horticulturae

MDPI

Review

Volume Estimation of Agricultural Products Using 2D Images: From Laboratory to Orchard

Quan Wei [1], Danying Lei [2], Ziwei Song [2], Wei Zhao [3], Fakun Wei [4] and Hua Yin [2,3,5,*]

[1] Ministry of Education Key Laboratory of Crop Physiology, Ecology and Genetic Breeding, Jiangxi Agricultural University, Nanchang 330045, China; quanw8781@163.com
[2] School of Computer and Information Engineering, Jiangxi Agricultural University, Nanchang 330045, China; jxau_24681017@stu.jxau.edu.cn (D.L.); songzw@jxau.edu.cn (Z.S.)
[3] School of Software, Jiangxi Agricultural University, Nanchang 330045, China; zhaowei@jxau.edu.cn
[4] Development Research Institute of Testing Certification Technology, Jiangxi General Institute of Testing and Certification, Nanchang 330052, China; weifakunsir@163.com
[5] Bioengineering and Technological Research Centre for Edible and Medicinal Fungi, Jiangxi Agricultural University, Nanchang 330045, China
* Correspondence: youki1234@163.com

Abstract

Accurate and non-destructive volume estimation of agricultural products is essential for precision agriculture, yet remains challenging when transitioning from controlled laboratory conditions to complex orchard environments. Although 2D image-based volume estimation methods provide a cost-effective and scalable solution, existing studies are fragmented and lack a unified perspective on their real-world applicability. This review presents a systematic synthesis of 2D image-based volume estimation methods, explicitly framed through the laboratory-to-orchard transition. We categorized existing volume estimation approaches according to the sensing modality into monocular RGB-based approaches and depth-assisted methods, and further reviewed them based on the image processing methods. A key finding is that high-precision geometric estimation can be achieved in laboratory environments, whereas deep learning and RGB-D fusion have driven a shift from conventional geometric modeling toward data-driven and hybrid learning frameworks in orchard settings. However, 2D image-based volume estimation remains fundamentally limited by scale ambiguity, severe occlusion, and sensitivity to illumination and background variability in real orchard environment. Overall, this review provides a unified perspective for understanding volume estimation methodology across environments and offers guidance for developing robust, scalable, and field-deployable volume estimation systems for real-world agricultural applications.

Keywords: volume estimation; orchard; 2D image; post-harvest processing

Academic Editor: Weikuan Jia

Received: 21 May 2026
Revised: 23 June 2026
Accepted: 23 June 2026
Published: 25 June 2026

1. Introduction

Measurement of agricultural product volume is a crucial indicator in modern agriculture [1]. For breeders, volume serves as an important phenotypic parameter that aids in selecting superior varieties [2]. For farmers, it accurately reflects crop growth status, facilitating scientific management [3]. Moreover, given the strong correlation between volume and weight, many factories sort, package, and price agricultural products based on volume, which significantly enhances processing efficiency [4]. Additionally, individuals who are on a diet or have specific nutritional restrictions can estimate the weight of the food they consume by its volume [5]. To sum up, accurately and rapidly measuring the volume of agricultural products not only supports yield prediction and growth monitoring, but also in supporting quality classification, market pricing, post-harvest logistics optimization, and variety selection [6]. Nevertheless, traditional methods, such as the water displacement method, are often destructive, time-consuming, and impractical for large-scale, high-throughput applications [7,8].

The advent of vision-based techniques has offered a promising alternative, enabling non-contact, efficient, and high-precision volume estimation [9]. These methods can be broadly categorized into 3D reconstruction-based approaches and 2D image-based approaches. Although 3D methods can capture detailed spatial morphology, their measurement precision depends heavily on reconstruction accuracy, rather than 3D volume calculation [10]. Furthermore, 3D techniques often entail higher computational complexity and hardware costs, limiting their practicality in field conditions where lighting, occlusion, and background complexity present significant challenges [11]. In contrast, 2D image-based methods establish statistical or geometric relationships between features in a 2D image and the actual volume. These methods are typically more computationally efficient, lower in cost, and easier to deploy [12,13]. Although 2D approaches may face challenges such as occlusion and perspective distortion, their balance of accuracy, simplicity, and robustness makes them particularly suitable for a wide range of agricultural applications [14,15]. Therefore, this review focuses specifically on 2D image-based volume measurement techniques and their transition from controlled laboratory settings to complex field environments.

The application of 2D image-based volume measurement significantly differs between laboratory conditions and on-site field conditions [16,17]. Since the shooting environment can be customized in industrial applications, these systems are also considered laboratory-like scenarios. In the laboratory, the environment is controllable, standardized setups with consistent lighting, fixed backgrounds, and stable camera angles facilitate the acquisition of high-quality images [18]. Most laboratory-methods often involve using one or multiple digital cameras to capture images from specific angles [19–21]. The resulting images are then processed through grayscale conversion, noise reduction, image segmentation, and feature extraction are performed to build measurement models that map these 2D features to the measured volume [22]. Under such controlled conditions, these methods typically exhibit high measurement accuracy and repeatability. However, field environments such as orchards and farmlands introduce numerous challenges, including variable natural lighting, complex and dynamic backgrounds, partial occlusion of targets, and even equipment vibration [23,24]. The core difficulty lies in ensuring measurement accuracy and robustness amidst these unpredictable field variables, which is a primary factor constraining the widespread adoption of vision-based volume measurement technology in practical agricultural scenarios.

Although 2D image-based volume estimation has been widely applied in agriculture, existing reviews have primarily focused on specific sensing modalities, computer vision techniques, or general agricultural imaging applications. To the best of our knowledge, no review has systematically examined the volume estimation methodological progression from controlled laboratory measurements to in-field and orchard-scale deployment, nor synthesized the technical challenges that limit the transferability of laboratory-developed approaches to real-world agricultural environments. Therefore, this review aims to fill this gap by providing a systematic overview of 2D image-based volume measurement for agricultural products. To clearly define the scope of this review, image-based volume estimation methods are categorized into two groups according to the type of visual information employed. The first category comprises monocular RGB methods, which estimate volume from a single RGB image using geometric models, statistical relationships, or deep learning regression. The second category consists of depth-assisted 2D methods that incorporate depth maps or RGB-D data to improve geometric representation while avoiding full 3D reconstruction. This taxonomy provides a clearer framework for understanding the evolution of image-based volume estimation methods from laboratory environments to real orchard scenarios.

2. Literature Search and Study Selection

The literature search strategy was conducted following the PRISMA 2020 guidelines to ensure transparency and reproducibility (Figure 1). Four electronic databases were queried, Web of Science Core Collection, Scopus, IEEE Xplore, and ScienceDirect. The search was performed on title, abstract, and keywords using a combination of Boolean operators. The search string was defined as: [“monocular vision” OR “single-view” OR “2D image” OR “RGB image”) AND (“volume estimation” OR “fruit volume” OR “size estimation” OR “biomass estimation” OR “phenotyping”) AND (“agriculture” OR “fruit” OR “crop” OR “orchard” OR “field”]. Only peer-reviewed journal articles published between 2000–2026 were included. Duplicate records were manually removed prior to screening. The selection process included three stages: (i) identification of records, (ii) title and abstract screening, and (iii) full-text eligibility assessment.

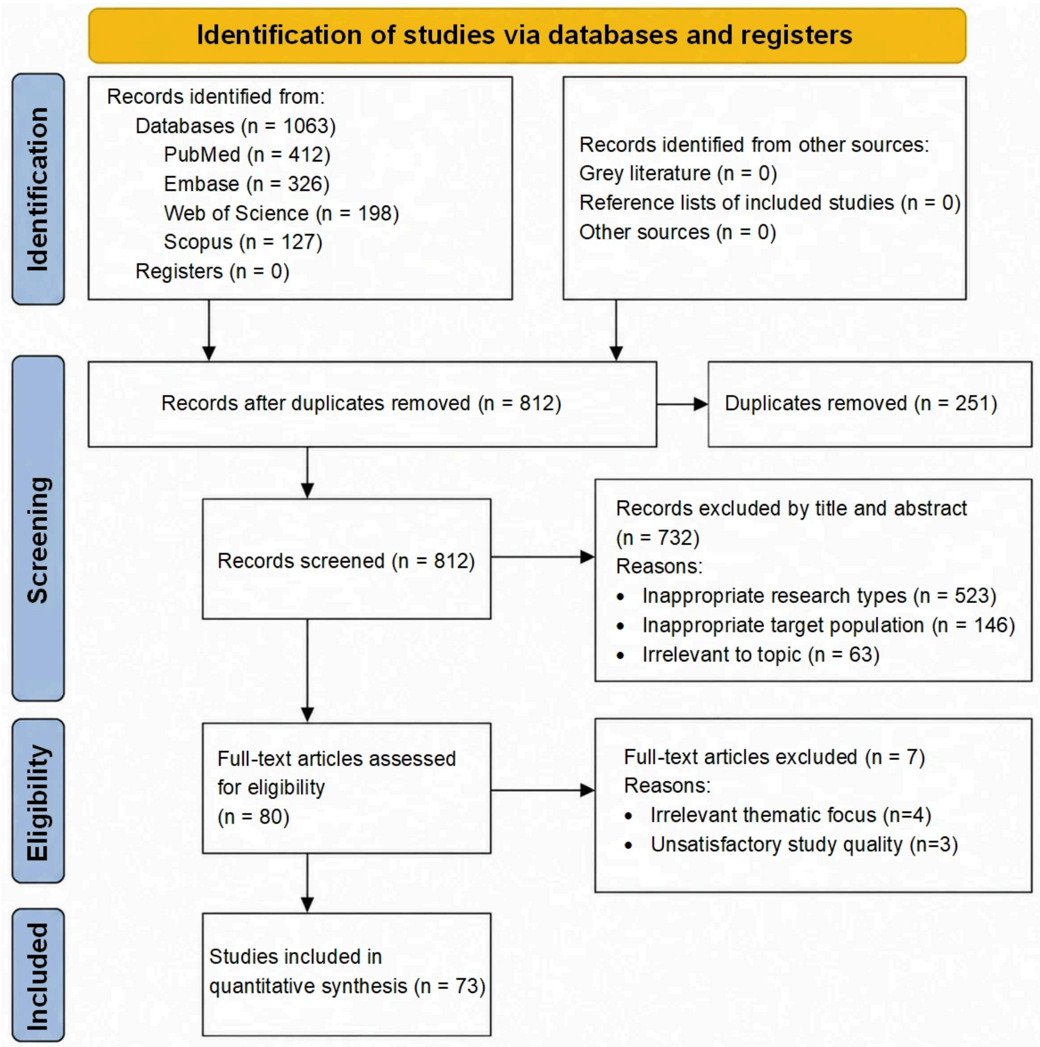

Figure 1. PRISMA 2020 flow diagram of study selection process. The diagram illustrates the identification, screening, eligibility assessment, and inclusion stages of the systematic review. Records were identified from database searching and other sources, followed by removal of duplicates and

irrelevant records. Remaining studies underwent title/abstract screening and full-text eligibility assessment, with reasons for exclusion recorded at each stage. A total of 73 studies were included in the quantitative synthesis.

3. RGB-Based Volume Estimation

3.1. Geometry-Based Volume Estimation Methods

Geometry-based methods represent the earliest and most widely adopted approaches for estimating the volume of agricultural products from monocular RGB images (Figure 2). These methods infer volume by establishing geometric relationships between image-derived measurements (e.g., projected area, contour dimensions, fruit diameter, and shape descriptors) and the target volumetric traits. Owing to their simplicity, interpretability, and low computational requirements, geometry-based approaches remain attractive for applications where imaging conditions are controlled and target objects exhibit relatively regular shapes. According to the level of geometric abstraction, existing monocular RGB-based methods can be broadly categorized into three groups: standard geometric approximation, slice-based accumulation methods, and boundary integration methods.

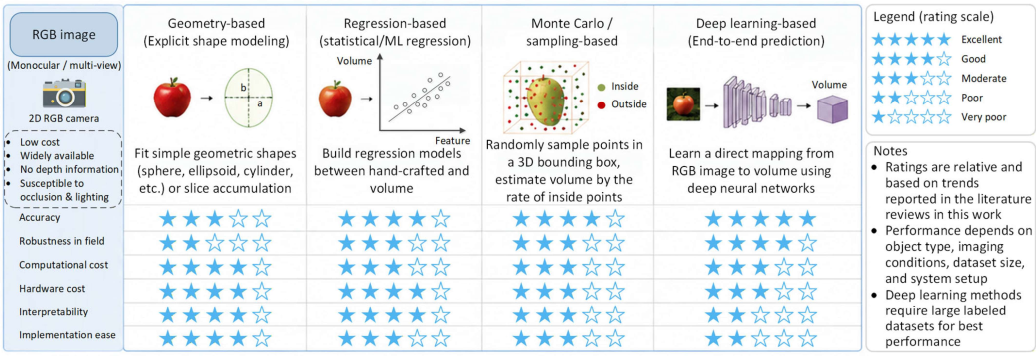

Figure 2. Comparative assessment of reviewed major agricultural product volume estimation methodologies using RGB images. Star ratings were assigned based on a qualitative assessment of the reviewed studies considering reported accuracy, computational efficiency, robustness, implementation complexity, and practical applicability. Ratings are intended for comparative overview only rather than quantitative benchmarking.

The image-acquisition devices mainly consist of a camera, a bracket, lighting equipment, a background board and a computer running the software (Figure 3) [24]. In the device, the camera is installed directly above the device, and the object to be measured is placed at the center of the camera's field of view to ensure the optimization of the shooting effect. To obtain more comprehensive image information, the improved system acquires more perspective images through two orthogonal cameras [25]. This method directly estimates volume from calibrated images or silhouette information without reconstructing a complete 3D surface, avoiding the errors caused by 3D reconstruction [26,27]. Furthermore, the lighting arrangement enhances image quality while reducing shadows or reflections. The background board is typically chosen to have a high contrast with the target object, facilitating accurate segmentation during image processing. Finally, the computer runs software that synchronizes image capture, stores the data, and may perform preliminary processing such as calibration, correction, and real-time visualization, ensuring that the acquired images are ready for subsequent analysis.

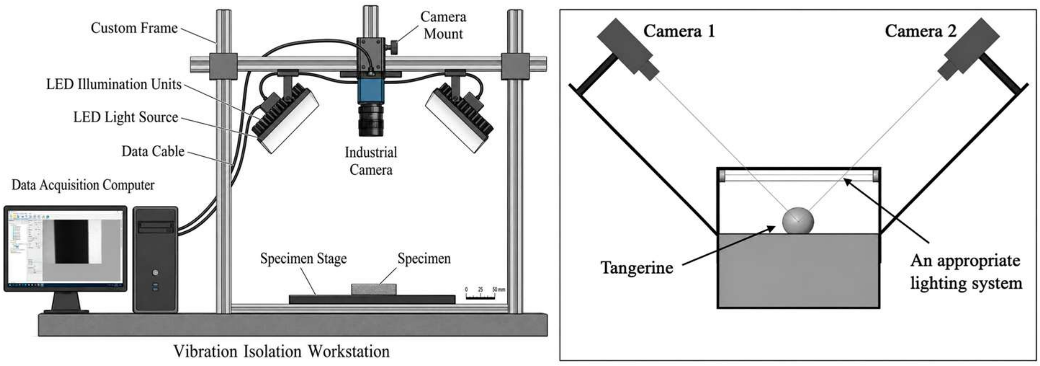

Figure 3. A schematic diagram of a classical monocular image acquisition platform. The (**left**) panel uses a monocular camera, while the (**right**) panel uses a single-camera multi-view imaging strategy to obtain more imaging information [25].

3.1.1. Standard Geometric Approximation Methods

Standard geometry-based methods are widely used for estimating the volume of agricultural products with relatively regular shapes. By approximating the object as a simple geometric solid such as a sphere, ellipsoid, or cylinder, researchers can calculate volume directly from measurable physical parameters like length, width, and height. Table 1 summarizes several typical applications of geometry-based volume estimation across different crops, highlighting the shape assumptions, measurement parameters, and achieved accuracy. For agricultural products with relatively regular shapes (e.g., apples and oranges), objects can be approximated as standard geometric solids, such as spheres, ellipsoids, or paraboloids. By extracting key geometric parameters (e.g., radius, diameters, or axial lengths) from images and applying analytical formulas, object volume can be efficiently estimated. Early studies predominantly adopted spherical or ellipsoidal approximations. For instance, Gokul et al. [28] approximated oranges as spheres using radius estimation from edge detection, while Koc et al. [29] and Rashidi and Gholami [30] employed ellipsoidal models for watermelons and kiwifruits using orthogonal views. However, such simplified models often lead to systematic underestimation due to discrepancies between idealized shapes and real fruit morphology. To improve parameter estimation, multi-view imaging strategies have been introduced. Du et al. [31] acquired images from four viewpoints and averaged geometric parameters, reducing variability and improving accuracy. Similarly, multi-camera systems, such as the dual-camera configuration proposed by Khojastehnazhand et al. [25], enable simultaneous acquisition of orthogonal views, enhancing efficiency and robustness.

A key limitation of standard geometric approximation is the assumption that a single geometric model can represent all samples within a category. In practice, agricultural products exhibit significant intra-class variability. To address this, Iqbal et al. [32] classified apples into spherical, ellipsoidal, and paraboloidal categories and applied corresponding models, achieving improved correlations with reference measurements. Further improvements have focused on enhancing model flexibility and generalizability. For example, G. Vivek Venkatesh et al. [33] reformulated paraboloidal shapes as special cases of ellipsoids and combined this with five-view imaging, achieving estimation errors within $\pm 10\%$ across multiple fruit types. In addition, it is worth noting that the inference times presented in Table 1 provide a comparative snapshot of algorithmic efficiency. However, their practical significance should be interpreted in the context of specific application requirements. In real-world scenarios such as high-speed conveyor-belt sorting, automated fruit and vegetable harvesting, and real-time disease monitoring, the suitability of a model depends not only on its predictive accuracy but also on its ability to satisfy throughput demands within available hardware constraints. Therefore, runtime performance should be evaluated alongside accuracy, throughput requirements, hardware limitations, and the intended deployment environment.

Table 1. Application cases for standard geometry-based methods.

Geometry	Product	Results	Runtime (s)	Reference
Ellipsoid	Orange	$R^2 = 0.946$	-	[25]
	Watermelon	$R^2 = 0.9591$	-	[29]
	Kiwi fruit	MAPE = 7.8%	-	[30]
	Apple	R = 0.97; Dev = ±8 cm^3	-	[32]
Sphere	Apple	R = 0.97; Dev = ±7 cm^3	-	[32]
Parabolic	Apple	R = 0.89; Dev = ±11 cm^3	-	[32]
Cylinder	Egg	ARE < 5%	0.31	[34]
	Egg, orange, peach, lemon	ARE < 10%	-	[35]
	Eggs, lemon, lime, tomato	ARE = 3.23~6.68%	-	[36]
	Tomato	ARE > 10%	-	[37]
Elliptical cylinder	Mango	RMSE = 2.9%; AP = 96.8%	0.45	[11,38]
Elliptical truncated cone	Orange	$R^2 = 0.9852$	-	[39]
	Lemon, lime, orange, tangerines	$R^2 =$ 0.985~0.959	-	[40]
	Oranges	$R^2 = 0.959$; MAPE = 2.6%	-	[25]
Elliptical cone	Cucumbers, carrots	R^2 = 0.981~0.982; ARE = 3.2025~3.421%	0.032~0.05	[41]
Boundary integral	Egg, orange, lemon	MAPE = 1.03%	0.20~0.25	[42]
	Potato, orange, tomato	P = 88.82~92.54% R^2 = 0.92~0.99;	≤1	[43]

Note: R^2, R, MAPE, Dev, ARE, RMSE, AP, and P represent the coefficient of determination, correlation coefficient, mean absolute percentage error, deviation, absolute relative error, root mean square error, accuracy percentage, and precision, respectively.

3.1.2. Slice-Based Accumulation Methods

For irregularly shaped agricultural products that cannot be adequately represented by a single geometric primitive, slice-based accumulation methods provide a flexible alternative. These approaches discretize the object into a series of thin slices along a predefined axis, approximate each slice using simple geometric shapes (e.g., cylinders, frustums, or elliptical frustums), and compute the total volume by summing the contributions of individual slices. Compared with single-shape approximation, this piecewise modeling strategy better captures local contour variations and improves estimation accuracy. Early slice-based methods modeled objects as stacks of equal-height frustums, often with slice thickness corresponding to a single pixel [35–37]. Although these approaches improved accuracy over standard geometric models, they were sensitive to imaging conditions such as shadows and occlusions, which can degrade contour extraction. Initial cylindrical approximations [44] tended to underestimate volume, whereas frustum-based models better captured tapering structures, though head and tail regions were sometimes neglected. To improve accuracy, Koc et al. [29] introduced a slice-based cylindrical model using average diameters, achieving results comparable to water displacement measurements. Subsequent work extended these methods to various agricultural products with consistent performance [30,34,45].

To overcome the limitation of assuming circular cross-sections, later studies introduced elliptical slice models. Khojastehnazhand et al. [25,39,40] utilized dual orthogonal cameras to extract major and minor axes, enabling elliptical frustum modeling with improved accuracy. Single-camera approaches have also been proposed, such as those by Moreno et al. [38] and The Mon et al. [11], which infer missing geometric parameters through empirical relationships. More advanced hybrid strategies further improve accuracy by modeling different regions of the object with distinct geometric primitives. For example, Cheng et al. [31] modeled terminal regions as spherical caps and central regions as frustums, while Huynh et al. [41] used elliptical cones for ends and elliptical frustums for the middle. Additionally, adaptive slice selection strategies have been introduced to balance computational efficiency and accuracy, where the number of slices is determined based on convergence criteria rather than fixed resolution.

For objects exhibiting rotational symmetry, boundary curve integration provides a more rigorous mathematical solution. In this approach, the object is modeled as a solid revolution, and volume is computed by integrating the function describing its boundary curve. Siswantoro et al. [46] proposed a framework using cubic spline interpolation to fit smooth boundary curves, followed by numerical integration. Compared with linear interpolation, cubic splines better approximate nonlinear contours while avoiding oscillations associated with higher-order polynomials. Experimental validation on eggs, lemons, and oranges demonstrated high accuracy, with an average relative error of 1.03% compared to water displacement measurements [42].

Collectively, geometry-based methods offer advantages in simplicity, interpretability, and computational efficiency. Standard geometric approximation is efficient but limited by model assumptions; slice-based accumulation improves flexibility and accuracy at the cost of increased computational complexity. Future research should focus on improving adaptive modeling strategies, enhancing robustness in complex imaging conditions, and developing automated, high-precision geometric feature extraction techniques to support real-world agricultural applications.

3.1.3. Boundary Integral-Based Methods

In contrast to slice-based methods that approximate volume using discrete cross-sections, boundary-integration methods estimate volume by fitting a continuous contour function and applying the calculus of solids of revolution. For axially symmetric agricultural products, the boundary profile is extracted and modeled as a continuous curve, and the volume is subsequently calculated by integrating the curve around the axis of symmetry.

As summarized in Table 2, Siswantoro and Asmawati [46] proposed a volume estimation framework using cubic spline interpolation. The method represents the cross-sectional boundary with piecewise cubic polynomials and calculates volume through integration of the fitted curve. Compared with piecewise linear approximations, cubic spline interpolation more accurately captures smooth nonlinear contours while avoiding the oscillation problems often associated with high-order polynomial fitting. However, the study was initially limited to methodological development and lacked validation using real agricultural products.

Subsequently, Siswantoro et al. [42] validated the approach on 50 axially symmetric samples, including eggs, lemons, and oranges. Compared with the water displacement method, the proposed framework achieved a mean absolute relative error of only 1.03%, with statistical analysis indicating no significant difference between the two measurement methods. Despite its high accuracy, the method remains restricted to axially symmetric objects, and its robustness requires further verification using larger and more diverse datasets.

To overcome this limitation, Jana et al. [47] developed a boundary-curve integration method for non-axially symmetric agricultural products. The longest axis of the object was first identified to divide the product into upper and lower regions. Boundary points were then extracted using the Moore neighborhood contour tracing algorithm, followed by polynomial fitting of the boundary curves. Volume was estimated by integrating the fitted functions. Experiments on 52 potato samples yielded an average accuracy of 92.54% relative to the water displacement method, with a correlation coefficient of 0.99. Nevertheless, the applicability of the method to other crop types remained largely unexplored. To further assess generalizability, Jana et al. [43] expanded the dataset to 77 agricultural products with varying shapes and symmetry characteristics, including potatoes, oranges, and tomatoes. The proposed method achieved average accuracies of 92.54%, 88.82%, and 89.02% for the three product categories, respectively, while the corresponding correlation coefficients between estimated and reference volumes were 0.99, 0.92, and 0.94. Although these results demonstrate the feasibility of the approach across different product types, the estimation accuracy remains moderate, and additional validation on larger and more diverse datasets is still required.

Table 2. Examples of statistical volume-estimation methods.

Methods	Product	Results	Reference
Monte Carlo	Apple, mango, tomato, *Pleurotus eryngii*	MAPE = 0.82~1%; CV = 0.16~0.18%	[7,9,48]
$Y = ax + c$ $Y = ax^2 + bx + c$ $Y = \frac{ax^2 + bx + c}{x + w}$	Elephant apple	R^2 = 0.98; RMSE = 10.73	[49]
$Y = ax^b$	Thai apple, strawberry, rosa roxburghii	R^2 = 0.83~0.97; MSE = 1.20	[50–52]

Note: MAPE, CV, R^2, RMSE, and MSE represent mean absolute percentage error, coefficient of variation, coefficient of determination, root mean square error, and mean square error, respectively.

Overall, boundary curve integration methods rely heavily on the fidelity of boundary representation. Their performance is primarily determined by two factors: accurate boundary extraction and appropriate curve-fitting strategies. Consequently, advances in contour detection and boundary modeling are critical for improving the accuracy, robustness, and broader applicability of this class of volume estimation methods.

3.2. Statistical-Based Volume Estimation Methods

Monte Carlo methods have been used to estimate the volume of irregular agricultural products without 3D reconstruction. Siswantoro et al. [9] developed a five-camera multi-view system integrating calibration, image acquisition, preprocessing, segmentation, and stochastic volume estimation. The method defines a 3D bounding box around the object, uniformly samples random points within it, and projects them onto all calibrated image planes. Points whose projections fall within the segmented object regions in every view are classified as internal, and object volume is estimated from the proportion of internal points to the total number of samples. Similarly, Luo et al. [7] proposed a depth camera-based approach combined with an improved circular disk method to measure the volume of *Pleurotus eryngii*, where the cap is regarded as an ellipsoid and then its volume is measured by Monte Carlo algorithm, and the total volume is obtained by summing the volumes of individual disks.

Experimental validation using a standard sphere (208.75 cm^3) demonstrated that estimation accuracy improves with increasing sampling density [53]. When 100,000 points were used, the absolute relative error (ARE) and coefficient of variation (CV) were reduced to 0.36% and 0.32%, respectively, while computation time remained below 0.8 s, indicating a favorable trade-off between efficiency and accuracy. However, due to the limited number of viewpoints, systematic bias may occur, as some external points can be incorrectly classified as internal when their projections overlap object regions in all views. This issue leads to volume overestimation and is closely related to object morphology.

To address this limitation, Siswantoro et al. [48] further evaluated the method on 150 irregular agricultural products, including apples, mangoes, and tomatoes, and introduced category-specific correction factors based on comparisons with water displacement measurements. With an optimal sampling density of 800,000 points, the corrected results achieved mean absolute relative errors of 1.00%, 0.97%, and 0.82% for apples, mangoes, and tomatoes, respectively, with coefficients of variation below 0.2%. Although these results demonstrate high accuracy and reliability, the dependence on empirical correction factors limits generalizability, particularly for heterogeneous or unseen products, suggesting the need for improved bias reduction strategies and more adaptive system designs.

3.3. Machine Learning-Based Volume Estimation Methods

3.3.1. Deep Learning-Based Crop/Fruit Size Estimation

The estimation of fruit phenotypic parameters in agricultural and orchard environments has been transformed by deep learning applied to 2D images. Attributes such as fruit count, size, shape, color, maturity stage, and surface defects are fundamental for yield prediction, quality grading, and harvest management. Traditional machine vision methods rely on hand-crafted features that often fail under variable natural illumination, partial occlusion, and dense canopy clutter. Convolutional neural networks (CNNs), region-based detectors such as Faster R-CNN [54], single-stage detectors such as YOLO [55], and instance segmentation architectures exemplified by Mask R-CNN have demonstrated promising performance in learning discriminative feature representations RGB images [56]. Studies conducted under a range of agricultural imaging conditions suggest that these approaches can effectively extract phenotypic traits from field-acquired images, although their transferability across different cultivars, species, sensors, and environmental conditions may require additional validation.

A large body of work focuses on fruit detection, counting, and in-field size estimation. In apple orchards, Bargoti and Underwood [57] employed patch-based CNN segmentation to detect fruits and map yield, illustrating that per-tree counts can be obtained with high accuracy. Sa et al. [58] proposed DeepFruits, a Faster R-CNN framework that successfully detected multiple fruit species under challenging lighting, providing the foundation for further phenotypic measurements. Koirala et al. [59] used YOLO for real-time mango detection and demonstrated that detection outputs could be integrated into yield forecasting models. For size estimation from 2D imagery, Apolo-Apolo et al. [60] combined UAV photography with Mask R-CNN to estimate citrus fruit size and total yield, showing that the apparent diameter measured in pixels can be converted to physical units when a reference object or camera calibration is available. Gené-Mola et al. [61] leveraged a multi-modal approach where the instance masks predicted by Mask R-CNN were used to derive the equatorial diameter of Fuji apples, achieving millimeter-level accuracy when fused with depth information, yet the 2D masks alone already captured the relevant size attributes. To handle dense clusters and occlusions, Häni et al. [62] systematically compared counting-by-detection and direct count regression paradigms for apple yield mapping, while Chen et al. [63] designed a data-driven counting pipeline that operated directly on whole images of oranges and apples, bypassing explicit detection and instead learning a mapping to the fruit number. Together, these studies demonstrate that 2D deep-learning pipelines can achieve accurate fruit counting and size-estimation performance on the evaluated datasets, highlighting their potential utility for orchard phenotyping and yield-estimation applications.

Beyond yield-related traits, deep learning has been extensively applied to assess fruit shape, maturity, and defect incidence from 2D images. Aquino et al. [64] used CNN segmentation to isolate individual grape berries in field images, from which berry number and size distributions were computed, enabling non-invasive grapevine yield component phenotyping. Santos et al. [65] applied Mask R-CNN for grape bunch detection and instance segmentation, then derived berry-level metrics including volume approximations, effectively turning a 2D segmentation map into shape- and size-related phenotypic descriptors. For maturity evaluation, color and textural changes encoded in RGB channels have been exploited by deep classifiers. Surface defects, which directly influence post-harvest quality, have been addressed by Fan et al. [66], who designed a CNN-based online inspection system to detect bruises, rots, and other anomalies on apples; the network learned defect features directly from 2D color images, outperforming traditional feature-engineering pipelines. Wan and Goudos [67] demonstrated that a single Faster R-CNN model could be extended to multi-class fruit detection while simultaneously predicting quality attributes, highlighting the flexibility of modern detection architectures for combined phenotypic tasks. In all these cases, the key advantage is that 2D sensors are inexpensive, universally available, and can be deployed on ground vehicles, unmanned aerial vehicles, or handheld devices, making high-throughput phenotyping feasible at a large scale.

The translation of pixel-level measurements into biologically meaningful parameters remains a challenge due to the loss of depth information in single 2D images. However, effective solutions have been developed through geometric scaling using known reference markers, camera-to-target distance estimates derived from structure-from-motion or single-image calibration, and direct regression approaches that implicitly learn the mapping from appearance to physical size. The review by Fu et al. [68] on consumer RGB-D cameras for fruit detection and localization discusses these strategies in depth and confirms that 2D cues, when properly calibrated, are often sufficient for phenotypic estimation. Kang and Chen [69] further emphasized that real-time detection speed can be achieved without sacrificing accuracy, an essential requirement for in-field vehicle-mounted phenotyping platforms. The comprehensive survey by Kamilaris and Prenafeta-Boldú [70] underscores that data augmentation, transfer learning, and the availability of large annotated datasets have been critical enablers for the successful application of 2D deep-learning models in agriculture. Future progress is expected to come from lightweight architectures optimized for edge devices, multi-view 2D imagery that resolves occlusions, and hybrid approaches that combine semantic segmentation with prior biological knowledge about fruit shape, ultimately improving the accuracy, scalability, and interpretability of phenotypic parameter estimation in orchard environments and suggesting the potential for enhanced performance across a broader range of deployment scenarios [71].

3.3.2. Phenotypic Parameter-Based Regression Methods

Regression-based methods have been widely adopted for indirect volume estimation by learning relationships between easily measurable features and object volume (Figure 4). For example, Fauzi et al. [72] applied the Mask R-CNN to obtain the width and length of apple, and then estimate the volume by using the width and height via 2D image captured. Building upon the feature extraction and detection frameworks discussed in the previous section on deep learning-based detection and sizing, these methods typically operate on the premise that accurate geometric or visual measurements can be translated into reliable volumetric predictions through statistical modeling. Ziaratban et al. [73] established regression models relating the maximum and mini-mum diameters of golden delicious apples to volume, achieving an R^2 of 0.9394. Similarly, Saikumar et al. [51] developed multiple regression models for elephant apple volume estimation using features such as length, width, perimeter, and projected area, where multivariate exponential regression achieved the best performance, with R^2 values up to 0.98 and relatively low prediction errors, demonstrating the effectiveness of combining multiple descriptors.

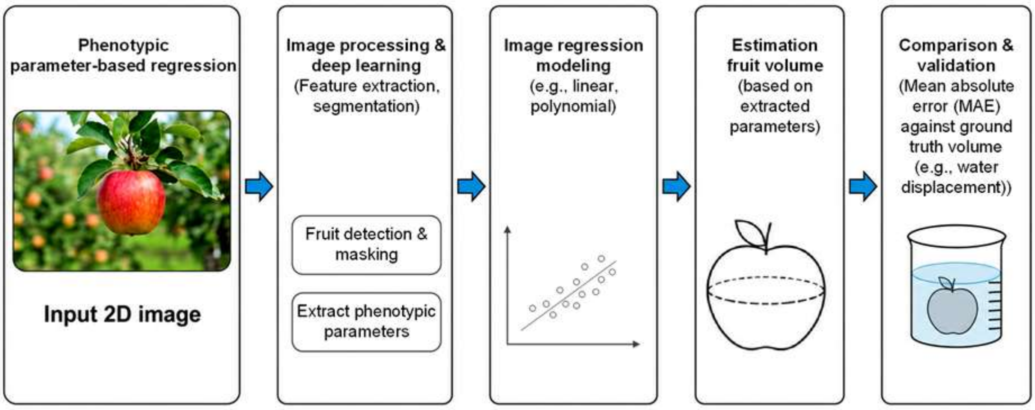

Figure 4. Framework of phenotypic parameter-based regression for volume estimation.

With the advancement of data-driven modeling, machine learning-based regression approaches have further enhanced the flexibility and nonlinear fitting capability of volume estimation methods. Soltani et al. [30] applied a multilayer perceptron (MLP) neural network to egg volume prediction, achieving an R^2 of 0.992, while Ziaratban et al. [73] reported an R^2 of 0.99995 for apple volume estimation using ANN models. However, reliance on limited geometric descriptors can lead to ambiguity when objects share similar dimensions but differ in shape. To address this, Gonzalez et al. [52] incorporated additional features such as color and texture using a hybrid PCA-LDA-ANN framework for passion fruit volume estimation, although performance (R^2 = 0.73) suggests that feature relevance remains critical.

Subsequent studies have further emphasized optimizing feature selection and model refinement. Siswantoro et al. [74] identified maximum diameter, minimum diameter, projected area, and perimeter as key predictors for egg volume, achieving an R^2 of 0.9738 and a mean absolute percentage error (MAPE) of 2.21% using a compact ANN model. Yin et al. [21] developed an exponential gaussian process regression to estimate the volume of multiple *Oudemansiella raphanipies*, achieving a coefficient of variation of 6.94% using 18 phenotypic parameters. In addition, Mansuri et al. [50] demonstrated that support vector machines (SVMs) can outperform traditional regression models, with multivariate linear SVM achieving R^2 values above 0.97 for Thai apple volume estimation.

Beyond single-object scenarios, regression-based methods have also been extended to more complex conditions. Nyalala et al. [75] developed a tomato volume estimation system capable of handling both isolated and clustered fruits. For single objects, multiple geometric features were used to train regression models, with radial basis function (RBF) SVM achieving the best performance (R^2 = 0.982). For clustered tomatoes, individual fruits were segmented using a concave-convex point detection algorithm prior to volume estimation. While the method demonstrated strong performance, it was validated on a single cultivar, and its generalizability across different varieties remains to be further investigated.

3.3.3. End-to-End Deep Learning for Volume Estimation

Recent advances in deep learning have enabled a paradigm shift from geometric modeling toward direct volume regression (Figure 5) [76,77]. Unlike geometry-based approaches that estimate volume through handcrafted measurements, shape fitting, or 3D reconstruction pipelines, end-to-end learning methods establish a direct mapping between image observations and object volume [78]. By bypassing geometric reconstruction pipelines, these approaches substantially reduce computational complexity and avoid error accumulation associated with segmentation, feature matching, point-cloud generation, registration, and surface fitting. Consequently, they have emerged as a promising alternative for high-throughput phenotyping, automated grading, and real-time agricultural monitoring.

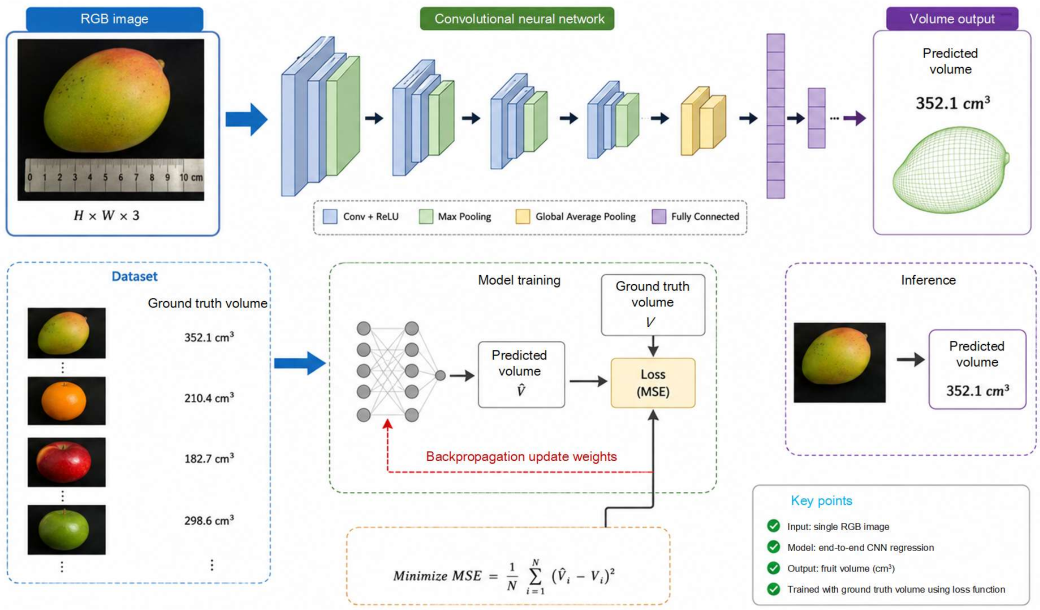

Figure 5. Framework of RGB image-based end-to-end deep learning methods for volume estimation.

One of the earliest attempts to formulate volume estimation as a direct learning problem was DeepVol proposed by Li and Han [78]. The framework combined an SSD-based fruit detector with a ResNet-based regression network to estimate fruit volume directly from a single RGB image. Unlike conventional image-based methods, DeepVol required neither camera calibration nor explicit geometric modeling, demonstrating that convolutional neural networks can learn latent visual representations associated with object volume from appearance cues alone. This study provided an important proof-of-concept for treating volume estimation as an end-to-end regression task and inspired subsequent research on data-driven volumetric inference. Following this direction, deep learning architectures have increasingly incorporated object detection and instance segmentation into volume estimation pipelines. For instance, Lüling et al. [79] employed a Mask R-CNN framework to segment cabbage heads from field images and estimated cabbage volume using image-derived morphological characteristics. Although geometric assumptions were incorporated during the final volume calculation, the study highlighted the effectiveness of deep neural networks in extracting volume-relevant structural information from RGB imagery under field conditions. Furthermore, Yi et al. [80] proposed an improved YOLOv9 framework for simultaneous fruit detection and volume estimation of multiple fruit species, including apples, pears, pomelos, and kiwifruits. By integrating volume regression directly into the detection pipeline and optimizing both the backbone and detection head, the model achieved accurate volume prediction under complex orchard environments with varying illumination and background conditions. The strong cross-species performance further suggests that deep neural networks can learn generalized morphological representations associated with fruit volume, highlighting the potential of single-stage detection-regression frameworks for practical, real-time volume estimation in agricultural applications.

Despite their simplicity, RGB-only approaches face inherent limitations because volume is fundamentally a three-dimensional attribute [81]. The projection from 3D space to a 2D image plane is intrinsically ill-posed, leading to scale ambiguity, where different real-world object sizes and depths can produce similar image projections in the absence of explicit metric information. This ambiguity is primarily caused by unknown camera-to-object distance, uncertain camera calibration parameters, and the lack of absolute depth cues, resulting in unavoidable information loss, particularly for object thickness and occluded surfaces [82,83]. Consequently, accurate volume inference from monocular images often requires additional constraints, such as known reference objects, fixed imaging geometry, or calibrated acquisition setups to recover metric scale. In practice, prediction performance therefore depends strongly on object shape regularity, imaging distance consistency, and the diversity and representativeness of the training data.

4. Depth-Assisted Image-Based Volume Estimation

Depth images provide direct geometric cues related to object shape and spatial structure while avoiding the computational complexity associated with full 3D reconstruction [84–86]. Instead of converting depth maps into point clouds or meshes, several studies have explored learning-based approaches that use depth images or RGB-D data as direct inputs for volumetric prediction [87,88]. In these frameworks (Figure 6), convolutional neural networks extract both appearance and depth features and learn an implicit mapping between image observations and object volume. Compared with RGB-only methods, depth-enhanced models generally improve robustness to illumination variation, texture heterogeneity, and viewpoint changes because geometric information is explicitly encoded in the input representation [89,90].

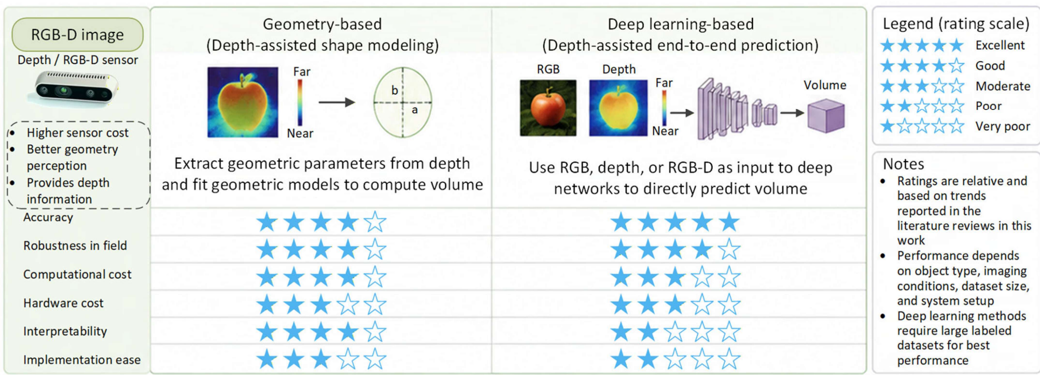

Figure 6. Comparison of deep learning-based volume estimation approaches using RGB-D images. Star ratings are assigned according to qualitative assessment of the reviewed studies considering accuracy, computational efficiency, robustness, implementation complexity, and practical applicability. Star ratings are assigned according to qualitative assessment of the reviewed studies considering accuracy, computational efficiency, robustness, implementation complexity, and practical applicability.

4.1. RGB-D-Based Geometric Modeling

The introduction of low-cost RGB-D sensors has significantly expanded the capability of image-based volume estimation in agriculture [68,91]. Unlike conventional RGB imaging, RGB-D systems simultaneously acquire color and depth information, enabling direct measurement of object geometry while preserving appearance characteristics [92]. The availability of affordable sensors such as Microsoft Kinect, Azure Kinect, and Intel

RealSense has stimulated extensive research on fruit and agricultural product phenotyping, where depth information is incorporated into geometric models to estimate size, shape, and volume with improved accuracy [93,94]. RGB-D sensing provides explicit spatial information that alleviates scale ambiguity inherent in monocular imaging and enables volumetric estimation without requiring complex multi-view reconstruction systems [95].

Most RGB-D-based geometric modeling approaches estimate volume by first extracting geometric descriptors from depth images and subsequently fitting predefined geometric models [7]. Typical descriptors include fruit diameter, major and minor axes, projected area, surface curvature, and depth-derived thickness measurements [96]. The extracted parameters are then incorporated into geometric formulations based on spheres, ellipsoids, spheroids, cylinders, or other parametric shapes to estimate object volume. Compared with purely image-based methods, RGB-D geometric modeling introduces direct depth measurements into the estimation process, thereby reducing uncertainty associated with hidden dimensions and improving robustness across varying imaging distances and illumination conditions [97].

Early studies primarily focused on fruit sizing using depth information as a geometric constraint [7]. Wang et al. developed an RGB-D-based framework for on-tree mango size estimation using a Kinect sensor [93]. After fruit detection and segmentation, depth information was employed to estimate the distance between the camera and the fruit, allowing accurate calculation of fruit dimensions from image measurements. The study demonstrated that RGB-D sensing substantially improved size estimation accuracy compared with conventional monocular approaches and established a practical foundation for subsequent volume estimation methods in orchard environments.

Subsequent research extended this concept from size estimation to volumetric characterization. Perez et al. [98] proposed a flexible multi-RGB-D sensor system for fruit measurement and classification in the agri-food industry. The system utilized synchronized RGB-D observations to estimate geometric attributes of fruits and infer volume-related characteristics for grading and classification. Experimental evaluations on multiple fruit species demonstrated that RGB-D-derived geometric measurements could provide reliable estimates of fruit size and production volume while maintaining low hardware cost and operational simplicity. Miranda et al. [96] investigated automatic algorithms for estimating apple size and weight using an Azure Kinect RGB-D camera. Their framework extracted major and minor geometric axes directly from RGB-D observations and integrated these measurements into allometric models for fruit characterization. Results showed that depth-derived geometric descriptors enabled accurate estimation of fruit dimensions and biomass-related traits, highlighting the potential of RGB-D geometric modeling for yield prediction and orchard management applications. For fruits with relatively regular morphology, simplified geometric models remain particularly effective. Fukuda et al. [99] proposed an RGB-D-based method for fruit diameter measurement aimed at monitoring fruit growth dynamics. Rather than reconstructing complete three-dimensional surfaces, the method estimated fruit diameter directly from object dimensions in RGB images combined with depth information at the fruit center. The resulting geometric measurements were subsequently used to characterize fruit enlargement and infer volumetric changes over time. This study demonstrated that accurate geometric estimation can be achieved from a single RGB-D observation without the computational burden associated with dense surface reconstruction.

Overall, RGB-D-based geometric modeling represents an important intermediate stage between traditional monocular geometry methods and more sophisticated three-dimensional reconstruction techniques. By directly integrating depth-derived geometric measurements into parametric volume models, these approaches achieve a favorable trade-off between accuracy, computational efficiency, and operational simplicity. As RGB-D sensors continue to improve in resolution, accuracy, and affordability, geometric modeling based on RGB-D observations is expected to remain a practical solution for high-throughput agricultural volume estimation, particularly in applications where real-time performance and low computational cost are critical.

4.2. Depth-Assisted Deep Learning

The emergence of RGB-D sensors has not only facilitated geometric modeling but has also enabled the development of depth-assisted deep learning approaches for volume estimation (Figure 7) [100]. Unlike RGB-based methods that rely solely on appearance cues, depth-assisted deep learning integrates geometric information from depth maps to learn implicit relationships between object morphology and volume. Rather than using explicit measurements and predefined shape assumptions, as in RGB-D geometric modeling, these methods treat volume estimation as a data-driven regression task and directly predict volume from RGB-D observations [100]. By jointly leveraging appearance and depth information, they can capture complex morphological variations that are difficult to describe with conventional geometric models, making them well suited for irregularly shaped agricultural products.

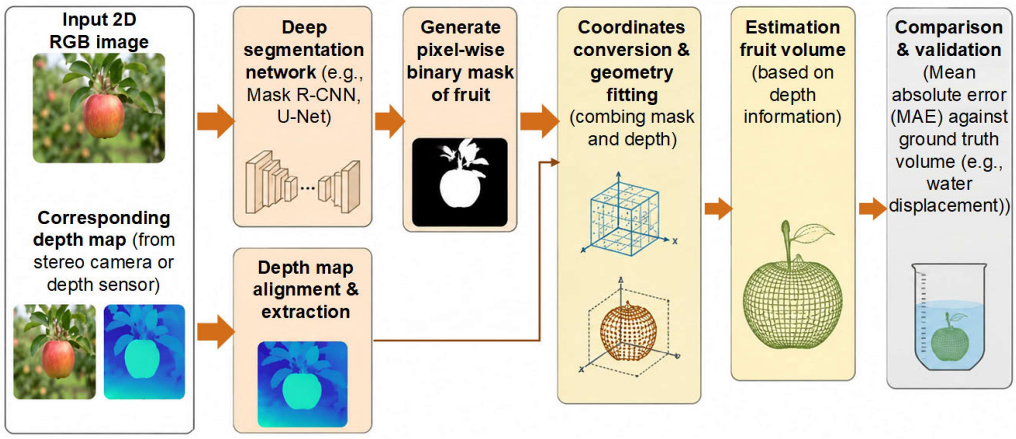

Figure 7. Framework of depth-assisted end-to-end deep learning methods for volume estimation.

One of the earliest studies demonstrating the potential of deep learning for direct volume estimation was DeepVol proposed by Li and Han [78]. They proposed an end-to-end framework that combined SSD-based fruit detection with a ResNet regression network to estimate volume directly from RGB images. Rather than reconstructing fruit geometry or extracting explicit descriptors, the model learned an implicit mapping between visual features and volume through supervised learning. Experimental results demonstrated accurate estimation across multiple fruit categories, providing early evidence that volumetric information can be inferred directly from image observations without explicit geometric modeling. Although the study relied on RGB imagery, it conceptually laid the foundation for subsequent RGB-D and depth-assisted volume regression approaches. Lo et al. [101] developed a deep learning framework for food volume estimation using a single depth map as input. Instead of reconstructing a complete three-dimensional model, the method employed depth-aware neural networks to synthesize missing views and infer volumetric properties directly from depth observations. The study demonstrated that depth maps contain rich geometric information closely related to object volume and that deep learning can effectively exploit these cues without relying on explicit geometric reconstruction. The results highlighted the potential of depth-assisted learning for volumetric estimation in situations where complete surface acquisition is unavailable. Kim et al. [102] developed a CNN-based framework for plum fruit detection and growth estimation using RGB-D images acquired under field conditions. The network integrated RGB texture information with depth-derived geometric features to estimate fruit growth-related traits. Although volume was not directly predicted, the study demonstrated that depth information significantly improved the estimation of morphological attributes closely associated with fruit size and volume. These findings suggest that RGB-D representations can provide valuable volumetric cues beyond those obtainable from RGB images alone.

Depth-assisted learning has also been widely adopted in plant phenotyping applications. For example, Ubbens et al. [103] incorporated depth information into convolutional neural networks for biomass estimation and structural trait prediction in crop canopies. By learning directly from RGB-D observations, the models achieved higher prediction accuracy than RGB-only counterparts, indicating that depth maps can effectively encode structural information relevant to plant volume and biomass. Similar findings were reported by Pound et al. [104], who demonstrated that depth-enhanced neural networks improve the estimation of three-dimensional plant traits under varying imaging conditions. These studies provide indirect evidence that depth-assisted learning can capture volumetric characteristics without requiring explicit reconstruction of plant architecture. Recent advances in RGB-D sensing and deep learning have further promoted the development of multimodal regression architectures. Hartley et al. [105] proposed GANana, a volumetric regression framework that combines synthetic fruit generation with domain adaptation to estimate fruit volume directly from image observations via deep feature learning, bypassing explicit geometric reconstruction and achieving robust performance across different imaging domains. Furthermore, Siripurapu et al. [106] employed Mask R-CNN to identify food items and used the monocular depth estimation model to generate depth maps from RGB images. These depth maps were subsequently utilized to estimate food volume information. Under orchard conditions, Devanna et al. [107] combined depth segmentation and depth-based clustering to automatically detect, count, and estimate the volume and weight of grape bunches using RGB-D data.

Despite these advantages, several challenges remain. The performance of depth-assisted deep learning models depends heavily on the availability of accurately labeled volumetric datasets, which are often difficult and expensive to obtain in agricultural environments. In addition, depth sensors are susceptible to measurement noise, occlusions, reflective surfaces, and varying illumination conditions, all of which may degrade prediction accuracy [68]. Furthermore, the black-box nature of deep neural networks limits interpretability and hinders the incorporation of biological knowledge into the estimation process [108]. Consequently, future research is expected to focus on hybrid frameworks that integrate deep learning with geometric constraints, physical priors, and plant growth knowledge. Such approaches may combine the flexibility of data-driven learning with the robustness and interpretability of mechanistic models, ultimately leading to more accurate and generalizable volume estimation systems for agricultural applications.

5. Challenges and Future Perspectives

5.1. Current Challenges

Despite substantial progress in image-based volume estimation, accurately deriving object volume from 2D images remains an inherently challenging problem. A major challenge arises from the loss of in-depth information. Monocular RGB images provide limited cues regarding object thickness, backside morphology, and internal shape variation [109]. Geometry-based methods attempt to compensate for this missing information through predefined shape assumptions, while statistical and machine learning approaches learn empirical relationships between image features and volume [76]. However, when agricultural products exhibit irregular shapes, asymmetry, deformation, or large morphological variability, image-derived features may no longer uniquely determine volume, resulting in increased estimation uncertainty.

The performance degradation of laboratory-developed volume estimation methods in orchard environments can be traced to the cumulative effects of environmental variability across the entire processing pipeline. Variable illumination alters fruit appearance through shadows, highlights, and color shifts, reducing the accuracy of image segmentation and object detection while introducing errors in contour extraction [110]. Camera motion and equipment vibration can introduce image blur and viewpoint inconsistency, reducing feature stability and degrading scale calibration [111]. For depth-assisted methods, orchard conditions introduce additional challenges, including noisy depth measurements caused by sunlight interference, reflective fruit surfaces, vegetation occlusion, and missing depth values near object boundaries [112]. These factors reduce the reliability of depth inference and geometric reconstruction, which subsequently affects volume estimation accuracy. Because errors introduced at early stages accumulate throughout the processing chain, the final volume prediction is often substantially less robust than that achieved under controlled laboratory conditions. This explains why methods reporting high accuracy in laboratory settings frequently experience significant performance degradation when deployed in real orchard environments.

Occlusion further complicates volume estimation, particularly in orchard environments [113]. Fruits are commonly partially hidden by leaves, branches, neighboring fruits, or supporting structures, preventing the acquisition of complete contour information [114,115]. Because volume estimation relies heavily on accurate shape representation, even partial occlusions may introduce substantial errors. This problem is particularly pronounced for geometry-based methods, where missing contour regions directly affect the extraction of dimensions, projected area, and shape descriptors, and for learning-based methods, where the network must infer hidden structures from incomplete observations. Studies conducted in clustered tomato canopies and commercial fruit orchards have shown that segmentation accuracy and volume prediction performance deteriorate substantially as occlusion severity increases, even when advanced detection and regression networks are employed [62,63]. The challenge therefore extends beyond object detection itself to the recovery of biologically realistic fruit geometry from partially visible observations. While modern object detection and segmentation networks have improved robustness under moderate occlusion, accurately recovering hidden geometry from incomplete visual observations remains an unresolved challenge.

Overall, the central challenge of 2D image-based volume estimation is not merely improving prediction accuracy, but overcoming the fundamental information gap between two-dimensional observations and three-dimensional object structure. Future advances will likely require the integration of geometric priors, multi-view observations, depth cues, and data-driven learning frameworks to reduce this ambiguity and achieve robust volume estimation under real-world agricultural conditions.

5.2. Future Prospects

The future development of agricultural product volumetric estimation based on 2D images is expected to overcome current challenges through multifaceted technological innovations. At the data acquisition level, the integration of multi-view, multispectral, and high-dynamic-range imaging is anticipated to enhance the capture of fruit morphology and surface features. Concurrently, strategies such as automated annotation, synthetic data generation, and simulation-based datasets can mitigate the difficulties of in-field labeling and the scarcity of samples. Furthermore, lightweight, high-stability mobile imaging systems and unmanned aerial platforms are likely to improve the efficiency and reproducibility of field data collection, thereby facilitating the translation of laboratory methodologies to in situ environments.

Rather than treating data-driven learning and geometric modeling as mutually exclusive approaches, an important future direction is their integration [116]. Incorporating physical constraints, projection geometry, and crop-specific geometric priors into deep learning frameworks can improve model interpretability, reduce reliance on large training datasets, and enhance the robustness of the mapping between 2D features and 3D volume under diverse field conditions. In addition, approaches leveraging attention mechanisms, multi-scale feature fusion, and end-to-end learning from images to volume can enhance model adaptability to occlusion, morphological variability, and lighting changes. The adoption of semi-supervised or self-supervised learning strategies may further improve generalization across diverse crops and environments, while reducing reliance on large-scale annotated datasets.

6. Conclusions

In conclusion, 2D image-based volume estimation has evolved from simple geometric approximation methods in controlled laboratory settings to increasingly sophisticated data-driven and depth-assisted frameworks capable of operating in complex orchard environments. Our review reveals that while laboratory studies have achieved high accuracy through standardized imaging conditions and explicit geometric modeling, the transfer of these approaches to real-world agricultural systems remains constrained by occlusion, illumination variability, scale ambiguity, and morphological diversity. Recent advances in deep learning and RGB-D sensing have substantially improved robustness under field conditions, yet generalizable and field-ready solutions remain elusive.

A central insight emerging from this review is that the future of agricultural volume estimation is unlikely to be driven by purely geometric or purely data-driven approaches alone. Instead, hybrid frameworks that integrate geometric constraints, depth information, biological priors, and learning-based representations offer the most promising pathway toward accurate, interpretable, and scalable volume estimation across diverse crops and environments. Equally important is the establishment of standardized datasets, evaluation protocols, and cross-environment benchmarking frameworks to enable rigorous comparison and reproducible progress.

Ultimately, volume estimation should be viewed not merely as a computer vision task, but as a cross-disciplinary agricultural sensing problem that requires the joint consideration of plant morphology, imaging physics, and machine learning. Bridging the gap between laboratory precision and orchard deployment will be essential for realizing high-throughput phenotyping, intelligent crop management, and next-generation digital agriculture. The advances synthesized in this review provide both a foundation and a roadmap for developing robust, scalable, and field-deployable volume estimation systems that can support future precision agriculture applications.

Author Contributions: Writing—review and editing, and writing—original draft preparation, Q.W.; writing—review and editing and validation, D.L.; formal analysis, Z.S.; methodology and validation, W.Z.; validation, F.W.; writing—review and editing, validation, funding acquisition, and conceptualization, H.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (62362039) and the Jiangxi Provincial Natural Science Foundation (No. 20242BAB25081).

Data Availability Statement: No new data were created or analyzed in this study.

Acknowledgments: Special thanks to the reviewers for their valuable comments.

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

References

1. Longchamps, L.; Tisseyre, B.; Taylor, J.; Sagoo, L.; Momin, A.; Fountas, S.; Manfrini, L.; Ampatzidis, Y.; Schueller, J.K.; Khosla, R. Yield sensing technologies for perennial and annual horticultural crops: A review. *Precis. Agric.* **2022**, *23*, 2407–2448. [CrossRef]
2. Meena, R.K.; Khan, R.; Rajawat, K.S. Evaluation of genetic variability, heritability and correlation among mandarin (Citrus reticulata Blanco) genotypes. *J. Exp. Agric. Int.* **2026**, *48*, 426–435. [CrossRef]
3. Saha, S.; Kucher, O.D.; Utkina, A.O.; Rebouh, N.Y. Precision agriculture for improving crop yield predictions: A literature review. *Front. Agron.* **2025**, *7*, 1566201. [CrossRef]
4. Mahawar, M.K.; Bibwe, B.; Jalgaonkar, K.; Ghodki, B.M. Mass modeling of kinnow mandarin based on some physical attributes. *J. Food Process Eng.* **2019**, *42*, e13079. [CrossRef]
5. Gibson, A.A.; Hsu, M.S.; Rangan, A.M.; Seimon, R.V.; Lee, C.M.; Das, A.; Finch, C.H.; Sainsbury, A. Accuracy of hands v. household measures as portion size estimation aids. *J. Nutr. Sci.* **2016**, *5*, e29. [CrossRef] [PubMed]
6. Opara, U.L.; Pathare, P.B. Bruise damage measurement and analysis of fresh horticultural produce—A review. *Postharvest Biol. Technol.* **2014**, *91*, 9–24. [CrossRef]
7. Luo, S.; Tang, J.; Peng, J.; Yin, H. A novel approach for measuring the volume of pleurotus eryngii based on depth camera and improved circular disk method. *Sci. Hortic.* **2024**, *336*, 113382. [CrossRef]
8. Moreda, G.; Ortiz-Cañavate, J.; García-Ramos, F.J.; Ruiz-Altisent, M. Non-destructive technologies for fruit and vegetable size determination—A review. *J. Food Eng.* **2009**, *92*, 119–136. [CrossRef]
9. Siswantoro, J.; Prabuwono, A.S.; Abdulah, A. Volume measurement of food product with irregular shape using computer vision and Monte Carlo method: A framework. *Procedia Technol.* **2013**, *11*, 764–770. [CrossRef]
10. Zhu, Y.; Cao, S.; Song, T.; Xu, Z.; Jiang, Q. 3D reconstruction and volume measurement of irregular objects based on RGB-D camera. *Meas. Sci. Technol.* **2024**, *35*, 125010. [CrossRef]
11. Mon, T.; ZarAung, N. Vision based volume estimation method for automatic mango grading system. *Biosyst. Eng.* **2020**, *198*, 338–349. [CrossRef]
12. Joyce, B.J.; Brown, G.E. Estimating the volume of biological structures from a single 2D image: Considering apparent cross-sectional area as an alternative to the ellipsoid method. *Evol. Ecol.* **2022**, *36*, 1061–1076. [CrossRef]
13. Piccinini, F.; Tesei, A.; Bevilacqua, A. Single-image based methods used for non-invasive volume estimation of cancer spheroids: A practical assessing approach based on entry-level equipment. *Comput. Methods Programs Biomed.* **2016**, *135*, 51–60. [CrossRef] [PubMed]
14. Sakthivel, N.; Adhisankaran, K.; Kabilan, M.; Balachandar, D.; Ragunath, K.; Raja, R.; Manivannan, V.; Vasumathi, V. A review on advancing agricultural practices using photogrammetric images. *Plant Sci. Today* **2025**, *12*, 1–16. [CrossRef]
15. Kirk, R.; Mangan, M.; Cielniak, G. Non-destructive soft fruit mass and volume estimation for phenotyping in horticulture. In *Proceedings of the International Conference on Computer Vision Systems*; Springer: Cham, Switzerland, 2021; pp. 223–233.
16. Islam, A.T.; Islam, A.S.; Islam, G.T.; Bala, S.K.; Salehin, M.; Choudhury, A.K.; Mahboob, M.G.; Dey, N.C.; Hossain, A. Monitoring wheat area using sentinel-2 imagery and In-situ spectroradiometer data in heterogeneous field conditions. *Discov. Agric.* **2024**, *2*, 52. [CrossRef]
17. Zumsteg, O.; Graf, N.; Haeusler, A.; Kirchgessner, N.; Storni, N.; Roth, L.; Hund, A. Fine-tuned vision transformers capture complex wheat spike morphology for volume estimation from RGB images. *Smart Agric. Technol.* **2026**, *14*, 102150. [CrossRef]
18. Roman, M.; Zahid, A.; Nafchi, A.M.; Sher, M. Development of an AI-enabled vision-based on-site rapid detection of nitrates and nitrites from drainage samples. *Water Air Soil Pollut.* **2026**, *237*, 469. [CrossRef]
19. Villacrés, J.; Vougioukas, S. Assessing a multi-camera system to enhance fruit visibility for robotic harvesting in a V-trellised apple orchard. *Comput. Electron. Agric.* **2024**, *224*, 109164. [CrossRef]
20. Boumaraf, I.; Djeffal, A.; Taleb-Ahmed, A. Feature-level multimodal deep learning for date fruit quality assessment. *J. Food Meas. Charact.* **2026**, *20*, 8353–8371. [CrossRef]
21. Yin, H.; Lei, D.; Xiong, A.; Yuan, L.; Chen, M.; Xu, Y.; Wang, Y.; Xiao, H.; Wei, Q. A pipeline for mushroom mass estimation based on phenotypic parameters: A multiple oudemansiella raphanipies model. *Agronomy* **2026**, *16*, 124. [CrossRef]
22. Huynh, T.T.; TonThat, L.; Dao, S.V. A vision-based method to estimate volume and mass of fruit/vegetable: Case study of sweet potato. *Int. J. Food Prop.* **2022**, *25*, 717–732. [CrossRef]
23. Fang, C.; Chen, H.; Li, L.; Luo, Z.; Liu, L.; Ban, Z. A novel adaptive zone-fusion network for precise waxberry semantic segmentation to improve automated-harvesting in a complex orchard environment. *Comput. Electron. Agric.* **2024**, *221*, 108937. [CrossRef]
24. Fu, H.; Li, T.; Feng, Q.; Chen, L. MT-WavYOLO: Bridging multi-task learning and 3D frustum fusion for non-destructive robotic harvesting of occluded orchard fruits. *Comput. Electron. Agric.* **2026**, *242*, 111335.

25. Khojastehnazhand, M.; Omid, M.; Tabatabaeefar, A. Determination of tangerine volume using image processing methods. *Int. J. Food Prop.* **2010**, *13*, 760–770. [CrossRef]
26. Wang, X.; Wang, C.; Liu, B.; Zhou, X.; Zhang, L.; Zheng, J.; Bai, X. Multi-view stereo in the deep learning era: A comprehensive review. *Displays* **2021**, *70*, 102102. [CrossRef]
27. Feng, S.; Wu, X.; Cao, J. A survey of multi-view stereo 3D reconstruction algorithms based on deep learning. *Digit. Signal Process.* **2025**, *165*, 105291. [CrossRef]
28. Gokul, P.R.; Raj, S.; Suriyamoorthi, P. Estimation of volume and maturity of sweet lime fruit using image processing algorithm. In *Proceedings of the 2015 International Conference on Communications and Signal Processing (ICCSP)*; IEEE: New York, NY, USA, 2015; pp. 1227–1229.
29. Koc, A.B. Determination of watermelon volume using ellipsoid approximation and image processing. *Postharvest Biol. Technol.* **2007**, *45*, 366–371. [CrossRef]
30. Rashidi, M.; Gholami, M. Determination of kiwifruit volume using ellipsoid approximation and image-processing methods. *Int. J. Agric. Biol. Eng.* **2008**, *10*, 375–380.
31. Du, C.; Sun, D. Estimating the surface area and volume of ellipsoidal ham using computer vision. *J. Food Eng.* **2006**, *73*, 260–268. [CrossRef]
32. Iqbal, S.M.; Gopal, A.; Sarma, A. Volume estimation of apple fruits using image processing. In *Proceedings of the 2011 International Conference on Image Information Processing*; IEEE: New York, NY, USA, 2011; pp. 1–6.
33. Vivek Venkatesh, G.; Iqbal, S.M.; Gopal, A.; Ganesan, D. Estimation of volume and mass of axi-symmetric fruits using image processing technique. *Int. J. Food Prop.* **2015**, *18*, 608–626.
34. Widiasri, M.; Santoso, L.P.; Siswantoro, J. Computer vision system in measurement of the volume and mass of egg using the disc method. In *Proceedings of the IOP Conference Series: Materials Science and Engineering*; IOP Publishing: Bristol, UK, 2019; p. 012050.
35. Sabliov, C.; Boldor, D.; Keener, K.; Farkas, B. Image processing method to determine surface area and volume of axi-symmetric agricultural products. *Int. J. Food Prop.* **2002**, *5*, 641–653. [CrossRef]
36. Wang, T.; Nguang, S. Low cost sensor for volume and surface area computation of axi-symmetric agricultural products. *J. Food Eng.* **2007**, *79*, 870–877. [CrossRef]
37. UluiŞik, S.; Yildiz, F.; Özdemİr, A.T. Image processing based machine vision system for tomato volume estimation. In *Proceedings of the 2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT)*; IEEE: New York, NY, USA, 2018; pp. 1–4.
38. Bermúdez, M.; Ballesteros Padilla, D.; Sánchez Torres, G. Image analysis for automatic feature estimation of the mangifera indica fruit. *Ing. Desarro.* **2013**, *31*, 84–104.
39. Khojastehnazhand, M.; Omid, M.; Tabatabaeefar, A. Determination of orange volume and surface area using image processing technique. *Int. Agrophysics* **2009**, *23*, 237–242.
40. Omid, M.; Khojastehnazhand, M.; Tabatabaeefar, A. Estimating volume and mass of citrus fruits by image processing technique. *J. Food Eng.* **2010**, *100*, 315–321. [CrossRef]
41. Huynh, T.; Tran, L.; Dao, S. Real-time size and mass estimation of slender axi-symmetric fruit/vegetable using a single top view image. *Sensors* **2020**, *20*, 5406. [CrossRef] [PubMed]
42. Siswantoro, J.; Asmawati, E.; Siswantoro, M.Z. A rapid and accurate computer vision system for measuring the volume of axi-symmetric natural products based on cubic spline interpolation. *J. Food Eng.* **2022**, *333*, 111139. [CrossRef]
43. Jana, S.; Parekh, R.; Sarkar, B. A de novo approach for automatic volume and mass estimation of fruits and vegetables. *Optik* **2020**, *200*, 163443. [CrossRef]
44. Tagliavini, M.; Veto, L.; Looney, N. Measuring root surface area and mean root diameter of peach seedlings by digital image analysis. *HortScience* **1993**, *28*, 1129–1130. [CrossRef]
45. Khojastehnazhand, M.; Mohammadi, V.; Minaei, S. Maturity detection and volume estimation of apricot using image processing technique. *Sci. Hortic.* **2019**, *251*, 247–251. [CrossRef]
46. Siswantoro, J.; Asmawati, E. A new framework for measuring volume of axisymmetric food products using computer vision system based on cubic spline interpolation. In *Proceedings of the 2016 2nd International Conference on Science in Information Technology (ICSITech)*; IEEE: New York, NY, USA, 2016; pp. 74–78.
47. Jana, S.; Parekh, R.; Sarkar, B. Volume estimation of non-axisymmetric fruits and vegetables using image analysis. In *Proceedings of the 2019 International Conference on Computing, Power and Communication Technologies (GUCON)*; IEEE: New York, NY, USA, 2019; pp. 628–633.
48. Siswantoro, J.; Prabuwono, A.S.; Abdullah, A.; Idrus, B. Monte carlo method with heuristic adjustment for irregularly shaped food product volume measurement. *Sci. World J.* **2014**, *2014*, 683048. [CrossRef]
49. Saikumar, A.; Nickhil, C.; Badwaik, L.S. Physicochemical characterization of elephant apple (*Dillenia indica* L.) fruit and its mass and volume modeling using computer vision. *Sci. Hortic.* **2023**, *314*, 111947. [CrossRef]
50. Mansuri, S.M.; Gautam, P.V.; Jain, D.; Nickhil, C. Computer vision model for estimating the mass and volume of freshly harvested Thai apple ber (*Ziziphus mauritiana* L.) and its variation with storage days. *Sci. Hortic.* **2022**, *305*, 111436. [CrossRef]

51. Xie, Z.; Wang, J.; Yang, Y.; Mao, P.; Guo, J.; Sun, M. Image processing based modeling for rosa roxburghii fruits mass and volume estimation. *Sci. Rep.* **2024**, *14*, 15507. [CrossRef] [PubMed]
52. Lee, D.H.; Cho, Y.; Choi, J.M. Strawberry volume estimation using smartphone image processing. *Hortic. Sci. Technol.* **2017**, *35*, 707–716. [CrossRef]
53. Siswantoro, J.; Prabuwono, A.S.; Abdullah, A. Volume measurement algorithm for food product with irregular shape using computer vision based on monte carlo method. *J. ICT Res. Appl.* **2014**, *8*, 1–17. [CrossRef]
54. Girshick, R. Fast R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision*; IEEE Computer Society: Washington, DC, USA, 2015; pp. 1440–1448.
55. Jiang, P.; Ergu, D.; Liu, F.; Cai, Y.; Ma, B. A Review of Yolo algorithm developments. *Procedia Comput. Sci.* **2022**, *199*, 1066–1073. [CrossRef]
56. Bharati, P.; Pramanik, A. Deep learning techniques—R-CNN to mask R-CNN: A survey. *Comput. Intell. Pattern Recognit. Proc. CIPR* **2019**, *2019*, 657–668. [CrossRef]
57. Bargoti, S.; Underwood, J.P. Image segmentation for fruit detection and yield estimation in apple orchards. *J. Field Robot.* **2017**, *34*, 1039–1060. [CrossRef]
58. Sa, I.; Ge, Z.; Dayoub, F.; Upcroft, B.; Perez, T.; McCool, C. Deepfruits: A fruit detection system using deep neural networks. *Sensors* **2016**, *16*, 1222. [CrossRef] [PubMed]
59. Koirala, A.; Walsh, K.B.; Wang, Z.; McCarthy, C. Deep learning for real-time fruit detection and orchard fruit load estimation: Benchmarking of 'MangoYOLO'. *Precis. Agric.* **2019**, *20*, 1107–1135. [CrossRef]
60. Apolo-Apolo, O.E.; Martínez-Guanter, J.; Egea, G.; Raja, P.; Pérez-Ruiz, M. Deep learning techniques for estimation of the yield and size of citrus fruits using a UAV. *Eur. J. Agron.* **2020**, *115*, 126030. [CrossRef]
61. Gené-Mola, J.; Vilaplana, V.; Rosell-Polo, J.R.; Morros, J.-R.; Ruiz-Hidalgo, J.; Gregorio, E. Multi-modal deep learning for Fuji apple detection using RGB-D cameras and their radiometric capabilities. *Comput. Electron. Agric.* **2019**, *162*, 689–698. [CrossRef]
62. Häni, N.; Roy, P.; Isler, V. A comparative study of fruit detection and counting methods for yield mapping in apple orchards. *J. Field Robot.* **2020**, *37*, 263–282.
63. Chen, S.W.; Shivakumar, S.S.; Dcunha, S.; Das, J.; Okon, E.; Qu, C.; Taylor, C.J.; Kumar, V. Counting apples and oranges with deep learning: A data-driven approach. *IEEE Robot. Autom. Lett.* **2017**, *2*, 781–788. [CrossRef]
64. Aquino, A.; Diago, M.P.; Millán, B.; Tardáguila, J. A new methodology for estimating the grapevine-berry number per cluster using image analysis. *Biosyst. Eng.* **2017**, *156*, 80–95. [CrossRef]
65. Santos, T.T.; De Souza, L.L.; Dos Santos, A.A.; Avila, S. Grape detection, segmentation, and tracking using deep neural networks and three-dimensional association. *Comput. Electron. Agric.* **2020**, *170*, 105247. [CrossRef]
66. Fan, S.; Li, J.; Zhang, Y.; Tian, X.; Wang, Q.; He, X.; Zhang, C.; Huang, W. On line detection of defective apples using computer vision system combined with deep learning methods. *J. Food Eng.* **2020**, *286*, 110102. [CrossRef]
67. Wan, S.; Goudos, S. Faster R-CNN for multi-class fruit detection using a robotic vision system. *Comput. Netw.* **2020**, *168*, 107036. [CrossRef]
68. Fu, L.; Gao, F.; Wu, J.; Li, R.; Karkee, M.; Zhang, Q. Application of consumer RGB-D cameras for fruit detection and localization in field: A critical review. *Comput. Electron. Agric.* **2020**, *177*, 105687. [CrossRef]
69. Kang, H.; Chen, C. Fast implementation of real-time fruit detection in apple orchards using deep learning. *Comput. Electron. Agric.* **2020**, *168*, 105108. [CrossRef]
70. Kamilaris, A.; Prenafeta-Boldú, F.X. Deep learning in agriculture: A survey. *Comput. Electron. Agric.* **2018**, *147*, 70–90. [CrossRef]
71. Chopin, J.; Laga, H.; Miklavcic, S.J. A new method for accurate, high-throughput volume estimation from three 2D projective images. *Int. J. Food Prop.* **2017**, *20*, 2344–2357. [CrossRef]
72. Fauzi, A.I.M.; Ibrahim, M.Z.B.; Saealal, M.S.B. Estimation of volume and weight of apple by using 2D contactless computer vision measuring method. In *Proceedings of the 2022 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*; IEEE: New York, NY, USA, 2022; pp. 1–6.
73. Ziaratban, A.; Azadbakht, M.; Ghasemnezhad, A. Modeling of volume and surface area of apple from their geometric characteristics and artificial neural network. *Int. J. Food Prop.* **2017**, *20*, 762–768.
74. Siswantoro, J.; Hilman, M.; Widiasri, M. Computer vision system for egg volume prediction using backpropagation neural network. In *Proceedings of the IOP Conference Series: Materials Science and Engineering*; IOP Publishing: Bristol, UK, 2017; p. 012002.
75. Nyalala, I.; Okinda, C.; Nyalala, L.; Makange, N.; Chao, Q.; Chao, L.; Yousaf, K.; Chen, K. Tomato volume and mass estimation using computer vision and machine learning algorithms: Cherry tomato model. *J. Food Eng.* **2019**, *263*, 288–298. [CrossRef]
76. Wariky, V.C.; Siswantoro, J.; Benarkah, N. Volume prediction of axisymmetric fruits using optimized convolutional neural networks with transfer learning strategy. *Int. J. Adv. Sci. Eng. Inf. Technol. (IJASEIT)* **2025**, *15*, 436–443. [CrossRef]
77. Onmankhong, J.; Poonpakdee, P.; Lapcharoensuk, R. Estimation of the weight and volume of lime (*Citrus aurantifolia* (Christm.) Swingle) fruit using computer vision based on traditional machine learning and deep learning. *Agronomy* **2024**, *14*, 2434. [CrossRef]

78. Li, H.; Han, T. DeepVol: Deep fruit volume estimation. In *Proceedings of the International Conference on Artificial Neural Networks*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 331–341.
79. Lüling, N.; Reiser, D.; Straub, J.; Stana, A.; Griepentrog, H.W. Fruit volume and leaf-area determination of cabbage by a neural-network-based instance segmentation for different growth stages. *Sensors* **2022**, *23*, 129. [CrossRef] [PubMed]
80. Yi, W.; Xia, S.; Kuzmin, S.; Gerasimov, I.; Cheng, X. RTFVE-YOLOv9: Real-time fruit volume estimation model integrating YOLOv9 and binocular stereo vision. *Comput. Electron. Agric.* **2025**, *236*, 110401. [CrossRef]
81. Sapkota, R.; Ahmed, D.; Churuvija, M.; Karkee, M. Immature green apple detection and sizing in commercial orchards using YOLOv8 and shape fitting techniques. *IEEE Access* **2024**, *12*, 43436–43452. [CrossRef]
82. Weiss, I.; Ray, M. Model-based recognition of 3D objects from single images. *IEEE Trans. Pattern Anal. Mach. Intell.* **2001**, *23*, 116–128. [CrossRef]
83. Deng, K.; Liu, A.; Zhu, J.-Y.; Ramanan, D. Depth-supervised nerf: Fewer views and faster training for free. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*; IEEE Computer Society: Washington, DC, USA, 2022; pp. 12882–12891.
84. Baak, A.; Müller, M.; Bharaj, G.; Seidel, H.-P.; Theobalt, C. A data-driven approach for real-time full body pose reconstruction from a depth camera. In *Proceedings of the 2011 International Conference on Computer Vision*; IEEE: Washington, DC, USA, 2011; pp. 1092–1099.
85. Karpushin, M.; Valenzise, G.; Dufaux, F. Local visual features extraction from texture+ depth content based on depth image analysis. In *Proceedings of the 2014 IEEE International Conference on Image Processing (ICIP)*; IEEE: New York, NY, USA, 2014; pp. 2809–2813.
86. Salazar, I.; Pertuz, S.; Martínez, F. Multi-modal RGB-D image segmentation from appearance and geometric depth maps. *TecnoLógicas* **2020**, *23*, 140–158. [CrossRef]
87. Huang, J.; Artemov, A.; Chen, Y.; Zhi, S.; Xu, K.; Nießner, M. Ssr-2d: Semantic 3d scene reconstruction from 2d images. *IEEE Trans. Pattern Anal. Mach. Intell.* **2024**, *46*, 8486–8501. [CrossRef] [PubMed]
88. Guo, Y.-X.; Tong, X. View-volume network for semantic scene completion from a single depth image. *arXiv* **2018**, arXiv:1806.05361.
89. Tan, P.X.; Hoang, D.-C.; Nguyen, A.-N.; Kamioka, E.; Duong, T.H.A.; Huynh, T.-M.; Nguyen, D.-M.; Ngo, D.-H.; Cao, M.-D.; Nguyen, T.-U. Depth-guided monocular object pose estimation for warehouse automation. *IEEE Access* **2025**, *13*, 110166–110184. [CrossRef]
90. Lee, W.; Park, N.; Woo, W. Depth-assisted real-time 3D object detection for augmented reality. In *Proceedings of the ICAT*; Virtual Reality Society of Japan: Shinjuku-ku, Tokyo, 2011; pp. 126–132.
91. Zhao, S.; Lin, H.; Jiang, Y.; Liu, J. A Review: Research Progress on Intelligent Technologies for Orchard Yield Monitoring. *Front. Plant Sci.* **2026**, *17*, 1879872. [PubMed]
92. Kurtser, P.; Lowry, S. RGB-D datasets for robotic perception in site-specific agricultural operations—A survey. *Comput. Electron. Agric.* **2023**, *212*, 108035. [CrossRef]
93. Wang, Z.; Walsh, K.B.; Verma, B. On-tree mango fruit size estimation using RGB-D images. *Sensors* **2017**, *17*, 2738. [CrossRef] [PubMed]
94. Miranda, J.C.; Arnó, J.; Gené-Mola, J.; Fountas, S.; Gregorio, E. AKFruitYield: Modular benchmarking and video analysis software for Azure Kinect cameras for fruit size and fruit yield estimation in apple orchards. *SoftwareX* **2023**, *24*, 101548. [CrossRef]
95. Vit, A.; Shani, G. Comparing RGB-D sensors for close range outdoor agricultural phenotyping. *Sensors* **2018**, *18*, 4413. [CrossRef] [PubMed]
96. Miranda, J.C.; Arnó, J.; Gené-Mola, J.; Lordan, J.; Asín, L.; Gregorio, E. Assessing automatic data processing algorithms for RGB-D cameras to predict fruit size and weight in apples. *Comput. Electron. Agric.* **2023**, *214*, 108302. [CrossRef]
97. Li, S.; Zanchin, A.; Marinello, F.; Guerrini, L. Three-dimensional analysis of fruit and vegetables: Sensors and methods. *Comput. Electron. Agric.* **2026**, *245*, 111497. [CrossRef]
98. Perez, R.M.; Cheein, F.A.; Rosell-Polo, J.R. Flexible system of multiple RGB-D sensors for measuring and classifying fruits in agri-food Industry. *Comput. Electron. Agric.* **2017**, *139*, 231–242. [CrossRef]
99. Fukuda, T.; Ishii, M.; Ito, H.; Dohsaka, K. Measurement of fruit diameter using RGB-D cameras for the purpose of fruit growth assessment. *Electron. Commun. Jpn.* **2026**, *146*, 202–209. [CrossRef]
100. Kim, J.-h.; Lee, D.-s.; Kwon, S.-k. Volume estimation method for irregular object using RGB-D deep learning. *Electronics* **2025**, *14*, 919. [CrossRef]
101. Lo, F.P.-W.; Sun, Y.; Qiu, J.; Lo, B. Food volume estimation based on deep learning view synthesis from a single depth map. *Nutrients* **2018**, *10*, 2005. [CrossRef] [PubMed]
102. Kim, E.; Hong, S.-J.; Kim, S.-Y.; Lee, C.-H.; Kim, S.; Kim, H.-J.; Kim, G. CNN-based object detection and growth estimation of plum fruit (Prunus mume) using RGB and depth imaging techniques. *Sci. Rep.* **2022**, *12*, 20796. [CrossRef] [PubMed]
103. Ubbens, J.R.; Stavness, I. Deep plant phenomics: A deep learning platform for complex plant phenotyping tasks. *Front. Plant Sci.* **2017**, *8*, 1190. [CrossRef] [PubMed]

104. Pound, M.P.; Atkinson, J.A.; Townsend, A.J.; Wilson, M.H.; Griffiths, M.; Jackson, A.S.; Bulat, A.; Tzimiropoulos, G.; Wells, D.M.; Murchie, E.H. Deep machine learning provides state-of-the-art performance in image-based plant phenotyping. *Gigascience* **2017**, *6*, gix083. [CrossRef]
105. Hartley, Z.K.; Jackson, A.S.; Pound, M.; French, A.P. GANana: Unsupervised domain adaptation for volumetric regression of fruit. *Plant Phenomics* **2021**, *2021*, 9874597. [CrossRef] [PubMed]
106. Siripurapu, I.D.; Mitra, A.; Mohanty, S.P.; Kougianos, E. iLog 3.0: Estimating food volume from 2D images using Mask R-CNN and monocular depth estimation. In *Proceedings of the 2025 IEEE Computer Society Annual Symposium on VLSI (ISVLSI)*; IEEE: Washington, DC, USA, 2025; pp. 1–6.
107. Devanna, R.P.; Romeo, L.; Reina, G.; Milella, A. Yield estimation in precision viticulture by combining deep segmentation and depth-based clustering. *Comput. Electron. Agric.* **2025**, *232*, 110025. [CrossRef]
108. Danilevicz, M.F.; Upadhyaya, S.R.; Batley, J.; Bennamoun, M.; Bayer, P.E.; Edwards, D. Understanding plant phenotypes in crop breeding through explainable AI. *Plant Biotechnol. J.* **2025**, *23*, 4200–4213. [CrossRef] [PubMed]
109. Li, G.; Zhi, Z.; Ling, B.W.-K. Globaldepth: A global information aggregation network for depth estimation. *IEEE Trans. Circuits Syst. II Express Briefs* **2024**, *71*, 3201–3205. [CrossRef]
110. Navarro, P.J.; Salas, E.; Rosique, F.; Ruszczak, B.; Egea-Cortines, M. Conditional generative adversarial network for segmentation and image reconstruction of grape berries in variable illumination conditions. *Expert Syst. Appl.* **2025**, *287*, 128076. [CrossRef]
111. Park, Y.; Kim, C.; Son, H.I. Fast and stable pedicel detection for robust visual servoing to harvest shaking fruits. *Comput. Electron. Agric.* **2024**, *220*, 108863. [CrossRef]
112. Bhattacharya, S.; Arunachalam, C.; Zhang, K.; Li, J.; Lu, R.; Li, Z. Detect, segment, cluster: Apple localization for robotic harvesting in complex orchards. *Smart Agric. Technol.* **2025**, *12*, 101642. [CrossRef]
113. Neupane, C.; Pereira, M.; Koirala, A.; Walsh, K.B. Fruit sizing in orchard: A review from caliper to machine vision with deep learning. *Sensors* **2023**, *23*, 3868. [CrossRef] [PubMed]
114. Blok, P.M.; van Henten, E.J.; van Evert, F.K.; Kootstra, G. Image-based size estimation of broccoli heads under varying degrees of occlusion. *Biosyst. Eng.* **2021**, *208*, 213–233. [CrossRef]
115. Gené-Mola, J.; Ferrer-Ferrer, M.; Gregorio, E.; Blok, P.M.; Hemming, J.; Morros, J.-R.; Rosell-Polo, J.R.; Vilaplana, V.; Ruiz-Hidalgo, J. Looking behind occlusions: A study on amodal segmentation for robust on-tree apple fruit size estimation. *Comput. Electron. Agric.* **2023**, *209*, 107854. [CrossRef]
116. Sukkuea, A.; Kurdthongmee, W.; Suwannarat, K. AGRO-YOLO-V: Hybrid instance segmentation and geometric modeling for single-view orange volume estimation. *IEEE Access* **2026**, *14*, 38088–38098. [CrossRef]