

A Modular UAV-Based Framework for Apple Detection and Yield Extrapolation from 3D Point Clouds

Kenny Nguyen

knguyen26@seattleu.edu
Seattle University
Seattle, USA

Nathan Tran

ntran9@seattleu.edu
Seattle University
Seattle, USA

Wan D. Bae

baew@seattleu.edu
Seattle University
Seattle, USA

Dhruv J. Patel

dpatel5@seattleu.edu
Seattle University
Seattle, USA

Amin Riazi

ariazi@seattleu.edu
Seattle University
Seattle, USA

Tony Nguyen

tnguyen49@seattleu.edu
Seattle University
Seattle, USA

J. Wesley Lauer

lauerj@seattleu.edu
Seattle University
Seattle, USA

Shayma Alkobaisi

shayma.alkobaisi@uaeu.ac.ae
United Arab Emirates University
Al Ain, United Arab Emirates

Abstract

Precision agriculture increasingly relies on unoccupied aerial vehicles (UAVs) and machine learning to automate fruit monitoring and improve sustainability. While 2D imaging approaches face challenges of occlusion, clustering, and calibration, UAV-derived 3D point clouds offer richer structural information. This paper presents a modular framework for automated apple detection, counting, and yield estimation. The framework integrates (1) color-based segmentation (HSV and ExR-LAB), (2) regression and clustering (DBSCAN) for apple count estimation, and (3) geometric approximations (minimum bounding box, sphere, and ellipsoid) for validation and volume modeling. Experiments on UAV imagery from a commercial orchard using human-annotated section-level counts as reference ground truth show that regression achieves high accuracy ($R^2 \approx 0.99$, relative error ≤ 0.03) for apple-count estimation, while clustering with geometric validation provides spatial localization and per-fruit morphology for downstream yield assessment. Beyond counting, per-fruit volume and harvest mass are estimated by approximating each apple with sphere or ellipsoid models and scaling to the field level using density and coverage ratios. Together, regression and clustering form a hybrid estimation strategy that delivers both accurate counts and indicative yield forecasts. The proposed open-source framework demonstrates strong potential for scalable, UAV-assisted orchard management.

CCS Concepts

- Computing methodologies → Neural networks; Point-based models;
- Applied computing → Agriculture.

Keywords

precision agriculture, UAV imagery, 3D point clouds, apple detection, yield estimation, spatial clustering, regression modeling

ACM Reference Format:

Kenny Nguyen, Dhruv J. Patel, Tony Nguyen, Nathan Tran, Amin Riazi, J. Wesley Lauer, Wan D. Bae, and Shayma Alkobaisi. 2026. A Modular UAV-Based Framework for Apple Detection and Yield Extrapolation from 3D Point Clouds. In *The 41st ACM/SIGAPP Symposium on Applied Computing (SAC '26)*, March 23–27, 2026, Thessaloniki, Greece. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3748522.3779768>

1 Introduction

Precision agriculture (PA) seeks to improve agricultural sustainability by integrating high-resolution monitoring, predictive analytics, and automated decision-making. Timely knowledge of crop conditions enables farmers to optimize management strategies, improve produce quality, increase profits, and reduce environmental impact [2, 6]. Among PA applications, automated fruit detection and yield estimation are particularly important, as they can reduce reliance on labor-intensive manual surveys while improving accuracy and scalability.

Unoccupied aerial vehicles (UAVs) have emerged as a cost-effective and flexible tool for PA. Compared to ground-based methods, UAVs provide efficient coverage of large commercial orchards, enabling frequent monitoring with minimal human labor. UAVs have been successfully deployed for orchard health monitoring and yield estimation [1, 8], utilizing RGB imagery, photogrammetry, and remote sensing. However, UAV-based approaches face challenges including lower resolution relative to ground images, variability in lighting conditions, and cluttered backgrounds of foliage and infrastructure. These factors complicate reliable detection of fruits, especially when they are occluded, clustered, or partially hidden.

Traditional methods for fruit yield estimation often rely on 2D imagery combined with calibration targets or depth cues. While effective at small scale, these approaches add complexity and cost, and they struggle in orchards where fruits overlap. Ground-based 3D sensing methods can provide more accurate spatial reconstructions



This work is licensed under a Creative Commons Attribution 4.0 International License.
SAC '26, Thessaloniki, Greece
© 2026 Copyright held by the owner/authors(s).
ACM ISBN 979-8-4007-2294-3/2026/03
<https://doi.org/10.1145/3748522.3779768>

[15], but they are typically time-consuming, labor-intensive, and difficult to scale across commercial orchards. Similarly, machine-learning models based on 2D object detection (e.g., YOLO) are constrained by limited labeled datasets, most of which focus on ground-level leaf images, reducing their applicability to UAV-derived data. These limitations highlight the need for UAV-based approaches that combine wide-area coverage and high-resolution sensors with robust fruit detection and yield estimation.

To address these challenges, we propose a modular framework for automated apple counting, detection, and yield estimation from UAV-derived 3D point clouds (a collection of points in 3-dimensional space). Particularly when developed using both ground-based and overhead imagery, high-resolution point clouds can contain multiple points for each piece of fruit. The challenge lies in filtering out the many non-fruit points and in estimating fruit counts and sizes even when many pieces are not fully resolved. Our framework integrates two complementary approaches: (1) regression modeling on red point distributions to predict total apple counts with high accuracy, and (2) clustering with spatial object approximation (e.g., minimum bounding box (MBB), sphere, or ellipsoid) to validate cluster geometry and estimate fruit size. By combining these methods, the hybrid framework links regression-based counts with clustering-based spatial analysis, providing a foundation for yield estimation.

Unlike conventional 2D object detectors that rely on RGB-frame bounding boxes, our framework leverages the full 3D structure of UAV-derived point clouds to detect apples that are clustered, occluded, or partially hidden. By combining spatial clustering with geometric approximation, it filters out foliage, branches, and trellis elements while recovering per-fruit geometry.

Our key contributions are:

- (1) We present a modular UAV-based framework that integrates color filtering, regression, clustering, and geometric approximation for apple detection and yield estimation.
- (2) We evaluate model performance against a virtual ground truth, section-level apple counts obtained by human annotation of the UAV-derived point clouds, showing that regression captures overall counts more accurately, while clustering provides spatial localization and per-fruit statistics.
- (3) We propose a hybrid strategy that combines regression's counting accuracy with clustering's interpretability to support future yield estimation, and release the framework as open-source to encourage broader adoption in precision agriculture.

The framework was developed in collaboration with a large fruit producer, demonstrating its potential relevance for commercial orchard management. Validation relied on two complementary references: *field-measured apple sizes* guided the geometric thresholds for identifying plausible fruit clusters, while *virtual ground truth* from human annotation of UAV-derived 3D point clouds served as the benchmark for evaluating regression and clustering models in the absence of harvest-weight data.

2 Related Work

Fruit detection and yield estimation have been widely studied across computer vision, remote sensing, and precision agriculture, yet

challenges remain in achieving accuracy under occlusion, variable lighting, and dense canopies. Existing approaches span from classical image-based methods to modern deep learning and 3D sensing frameworks. This section reviews prior work in two key areas, fruit count estimation and fruit yield estimation, including their evolution, limitations, and relevance to our proposed UAV-based 3D framework.

2.1 Fruit Count Estimation

Accurate fruit counting is a key step for yield monitoring, yet remains challenging due to occlusion, clustering, and variability in fruit appearance across growth stages. Early studies relied on classical computer vision and machine learning, while recent works increasingly adopt deep learning for improved robustness.

Classical approaches combined handcrafted features with shape- or color-based detection. For example, [14] integrated Support Vector Machines (SVMs), the Hough Transform, and spatial masks to detect oranges, achieving 97% accuracy and a 5% error margin against manual counts. Similarly, [23] applied color segmentation (RGB/HSI) with connected component analysis for apple counting, achieving $R^2 = 0.85$ and RMSE of 20 fruits per tree during ripening.

Deep learning has enabled more flexible and scalable solutions. [22] proposed MangoYOLO with Kalman filtering and the Hungarian Algorithm for tracking, achieving 62% consistency with actual harvest counts, outperforming dual-view imaging (40%). For tomatoes, [4] combined YOLOv5 with Deep SORT, reporting high accuracy for green tomatoes (99%) but lower for flowers (50%), reflecting challenges of small object detection. These examples illustrate how integrating detection with temporal tracking addresses occlusion and continuity.

Collectively, prior work demonstrates a spectrum of approaches from classical vision to deep learning pipelines. While YOLO-based methods achieve strong results in 2D imagery, they often rely on high-quality labeled datasets and do not directly leverage 3D spatial structure, which limits performance in orchard environments with occlusions and overlapping fruit.

2.2 Fruit Yield Estimation

Fruit yield prediction extends beyond counting, requiring integration of spatial, temporal, and environmental factors. A wide range of machine learning (ML) models have been explored, from regression and fuzzy logic to ensemble methods and neural networks.

Classical approaches include fuzzy logic with decision trees for tomato ripeness estimation [7] and SVMs predicting apple yield using fruit count and canopy features in YCbCr images [11]. Regression-based models also remain common: [19] applied Multiple Linear Regression on UAV-derived NDVI and canopy volume for olive orchards ($R^2 = 0.60$), while [3] showed Random Forest outperformed the mechanistic CASA model for apple yield estimation ($R^2 = 0.71$ vs. 0.57).

Recent studies highlight the advantages of advanced ML. Hybrid CNN-LSTM models have predicted rice yields with $R^2 = 0.86$ [12], while XGBoost achieved $R^2 = 0.94$ for wild blueberry yield [17]. Gradient Boosting, Random Forest, and Partial Least Squares Regression have been tested on UAV and ground data, with PLSR

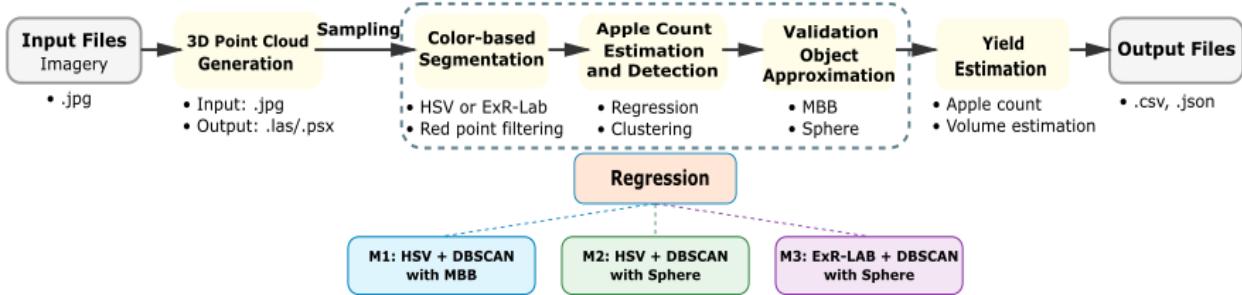


Figure 1: Apple detection and volume estimation pipeline. The framework is modular and supports multiple configurations

reaching a MAPE of 23.45% for tree-level yield [21]. Deep learning also shows strong potential: CNN-LSTM ensembles captured temporal dependencies for fruit yield forecasting with $R^2 = 0.85$ [18], and BPNNs integrated vegetation indices with UAV-derived morphological features, reaching R^2 between 0.83 and 0.88 [20]. For corn yield, deep learning surpassed SVM and tree-based methods, demonstrating scalability to large datasets [13].

Overall, yield estimation research demonstrates that ML approaches, particularly ensemble and deep learning models, outperform traditional mechanistic or regression models by capturing nonlinear relationships, temporal dynamics, and spatial variability. UAV and satellite imagery further expand these capabilities, enabling scalable, data-driven forecasting across diverse crops.

Recent work has attempted to explicitly link fruit counting with yield estimation by combining filtering, clustering, and geometric modeling. Authors in [10] proposed a UAV-based pipeline that applies HSV color filtering, DBSCAN clustering, and sphere fitting to detect apples and estimate yield. To handle oversized clusters, they introduced a secondary K-means step to split groups more than twice the average cluster size. While conceptually promising, such two-stage clustering remains highly sensitive to hyperparameters and consistent performance gains are difficult to demonstrate. Our framework explores similar ideas, but we found that the effect of hyperparameter tuning outweighed that of post-processing refinements, leading us to exclude the secondary clustering step from our main configurations.

3 Methods

Our primary goal is to accurately estimate apple counts from UAV-derived 3D point clouds, providing a foundation for yield estimation. We propose a modular framework that integrates two complementary strategies: (1) regression that predicts apple counts directly from the number of filtered red points (2) clustering with spatial object approximation for grouping red points into candidate apple clusters and deriving indicative volume and mass estimates.

3.1 System Overview

The proposed framework follows a five-phase pipeline, illustrated in Figure 1. The phases are:

- (1) 3D point cloud generation: UAV imagery of orchard sections is processed using photogrammetry software to generate dense 3D point clouds.
- (2) Color-based segmentation: Red points consistent with apple skin are isolated using a filtering method.
- (3) Apple count estimation and detection: Two complementary approaches are applied: regression modeling and clustering-based detection.
- (4) Validation and object approximation: Clusters are validated using geometric approximations (e.g., MBB, sphere). This step filters out false positives such as branches or support structures and provides per-apple size estimates.
- (5) Yield estimation: Apple volumes are estimated from validated approximated objects of clusters using sphere or ellipsoid models. These per-apple volumes can be scaled to orchard-level yield by applying density values and field coverage ratios.

This modular design supports multiple configurations, including M1: HSV + DBSCAN with MBB, M2: HSV + DBSCAN with Sphere, and M3: ExR-LAB + DBSCAN with Sphere. In addition, regression-based count estimation can be combined with geometric statistics from clustering to achieve both high-accuracy fruit counts and realistic yield estimation.

3.2 3D Point Cloud Generation

High-resolution 3D point clouds were reconstructed from UAV imagery (see Section 4.1) using Agisoft Metashape Professional (version 2.1.2.18548). A total of 80 JPEG images captured by the DJI Mavic 3M camera system were aligned using the “Highest” accuracy setting with both generic and reference preselection enabled. The key-point and tie-point limits were set to 40,000 and 4,000, respectively, producing a sparse cloud of 63,212 tie points after alignment. Misaligned and spurious points were manually removed to improve geometric accuracy.

Dense reconstruction was then carried out at the “Ultra High” quality setting with “Moderate” depth filtering, yielding a final model of approximately 2.59 million points. The project was georeferenced to the WGS 84 (EPSG:4326) coordinate system using on-board UAV GPS data. These point clouds provided canopy and fruit structures at sufficient resolution to support subsequent filtering, clustering, and object approximation. An example raw UAV image

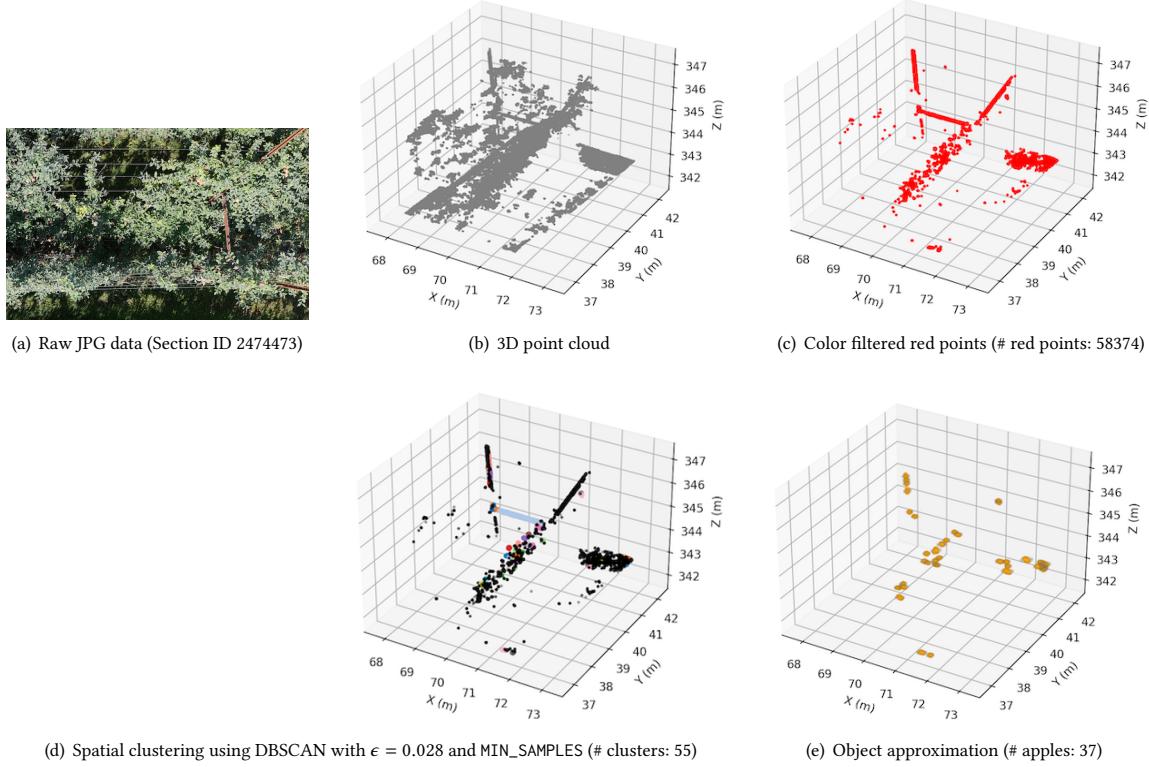


Figure 2: An example of 3D point cloud, HSV filtering, clustering and apple detection using MBB (M1)

and the corresponding reconstructed point cloud for a sampled orchard section are shown in Figure 2 (a) and (b). All point-cloud coordinates were re-projected to a local metric coordinate system prior to any metric-based operations such as DBSCAN ϵ , MBB side lengths, or space-diagonal.

3.3 Color-based Segmentation

Color-based segmentation was applied as the first step to reduce the search space by filtering out foliage and background points, leaving only apple-like candidates. Two methods were integrated into the framework:

- HSV filtering: Operates in the hue–saturation–value color space, separating chromatic and brightness components. This representation provides broad robustness to orchard lighting variation and preserves most apple-skin candidates.
- ExR-LAB filtering: A two-stage approach that first applies the Excess Red (ExR) index and then refines the selection with constraints in the LAB color space. This method is more selective, effectively suppressing background foliage and non-fruit structures, though it tends to yield fewer candidate points.

Parameter ranges for both methods are summarized in Table 1. Figure 2 (c) illustrates an example of red points obtained after filtering. The comparative performance of HSV and ExR-LAB is evaluated in Section 4.3.

3.4 Apple Count Estimation and Detection

Our framework addresses apple count estimation through two complementary strategies: (1) regression modeling based on the number of red points, and (2) spatial clustering with object approximation. Both approaches were evaluated against the human-annotated section-level counts established as virtual ground truth (Section 4.1), which serve as the reference for assessing model performance.

3.4.1 Regression modeling. Regression models predict section-level apple counts directly from the number of filtered red points (Table 3). This approach is computationally efficient and exploits the strong empirical correlation between red-point density and visible apples in each section. However, it provides only global counts without localizing apples or characterizing their geometry, volume, or mass. While regression generally produces more accurate count estimates than clustering, it cannot by itself support spatial analysis or yield estimation.

3.4.2 Clustering and object approximation. To recover spatial structure, we group red points into candidate clusters using Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [5]. DBSCAN parameters are tuned to reflect apple-scale density, with ϵ controlling the neighborhood radius and MIN_SAMPLES the minimum number of neighbors. Figure 2 (d) illustrates the resulting clusters.

Algorithm 1 M1: Apple Count Estimation from 3D Point Cloud using HSV + DBSCAN with MBB Approximation

```

1: Input: 3D point cloud  $\mathcal{P} = \{(x_i, y_i, z_i, c_i)\}_{i=1}^N$  with RGB color  $c_i$ 
2: Parameters: HSV thresholds  $H_{\text{low\_wrap}}, H_{\text{high\_wrap}}, S_{\text{min}}, S_{\text{max}}, V_{\text{min}}, V_{\text{max}}$ ; DBSCAN ( $\epsilon, \text{MIN\_SAMPLES}$ ); cube-likeness ratio threshold  $\text{CUBE\_RATIO\_MIN}$ ; diameter window  $[D_{\text{min}}, D_{\text{max}}]$ 
3: Output: Set of accepted apple clusters  $\mathcal{A}$  with MBBs
4: // Color filtering to keep “red” points
5:  $\mathcal{R} \leftarrow \emptyset$ 
6: for each point  $(x_i, y_i, z_i, c_i) \in \mathcal{P}$  do
7:   Convert  $c_i$  (RGB) to  $(H_i, S_i, V_i)$  in HSV
8:   if  $(H_i > H_{\text{high\_wrap}} \vee H_i < H_{\text{low\_wrap}})$  and  $S_{\text{min}} \leq S_i \leq S_{\text{max}}$  and  $V_{\text{min}} \leq V_i \leq V_{\text{max}}$  then
9:      $\mathcal{R} \leftarrow \mathcal{R} \cup \{(x_i, y_i, z_i)\}$ 
10:  end if
11: end for
12: //  $\mathcal{R}$  is the red-point set (apple-skin candidates)
13: // Cluster red points with DBSCAN
14:  $C \leftarrow \text{DBSCAN}(\mathcal{R}; \epsilon, \text{MIN\_SAMPLES})$ 
15: // returns clusters  $\{C_1, \dots, C_K\}$ 
16:  $\mathcal{A} \leftarrow \emptyset$ 
17: for each cluster  $C \in C$  do
18:   Compute MBB side lengths  $d_x, d_y, d_z$  and space-diagonal  $D = \sqrt{d_x^2 + d_y^2 + d_z^2}$ 
19:   // Cube-likeness and size tests
20:   Let  $s_{\text{min}} = \min(d_x, d_y, d_z)$ ,  $s_{\text{max}} = \max(d_x, d_y, d_z)$ 
21:   if  $s_{\text{min}}/s_{\text{max}} \geq \text{CUBE\_RATIO\_MIN}$  and  $D_{\text{min}} \leq D \leq D_{\text{max}}$  then
22:      $\mathcal{A} \leftarrow \mathcal{A} \cup \{(C, \text{MBB}(C))\}$ 
23:   end if
24: end for
25: // Apple counting: one accepted cluster  $\Rightarrow$  one apple
26:  $\text{APPLE\_COUNT} \leftarrow |\mathcal{A}|$ 
27: return  $\mathcal{A}$  and  $\text{APPLE\_COUNT}$ 

```

Not all clusters correspond to apples: some are too small (noise) or too large (merged fruit). To validate clusters, each is approximated with an axis-aligned MBB with side lengths (d_x, d_y, d_z) and space diagonal D :

- (1) *Cube-likeness (ratio-based).* Let $s_{\text{min}} = \min\{d_x, d_y, d_z\}$ and $s_{\text{max}} = \max\{d_x, d_y, d_z\}$. A cluster is considered apple-like only if $s_{\text{min}}/s_{\text{max}} \geq \text{CUBE_RATIO_MIN}$ (e.g., 0.5). This rejects highly elongated clusters.
- (2) *Size window.* The space diagonal D (MBB body diagonal) must fall within $D_{\text{min}} - D_{\text{max}}$, guided by field-measured apple sizes (Section 4.1) and empirical tuning.

Each accepted MBB corresponds to one apple. Figure 2 (e) shows accepted clusters after filtering. Algorithm 1 summarizes the procedure for M1 (HSV + DBSCAN with MBB). A summary of accepted and rejected cluster statistics is provided in Table 4.

Remark: The framework is modular: color filtering can use HSV or ExR-LAB, and object approximation can use MBB, sphere, or ellipsoid. M1–M3 represent specific instantiations of this pipeline. While clustering-based approaches achieve only moderate performance in recovering section-level counts (best relative error ≈ 0.21 for M1), they provide spatial localization and per-fruit geometric attributes essential for subsequent volume and mass based yield extrapolation.

3.4.3 Hybrid Estimation Framework. Regression models establish section-level counts by leveraging the strong red-point-to-apple relationship, while clustering with geometric validation recovers per-apple statistics and spatial distribution. We integrate the two methods in a hybrid framework:

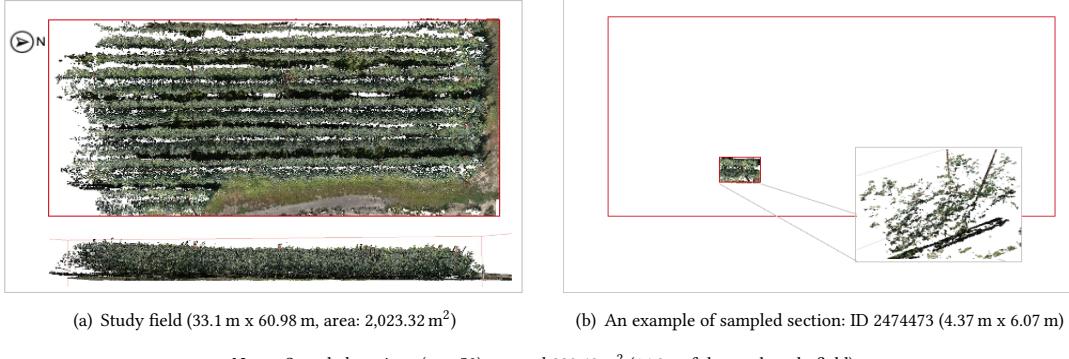
- (1) Count estimation: Regression models anchor the total number of apples in each section.
- (2) Geometric context: Clustering provides per-fruit attributes (size, volume, mass) and spatial distribution.
- (3) Yield extrapolation: The mean per-apple volume and mass from clustering are combined with regression-based counts to extrapolate section- and field-level yields.

This hybrid approach capitalizes on the complementary strengths of the two methods: regression for consistent section-level counts and clustering for geometric and spatial context. Together, they form a flexible foundation for UAV-based apple detection and exploratory yield estimation.

3.5 Apple Volume and Mass Estimation

Beyond apple counting, we estimate apple volume and harvest yield. Each accepted MBB is approximated by either:

- Sphere: The largest inscribed sphere with radius based on the smallest MBB side.

**Figure 3: Test site and an example of sampled sections**

- Ellipsoid: A prolate ellipsoid with semi-axes (a, b, c) , with $a = b$ and elongation factor $\alpha \in [1.1, 1.2]$ for Fuji apples.

Total volume is obtained by summing per-apple volumes across sampled sections, scaling by the field coverage ratio, and multiplying by average apple density ($\rho = 850\text{--}950 \text{ kg/m}^3$ [15]). This yields an estimate of harvest mass. Algorithm 2 summarizes the procedure.

Algorithm 2 Apple Volume & Mass Estimation

- 1: **Input:** Accepted MBBs of apple clusters
 - 2: **Parameters:** Choice of shape model (sphere or ellipsoid), apple density ρ , coverage ratio γ
 - 3: **Output:** Per-apple volumes, total estimated volume, and total mass of apples
 - 4: **for** each MBB in the accepted set **do**
 - 5: Estimate apple volume using the selected model (sphere or ellipsoid)
 - 6: **end for**
 - 7: Compute total volume of apples in sampled sections by summing per-apple volumes
 - 8: Scale the total volume to the field level using the coverage ratio γ
 - 9: Multiply the field-scale volume by density ρ to obtain total mass estimate
 - 10: **return** Per-apple volumes, sample and field-scale totals, and mass estimates
-

4 Experiments

4.1 Study Site and Data Collection

Data collection was carried out in a commercial Fuji apple orchard located in USA. UAV flights were conducted on September 1, 2024, under clear skies and calm wind conditions. The orchard employed a V-shaped trellis system, which provided structured rows suitable for aerial and ground-based imaging.

Imagery was acquired using a DJI Mavic 3M at three different altitudes and viewing angles. Two flights were conducted at 12 m

Table 1: Parameter settings and values used in experiments

Method	Parameter	Values
HSV	H (hue)	$H < 0.10$ or $H > 0.96$
	S (saturation)	$0.20 \leq S \leq 0.80$
	V (value)	$0.50 \leq V \leq 1.00$
ExR-LAB	Red threshold (ExR)	0.20
	a^* (LAB)	$10 \leq a^* \leq 60$
DBSCAN	Neighbor radius (ε)	0.028
	MIN_SAMPLES	136
MBB approx.	Space diagonal (D)	$4 \leq D \leq 20 \text{ cm}$
	CUBE_RATIO_MIN	≥ 0.50
Sphere approx.	Radius (R)	$3 \leq R \leq 10 \text{ cm}$

altitude, one with a nadir camera angle and the other at an oblique angle (80° from horizontal). Flight lines were oriented perpendicular to the rows for nadir imagery and parallel to the rows for oblique imagery. Additional low-altitude images were collected manually at 2–3 m above the canopy with a 45° oblique angle. The ground resolution of the 12 m flights was 0.55 cm. Figure 3 (a) shows the test site. Ground-based imagery was also acquired by walking the UAV through orchard rows at approximately 1 m above ground, alternating between left- and right-facing perspectives. A higher-altitude (30 m) flight with a 75° oblique angle and pixel resolution of 1.4 cm was conducted for orthophotograph generation but not used in point-cloud reconstruction.

The dataset for modeling consisted of 50 individual 3D point-cloud samples, each representing a section of the orchard canopy (see an example in Figure 3 (b)). Each point was defined by 3D spatial coordinates and RGB color values.

Field size reference for geometric thresholds. To guide geometric filtering of clusters (e.g., the MBB space-diagonal range [D_{\min}, D_{\max}]), we collected field measurements of apple size. Forty-seven apples were measured within a ~ 15 m segment of a single row, using trellis wires to define zones; measurements were taken between the third and fourth horizontal wires (approximately 2–2.6 m above ground) to the nearest centimeter. The sampled apples

Table 2: Filtering performance of HSV and ExR-LAB methods (3D point cloud red-point preservation)

Color filter	Study field (260,430,642 pts, 2,023.32 m ²)	Sampled sections (52,402,836 pts, 339.49 m ²)	
	Filtered red pts	Filtered red pts	% of study-field red points
HSV	2,593,894	473,065	18.23%
ExR-LAB	589,343	145,552	24.70%

had an average width of 6.7 cm ($SD = 0.9$ cm) and an average height of 7.1 cm ($SD = 1.0$ cm), providing biologically realistic bounds for D_{\min} and D_{\max} to retain apple-sized clusters while excluding clusters that were too small or too large to represent individual fruit.

Evaluation ground truth for apple counts. For supervised evaluation of apple-count performance, we established a section-level virtual ground truth. Four trained raters independently counted all visible apples in each of the 50 sampled sections using the 3D point-cloud visualizations; discrepancies were resolved through discussion to reach a consensus. These human-annotated section-level counts, totaling 759 apples over 339.49 m² of sampled area, serve as the reference against which all regression- and clustering-based apple-count models were evaluated.

4.2 Experimental Setup

Experiments were implemented in Python 3.9.13, integrating Open3D and scikit-learn for 3D processing and clustering, and TensorFlow for regression modeling. The dataset and source code are publicly available in a GitHub repository [16], supporting reproducibility and future research.

Model performance for apple counting was evaluated against the section-level virtual ground truth defined in Section 4.1, and geometric thresholds such as the MBB space-diagonal range [D_{\min}, D_{\max}] were guided by the field-measured apple-size reference reported in the same section.

Regression models, such as decision tree (DT), random forest (RF), K-nearest neighbor (KNN), gradient boosting (GB), and neural networks (NN), were trained on red-point distributions extracted by color filtering. A 70/30 train–test split was performed at the section level (avoiding data leakage), with random seed set to 42 for reproducibility. Model performance was assessed against the virtual ground truth using relative error (RE), root-mean-squared error (RMSE), and the coefficient of determination (R^2). RE is reported as the mean absolute error percentage across test sections.

Hyperparameters for HSV and ExR-LAB filtering, DBSCAN clustering, and geometric approximations were tuned empirically to balance sensitivity and robustness under orchard conditions. Thresholds for the MBB space diagonal D were adjusted to remain consistent with field-measured apple sizes while excluding clusters that were too small or too large to represent individual fruit. A summary of hyperparameters is provided in Table 1.

4.3 Color Filtering

We evaluated the performance of HSV and ExR-LAB filtering methods introduced in Section 3.3. Table 2 summarizes the results across the dataset. HSV filtering retained a larger proportion of points overall, ensuring that most apple-skin candidates were preserved, but also introduced more false positives from leaves and branches.

In contrast, ExR-LAB achieved higher precision by reducing background noise, though at the expense of discarding some true apple points under challenging lighting conditions.

These complementary characteristics are important for downstream tasks. HSV filtering provided a stronger foundation for regression-based apple count estimation, where retaining sufficient apple-related points is critical for predictive accuracy. ExR-LAB filtering, by contrast, was more effective for clustering-based detection, where minimizing false positives improves geometric validation and reduces spurious apple candidates. In our experiments, HSV was found to perform better overall for regression and clustering in M1 (HSV + DBSCAN + MBB), while ExR-LAB supported more precise cluster validation in M2 and M3 (see Table 5).

Table 3: Regression performance on apple count estimation using color-filtered 3D point cloud data

Color filter	Regression	RE	RMSE	R^2
HSV	DT	0.036	1.133	0.969
	RF	0.033	0.744	0.988
	KNN	0.106	2.155	0.898
	GB	0.036	1.093	0.974
	NN	0.022	0.417	0.996
ExR-LAB	DT	0.062	1.186	0.961
	RF	0.060	1.260	0.967
	KNN	0.151	2.293	0.892
	GB	0.053	1.278	0.966
	NN	0.026	0.350	0.997

Notes: Best-performing values for each color filter are shown in bold.

4.4 Apple Count Estimation and Detection

Apple count estimation was performed using two complementary strategies: regression modeling on red point distributions and clustering with geometric validation. Together, these approaches balance predictive accuracy with interpretable object-level detection.

4.4.1 Regression modeling. Regression models used a single predictor, the total number of color-filtered red points per sampled section, capturing the relationship between red-point abundance and apple counts. Table 3 summarizes the performance of five regression models trained on red point distributions filtered by HSV and ExR-LAB. Across both color filtering methods, regression achieved consistently high accuracy, with NN performing best ($R^2 \approx 0.996$ –0.997, $RE \leq 0.026$). RF and GB also produced strong results ($R^2 > 0.96$), while KNN underperformed relative to other methods. These results demonstrate that regression is a reliable strategy for predicting total apple counts directly from point distributions, provided that sufficient candidate points are retained.

Table 4: Mean dimensions of MBBs for accepted and rejected clusters

Group	x mean (m)	y mean (m)	z mean (m)	x:y:z ratio (x = 1)	MBB D mean (cm)
Rejected (< min threshold)	6.2	4.4	3.1	1:0.752:0.556	3.6
Accepted	7.5	6.0	4.4	1:0.811:0.623	6.0
Rejected (> max threshold)	67.3	25.8	14.6	1:0.475:0.255	35.9

Table 5: Apple count estimation: accuracy on sampled sections and total estimated apples

Color	Model	Cluster	Obj. approx.	50 sampled sections (339.49 m ²)				Study field (2,023.32 m ²)
				RE	RMSE	R ²	Estimated apples	
HSV	M1	DBSCAN	MBB	0.21	3.49	0.75	698 (RE = 0.0804)	3,828*
	M2	DBSCAN	Sphere	0.23	4.08	0.66	657 (RE = 0.1344)	3,603*
	M3	DBSCAN	Sphere	0.22	4.64	0.58	694 (RE = 0.0856)	2,809**

Notes: Ground-truth total for sampled sections is 759 apples.

* Field-level estimates scaled using HSV red-point sampling ratio of 18.23%.

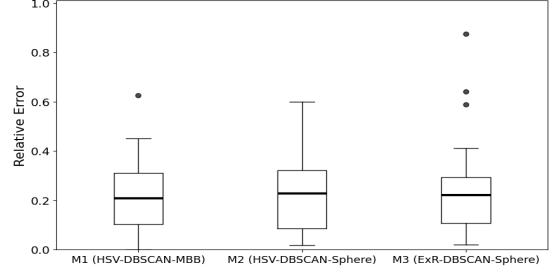
** Field-level estimates scaled using ExR-LAB red-point sampling ratio of 24.60%.

4.4.2 MBB size and rejection thresholds. To refine clustering-based detection, we analyzed the geometric characteristics of MBBs. Table 4 reports mean values of side lengths (x , y , z) and overall space diagonal D for accepted clusters, as well as those rejected for being too small or too large. Accepted MBBs had a mean D of approximately 6 cm, consistent with ground-truth apple sizes (Section 4.1), whereas rejected clusters deviated substantially in scale or aspect ratio. Clusters below the minimum threshold often represented noise or partial points, while those exceeding the maximum threshold corresponded to foliage or structural elements in the trellis. This filtering step thus improves robustness by discarding spurious detections before volume or count estimation.

4.4.3 Clustering with object approximations. Beyond regression, spatial localization was achieved using DBSCAN clustering combined with geometric approximations. Three configurations were evaluated: M1 (HSV + DBSCAN + MBB), M2 (HSV + DBSCAN + Sphere), and M3 (ExR-LAB + DBSCAN + Sphere). Results are shown in Table 5. M1 achieved the best performance ($R^2 = 0.75$, RE = 0.21, RMSE = 3.49), indicating that MBBs provided a reasonable approximation of per-fruit geometry. M2 and M3 yielded lower accuracy ($R^2 = 0.66$ and 0.58, respectively), reflecting the sensitivity of clustering and object fitting to color filtering and geometric assumptions. Figure 4 illustrates the distribution of relative errors, with M1 showing tighter bounds compared to M2 and M3.

Scaling to total counts, sampled sections contained 657–698 apples, and extrapolations to the study field ranged from 2,809 (M3) to 3,828 (M1). Although less accurate than regression-based models (Table 3), clustering provides spatial localization and per-fruit geometry that regression alone cannot. This complementarity supports our hybrid estimation strategy, that balances predictive accuracy with interpretable object-level validation. In addition, clustering-derived geometry directly supports the volume and mass estimation.

4.4.4 Hybrid estimation strategy. Regression and clustering provide complementary strengths for apple count estimation. Regression on red point distributions achieved near-perfect accuracy

**Figure 4: Comparison of relative errors across clustering configurations M1–M3.**

($R^2 \approx 0.99$), ensuring reliable overall yield prediction. In contrast, clustering with geometric validation offered spatial localization and per-fruit geometry, though with moderate accuracy ($R^2 = 0.58$ –0.75) due to sensitivity to filtering and object fitting. By integrating these two approaches, the framework balances predictive precision with interpretable, object-level validation. This hybrid strategy not only improves robustness to orchard variability but also establishes a foundation for the volume and mass estimation presented in Section 4.5.

4.5 Apple Yield Estimation

Following apple counts and shape approximation, the framework extends to yield estimation by first estimating per-fruit volumes and then aggregating them to the orchard scale. Accepted clusters identified in Section 3.4 were fitted with geometric models (sphere and ellipsoid) to estimate individual fruit volumes. These volumes were then summed to obtain total fruit volume, which was converted to mass using a reported apple density range of 850–950 kg/m³ [15].

Figure 5 shows density-based histograms of estimated per-fruit volumes for accepted clusters using three geometric models: (a) MBBs, (b) spheres, and (c) ellipsoids. In each panel, the orange curve represents a kernel density estimate (KDE) summarizing the underlying distribution. All three methods exhibit broadly similar

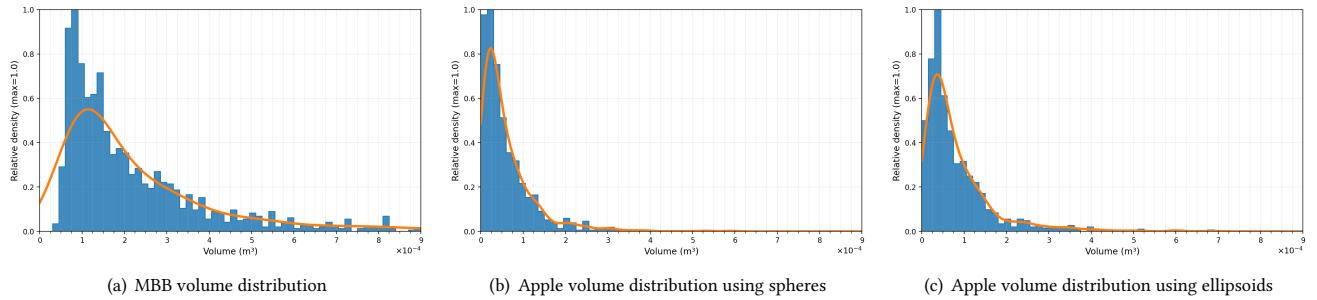


Figure 5: Volume distributions of accepted apple clusters using three geometric models based on M1 detections.

Table 6: Total apple volume and mass estimation using clustering- and regression-based methods

Scale	Method	Volume approximation	Volume (m^3)		Mass (kg)	
			Min	Max	Min	Max
Sampled sections	M1	Sphere	0.0428	36.38	40.66	
		Ellipsoid	0.0573	48.71	54.44	
	Regression (NN + M1 mean vol.)	Sphere	0.0455	38.67	43.22	
		Ellipsoid	0.0609	51.78	57.87	
Field-level extrapolation	M1	Ellipsoid	0.3143	257.73	298.59	
		Hybrid (Sphere+Ellipsoid)	0.2918	248.03	277.21	
	Regression (NN + M1 mean vol.)	Ellipsoid	0.3340	283.90	317.30	

Notes: Field-level extrapolation using HSV red-point sampling ratio of 18.23%.

right-skewed distributions across the dataset. As expected, MBBs in panel (a) tend to overestimate volumes because their cubic shape encloses more empty space, while ellipsoids in panel (c) yield tighter fits that better reflect fruit-like morphology. Spheres in panel (b) provide intermediate estimates, balancing compactness with simplicity of representation. These complementary behaviors illustrate how different geometric models capture fruit morphology at varying levels of approximation and guided the choice of ellipsoids for the yield estimates reported in Table 6.

Table 6 summarizes the aggregated yield estimates. For clustering-based approaches, MBBs produced inflated totals (141–158 kg within the sampled sections), while spheres and ellipsoids yielded more conservative estimates of 36–41 kg and 49–54 kg, respectively. Ellipsoids provided the most consistent values given their tighter geometric fit. To extrapolate to field-level yield, we applied the HSV red-point sampling ratio of 18.23% (and 24.60% for ExR–LAB where relevant). In addition, regression-based estimation was performed by combining neural network apple count predictions ($R^2 = 0.997$, relative error = 0.026) with the mean per-fruit volume from M1. This hybrid strategy integrates the accuracy of regression counts with the geometric realism of clustering-based volumes, producing scalable yield estimates at the orchard level.

Extrapolating to the orchard level using the sampling ratio, clustering with ellipsoids produced 258–299 kg, while regression-based hybrid approaches estimated 248–317 kg depending on the volume model. The close agreement across these methods shows the robustness of the hybrid regression–clustering strategy. Although

ground-truth harvest weights were not available for direct validation, the consistency of per-fruit distributions (Figure 5) and convergence of independent estimation methods suggest that UAV-derived point clouds can support reliable, scalable yield forecasting in orchard settings.

5 Discussion

During experiments, some apple-like clusters were rejected because their bounding boxes exceeded the maximum allowable size threshold ($D_{\max} = 20$ cm in our setup). These large clusters often represent merged apples or partially occluded fruit groups that DBSCAN failed to separate. Figure 6 (a) illustrates such cases, where the resulting MBBs covered multiple apples and were therefore excluded from count estimation, while similar patterns in Figure 5 (b) could instead reflect irregular structures. Such ambiguity highlights the need for further analysis and refinement.

Even with 3D reconstructions, occlusion remains a persistent limitation in dense canopies, particularly where overlapping fruit clusters and foliage reduce point density. A related limitation is that human-annotated counts may *underestimate* the true number of apples, especially when green or partially occluded fruit are difficult to distinguish from foliage in the point clouds. This suggests that the virtual ground truth may be conservative in dense conditions and that future work should incorporate improved approaches for identifying non-red fruit, such as color normalization, ripeness-aware filtering, or multi-date image capture.

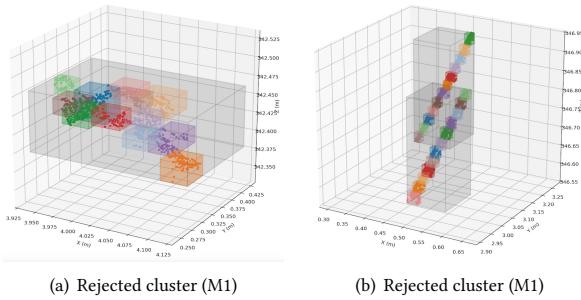


Figure 6: Rejected large MBBs and K-means clustering (M1)

One possible strategy to recover apples from these rejected clusters is a second-stage clustering step (e.g., K-means [9]), applied when a cluster exceeds D_{\max} . In preliminary tests, this refinement slightly improved M3 (ExR-LAB + DBSCAN + Sphere) but reduced accuracy in M1 and M2, indicating that its benefit is highly sensitive to DBSCAN and K-means hyperparameters. Future work will explore adaptive cluster-splitting strategies that can be integrated robustly into the pipeline.

We also note that the kernel density curves in Figure 5 exhibit a pronounced right-tail, indicating a small number of unusually large accepted clusters. Although these clusters meet the current D -based acceptance criteria, they likely reflect merged or irregular fruit groupings. This highlights the need for further refinement of geometric validation and cluster-splitting to address such outliers.

These findings are broadly consistent with those of [10], who combined DBSCAN with K-means to refine large clusters and reported R^2 values of 0.41–0.53 for apple counts and 0.56–0.76 for yield estimation at different UAV altitudes. In our case, regression on red-point distributions provided higher section-level agreement with the virtual ground-truth counts (up to $R^2 \approx 0.99$), while clustering with geometric approximation achieved moderate agreement ($R^2 \approx 0.58$ –0.75) but contributed essential spatial and size information. Because no harvest-weight data were available, we regard the derived volume- and mass-based yield as indicative rather than validated, underscoring the need for future field-level calibration.

6 Conclusion

We presented a modular UAV-based framework for apple detection and counting from 3D point clouds, with an initial extrapolative step toward yield estimation. The framework integrates color-based segmentation, regression on red-point distributions, and DBSCAN clustering with geometric approximations to capture both overall apple counts and per-fruit spatial patterns. Model performance for apple counting was evaluated against a virtual ground truth, section-level counts obtained by human annotation of UAV-derived point clouds, providing a consistent reference in the absence of harvest data. Volume and mass estimates were extrapolated from geometric models but remain unvalidated against field measurements and should be treated as indicative. Future work will focus on improving cluster separation in dense fruit regions, validating yield estimates with harvest data, and extending the framework to time-series UAV observations for tracking fruit growth.

References

- [1] Z. A. Ali, C. Yang, A. Israr, and Q. Zhu. 2023. A Comprehensive Review of Scab Disease Detection on Rosaceae Family Fruits via UAV Imagery. *Drones* 7, 2 (2023), 97. <https://doi.org/10.3390/drones7020097>
- [2] G. S. Avellar, G. A. Pereira, L. C. Pimenta, and P. Iscold. 2015. Multi-UAV Routing for Area Coverage and Remote Sensing with Minimum Time. *Sensors* 15, 11 (2015), 27783–27803. <https://doi.org/10.3390/s151127783>
- [3] X. Bai, Z. Li, W. Li, Y. Zhao, M. Li, H. Chen, S. Wei, Y. Jiang, G. Yang, and X. Zhu. 2021. Comparison of machine-learning and casa models for predicting apple fruit yields from time-series planet imageries. *Remote Sensing* 13, 16 (2021), 3073.
- [4] Y. Egi, M. Hajyzadeh, and E. Eyceyurt. 2022. Drone-computer communication based tomato generative organ counting model using YOLO V5 and deep-sort. *Agriculture* 12, 9 (2022), 1290.
- [5] M. Ester, H. Kriegel, J. Sander, and X. Xu. 1996. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)*, 226–231.
- [6] J. Gené-Mola, V. Vilaplana, J. R. Rosell-Polo, J.-R. Morros, J. Ruiz-Hidalgo, and E. Gregorio. 2019. Multi-modal deep learning for Fuji apple detection using RGB-D cameras and their radiometric capabilities. *Computers and Electronics in Agriculture* 162 (2019), 689–698. <https://doi.org/10.1016/j.compag.2019.05.020>
- [7] N. Goel and P. Sehgal. 2015. Fuzzy classification of pre-harvest tomatoes for ripeness estimation—An approach based on automatic rule learning using decision tree. *Applied Soft Computing* 36 (2015), 45–56.
- [8] Antoine L. Harfouche, Farid Nakhlé, Antoine H. Harfouche, Orlando G. Sardella, Eli Dart, and Daniel Jacobson. 2023. A primer on artificial intelligence in plant digital phenomics: embarking on the data to insights journey. *Trends in Plant Science* 28, 2 (2023), 154–184. <https://doi.org/10.1016/j.tplants.2022.08.021>
- [9] J. A. Hartigan and M. A. Wong. 1979. Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 28, 1 (1979), 100–108. <https://doi.org/10.2307/2346830>
- [10] M. Hobart, M. Pflanz, N. Tsoulias, C. Weltzien, M. Kopetzky, and M. Schirrmann. 2025. Fruit Detection and Yield Mass Estimation from a UAV Based RGB Dense Cloud for an Apple Orchard. *Drones* 9, 1 (2025), 60. <https://doi.org/10.3390/drones9010060>
- [11] C. Hong, L. Damerow, M. Blanke, and S. Yurui. 2015. Early Yield Estimation of 'Gala' Apple Trees Using Image Processing Combined with Support Vector Machine. *Nongye Jixie Xuebao/Transactions of the Chinese Society of Agricultural Machinery* 46, 3 (2015).
- [12] S. Jeong, J. Ko, and J. Yeom. 2022. Predicting rice yield at pixel scale through synthetic use of crop and deep learning models with satellite data in South and North Korea. *Science of The Total Environment* 802 (2022), 149726.
- [13] N. Kim and Y. Lee. 2016. Machine Learning Approaches to Corn Yield Estimation Using Satellite Images and Climate Data: A Case of Iowa State: A Case of Iowa State. *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography* 34, 4 (2016), 383–390.
- [14] W. Maldonado Jr and J. Barbosa. 2016. Automatic green fruit counting in orange trees using digital images. *Computers and Electronics in Agriculture* 127 (2016), 572–581.
- [15] J. C. Miranda, J. Gené-Mola, M. Zude-Sasse, N. Tsoulias, A. Escolà, J. Arnó, J. R. Rosell-Polo, R. Sanz-Cortiella, J. A. Martínez-Casasnovas, and E. Gregorio. 2023. Fruit sizing using AI: A review of methods and challenges. *Postharvest Biology and Technology* 206 (2023), 112587. <https://doi.org/10.1016/j.postharbiov.2023.112587>
- [16] K. Nguyen, D. J. Patel, T. Nguyen, N. Tran, A. Riazi, J. W. Lauer, W. D. Bae, and S. Alkobaisi. [n. d.]. Apple Detection Framework. <https://github.com/baew-seattle/Apple-Detection-and-Yield-Extrapolation>, 2025.
- [17] E. Y. Obsie, H. Qu, and F. Drummond. 2020. Wild blueberry yield prediction using a combination of computer simulation and machine learning algorithms. *Computers and Electronics in Agriculture* 178 (2020), 105778.
- [18] I. Okwuchi. 2020. *Machine learning based models for fresh produce yield and price forecasting for strawberry fruit*. Master's thesis. University of Waterloo.
- [19] D. Stateras and D. Kalivas. 2020. Assessment of olive tree canopy characteristics and yield forecast model using high resolution UAV imagery. *Agriculture* 10, 9 (2020), 385.
- [20] G. Sun, X. Wang, H. Yang, and X. Zhang. 2020. A canopy information measurement method for modern standardized apple orchards based on UAV multimodal information. *Sensors* 20, 10 (2020), 2985.
- [21] V. Vijayakumar, Y. Ampatzidis, and L. Costa. 2023. Tree-level citrus yield prediction utilizing ground and aerial machine vision and machine learning. *Smart Agricultural Technology* 3 (2023), 100077.
- [22] Z. Wang, K. Walsh, and A. Koirlala. 2019. Mango fruit load estimation using a video based MangoYOLO—Kalman filter—hungarian algorithm method. *Sensors* 19, 12 (2019), 2742.
- [23] R. Zhou, L. Damerow, Y. Sun, and M. M. Blanke. 2012. Using colour features of cv.'Gala'apple fruits in an orchard in image processing to predict yield. *Precision Agriculture* 13 (2012), 568–580.