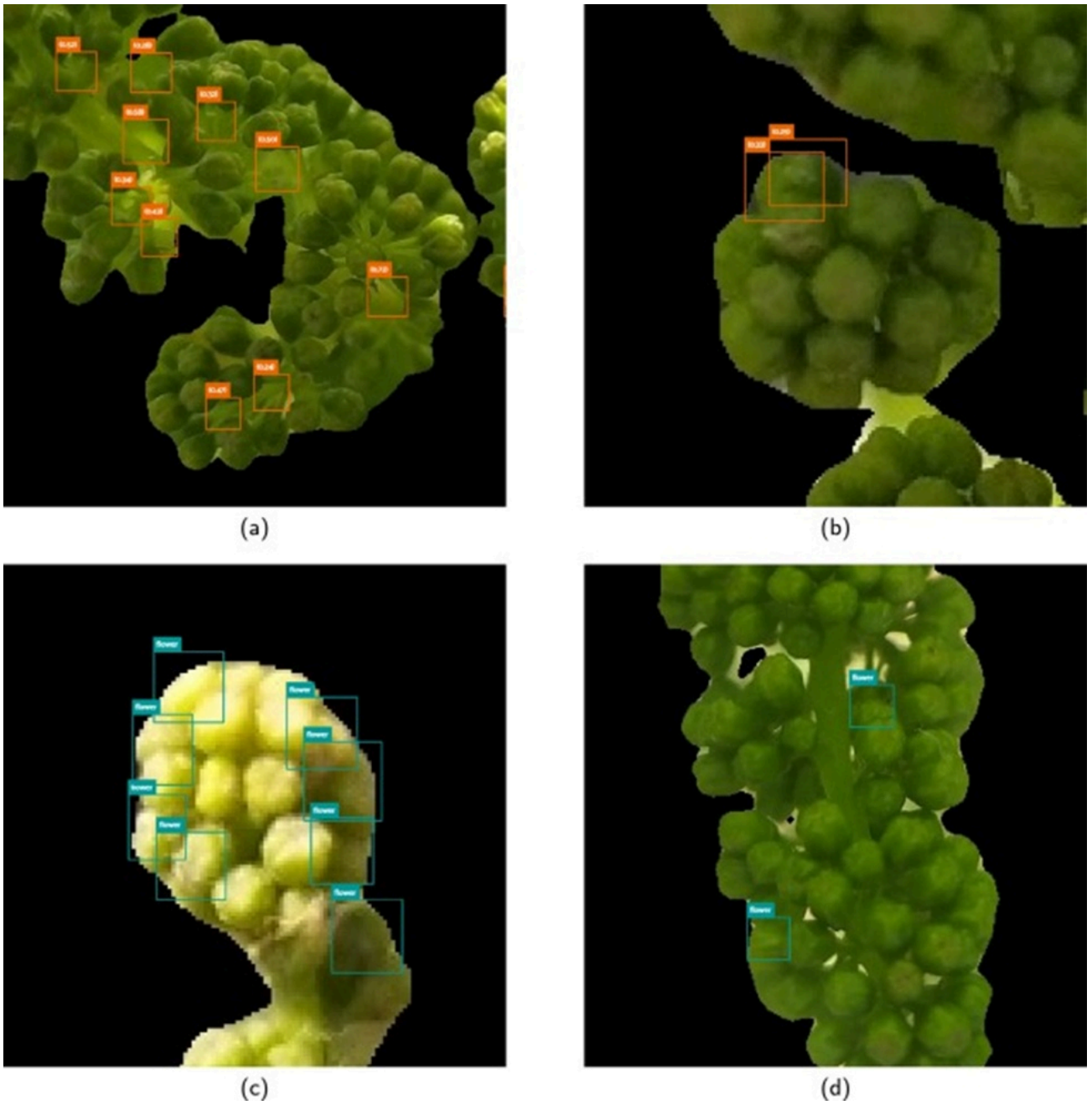


Figure 10. Occurrence of False Positives stemming from misidentified peduncle fragments (YOLOv5n) (a) or duplicate detections of already-counted blooms (YOLOv8n) (b), and False Negatives arising from the failure to capture diminutive flowers (YOLOv9t) (c) or flowers occluded by neighbours (YOLOv10n) (d). Orange bounding boxes signal model predictions; blue bounding boxes signal ground truth annotations.

When held up against the wider body of published work, the figures achieved in this study appear decidedly promising. They outclass the outcomes reported by Grimm *et al.* [20], who documented an F1-Score of 86.7% upon testing a DL

© 2023 The Author(s). | Published by Agro Intelligence Research Press

framework grounded in Fully Convolutional Networks (FCN) across 15 tiled image patches, and those of Palacios *et al.* [25], who deployed a SegNet architecture paired with contour-filtering post-processing on a batch of 24 full-resolution frames, attaining an F1-Score of 72.9%. The sole results that exceed those presented here belong to Rahim *et al.* [23], who initially trialled three DL architectures on 204 images of inflorescences—both still attached to the vine and excised—photographed against a black artificial background. Within that study, the patch-based Mask R-CNN configuration achieved the strongest results, with an F1-Score hovering around 98.0% for the Shiraz cultivar. That same model was later adapted to in-field conditions and, with the aid of varied pre-processing parameter configurations, continued to deliver top-tier results, clinching an F1-Score of 96.9% when assessed across 100 images representing multiple cultivars [27].

For the most part, DL-based approaches to flower detection incorporate some variety of pre-processing, whether contrast manipulation or resolution scaling. Another salient trend is that certain research groups have elected to frame flower detection as a segmentation problem, a choice that tends to yield tighter RoI demarcation and cleaner separation from background disturbances and extraneous vegetative elements. Even with these efforts, the field continues to suffer from a lack of publicly accessible, annotated datasets on grapevine flowers. This shortfall inevitably invites scepticism regarding the claimed robustness of particular models, since they often undergo validation on only a tiny handful of frames—an unsurprising state of affairs given the punishing annotation workload, which for grapevine blossoms can consume as much as 35 minutes per image, according to the estimate provided by Grimm *et al.* [20].

When one considers that the pipeline constructed for this study forgoes any pre-processing, treats the challenge as an object detection task, and has been rigorously evaluated across a substantial corpus of images spanning multiple grape cultivars, the results take on considerable significance. They offer compelling evidence of the solid aptitude that compact, architecturally uncluttered DL models possess for the demanding work of blossom detection.

Marker-controlled Watershed algorithm

A more stripped-down route relative to DL pipelines, namely the “Iterative H-minima Based Marker-Controlled Watershed” proposed by Koyuncu *et al.* [41], was evaluated for resolving solitary grapevine blossoms. As noted earlier in section 3.2.3, the method’s reliance on four externally supplied settings meant that a range of parameter configurations had to be explored. **Table 6** catalogues the metric scores that resulted. From the full grid of attempted pairings, the combination tarea = {20} coupled with dsize = {7} rose to the top, delivering an F1-Score of 72.3%. A wider trend also surfaces: dialling up both tarea and dsize tends to push Precision upward while simultaneously dragging Recall downward. With larger values, the algorithm grows more selective about what qualifies as a positive detection, gravitating toward larger-scale, visually dominant elements. The net effect is cleaner Precision because spurious detections are largely suppressed, yet a drop in Recall because diminutive or less conspicuous genuine regions get discarded or eroded **(Figure 11)**.

Table 6. Single-flower detection performance on the test set achieved by the Iterative H-minima Based Marker-Controlled Watershed approach across varying parameter settings.

Parameters				F1-Score	Recall	Precision
tarea	dsize	p	a			
t5	d5	2	30	0.550	0.906	0.394
t5	d7	2	30	0.704	0.857	0.597
t5	d10	2	30	0.721	0.655	0.801
t5	d13	2	30	0.554	0.401	0.895
t10	d5	2	30	0.590	0.883	0.443
t10	d7	2	30	0.719	0.828	0.635

© 2023 The Author(s). | Published by Agro Intelligence Research Press

t10	d10	2	30	0.714	0.634	0.818
t10	d13	2	30	0.549	0.394	0.902
t20	d5	2	30	0.609	0.774	0.502
t20	d7	2	30	0.723	0.770	0.682
t20	d10	2	30	0.687	0.585	0.832
t20	d13	2	30	0.528	0.372	0.908
t30	d5	2	30	0.553	0.586	0.524
t30	d7	2	30	0.689	0.676	0.704
t30	d10	2	30	0.652	0.532	0.842
t30	d13	2	30	0.507	0.351	0.913

© 2023 The Author(s). | Published by Agro Intelligence Research Press

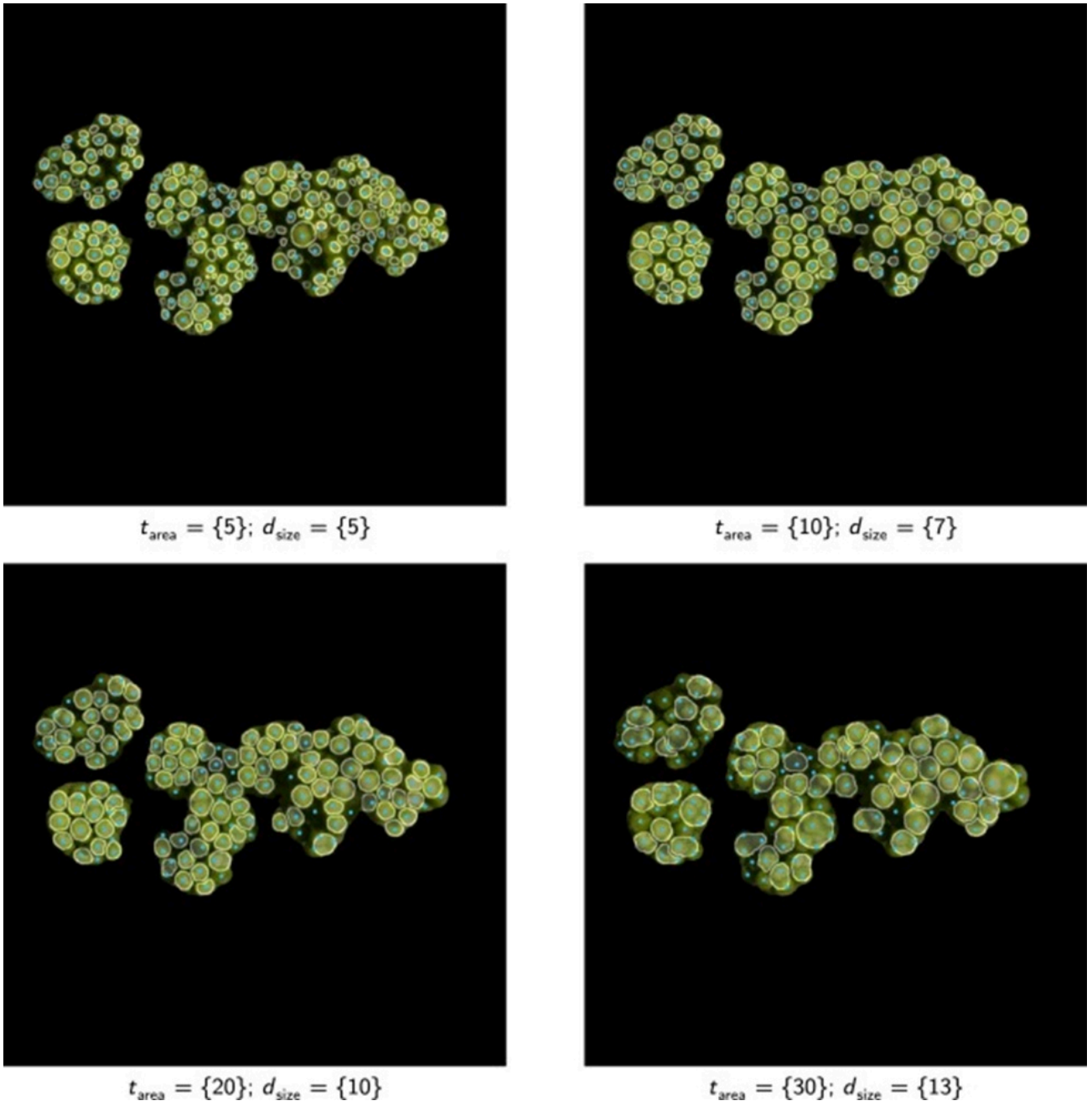


Figure 11. Progression of flower detections produced by the “Iterative H-minima Based Marker-Controlled Watershed” procedure as the tarea and dsize parameters are stepped upward. White-filled masks indicate algorithm-generated regions; blue keypoints indicate the centre of each hand-labelled ground truth blossom.

Juxtaposed with the DL-based systems, the Watershed pipeline fell well short, a gap starkly reflected in the F1-Score readings. Techniques of this kind are exceptionally brittle with respect to the spatial extent of the RoI, because the source photographs were taken at various standoff distances, blossom dimensions vary even within a single cluster, and additional systematic variation occurs between cultivars, across years, and throughout developmental stages. For these reasons, such an approach may not be the most durable or reproducible strategy, particularly since the governing parameters must be chosen manually. Still, the algorithm produced passable results compared with other classical flower detection routines reported in the literature. It outperformed the outcomes published by Palacios *et al.* [25], who likewise pitted DL models (SegNet-VGG19) against a Watershed method that scored an F1 of 70.80%. It nearly matched the results of Rudolph *et al.* [21], who applied the Circular Hough Transform to blossom extraction and secured an F1-Score

© 2023 The Author(s). | Published by Agro Intelligence Research Press

of 75.2%. The only classical methods that sit clearly above it are the extended H-maxima transform framework, originally put forward by Diago *et al.* [8] and subsequently reworked in numerous adaptations across the field [12-14], which has yielded F1-Scores in the 84% to 88% band, and the SURF key-point detection strategy adopted by Liu *et al.* [15], which remains the most durable and highest-performing classical pipeline on record, having reached an F1-Score of 93%.

Even though certain papers boast stronger numbers, performing a full and fair side-by-side assessment is hampered by the reality that those algorithms were largely validated on an artificial backdrop (black card), often on inflorescences already clipped from the vine, and over a very modest image corpus. Diago *et al.* [8] and Aquino *et al.* [14], for example, applied their methods on only 15 and 40 frames, respectively. Seen through this lens, the figures produced in the present study provide a benchmark to orient future refinements. The fact that the evaluation was conducted on imagery that faithfully reproduces the tangled reality of field conditions—embracing the authentic diversity of cluster shapes and flower dimensions, along with the inconsistencies of capture circumstances (mixed distances and natural lighting)—and did so with no pre-processing or adaptive correction stages, means the scores remain relevant. They provide a credible foundation and suggest that substantial performance gains are attainable through continued development and tuning.

Flower estimation

Encouraged by the strong detection metrics, the DL architectures were further scrutinized for their capacity to estimate grapevine blossom abundance using three core indicators: R^2, MAPE, and RMSE. Taking all metrics together, YOLOv8n occupied the leading position across the board—boasting the peak R^2 of 95.9% along with the smallest MAPE (5.245) and RMSE (14.016)—highlighting its enhanced aptitude for reliable flower quantification. The other candidates, though generally performing adequately, showed moderately reduced accuracy and inflated error rates, with YOLOv10n showing a particularly pronounced drop-off.

Figure 12 translates these numerical patterns into a more directly graspable format through scatter plots that place the actual human-counted blossom totals from the test set (ground truth) side by side with the numbers forecast by each model. The plots for YOLOv5n, YOLOv8n, and YOLOv9t reveal that the predicted points cluster snugly around the diagonal line of perfect agreement, consistent with their elevated R^2 values and indicating a well-fitted model with modest predictive scatter. YOLOv8n, most notably, shows the tightest grouping, further backing the quantitative evidence of its forecasting superiority. The YOLOv10n scatter plot, in contrast, displays substantially wider dispersion, with a considerable number of points veering away from the ideal-fit diagonal. This visual signature corroborates the comparatively elevated MAPE and RMSE statistics for that model, implying that while YOLOv10n can still produce broadly sensible estimates, its precision meaningfully lags behind the others.

© 2023 The Author(s). | Published by Agro Intelligence Research Press

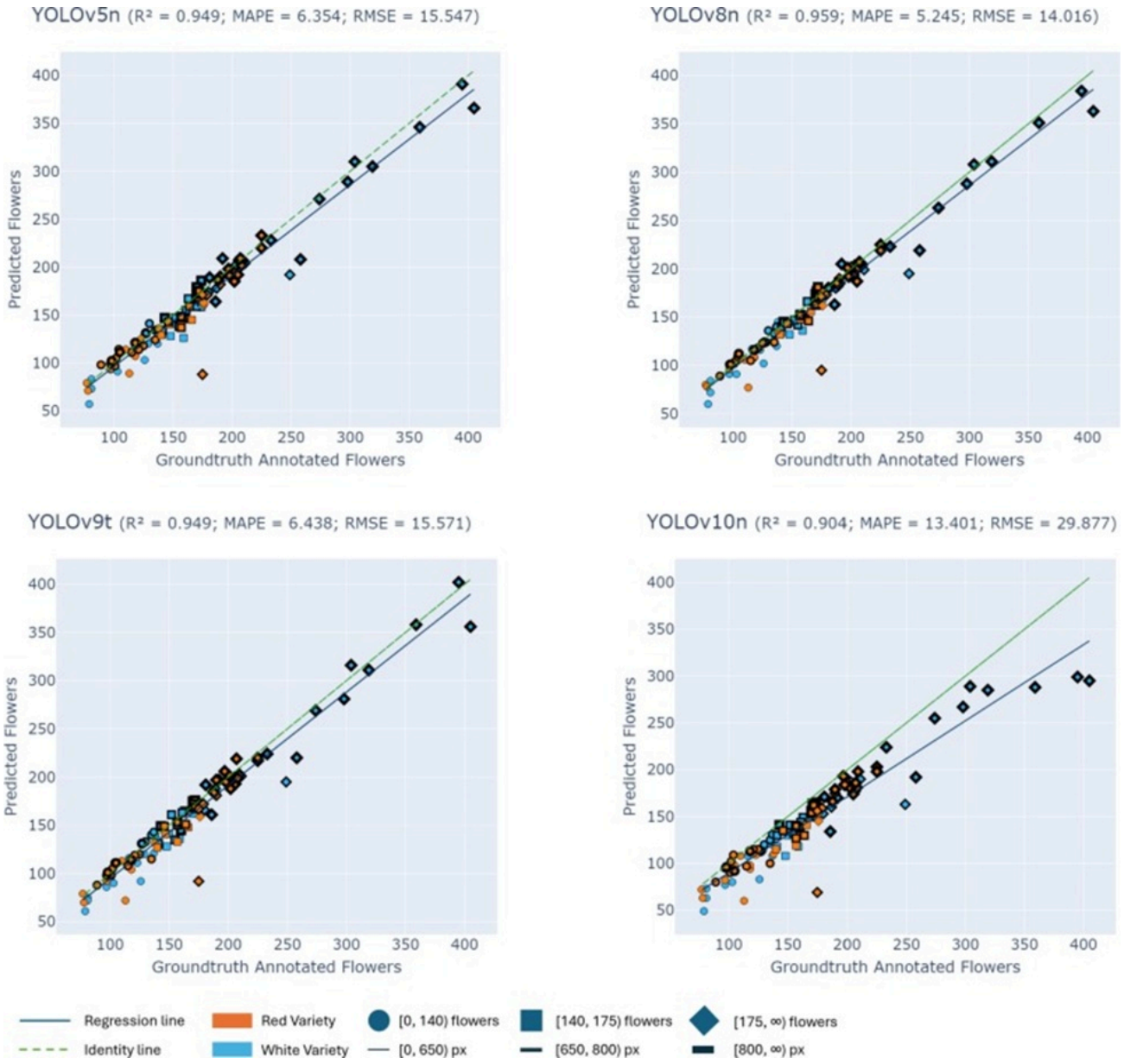


Figure 12. Scatter comparison between the visually countable grapevine blossoms per inflorescence and the corresponding model-predicted flower counts across the test set images.

Looking at the bigger picture, although both overcounting and undercounting incidents can be spotted among the outputs of all four architectures, the YOLOv10n variant seems to exhibit a greater tendency toward—and more severe instances of—undercounting. This pattern can be inferred from the concentration of data points sitting beneath the equality line, especially pronounced at higher actual flower densities, suggesting this model may encounter genuine difficulty when faced with inflorescences that carry large blossom populations. Underestimation is, more broadly, the dominant error direction, a natural consequence of the models' greater vulnerability to producing FN rather than FP, stemming from missed detection of tiny or mutually hidden buds and the challenge of reliably picking out blooms under blur or dim illumination. Curbing undercounts driven by FN within object detection frameworks calls for boosting the richness and variety of training data, recalibrating operational knobs such as the confidence cutoff, and possibly adopting ensemble strategies or custom loss formulations to improve detection thoroughness.

© 2023 The Author(s). | Published by Agro Intelligence Research Press

The weaker performance of YOLOv10n was further substantiated by the ANOVA outcomes **(Table 7)**. This analysis probed how model identity, differing inflorescence morphologies—captured through cluster span and blossom count bins—and cultivar colour grouping (white versus red) shape error magnitudes. A robust main effect of model choice on error rates was detected ($P < 0.001$), with YOLOv10n showing a markedly elevated MAPE (13.401), a finding supported by the Duncan post-hoc test. Examining the role of cultivar type revealed no meaningful MAPE differences between white and red varieties for any of the models studied ($P > 0.127$). Likewise, the analysis uncovered no significant performance shifts across the binned flower-number strata ($P > 0.387$) nor across the inflorescence-size strata ($P > 0.859$) when examining each model individually. However, regarding inflorescence size specifically, although the interaction term between model and size class did not reach significance, within-model probing showed that size significantly conditions prediction accuracy for YOLOv5n ($P = 0.019$), YOLOv9t ($P = 0.015$), and YOLOv10n ($P = 0.009$). Across these three, the smallest inflorescence category (1) was associated with notably higher average errors. What distinguishes YOLOv8n is the complete lack of statistically significant variation across inflorescence size categories ($P = 0.111$), indicating that this model maintains comparatively steady performance regardless of cluster size.

Table 7. ANOVA-driven statistical comparison of mean discrepancies between observed and predicted blossom counts for the tested DL models, cultivar groupings, inflorescence morphology categories, and their combined effects.

Empty cell	YOLOv10n	YOLOv9t	YOLOv8n	YOLOv5n	$F_{interaction}$ (*P*-value)
Models (n = 107)	13.401^{B}	6.483^{A}	5.245^{A}	6.354^{A}	$F_{model} < 0.001$
Varieties					
White (n = 54)	14.292	6.695	5.750	6.893	$F_{Interaction\ model/variety} > 0.127$
Red (n = 53)	12.494	6.177	4.730	5.605	
$F_{variety}$ (P-value)	0.323	0.703	0.418	0.401	
Flowers number					
[0,140) (n = 38)	13.736	7.356	5.922	6.716	$F_{Interaction\ model/number} > 0.387$
[140,175) (n = 31)	12.314	5.537	4.458	6.409	
[175,∞) (n = 38)	13.954	6.257	5.201	5.947	
F_{number} (P-value)	0.745	0.555	0.650	0.881	
Inflorescence size					
[0,650) (n = 37)	17.028^{b}	8.887^{b}	6.869	8.404^{b}	$F_{Interaction\ model/size} > 0.859$
[650,800) (n = 35)	12.516^{ab}	6.153^{ab}	5.133	6.575^{ab}	
[800,∞) (n = 37)	10.536^{a}	4.245^{a}	3.72	4.107^{a}	
F_{size} (P-value)	0.009	0.015	0.111	0.019	

ANOVA F: the p-value tied to the Fisher test performed within the ANOVA; means with a p-value below 0.05 are treated as statistically distinguishable. Means that share a common letter label, whether within rows (uppercase) or within columns (lowercase), are not significantly different according to the Duncan test ($P = 0.05$).

These findings throw into sharp relief the degree to which the selection of a particular model architecture dictates error levels. By contrast, natural variation in inflorescence configuration and the split between red and white grape lineages appear to leave no statistically detectable footprint on the performance gaps among the models. This observation

© 2023 The Author(s). | Published by Agro Intelligence Research Press

resonates with earlier reports in the literature demonstrating that variety- and morphology-agnostic models can hold their own against, or even outperform, models deliberately tuned to specific cultivars or shape categories [12, 14, 23, 25, 27].

The set of models benchmarked here, above all YOLOv8n, compares favourably with the broader body of published work. They exceed the correlation strength reported by Pahalawatta *et al.* [24] and Rudolph *et al.* [21], who reported R^2 values of 81.5% and 93.0%, respectively. They match the results of Millan *et al.* [12] and Rahim *et al.* [23], both of whom reported R^2 values of around 96.0%. The YOLOv8 model was surpassed solely by the Mask R-CNN pipeline fielded by Rahim *et al.* [27], which attained an R^2 of 98.1%. Beyond this, several earlier studies have tied the visible flower count within a single image—whether estimated via models (linear or nonlinear) or derived from manual annotation—to the absolute flower complement per inflorescence determined through destructive hand counting, with R^2 readings exceeding 90%. Given the precedent set by prior work and the strong performance achieved here, the body of evidence suggests that the models under study hold real promise as dependable proxies for total flower abundance and, by extension, as functional instruments for early-season yield estimation in viticulture.

Conclusion

This study highlights the considerable potential of CV technologies for automated sensing and quantification of grapevine yield components. A processing pipeline fusing DL architectures with foundational image analysis algorithms was assembled to both localize inflorescences and estimate the number of individual flowers they contain, leveraging a purpose-built, publicly shared dataset of grapevine inflorescence images. Modern YOLO variants handle the dual tasks of inflorescence detection and pixel-wise masking; in the subsequent step, where the goal shifts to isolating and tallying constituent flowers, those same architectures were critically evaluated against a classical iterative Watershed image processing pipeline.

The investigation confirms the effectiveness of DL-based approaches for both demarcating grapevine inflorescences and for detecting and enumerating grapevine flowers. YOLOv8n and YOLOv5n each delivered standout results in inflorescence localization and segmentation, achieving F1 scores of 95.9% (Box) and 98.6% (Mask), respectively. YOLOv8n further proved its mettle in blossom detection, achieving an F1-Score of 91.4%, and in flower quantification, where it produced a tight alignment between detected flower counts and the actual number of blossoms observable within each inflorescence (R^2 = 95.9%). A complementary statistical evaluation indicated that grape cultivar and inflorescence architecture had only a modest bearing on model performance.

The on-the-ground significance of these findings is substantial, placing into the hands of viticulturists a tool suited to more streamlined, evidence-based vineyard oversight. The capability to execute real-time, non-invasive yield forecasts opens the door to more precise harvest planning and tighter quality management, ultimately strengthening the economic footing of vineyard operations. Looking ahead, research should prioritize i) enlarging the dataset to capture a broader cross-section of vineyard environments and grape cultivars, ii) further polishing these models and retooling them for the detection and counting of inflorescences and flowers on a per-vine basis, iii) validating their estimation outputs against physical manual censuses of grapevine inflorescences and flowers, and iv) porting the models onto portable devices for field trials under genuine vineyard operating conditions.

Acknowledgements

The authors would like to acknowledge the scholarship number “2022.09726.BD”, funded by National Funds through the Portuguese funding agency, FCT (Fundação para a Ciência e Tecnologia).

Conflict of interest

None

© 2023 The Author(s). | Published by Agro Intelligence Research Press

Financial support

This work was co-financed by “Component 5: Capitalization and Business Innovation”, integrated in the Resilience Dimension of the Recovery and Resilience Plan within the scope of the Recovery and Resilience Mechanism (MRR) of the European Union (EU), framed in the NextGenerationEU, for the period 2021–2026 within the project “Wine4cast – Space-time prediction of wine productivity for multi-actor usability: integration of remote optical-photonic sensors, artificial intelligence and climate scenarios” (ref: PRR-C05-i03-I-000071).

Ethics statement

None

Received: 09 Jan 2022 Revised: 16 Mar 2022 Accepted: 01 Apr 2022
Published online: 25 May 2023

Rights and permissions

Open Access The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at https://creativecommons.org/licenses/by-nc-sa/4.0/.

References

1. Mekouar MA. Food and Agriculture Organization of the United Nations (FAO). Yearbook Int Environ Law. 2020;31:326–40.
2. Marín D, Armengol J, Carbonell-Bejerano P, Escalona J, Gramaje D, Hernández-Montes E, et al. Challenges of viticulture adaptation to global change: tackling the issue from the roots. Aust J Grape Wine Res. 2021;27:8–25.
3. Syal A, Garg D, Sharma S. A survey of computer vision methods for counting fruits and yield prediction. Int J Comput Sci Eng. 2013;2:346–50.
4. Laurent C, Oger B, Taylor JA, 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.https://www.sciencedirect.com/science/article/pii/S1161030121001106.
5. Carrillo E, Matese A, Rousseau J, Tisseyre B. Use of multi-spectral airborne imagery to improve yield sampling in viticulture. Precis Agric. 2016;17:74–92.
6. Clingeleffer PR, Martin S, Dunn G, Krstic M. Crop development, crop estimation and crop control to secure quality and production of major wine grape varieties: a national approach, final report to grape and wine research. 2001.
7. Arnó J, Casasnovas M, Ribes-Dasi M, Rosell-Polo J. Precision viticulture. Research topics, challenges and opportunities in site-specific vineyard management. Span J Agric Res. 2009;7.
8. Diago M-P, Sanz García A, Millan B, Blasco J, Tardaguila J. Assessment of flower number per inflorescence in grapevine by image analysis under field conditions. J Sci Food Agric. 2014;94. https://doi.org/10.1002/jsfa.6512.

9. Tardaguila J, Stoll M, Gutiérrez S, Proffitt T, Diago MP. Smart applications and digital technologies in viticulture: a review. Smart Agric Technol. 2021;1:100005. https://doi.org/10.1016/j.atech.2021.100005.

10. Mohimont L, Alin F, Rondeau M, Gaveau N, Steffenel LA. Computer vision and deep learning for precision viticulture. Agronomy. 2022;12. https://www.mdpi.com/2073-4395/12/10/2463

11. Farjon G, Huijun L, Edan Y. Deep-learning-based counting methods, datasets, and applications in agriculture – a review. 2023. https://arxiv.org/abs/2303.02632

12. Millan B, Aquino A, Diago M-P, Tardaguila J. Image analysis-based modelling for flower number estimation in grapevine. J Sci Food Agric. 2016;97.

13. Aquino A, Millan B, Gaston D, Diago M-P, Tardaguila J. vitisflower®: development and testing of a novel Android-smartphone application for assessing the number of grapevine flowers per inflorescence using artificial vision techniques. Sensors. 2015;15:21204–18. https://www.mdpi.com/1424-8220/15/9/21204

14. Aquino A, Millan B, Gutiérrez S, Tardáguila J. Grapevine flower estimation by applying artificial vision techniques on images with uncontrolled scene and multi-model analysis. Comput Electron Agric. 2015;119:92–104. https://www.sciencedirect.com/science/article/pii/S016816991500321X

15. Liu S, Li X, Wu H, Xin B, Tang J, Petrie PR, et al. A robust automated flower estimation system for grape vines. Biosyst Eng. 2018;172:110–23. https://www.sciencedirect.com/science/article/pii/S1537511017304610

16. Liu S, Li X, Wu H, Xin B, Tang J, Petrie PR, et al. A robust automated flower estimation system for grape vines. Biosyst Eng. 2018;172:110–123. https://www.sciencedirect.com/science/article/pii/S1537511017304610

17. Schmidhuber J. Deep learning in neural networks: an overview. Neural Netw. 2015;61:85–117.

18. Kamilaris A, Prenafeta-Boldú FX. Deep learning in agriculture: A survey. Comput Electron Agric. 2018;147:70–90.

19. Victorino G, Braga R, Santos-Victor J, Lopes CM. Yield components detection and image-based indicators for non-invasive grapevine yield prediction at different phenological phases. OENO One. 2020.

20. Grimm J, Herzog K, Rist F, Kicherer A, Töpfer R, Steinhage V. An adaptable approach to automated visual detection of plant organs with applications in grapevine breeding. Biosyst Eng. 2019;183:170–83. https://www.sciencedirect.com/science/article/pii/S1537511018311838

21. Rudolph R, Herzog K, Toepfer R, Steinhage V. Efficient identification, localization and quantification of grapevine inflorescences and flowers in unprepared field images using fully convolutional networks. Vitis. 2019;58:95–104.

22. Khokher MR, Liao Q, Smith AL, Sun C, Mackenzie D, Thomas MR, et al. Early yield estimation in viticulture based on grapevine inflorescence detection and counting in videos. IEEE Access. 2023;11:37790–808.

23. Rahim U, Mineno H, Tomoyoshi U. Comparison of grape flower counting using patch-based instance segmentation and density-based estimation with convolutional neural networks comparison of grape flower counting using patch-based instance segmentation and density-based estimation with convolutional neural networks, 2021.

24. Pahalawatta K, Fourie J, Parker A, Carey P, Werner A. Detection and classification of opened and closed flowers in grape inflorescences using mask R-CNN. 2020 35th International Conference on Image and Vision Computing New Zealand (IVCNZ). 2020:1–6.

© 2023 The Author(s). | Published by Agro Intelligence Research Press

25. Palacios F, Bueno G, Salido J, Diago MP, 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. https://www.sciencedirect.com/science/article/pii/S0168169920315428

26. Jaramillo J, Vanden Heuvel J, Petersen KH. Low-cost, computer vision-based, prebloom cluster count prediction in vineyards. Front Agron. 2021;3. https://www.frontiersin.org/articles/ https://doi.org/10.3389/fagro.2021.648080.

27. Rahim U, Utsumi T, Mineno H. Deep learning-based accurate grapevine inflorescence and flower quantification in unstructured vineyard images acquired using a mobile sensing platform. Comput Electron Agric. 2022;198:107088.

28. Buayai P, Yok-In K, Inoue D, Leow CS, Nishizaki H, Makino K, et al. End-to-end inflorescence measurement for supporting table grape trimming with augmented reality. 2021 International Conference on Cyberworlds (CW). 2021:101–8.

29. Du W, Zhu Y, Li S, Liu P. Spikelets detection of table grape before thinning based on improved Yolov5s and Kmeans under the complex environment. Comput Electron Agric. 2022;203:107432. https://www.sciencedirect.com/science/article/pii/S0168169922007402

30. Meier U. Growth Stages of Mono- and Dicotyledonous Plants. Blackwell Wissenschafts-Verlag. 1997.

31. OIV. The 3rd Edition of the OIV Descriptor List for Grape Varieties and Vitis Species. International Organisation of Vine and Wine. 2023.

32. Bakhtiarnia A, Zhang Q, Iosifidis A. Efficient high-resolution deep learning: a survey. arXiv:2207.13050. 2022.

33. Buslaev A, Iglovikov VI, Khvedchenya E, Parinov A, Druzhinin M, Kalinin AA. Albumentations: Fast and flexible image augmentations. Information. 2020;11. https://www.mdpi.com/2078-2489/11/2/125.

34. Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: unified, real-time object detection. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016:779–88.

35. Jocher G. Ultralytics yolov5. 2020. https://github.com/ultralytics/yolov5. https://doi.org/10.5281/zenodo.3908559.

36. Jocher G, Chaurasia A, Qiu J. Ultralytics yolov8. 2023. https://github.com/ultralytics/ultralytics

37. Lin T-Y, Maire M, Belongie S, Bourdev L, Girshick R, Hays J, et al. Microsoft coco: common objects in context. arXiv:1405.0312. 2015.

38. Wang C-Y, Liao H-YM. Yolov9: learning what you want to learn using programmable gradient information. 2024.

39. Wang A, Chen H, Liu L, Chen K, Lin Z, Han J, et al. YOLOv10: real-time end-to-end object detection. arXiv preprint arXiv:2405.14458. 2024.

40. MATLAB. Version 7.10.0 (R2010a). The MathWorks Inc. Natick, Massachusetts. 2010.

41. Koyuncu C, Akhan E, Ersahin T, Cetin-Atalay R, Gunduz-Demir C. Iterative h-minima-based marker-controlled watershed for cell nucleus segmentation. Cytometry A. 2016;89.

© 2023 The Author(s). | Published by Agro Intelligence Research Press