Applied Information System and Management (AISM)
Volume 9, (1) 2026, p. 147–158
P-ISSN: 2621-2536; E-ISSN: 2621-2544; DOI: 10.15408/aism.v9i1.50286

A Systematic Review of Deep Learning and Computer Vision Methods for Accurate Object Volume Measurement

Muhamad Achya Arifudin[1*], Kusrini[2], Andi Sunyoto[3], Ferry Wahyu Wibowo[4]

***Abstract*—Precise and efficient object volume measurement is crucial across diverse industrial domains, including logistics, manufacturing, and agriculture, where traditional methods are often labor-intensive and error-prone. DL and CV technologies offer a compelling alternative, providing greater accuracy, speed, and safety through non-contact, real-time, and adaptive solutions. To address existing knowledge gaps, this SLR is believed to be the first to integrate and analyze the three critical dimensions of DL holistically- and CV-based volume measurement: core methodologies, practical applications across various industries, and outstanding challenges, thereby providing a unified and comprehensive understanding of the state of the art that previous, more fragmented reviews have failed to deliver. Regarding methodology, dominant DL techniques include CNNs, Mask R-CNN, and U-Net for segmentation; GANs for 3D model generation; and PointNet/voxel networks for 3D data processing, with sensor integration impacting model architecture. Applications span agriculture, logistics, manufacturing, and construction, demonstrating high accuracy with error rates as low as 0.75% and MAPE typically ranging from 3.2% to 5%. Challenges involve occlusions, diverse environmental conditions, data scarcity, and computational costs. Prioritized research directions include lightweight models, multi-task learning, improved generalization, greater robustness, and explainable AI. Overall, this SLR comprehensively synthesizes current DL and CV methodologies, their practical applications, and future research directions in object volume measurement.**

***Index Term*—3D reconstruction, computer vision, deep learning, object segmentation, volume measurement.**

I. Introduction

Precise and efficient estimation of object volumes is crucial across diverse industrial domains, including logistics, manufacturing, agriculture, and quality control [1]. Traditional methods, which typically involve manual measurements with specialized equipment such as calipers and laser scanners, remain prevalent [2]. Nevertheless, these techniques are frequently labor-intensive, time-consuming, and susceptible to human error, especially for irregularly shaped objects [3]. Furthermore, they often present safety hazards to operators. In contrast, automated systems employing deep learning and computer vision provide a compelling alternative, offering improved accuracy, speed, and safety in various applications [4], [5].

Over the past decade, the field of computer vision (CV) has experienced rapid development alongside advances in deep learning (DL) [6]. This phenomenon is demonstrated by remarkable progress in processing, analyzing, and extracting geometric information from visual data, including videos and images, using DL techniques such as Convolutional Neural Networks (CNN), which have significantly enhanced capabilities for accurate object recognition, segmentation, and subsequent volume calculation, even for complex geometries [2], [4]. The integration of DL into volume measurement applications offers the potential for non-contact, fast, real-time, and highly adaptive solutions across various environments and objects [7].

Many individual studies have explored the use of specific techniques such as stereoscopic-based 3D reconstruction, Structured Light, or Single-Image Depth Estimation powered by DL for volume estimation [3], [6]. Due to the rapid growth of models and methodologies, there is significant fragmentation of knowledge regarding their practical applications, challenges, and limitations. The first fragmentation concerns the core methodology, which consists of DL models and CV sensors: which combination is most effective in the context of volume measurement (e.g., RGB cameras, RGB-D depth sensors, or LiDAR), and which specific deep learning architectures, such as Mask R-CNN with amodal heads, are most computationally efficient for real-time deployment while maintaining high accuracy, particularly in challenging scenarios such as occluded objects [8]. The second fragmentation concerns the practical application of the technology, specifically how it has been successfully implemented in real-world applications such as measuring crop yields, determining the dimensions of logistics packages, and measuring the volume of bulk materials [9]. The third fragmentation concerns the challenges and limitations of these

Received: 2 March 2026; Revised: 21 March 2026; Accepted: 5 April 2026.
*Corresponding author
[1]Muhamad Achya Arifudin, Dept. Informatika, Universitas Informatika dan Bisnis Indonesia (e-mail: achyaarifudin@unibi.ac.id).
[2]Kusrini, Dept. of Informatics Doctorate, University Amikom Yogyakarta, Indonesia (e-mail: kusrini@amikom.ac.id)
[3]Andi Sunyoto, Dept. of Informatics Doctorate, University Amikom Yogyakarta, Indonesia (e-mail: andi@amikom.ac.id)
[4]Ferry Wahyu Wibowo, Dept. of Informatics Doctorate, University Amikom Yogyakarta, Indonesia (e-mail: ferry.w@amikom.ac.id)

models and technologies, including operational constraints, accuracy issues under non-ideal conditions (e.g., lighting, occlusion), and scalability limitations.

Based on the above, this Systematic Literature Review (SLR) aims to bridge this gap. In this SLR, the latest DL and CV methodologies used in volume measurement are identified and synthesized. Furthermore, the implementation and practical applications of these technologies in various domains are analyzed and classified. Finally, remaining challenges are highlighted, and future research directions are proposed to encourage further development of these technologies. Unlike previous reviews limited to specific application domains or single sensor types, the unique contribution of this SLR lies in its holistic cross-domain synthesis, directly mapping how sensor integration influences deep learning architectures and evaluating their practical performance metrics across various industries. This SLR is therefore expected to provide a comprehensive guide for researchers, industry practitioners, and engineers seeking to harness the potential of these technologies for advanced volume measurement solutions.

While existing surveys have covered specific aspects of deep learning in computer vision, they often suffer from limitations such as focusing exclusively on a single application domain (e.g., agriculture) or overlooking the crucial impact of sensor integration on architectural design. As a result, prior reviews have failed to provide the comprehensive cross-domain synthesis necessary to understand the universal limitations and generalizability of current methodologies. This fact has led to significant fragmentation of knowledge, which this SLR seeks to overcome.

II. Related Work

Several literature reviews have discussed dimension measurement based on computer vision and deep learning. Some have discussed deep learning and deep-learning-based dimensionality measurement, including deep learning for depth estimation, which is the basis for 3D dimensional measurement [3], and others focus on object detection using the YOLO algorithm in specific applications, such as agriculture, analyzing its modifications and integrations to address domain-specific challenges [4].

Several literature reviews focus on deep learning for depth estimation, which is a fundamental component of 3D measurement. For instance, [10], [11], [12], [13], [14], and [15] provide comprehensive overviews of deep-learning-based monocular depth estimation methods. These reviews categorize various approaches, discuss datasets, highlight key aspects of state-of-the-art techniques, and trace the evolution from traditional geometry-based methods to modern deep learning models. Their analyses often cover the challenges, input data requirements, and training methodologies for accurately predicting depth from single or multiple images, laying the foundation for subsequent 3D measurements.

Beyond general depth estimation, other reviews address specific applications and broader scopes of dimensional measurement. Reference [16] synthesize the integration of deep learning into optical metrology, a field inherently concerned with precise dimensional measurement, including length and 3D shapes. Then [17] survey methods and datasets specifically for monocular 3D object detection, directly contributing to the derivation of volume, length, and area from images. While the review by [18] focuses on diagnostic accuracy in medical imaging, it remains relevant given the critical role of dimensional measurements of anatomical structures in that domain. Furthermore, [19] and [20] conduct systematic literature reviews on computer vision for livestock weight estimation and on public computer vision datasets for precision livestock farming, respectively, both of which involve direct dimensional analysis for agricultural purposes.

Collectively, these reviews provide a rich overview of deep learning and computer vision applications in dimensional measurement, ranging from foundational depth estimation techniques to specialized applications in optical metrology, medical imaging, and agriculture. They not only identify key methodologies and datasets but also discuss inherent challenges such as accuracy under non-ideal conditions, generalization across diverse scenes, and the need for robust metric information recovery. While these syntheses are crucial for understanding the current state of the field, they remain highly fragmented. Prior reviews excel in their specific domains, but none systematically connect methodological choices, such as sensor selection and architectural design, with their practical performance across diverse applications, including agriculture, logistics, and healthcare. Addressing this gap, this SLR provides the first holistic, cross-domain synthesis that explicitly links specific DL and CV methodologies to measurable performance outcomes and persistent challenges. Table 1 summarizes the differences between this SLR and the previous reviews.

Table 1.
Summary of Differences Between This SLR and Previous Reviews

Comparative Dimensions	Other Review	This Review
Topic focus	Depth estimation, optical metrology integration, object detection, diagnostic/medical imaging, weight estimation, and public computer vision datasets for precision livestock farming	Focus on measuring the volume of specific objects (quantitative).
Technique covered	Primarily, current deep learning methods, traditional and current deep learning, current computer vision and machine learning methods, and current computer vision and deep learning applications	Comparing and synthesizing state-of-the-art deep learning methods
Application scope	General 3d dimensional measurement, precise dimensional measurement (length, 3d shapes), deriving volume, length, and area from images, measurements of anatomical structures, and agriculture.	Systematically reviewing methodologies, applications, and future challenges across various domains.
Methodology	Narrative survey, systematic literature review	Systematic literature review (SLR) based on a strict protocol

Applied Information System and Management (AISM)
Volume 9, (1) 2026, p. 147–158
P-ISSN: 2621-2536; E-ISSN: 2621-2544; DOI: 10.15408/aism.v9i1.50286

III. Research Method

A. Research Questions

Developing a set of research questions is crucial to guide this SLR. These questions shape the search strategy, data extraction, and synthesis of findings, ensuring a focused, comprehensive, and relevant analysis of the literature. The following research questions were proposed:

- RQ1 (methodology): Which DL and CV methodologies are the most dominant and effective in measuring object volume, and what is the impact of sensor integration on model architecture design?
- RQ2 (practical applications and domains): In which industries have DL and CV-based volume measurement applications been implemented, and how have these solutions performed based on accuracy and efficiency metrics?
- RQ3 (limitations and future directions): What are the main challenges and limitations currently faced by CV and DL-based volume measurement systems? What are the recommended and prioritized research directions?

B. Search Strategy and Selection Criteria

A robust search strategy is essential for identifying pertinent studies that meet the criteria outlined in the research questions across the vast academic landscape [21], [22]. Defining inclusion and exclusion criteria is a crucial next step [22], [23], ensuring only the most relevant and high-quality studies are selected to enhance the review's rigor and validity [24].

The first criterion for included documents is that the research explicitly discusses methods for measuring, estimating, or determining the volume/dimensions of objects using deep learning or computer vision techniques. This criterion ensures that the document is directly relevant to the research's primary objective. Additionally, the research must meet the core technological requirements of this SLR, requiring at least one DL technique (such as CNN, U-Net, or a 3D reconstruction network) applied to CV data (such as RGB imagery, RGB-D depth imagery, or LiDAR point clouds). Furthermore, to ensure the quality and validity of the research results, included outputs must be journal articles, major conference proceedings, or theses/dissertations from reputable institutions [22]. This SLR also uses full-text articles to enable comprehensive data synthesis [22]. Finally, all included studies must have been published between January 2020 and December 2025 to ensure the currency and relevance of the findings [22], [25]. A few articles from early 2026 were also included if they were available as "early access" or "online first" at the time of the search. It is also important to exclude certain documents, such as studies that use only non-visual volume measurement techniques, studies that use CV without integrating DL, and publications with substantial content overlap. This exclusion ensures the review focuses on technologies relevant to CV and DL, and that each contribution is counted only once.

Another crucial component of an SLR is a search strategy to ensure the review is comprehensive and replicable [22]. This strategy includes selecting databases for literature searches and developing keyword combinations for effective searches [26]. This SLR uses the following databases: Scopus, which has broad coverage of journal articles including those in engineering and computer science [23], [27]; IEEE Xplore, which is highly relevant to topics involving the technical implementation of DL and CV; the Web of Science, known for its multidisciplinary coverage and citation indexing capabilities [25], [28]; and Google Scholar as a secondary source for specific searches [29], [30].

The next step is to formulate the search string to retrieve relevant articles. This SLR identifies three core concepts: volume measurement, deep learning, and computer vision/visual sensors. To retrieve literature on volume measurement, the keywords "volume measurement OR volume estimation OR 3D measurement OR dimensioning OR size estimation" are used. To retrieve literature on deep learning, the keywords "deep learning OR convolutional neural network OR CNN OR neural network OR deep neural network OR DNN" are used. To retrieve literature on computer vision/visual sensors, the keywords "computer vision OR machine vision OR image processing OR visual sensor OR RGB-D OR point cloud OR photogrammetry OR stereoscopy" are used. The fields searched are 'Title', 'Abstract', and 'Keywords' from each database to maximize relevance. These keyword combinations were deliberately crafted to align with the research questions: the methodological terms (e.g., "CNN," "point cloud") ensure comprehensive coverage for RQ1. In contrast, the volume and dimensioning terms capture the diverse practical applications and performance metrics required to answer RQ2 and RQ3. The results were filtered by year, with a search limit of 2020 to 2026, to retrieve the most recent research. Figure 1 shows the distribution of studies included in this SLR by year.

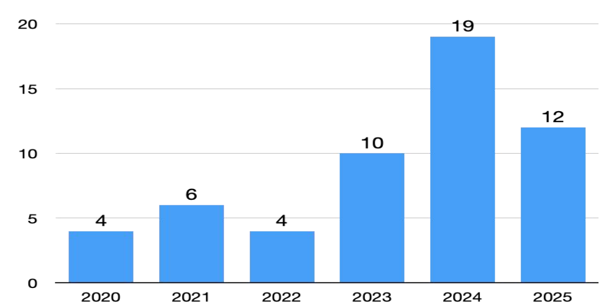

Fig. 1. Distribution of research by year.

After establishing a search strategy, the next step was to design an article selection protocol and data extraction procedure. This protocol is illustrated using a PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) flowchart [31]. This approach ensures transparency and reproducibility throughout the study selection process, aligning with established guidelines for systematic reviews [22]. Following these steps, starting with identifying and searching using various keywords, removing duplicates, and screening, the process involves meticulously evaluating each retrieved study against the predefined eligibility criteria to determine its suitability for inclusion [22]. Only articles that pass these stages are included in the final set for analysis and data extraction. Figure 1 shows the number of studies per year on object dimensional measurement using computer vision and deep learning. This trend highlights the growing academic interest and advancements in these interdisciplinary fields for accurate volumetric assessments. This rigorous selection process, guided by the PRISMA methodology, ensures the inclusion of only the most relevant and high-quality studies for comprehensive data synthesis and analysis [22], [32]. Figure 2 shows the literature search and selection process using the PRISMA framework.

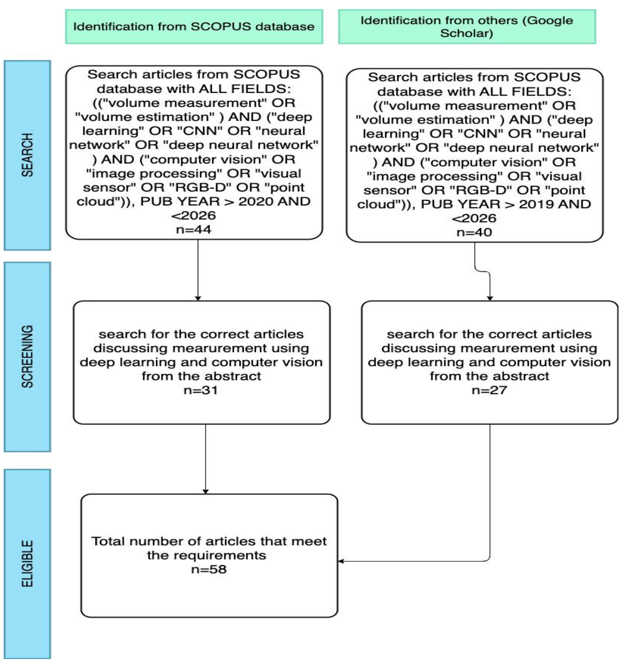

Fig. 2. The literature search and selection process using the PRISMA framework.

IV. RESULT

Figure 3 presents a clear understanding of the conceptual workflow for a typical volume measurement system, derived from the literature. This workflow illustrates the sequential stages from data acquisition using various sensors to the final volumetric output, serving as a technical foundation for the subsequent analysis of the research questions.

A. Research Question 1

Many deep learning architectures are used to measure or estimate the volume of objects. Most of the reviewed studies use a Convolutional Neural Network (CNN) architecture due to its effectiveness in extracting features from visual data [3]. This architecture extracts basic features such as edges, corners, and textures, which are crucial for understanding the geometric properties of objects within an image [10], or as integral components of more complex systems.

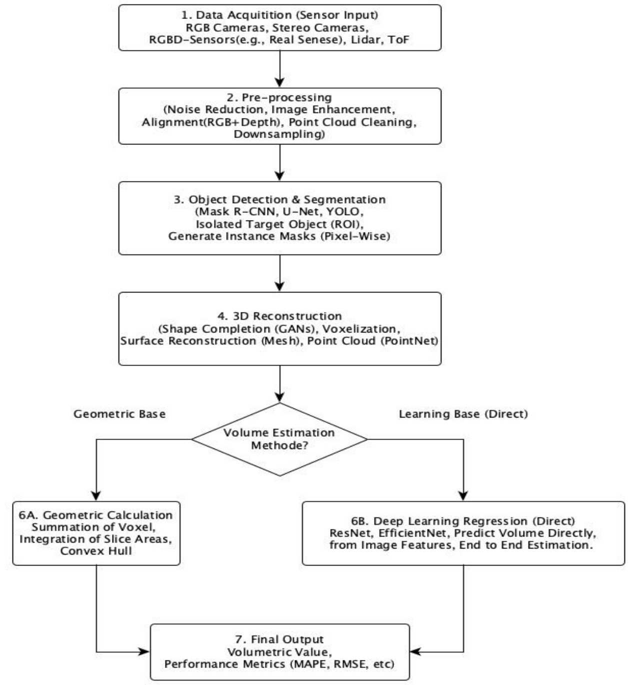

Fig. 3. Conceptual workflow of DL- and CV-based volume measurement systems.

Specialized segmentation networks, such as Mask R-CNN [33], [34], [35] and U-Net-inspired architectures [36], [37], are widely used to precisely delineate object boundaries in 2D images or 3D point clouds, which is an important precursor to volume calculations [3]. To generate and reconstruct objects while preserving their shape, Generative Adversarial Networks (GANs) are used, enabling the creation of 3D models from limited 2D data, which can then be used to estimate volume [38], [39]. For technologies that require robust 3D reconstruction and analysis, architectures designed for point clouds, such as PointNet and its derivatives, or voxel-based networks, are increasingly important, especially for the direct handling of 3D spatial data [35], [40].

Beyond merely cataloging these architectures, this synthesis reveals how they are implemented: their dominance is rarely based on a single model. Standard CNNs primarily serve as feature extractors in the initial stage. This feature extraction capacity is the primary driver of their popularity, as it eliminates the need for manual feature engineering. These features are subsequently utilized by specialized models, most notably Mask R-CNN, to generate precise instance segmentation masks. This two-stage approach, comprising feature extraction followed by precise segmentation, is dominant because it provides the critical, isolated region of interest (ROI) necessary for accurate geometric or depth-based volume calculations.

To determine the effectiveness of a methodology for measuring object volume, it is often assessed by its ability to accurately reconstruct 3D shapes or estimate volumes directly

Applied Information System and Management (AISM)
Volume 9, (1) 2026, p. 147–158
P-ISSN: 2621-2536; E-ISSN: 2621-2544; DOI: 10.15408/aism.v9i1.50286

from visual data [3]. One approach involves analyzing the combination of color and depth images, which can be a highly effective strategy for pixel-level volume estimation, where deep learning directly learns to estimate the volume of elements from the smallest pixel units [3]. By summing the volumes of these elements, the total volume of an irregular object can be accurately determined, eliminating the need for explicit 3D modeling of complex shapes [3]. Another strategy is to use precise object segmentation to isolate the target object, followed by geometric calculations based on depth information (e.g., from RGB-D images) or regression models that directly predict the volume of the segmented shape [3], [7], [33]. The techniques also include 3D reconstruction from multi-view images or depth scans, where the reconstructed 3D model (e.g., point cloud or mesh) serves as a basis for accurate volume calculations, particularly for objects with complex geometries [40], [41]. The integration of depth estimation capabilities, often produced from specialized deep ordinal regression networks or encoder-decoder CNNs, further improves the accuracy of these volume measurement systems [42], [43].

Table 2.
General Models Often Used in Research

Architecture	Articles
CNN	[36], [44], [45], [46], [47], [48]
Mask R-CNN	[8], [33], [34], [35], [41]
U-Net	[10], [36]
GAN	[10], [38], [39]
PointNet	[6], [35], [37], [40], [49], [50]
Voxel	[51], [52]

To measure an object's volume, sensor-generated images are required as input. Various sensors are commonly used in research to measure object volume, each offering a different type of data that influences subsequent model design and processing [3]. Commonly used sensors include RGB cameras, which produce standard color images and are often paired with deep learning models for object detection and segmentation [34], [53], [54]. RGB-D cameras provide both color and depth information, which significantly enhances the accuracy of 3D reconstruction and volume estimation [3], [38]. LiDAR sensors provide precise 3D point cloud data, which is crucial for detailed spatial reconstruction [10], [55], [56]. Active sensors, such as LiDAR and time-of-flight cameras, directly capture accurate 3D information but can be expensive and struggle to obtain depth for small or close-range objects due to their sensitivity limitations. Stereo cameras capture multiple views for passive depth estimation [10], [37], [57]. In addition, specialized sensors, such as multispectral cameras [58] and radar [49], are integrated with platforms, such as uncrewed aerial vehicles (UAVs) equipped with cameras, for large-scale environmental sensing [43], [59]. The selection of input sensors affects the capabilities and requirements of the computational model used to estimate volume [3]. For instance, depth sensors are particularly effective at capturing volumetric data, yet they often struggle with highly reflective surfaces or transparent objects, necessitating compensatory algorithmic approaches.

Beyond merely cataloging these hardware and software options, a critical comparative synthesis reveals the inherent trade-offs that dictate why certain architecture-sensor pairings succeed or fail in specific contexts. Standard RGB cameras offer high accessibility and rich texture information but critically lack inherent 3D depth, making them prone to failure in complex occlusion scenarios. Conversely, active sensors like LiDAR and time-of-flight cameras provide highly accurate 3D spatial data but incur high computational costs and are sensitive to environmental conditions, often failing when faced with transparent or highly reflective surfaces. Consequently, specific pairings dominate by application domain: RGB-D sensors paired with Mask R-CNN or U-Net excel in controlled indoor logistics for precise parcel segmentation, whereas LiDAR coupled with PointNet is the preferred combination for large-scale outdoor volume measurements in agriculture and construction.

As noted earlier, integrating multiple sensors affects the model architecture, requiring researchers to adapt deep learning models to process and combine these unique data streams effectively. For example, if depth data from an RGB-D camera or a stereo vision system is available, the selected architecture is typically one that can directly estimate pixel-level volume or achieve higher-accuracy 3D reconstruction [3]. This integration often leads to models that process color and depth channels simultaneously, sometimes using separate network branches that are combined at a later stage [3]. If LiDAR or structured light scanning is used, an architecture designed for point cloud data is required, enabling direct processing of 3D spatial information [35], [40], [56]. Furthermore, multimodal sensor setups (e.g., combining RGB with LiDAR or radar) require advanced fusion architectures that can effectively integrate heterogeneous data, compensate for individual sensor limitations, and improve robustness under varying environmental conditions [10], [49]. These integration techniques often involve multiple strategies, ranging from early- to late-and feature-level integration, each defining a specific network structure to enable comprehensive and accurate volume estimation [3].

Table 3
Various Sensors That are Often Used in Research

Input Sensor	Articles
RGB-camera	[7], [34], [49], [53], [54], [57], [60], [61], [62], [63], [64], [65], [66], [67], [68]
RGB-D camera	[3], [38], [41], [69], [70], [71]
LIDAR	[49], [54], [55], [56]
Stereo camera	[8], [37], [40], [44], [47], [57], [72]
UAV and camera	[35], [45], [59], [73]
Binocular cameras	[6], [74]
Intel RealSense	[36], [75]

Multispectral camera	[58], [68]
RADAR	[49], [54]
Single 2D image	[10], [43], [52], [76]
Stereo images	[10], [48]
Structured light	[61], [77]
Others	[7], [10], [34], [36], [49], [54], [56], [57], [58], [61], [66], [68], [69], [72], [75], [78], [79], [80], [81], [82], [83], [84], [85], [86], [87]

B. Research Question 2

Deep learning and CV-based volume measurement applications have been widely implemented across various industries, from manufacturing and logistics to healthcare and environmental monitoring [3]. This is driven by the need for greater accuracy, speed, and safety compared to traditional manual methods [7]. These advanced solutions offer non-contact, real-time, and adaptive capabilities, making them suitable for a wide range of environments and object types [3], [56].

In agriculture, DL and CV are used for tasks such as measuring crop yields, estimating weed volume, and characterizing plant structure. Systems can measure fruit volume individually or across entire plants, which is crucial for precision farming and predicting crop yields [3]. In logistics, automated volume estimation is crucial for managing merchandise, optimizing packaging, and calculating shipping costs. These systems typically integrate laser scanners and cameras with deep learning to measure package and container dimensions [3], [7]. In manufacturing, applications range from quality control through precise component volume verification to optimizing material usage and inventory management [3]. DL and CV are also applied in construction to estimate the volume of materials, such as soil in an excavator bucket, improving efficiency and safety on construction sites [3], [7]. This technology is also used in food science and nutrition to predict food volume from a single depth map, especially for dietary assessment [51]. Furthermore, these techniques are also applied in research and development beyond the specific industries mentioned above, including general vision-based estimation of various static objects and specialized tasks such as estimating coral volume from multi-view images [3].

In various studies, models have been developed for various object types. The first and most frequently studied type is simple geometric objects, which are often used to validate the fundamental accuracy of new volume estimation methods [3]. Some studies aim to measure the volume of well-defined items, such as merchandise boxes [3], [47]; even food items are sometimes modeled as geometric shapes, such as prisms, ellipsoids, cylinders, and rectangles, for estimation purposes [3], [51]. Most research focuses on irregular objects, which naturally pose greater challenges for traditional measurement techniques due to their complex, non-uniform surfaces. Examples include irregular shapes such as pyramids, cylinders, rods, concave shapes, U-shapes, and irregular planar shapes [3]. Food items, such as "ham" studied for automatic surface area and volume prediction, also fall into this category due to their inherent variability and non-standard shapes [3], [41]. Natural objects such as "logs in a paper mill" are also considered irregular [3]. Various studies have been conducted to improve the ability of DL and CV methods to address the challenges of measuring objects without a simple mathematical description [3]. The third category is bulk materials, which represent another important class, especially in industrial and environmental applications where large amounts of loose substances need to be quantified. This category includes materials like gravel, sand, or agricultural produce, where accurate volume measurement is critical for inventory management and process optimization, including "soil in the excavator bucket" in construction [47] and "piles of material" or "stockpiles" found in open-pit mines [7], [35]. Accurate measurement of bulk materials is critical for inventory management, operational efficiency, and environmental monitoring in these sectors.

Assessing the accuracy and efficiency metrics of DL and CV-based object-volume measurement models is equally important for their development. For absolute accuracy, the average error rate in cubic centimeters (cm^3) can be used, as it is directly measurable. For example, a deep learning method for estimating the volume of irregular objects using RGB-D images achieved an error of 3.40% for an object with a volume of 68.40 cm^3 (approximately 2.32 cm^3), substantially lower than the 8.54% error of conventional geometric measurements [3]. Similarly, other studies report low error percentages, such as 0.75%, 2.76%, and 3.20%, indicating precise volume estimations for various object types [7]. For a larger object with an actual volume of 420.29 cm^3, the deep learning approach yields an error of 2.17% (about 9.13 cm^3), outperforming the geometric method, which has an error of 6.19% [3]. These findings underscore deep learning's ability to predict absolute volumes more precisely, especially for objects with complex surfaces [3]. The Mean Absolute Percentage Error (MAPE) is also frequently used as an evaluation metric [3]. MAPE represents the average absolute prediction error as a percentage of the actual value; a lower MAPE indicates better model accuracy [10]. For example, in coral volume estimation, the MAPE for volume prediction ranged from 26.24% with 100 samples to 12.40% with 3,200 samples, indicating improved relative accuracy with a larger dataset [46]. Another study evaluating an excavator bucket fill factor estimation reported an overall average accuracy of 95.23%, corresponding to a MAPE of around 5% across six environmental conditions [47]. In food volume estimation, one system achieved an overall MAPE of 10.5%, with higher MAPE observed for food categories with fewer and weaker features [88].

Other metrics such as MAE, RMSE, and MAPE are also frequently used to evaluate DL and CV-based object volume measurements [3]. MAE describes the average absolute difference between the predicted and actual values [10]. Several studies have reported MAEs of 0.61 cm^3 and 0.49 cm^3, demonstrating precise quantitative deviation from true volumes [7]; MAE for tomato fruit dimension prediction reached approximately 0.77 mm, and for an object volume study, it was

Applied Information System and Management (AISM)
Volume 9, (1) 2026, p. 147–158
P-ISSN: 2621-2536; E-ISSN: 2621-2544; DOI: 10.15408/aism.v9i1.50286

0.046 cm^3 [3]. RMSE provides an estimate of the average deviation of measurements from the true value. For instance, RMSE values for various geometric shapes such as triangular prisms, rectangular prisms, and square prisms range from 0.44 cm^3 to 1.01 cm^3, further confirming the precision of the proposed methods [3]. In practical terms, a lower MAE/RMSE (e.g., below 1 cm^3) implies high precision, suggesting these systems are suitable for applications where millimeter-level accuracy matters, such as quality control for small components or sizing small agricultural items. Conversely, a higher absolute error might still be acceptable for large objects (e.g., shipping crates) as long as the relative percentage error remains low. For apple measurement, an RMSE of 0.77 mm was reported, while a study on concrete spalling depth reported an average error of −0.47 mm with a standard deviation of 1.92 mm, outperforming time-of-flight approaches [7]; other measurements for the same fruit were reported as 9.60 mm (14%) at a distance of 1.0 m, 10.50 mm (16%) at a distance of 1.5 m, and 10.00 mm (15%) across all data [69]. For the development of a model for measuring tomato fruit, values of 2.9 mm for fruit width [33], 3.4 mm for fruit length [33], 2.957 mm for fruit width [33], and 3.412 mm for fruit length [33] were reported. In addition to the metrics above, MAPE is the average relative error expressed as a percentage, representing the difference between the measured and actual volumes. In various studies, MAPE values of 5% [47], 4% [8], and 3.2% were reported across different applications, underscoring the consistent performance of deep learning models in volumetric analysis [7]. Analysis of this range confirms that the lower end (near 3.2%) is typically achieved for simple geometric objects with clearly defined edges and low occlusion. In comparison, the higher end (5% and above) reflects the greater difficulty in accurately measuring complex, organic, or irregular shapes (e.g., specific agricultural produce), where self-occlusion and surface texture vary significantly, challenging the models' generalization capabilities.

C. *Research Question 3*

While CV- and DL-based volume measurement systems offer significant advantages over traditional methods, they still encounter several critical challenges and limitations that hinder their widespread applicability and accuracy. A recurring challenge in existing literature is the difficulty in handling occlusions and complex object geometries [46], [51]. Current approaches often require extensive strategies such as view synthesis or point cloud completion to mitigate issues arising from occluded or overlapping objects, which can still lead to inaccuracies, particularly for irregularly shaped objects [10], [51].

Another significant limitation is reduced accuracy and robustness under diverse and non-ideal environmental conditions [47]. Many methods achieve high accuracy in controlled settings but struggle with transparent or reflective surfaces, varying lighting, and shadows [10]. This susceptibility to environmental factors, coupled with potential device dependency [3], limits the generalization of these systems to real-world outdoor or industrial environments where conditions are rarely uniform [47]. Furthermore, data scarcity and generalization challenges are prevalent, as models often underperform on data not encountered during training, highlighting the resource-intensive nature of dataset annotation for supervised learning [49]. The computational cost of sophisticated models, particularly vision transformers, also remains a barrier to real-time deployment and broader adoption [49]. Lastly, the lack of explainability in deep learning models, often referred to as their "black-box nature," makes interpretation and validation difficult in critical applications [10].

To address these knowledge gaps, several prioritized research directions are proposed. Future efforts should focus on developing lightweight and computationally efficient models, potentially through optimizing self-attention mechanisms in vision transformers or integrating invertible downsampling techniques [49]. Multi-task learning is another promising direction, aiming to create unified architectures that can simultaneously perform tasks such as 3D object detection, segmentation, and tracking, thereby improving efficiency [49]. Improving generalization is crucial, requiring research into unsupervised learning methods or more efficient data annotation strategies to mitigate data scarcity and model underperformance on novel data [49]. There is also a critical need for systems that demonstrate improved robustness across varying environmental conditions [47], accurately estimate depth for complex materials such as transparent or reflective surfaces, and function effectively from diverse viewpoints (e.g., aerial, wide-angle, and ground-level) [10]. Finally, research into explainable AI for deep learning models is essential to foster trust and facilitate wider adoption in domains requiring transparent decision-making [10]. Pursuing these research directions would significantly advance the practicality and reliability of CV- and DL-based volume measurement systems.

V. Conclusions

Convolutional Neural Networks (CNNs) are widely used in computer vision due to their effectiveness in feature extraction for object volume measurement. For accurate volume estimation, which typically begins with object segmentation, models such as Mask R-CNN and U-Net are widely used to delineate precise object boundaries. Generative Adversarial Networks (GANs) can generate 3D models from limited 2D images to support 3D understanding, while PointNets and voxel-based networks directly process 3D data for shape reconstruction and volume estimation. Common volume calculation methods combine segmentation with color and depth information, followed by geometry-based computations or regression. For reliable 3D reconstruction, multi-view images or depth data are often used, enabling 3D models to be built from images. Furthermore, advanced deep learning and zero-shot depth estimation models can further improve accuracy by generating high-quality depth maps. Sensor selection is also crucial, as RGB, RGB-D, LiDAR, stereo, and specialized sensors provide diverse data types, and multimodal integration improves robustness and performance.

DL and CV-based volume measurement systems are widely applied in industries such as manufacturing, logistics, healthcare, and environmental monitoring, driven by the need for greater accuracy, speed, and safety. In agriculture, these systems are used for measuring crop yield, estimating weed volume, and characterizing plant structure. The logistics sector leverages these systems for efficient inventory management, package optimization, and accurate shipping cost calculations. Manufacturing and food science utilize advanced methods for quality control and dietary assessment, respectively, showcasing high accuracy and efficiency. Error rates can be as low as 0.75% to 3.40%, with Mean Absolute Percentage Error (MAPE) ranging from 3.2% to 5%, indicating precise deviations in quantitative measurements across applications.

DL-based volume measurement systems encounter challenges with occlusions and irregular shapes that compromise accuracy. Many methods struggle under real-world conditions, including varying lighting and reflective surfaces. Furthermore, standard RGB cameras have limitations due to their lack of inherent 3D data, requiring reliance on specialized depth-sensing equipment. Data scarcity and high computational demands limit the development of high-quality datasets and real-time applications, while the opaque nature of deep learning complicates model interpretation and validation.

Consequently, current research emphasizes the development of lightweight and efficient models to support real-time performance. Researchers are also focusing on improving generalization through unsupervised learning or reduced annotation requirements. Finally, explainable AI is being promoted to improve model transparency, reliability, and user trust.

The synthesis of the research findings reveals three key messages: (1) the most effective volume measurement systems use a cascaded approach, integrating 2D segmentation models such as Mask R-CNN with specialized 3D techniques; (2) high accuracy (MAPE < 4%) is achievable, but primarily for rigid, simple objects, as complex shapes pose significant challenges; and (3) there is a trend towards direct volume regression models and multimodal sensor fusion (RGB-D/LiDAR) to improve robustness, despite the increased computational complexity.

Despite adhering to a rigorous protocol, this SLR has limitations, including a search restricted to English-language publications from specific databases, potentially omitting relevant studies published in other languages. Additionally, due to the rapidly evolving nature of deep learning, recent state-of-the-art models may not be fully represented. Finally, direct comparisons across studies are complicated by the absence of standardized datasets and evaluation protocols across different application domains, such as agriculture and logistics.

REFERENCES

[1] F. S. Konstantakopoulos, E. I. Georga, and D. I. Fotiadis, "A review of image-based food recognition and volume estimation artificial intelligence systems," *IEEE Reviews in Biomedical Engineering* , vol. 17. Institute of Electrical and Electronics Engineers, pp. 136–152, Jun. 05, 2023. doi: 10.1109/rbme.2023.3283149.

[2] Z. Yu *et al.* , "Single-image estimation of tree volume via pixel-mapped 3D reconstruction: A low-cost solution using deep learning and curvature segmentation," *The Science of The Total Environment*, vol. 1002, Art. no. 180420, Sep. 2025, doi: 10.1016/j.scitotenv.2025.180420.

[3] F. S. Konstantakopoulos, E. I. Georga, and D. I. Fotiadis, "A review of image-based food recognition and volume estimation artificial intelligence systems," *IEEE Reviews in Biomedical Engineering* , vol. 17, pp. 136–152, Jun. 05, 2023. doi: 10.1109/rbme.2023.3283149.

[4] Z. Yu *et al.*, "Single-image estimation of tree volume via pixel-mapped 3D reconstruction: A low-cost solution using deep learning and curvature segmentation," *The Science of The Total Environment*, vol. 1002, Art. no. 180420, Sep. 2025, doi: 10.1016/j.scitotenv.2025.180420.

[5] J.-H. Kim, D. Lee, and S. Kwon, "Volume estimation method for irregular object using RGB-D deep learning," *Electronics*, vol. 14, no. 5, Art. no. 919, Feb. 2025, doi: 10.3390/electronics14050919.

[6] C. Badgujar, A. Poulose, and H. Gan, "Agricultural object detection with you only look once (YOLO) algorithm: A bibliometric and systematic literature review," *Computers and Electronics in Agriculture*, vol. 223, Art. no. 109090, May 2024, doi: 10.1016/j.compag.2024.109090.

[7] Y. Ma, J. Yin, F. Huang, and Q. Li, "Surface defect inspection of industrial products with object detection deep networks: a systematic review," *Artificial Intelligence Review*, vol. 57, no. 12, Art. no. 326, Oct. 17, 2024. doi: 10.1007/s10462-024-10956-3.

[8] Y. Zhao, H. Qin, L. Xu, H. Yu, and Y. Chen, "A review of deep learning-based stereo vision techniques for phenotype feature and behavioral analysis of fish in aquaculture," *Artificial Intelligence Review*, vol. 58, no. 1. Art. no. 332, Nov. 07, 2024. doi: 10.1007/s10462-024-10960-7.

[9] H. Dow, M. Perry, S. Pennada, R. J. Lunn, and S. Pytharouli, "3D reconstruction and measurement of concrete spalling using near-field photometric stereo and YOLOv8," *Automation in Construction*, vol. 166, Art. no. 105633, Jul. 2024, doi: 10.1016/j.autcon.2024.105633.

[10] J. Gené-Mola *et al.*, "Looking behind occlusions: A study on amodal segmentation for robust on-tree apple fruit size estimation," *Computers and Electronics in Agriculture*, vol. 209, Art. no. 107854, Apr. 2023, doi: 10.1016/j.compag.2023.107854.

[11] W. Guan *et al.*, "A bucket fill factor estimation method in construction environments by fusing deep learning and machine vision," *Journal of Cleaner Production*, vol. 426, Art. no. 139094, Sep. 2023, doi: 10.1016/j.jclepro.2023.139094.

[12] U. Rajapaksha, F. Sohel, H. Laga, D. Diepeveen, and M. Bennamoun, "Deep learning-based depth estimation methods from monocular image and videos: A comprehensive survey," *ACM Computing Surveys*, vol. 56, no. 12, Art. no. 327, Jul. 2024, doi: 10.1145/3677327.

[13] A. Rohan, M. J. Hasan, and A. Petrovski, "A systematic literature review on deep learning-based depth estimation in computer vision," *arXiv (Cornell University)*, Jan. 2025, doi: 10.48550/arxiv.2501.05147.

[14] F. Khan, S. Salahuddin, and H. Javidnia, "Deep learning-based monocular depth estimation methods—a state-of-the-art review," *Sensors*, vol. 20, no. 8. Art. no. 2272, Apr. 16, 2020. doi: 10.3390/s20082272.

[15] A. Masoumian, H. A. Rashwan, J. Cristiano, M. S. Asif, and D. Puig, "Monocular depth estimation using deep learning: a review," *Sensors*, vol. 22, no. 14. Art. no 5353, Jul. 18, 2022. doi: 10.3390/s22145353.

[16] Y. Ming, X. Meng, C. Fan, and H. Yu, "Deep learning for monocular depth estimation: A review," *Neurocomputing*, vol. 438, pp. 14–38, Jan. 08, 2021. doi: 10.1016/j.neucom.2020.12.089.

[17] J. Zhang, Y. Wu, and H. Jiang, "Survey on monocular metric depth estimation," *Computers*, vol. 14, no. 11, Art. no 502, Nov. 2025, doi: 10.3390/computers14110502.

[18] C. Zuo *et al.*, "Deep learning in optical metrology: A review," *Light Science & Applications* , vol. 11, no. 1., Art. no. 39, Feb. 23, 2022. doi: 10.1038/s41377-022-00714-x.

P-ISSN: 2621-2536; E-ISSN: 2621-2544; DOI: 10.15408/aism.v9i1.50286

[19] S. Kim and Y. Hwang, "A survey on deep learning based methods and datasets for monocular 3D object detection," *Electronics*, vol. 10, no. 4, Art. no. 517, Feb. 2021, doi: 10.3390/electronics10040517.

[20] R. Aggarwal *et al.*, "Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis," *npj Digital Medicine*, vol. 4, no. 1. Nature Portfolio, Art. no 65, Apr. 07, 2021. doi: 10.1038/s41746-021-00438-z.

[21] R. Dohmen, C. Catal, and Q. Liu, "Computer vision-based weight estimation of livestock: a systematic literature review," *New Zealand Journal of Agricultural Research*, vol. 65, pp. 227-248, Jan. 2021, doi: 10.1080/00288233.2021.1876107.

[22] A. Bhujel, Y. Wang, Y. Lu, D. Morris, and M. Dangol, "Public computer vision datasets for precision livestock farming: A systematic survey," *arXiv*, Jun. 2024, doi: 10.48550/arXiv.2406.10628.

[23] T. van Klompenburg, A. Kassahun, and C. Catal, "Crop yield prediction using machine learning: A systematic literature review," *Computers and Electronics in Agriculture*, vol. 177, Art. no. 105709, Aug. 2020, doi: 10.1016/j.compag.2020.105709.

[24] I. Dawar, A. K. Srivastava, M. Singal, N. Dhyani, and S. K. Rastogi, "A systematic literature review on municipal solid waste management using machine learning and deep learning," *Artif. Intell. Rev.*, vol. 58, no. 6, Art. no. 585, Jun. 2025, doi: 10.1007/s10462-025-11196-9.

[25] S. Dodampegama, L. Hou, E. Asadi, G. Zhang, and S. Setunge, "Revolutionizing construction and demolition waste sorting: Insights from artificial intelligence and robotic applications," *Resources Conservation and Recycling* , vol. 202, Art. no. 107375, Dec. 2023, doi: 10.1016/j.resconrec.2023.107375.

[26] Z. Li, X. Xu, X. Tao, Y. Gao, and X. Rao, "Process, challenges and solutions of fruit 3D reconstruction: A review," *Computers and Electronics in Agriculture*, vol. 239, Art. no. 111068, Oct. 2025, doi: 10.1016/j.compag.2025.111068.

[27] O. Abayomi-Alli, R. Damaševičius, A. Qazi, M. Adedoyin-Olowe, and S. Misra, "Data augmentation and deep learning methods in sound classification: A systematic review," *Electronics*, vol. 11, no. 22. Multidisciplinary Digital Publishing Institute, Art. no. 3795, Nov. 18, 2022. doi: 10.3390/electronics11223795.

[28] H. Liu, H. Su, L. Sun, and D. Dias-da-Costa, "State-of-the-art review on the use of AI-enhanced computational mechanics in geotechnical engineering," *Artificial Intelligence Review*, vol. 57, no. 8, Art. no. 212, Jul. 2024, doi: 10.1007/s10462-024-10836-w.

[29] K. Mostafa and T. Hegazy, "Review of image-based analysis and applications in construction," *Automation in Construction*, vol. 122, Art. no. 103516, Dec. 2020, doi: 10.1016/j.autcon.2020.103516.

[30] A. Upadhyay *et al.*, "Deep learning and computer vision in plant disease detection: a comprehensive review of techniques, models, and trends in precision agriculture," *Artif. Intell. Rev.*, vol. 58, no. 3, Art. no. 278, Jan. 2025, doi: 10.1007/s10462-024-11100-x.

[31] N. Marín-Jiménez *et al.*, "Criterion-related validity of field-based methods and equations for body composition estimation in adults: A systematic review," *Curr. Obes. Rep.*, vol. 11, no. 4, pp. 336–364, Dec. 2022, doi: 10.1007/s13679-022-00488-8.

[32] F. A. Hidajat, "Augmented reality applications for mathematical creativity: A systematic review," *J. Comput. Educ.*, vol. 11, no. 4, pp. 991–1018, Dec. 2024, doi: 10.1007/s40692-023-00287-7.

[33] A. Sameh, M. Rostami, M. Oussalah, R. Korpelainen, and V. Farrahi, "Digital phenotypes and digital biomarkers for health and diseases: A systematic review of machine learning approaches utilizing passive non-invasive signals collected via wearable devices and smartphones," *Artif. Intell. Rev.*, vol. 58, no. 2, Art. no. 173, Feb. 2025, doi: 10.1007/s10462-024-11009-5.

[34] M. Gusenbauer and S. P. Gauster, "How to search for literature in systematic reviews and meta-analyses: A comprehensive step-by-step guide," *Technological Forecasting and Social Change*, vol. 212, Art. no. 123833, Dec. 2024, doi: 10.1016/j.techfore.2024.123833.

[35] J. Lee, "Artificial intelligence approach for tomato detection and mass estimation in precision agriculture," *Sustainability*, vol. 12, no. 21, Art. no. 9138, Oct. 2020, doi: 10.3390/su12219138.

[36] P. M. Blok, "Image-based size estimation of broccoli heads under varying degrees of occlusion," *Biosystems Engineering*, vol. 208, Art. no. 213, 2021, doi: 10.1016/j.biosystemseng.2021.06.001.

[37] A. Ellinger, "Automatic segmentation of bulk material heaps using color, texture, and topography from aerial data and deep learning-based computer vision," *Remote Sensing*, vol. 15, Art. no. 1, 2023, doi: 10.3390/rs15010211.

[38] M. M. Waqar, "A dual-segmentation framework for the automatic detection and size estimation of shrimp," *Sensors*, vol. 25, Art. no. 18, 2025, doi: 10.3390/s25185830.

[39] J. L. Louëdec and G. Cielniak, "3D shape sensing and deep learning-based segmentation of strawberries," *Computers and Electronics in Agriculture*, vol. 190, Art. no. 106374, Oct. 2021, doi: 10.1016/j.compag.2021.106374.

[40] G. Chen, Z.-H. Mao, M. Sun, K. Liu, and W. Jia, "Shape-preserving generation of food images for automatic dietary assessment," *arXiv*, Aug. 2024, doi: 10.48550/arXiv.2408.13358.

[41] Z. Shao, G. Vinod, J. He, and F. Zhu, "An end-to-end food portion estimation framework based on shape reconstruction from monocular image," in *2022 IEEE International Conference on Multimedia and Expo (ICME)*, Jul. 2023, Art. no. 942, doi: 10.1109/icme55011.2023.00166.

[42] D. Cui, P. Liu, Y. Liu, Z. Zhao, and F. Jiang, "Automated phenotypic analysis of mature soybean using multi-view stereo 3D reconstruction and point cloud segmentation," *Agriculture*, vol. 15, no. 2, Art. no. 175, Jan. 2025, doi: 10.3390/agriculture15020175.

[43] Y. Gan, S. Liong, and Y. Huang, "Automatic surface area and volume prediction on ellipsoidal ham using deep learning," *arXiv*, Mar. 2022, doi: 10.48550/arXiv.1901.04947.

[44] C. C. Lo, "Depth estimation from monocular images and sparse radar using deep ordinal regression network," in *Proc. 2021 IEEE Int. Conf. Image Process. (ICIP)*, Anchorage, AK, USA, 2021, pp. 3343–3347, doi: 10.1109/ICIP42928.2021.9506550.

[45] M. Hu *et al.*, "Metric3D v2: A versatile monocular geometric foundation model for zero-shot metric depth and surface normal estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 46, no. 12, pp. 8352–8368, Aug. 2024, doi: 10.1109/tpami.2024.3444912.

[46] G. Böer *et al.*, "A deep-learning based pipeline for estimating the abundance and size of aquatic organisms in an unconstrained underwater environment from continuously captured stereo video," *Sensors*, vol. 23, no. 6, Art. no. 3311, Mar. 2023, doi: 10.3390/s23063311.

[47] A. M. Elsheshtawy *et al.*, "Artificial intelligence and machine learning potential as tools for georeferencing and features detection using UAV imagery," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. XLVIII-1/W2-2023, pp. 1887–1894, Dec. 2023, doi: 10.5194/isprs-archives-xlviii-1-w2-2023-1887-2023.

[48] D. E. Farchione, R. Idoughi, and P. Wonka, "No mesh, no problem: Estimating coral volume and surface from sparse multi-view images," *arXiv*, Sep. 2025, doi: 10.48550/arXiv.2509.11164.

[49] J. Lu, "A neural network–based approach for fill factor estimation and bucket detection on construction vehicles," *Computer Aided Civil and Infrastructure Engineering*, vol. 36, Art. no. 12, 2021, doi: 10.1111/mice.12675.

[50] M. S. Dizaji, "Deep learning-based visual measurement extraction within an adaptive digital twin framework from limited data using transfer learning," *arXiv*, Oct. 2024, doi: 10.48550/arXiv.2410.05403.

[51] S. Y. Alaba, A. C. Gürbüz, and J. E. Ball, "Emerging trends in autonomous vehicle perception: Multimodal fusion for 3D object detection," *World Electr. Veh. J.*, vol. 15, no. 1, Art. no. 20, Jan. 2024, doi: 10.3390/wevj15010020.

[52] A. El-Dawy, A. El-Zawawi, and M. El-Habrouk, "MonoLite3D: Lightweight 3D object properties estimation," *arXiv*, Mar. 2025, doi: 10.48550/arXiv.2503.02201

[53] W. Tay, "Current developments in digital quantitative volume estimation for the optimisation of dietary assessment," *Nutrients*, vol. 12, Art. no. 4, 2020, doi: 10.3390/nu12041167.

[54] Z. Xu, A. Wang, F. Hou, and G. Zhao, "Three-dimensional reconstruction of industrial parts from a single image," *Visual Computing for Industry Biomedicine and Art*, vol. 7, no. 1, Art. no. 7, Mar. 2024, doi: 10.1186/s42492-024-00158-7.

[55] J. Yang and C.-H. Lee, "Real-time data-driven method for bolt defect detection and size measurement in industrial production," *Actuators*, vol. 14, no. 4, Art. no. 185, Apr. 2025, doi: 10.3390/act14040185.

[56] D. Kim, "Toward co-robotic construction: Visual site monitoring & hazard detection to ensure worker safety," Ph.D. dissertation, Dept. Civil Environ. Eng., Univ. Michigan, Ann Arbor, MI, USA, 2021.
[57] S. Jin *et al.*, "Non-destructive estimation of field maize biomass using terrestrial lidar: an evaluation from plot level to individual leaf level," *Plant Methods*, vol. 16, no. 1, Art. no. 69, May 2020, doi: 10.1186/s13007-020-00613-5.
[58] E. Lielāmurs *et al.*, "A distributed time-of-flight sensor system for autonomous vehicles: Architecture, sensor fusion, and spiking neural network perception," *Electronics*, vol. 14, no. 7, Art. no. 1375, Mar. 2025, doi: 10.3390/electronics14071375.
[59] B. Gonzalez, "Automated food weight and content estimation using computer vision and AI algorithms," *Sensors*, vol. 24, Art. no. 23, 2024, doi: 10.3390/s24237660.
[60] B. Höke, "Determining and evaluating new store locations using remote sensing and machine learning," *Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 29, Art. no. 3, 2021, doi: 10.3906/elk-2005-202.
[61] F. Jafari, "Vision based stockpile inventory measurement using uncrewed aerial systems," *Ain Shams Engineering Journal*, vol. 16, Art. no. 2, 2025, doi: 10.1016/j.asej.2024.103251.
[62] A. Karami, F. Menna, and F. Remondino, "Investigating 3D reconstruction of non-collaborative surfaces through photogrammetry and photometric stereo," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. XLIII-B2-2021, pp. 519–526, Jun. 2021, doi: 10.5194/isprs-archives-xliii-b2-2021-519-2021.
[63] W. Cui *et al.*, "Geometrical quality inspection in 3D concrete printing using AI-assisted computer vision," *Mater. Struct.*, vol. 58, no. 2, Art. no. 37, Feb. 2025, doi: 10.1617/s11527-025-02594-0.
[64] Z. Cai, J. Yang, T. Wang, H. Huang, and Y. Guo, "3D reconstruction of buildings based on 3D Gaussian splatting," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. XLVIII-4/W10-2024, pp. 37–43, May 2024, doi: 10.5194/isprs-archives-xlviii-4-w10-2024-37-2024.
[65] F. Wolling, G. Pal, L. Larcher, and F. Michahelles, "Smart trash can: Image classification and segmentation for waste quality management in the home," in *Proc. 14th Int. Conf. Internet Things (IoT '24)*, Oulu, Finland, 2024, pp. 65–72, doi: 10.1145/3703790.3703798.
[66] C. N. S. Silva, "A scalable open-source framework for machine learning-based image collection, annotation and classification: A case study for automatic fish species identification," *Sustainability Switzerland* , vol. 14, Art. no. 21, 2022, doi: 10.3390/su142114324.
[67] F. Masykur, K. Adi, and O. D. Nurhayati, "Measuring agricultural area using YOLO object detection and ArUco markers," *Ing. Syst. Inf.*, vol. 29, no. 1, pp. 95–104, Feb. 2024, doi: 10.18280/isi.290111.
[68] E. Adán and A. Adán, "Computer vision for glass waste: Technologies and sensors," *Sensors*, vol. 25, no. 21, Art. no. 6634, Oct. 2025, doi: 10.3390/s25216634.
[69] R. Sapkota, D. Ahmed, and M. Karkee, "Immature green apple detection and sizing in commercial orchards using YOLOv8 and shape fitting techniques," *IEEE Access*, vol. 12, pp. 43436–43450, 2024, doi: 10.1109/ACCESS.2024.3378261.
[70] H.-S. Choi, J.-K. Park, S. H. Park, and T. Lee, "NeRF-based 3D reconstruction pipeline for acquisition and analysis of tomato crop morphology," *Front. Plant Sci.*, vol. 15, Art. no. 1439086, Oct. 2024, doi: 10.3389/fpls.2024.1439086.
[71] G. Bortolotti, G. Pergher, and R. Ziosi, "A computer vision system for apple fruit sizing by means of low-cost depth camera and neural network application," *Precis. Agric.*, vol. 25, no. 3, pp. 1146–1175, Jun. 2024, doi: 10.1007/s11119-024-10139-8.
[72] S. Sridharan, O. Ulutan, S. N. T. Priyo, S. Rallapalli, and M. Srivatsa, "Object localization and size estimation from RGB-D images," in *Proc. 2018 IEEE 4th Int. Conf. Multimedia Big Data (BigMM)*, Xi'an, China, 2018, pp. 1–5, doi: 10.1109/BigMM.2018.8499092.
[73] C. Dumery *et al.*, "Counting stacked objects," *arXiv*, Nov. 2024, doi: 10.48550/arXiv.2411.19149.
[74] O. Mirbod, M. Karkee, and Q. Zhang, "On-tree apple fruit size estimation using stereo vision with deep learning-based occlusion handling," *Biosyst. Eng.*, vol. 226, pp. 27–42, Feb. 2023, doi: 10.1016/j.biosystemseng.2022.12.008.
[75] H. He, X. Xu, T. Chen, and P. Lu, "Volume measurement of sand carrier using UAV-based mapping," *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. V-3-2020, pp. 19–26, Aug. 2020, doi: 10.5194/isprs-annals-v-3-2020-19-2020.
[76] Y. Lin, P. Wang, Z. Wang, S. Ali, and L. Mihaylova, "Towards automated remote sizing in hot steel manufacturing with image registration and fusion," *J. Intell. Manuf.*, vol. 35, no. 7, pp. 3173–3189, Oct. 2024, doi: 10.1007/s10845-023-02251-9.
[77] D. Ahmed, R. Sapkota, M. Churuvija, and M. Karkee, "Machine vision-based crop-load estimation using YOLOv8," *IEEE Access*, vol. 11, pp. 110904–110915, 2023, doi: 10.1109/ACCESS.2023.3322971.
[78] H. Liu *et al.*, "ProtoCar: Learning 3D vehicle prototypes from single-view and unconstrained driving scene images," in *Proc. AAAI Conf. Artif. Intell.*, vol. 39, no. 5, 2025, Art. no. 32581, doi: 10.1609/aaai.v39i5.32581.
[79] S. Martínez-Pellitero, E. Cuesta, S. Giganto, and J. Barreiro, "New procedure for qualification of structured light 3D scanners using an optical feature-based gauge," *Opt. Lasers Eng.*, vol. 110, pp. 193–206, Nov. 2018, doi: 10.1016/j.optlaseng.2018.06.002.
[80] M. F. Vasiloglou *et al.*, "Assessing mediterranean diet adherence with the smartphone: The medipiatto project," *Nutrients*, vol. 12, no. 12, Art. no. 3763, Dec. 2020, doi: 10.3390/nu12123763.
[81] H. E. Pang and M. Biljecki, "3D building reconstruction from single street view images using deep learning," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 112, Art. no. 102859, Aug. 2022, doi: 10.1016/j.jag.2022.102859.
[82] S. E. Ghazouali, "Traceable metrology chain for in-line 3D inspection of complex mechanical parts using industrial robot," Ph.D. dissertation, Dept. Mech. Eng., Univ. Paris-Saclay, Gif-sur-Yvette, France, 2022. [Online]. Available: https://theses.hal.science/tel-03994528.
[83] F. Lombardo, M. Crescentini, and M. Tartagni, "Machine learning and data augmentation methods for multispectral capacitance images of nanoparticles with nanoelectrodes array biosensors," *Eng. Appl. Artif. Intell.*, vol. 127, Art. no. 107246, Jan. 2024, doi: 10.1016/j.engappai.2023.107246.
[84] V. V. E. Nogueira, L. F. Barca, and T. C. Pimenta, "A cost-effective method for automatically measuring mechanical parts using monocular machine vision," *Sensors*, vol. 23, no. 13, Art. no. 5994, Jun. 2023, doi: 10.3390/s23135994.
[85] S. Meng, S. Zhang, X. Liang, and J. Hu, "Automatic extraction of scale information for interactive measurement of anything in microscopy images," *SSRN Electron. J.*, Jan. 2025, doi: 10.2139/ssrn.5137009.
[86] S. Jiang, J. Hua, and Z. Han, "Query quantized neural SLAM," *arXiv*, Dec. 2024, doi: 10.48550/arXiv.2412.16476.
[87] D. Wang, Y. Dong, Y. Pan, and R. Ma, "Machine vision-based monitoring methodology for the fatigue cracks in U-rib-to-deck weld seams," *IEEE Access*, vol. 8, pp. 94204–94218, 2020, doi: 10.1109/ACCESS.2020.2995276.
[88] M. Wang, Q. Sun, C. Gao, Z. Ren, and W. Dai, "A three-dimensional vision measurement method based on double-line combined structured light," *Sci. Rep.*, vol. 13, no. 1, Art. no. 18451, Oct. 2023, doi: 10.1038/s41598-023-46176-y.
[89] H. S. Kim, "3D measurement using a single image for smart manufacturing of microscopic products in a ceramic powder pressing process," *Int. J. Precis. Eng. Manuf.-Green Technol.*, vol. 10, no. 1, pp. 117–128, Jan. 2023, doi: 10.1007/s40684-022-00434-y.
[90] F. S. Konstantakopoulos, P. C. Grammatikopoulos, S. A. Chatzichristofis, and Y. S. Boutalis, "An automated image-based dietary assessment system for Mediterranean foods," *IEEE Open J. Eng. Med. Biol.*, vol. 4, pp. 45–51, 2023, doi: 10.1109/OJEMB.2023.3266135.
[91] H. Wang, J. Geng, J. He, and H. Liao, "RECISTSup: Weakly-supervised lesion volume segmentation using RECIST measurement," *IEEE Trans. Med. Imag.*, vol. 41, no. 7, pp. 1655–1667, Jul. 2022, doi: 10.1109/TMI.2022.3149168.
[92] F. Suo, "Fish keypoints detection for ecology monitoring based on underwater visual intelligence," in *Proc. 16th Int. Conf. Control, Automat., Robot. Vision (ICARCV)*, Shenzhen, China, 2020, pp. 542–547, doi: 10.1109/ICARCV50220.2020.9305424.
[93] B. Ilo *et al.*, "Real-time rice milling morphology detection using hybrid framework of YOLOv8 instance segmentation and oriented bounding boxes," *Electronics*, vol. 14, no. 18, Art. no. 3691, 2025, doi: 10.3390/electronics14183691.
[94] J.-S. Lee, S.-H. Lee, and J.-H. Park, "Automation of trimming die design inspection by zigzag process between AI and CAD domains," *Eng. Appl. Artif. Intell.*, vol. 127, Art. no. 107283, Jan. 2024, doi: 10.1016/j.engappai.2023.107283.
[95] Y.-K. Lee, S. K. Panda, W. Wang, and M. K. Jawed, "Measure anything: Real-time, multi-stage vision-based dimensional measurement using Segment Anything," *arXiv*, Dec. 2024, doi: 10.48550/arXiv.2412.03472.

Applied Information System and Management (AISM)
Volume 9, (1) 2026, p. 147–158
P-ISSN: 2621-2536; E-ISSN: 2621-2544; DOI: 10.15408/aism.v9i1.50286

Table 4
Performance reports of models frequently used in research

Reasearch	Domain	Architecture	MAE	RMSE	MAPE	Accuracy	Precicion
[64]	Fisheries	Transfer learning, EfficientNet-Lite, MobileNetV2, ResNet50, EfficientDet-Lite	-	-	-	0.91	0.48
[57]	Food	YOLOv8, Mask R-CNN and FCN ,	-	15.230 - 29.803	-	0.93	mAP=0.873
[33]	Agriculture	DCNN	2.380 mm (width) and 2.580 for length	2.957 (width) and 3.412 (length)	4.114% (width) and 3.636% (length)	90.91%	-
[35]	Construction	Mask-RCNN, RestNet-50	-	-	-	0.51 - 0.81 for object detection, 0.90 - 0.91 for pixelwise segmentation	0.43-0.83
[58]	Urban Planning	CNN, R-CNN, Mask R-CNN	-	-	14%-1500%	16.4% (area) and 12.7% (height)	False Positive: 6.0% - 6.3% False Negative: 1.5% - 4.6%
[42]	Autonomous vehicle	ResNet-101	0.106-0.118	5.082-7.122	-	-	-
[36]	Aquaculture	RTMDet-m, ResNet-50, DeepLabV3, MobileNetV2	10.23 mm - 50.67 (shrimp)	12.97 mm-63.34 mm	-	0.631-0.960	96%
[51]	Food and Health	CNN	9%	-	7%-10%	98% and 92% (volume)	-
[69]	Agriculture	YOLO	2.8 mm - 7.2 mm	16% - 22%	-	94%	0.87
[51]	Food and Health	CNN	9%	-	7%-10%	-	-
[69]	Agriculture	YOLO	6% - 13%	16% - 22%	6% - 15%	94%	0.87
[87]	Manufacture	ResNet50, ResNet101, ResNext50	0.06 - -.208	0.156-2.697	-	0.953 - 0.957	-
[79]	Reconstruction	Pixel2Mesh, SphereInit, VoxMesh, Graph Convolution Networks,	Volume estimation: -6.142% - (-36,167%)	-	-	0.34911	0.9387
[89]	Medical	ResNet-18	-	-	-	0.61 ± 0.04 to 0.83 ± 0.03	-
[88]	Food and Health	EfficientNet-B2	-	-	5.1%-21.3%	83.8%-97.6%	-
[81]	Nanoparticles	MobileNet-V3 Small, Tiny Mobile Net, XGBoost	-	-	6%- 33%	-	-
[67]	Agriculture	YOLO	1.66 mm- 7.8 mm	2.35 mm-9.65 mm	6.15 mm-29.48 mm	0.9-0.94	-
[59]	Logistic	1DCNN network	< 3.5 %	-	<3 % - 9 %	higher than 98 %	consistency and repeatability
[47]	Manufacture	ResNet-18	-	-	-	95.23%	92.62%
[72]	Agriculture	Faster ReCNN, Mask ReCNN	1.1 mm-4.2 mm	-	< 10%	-	-
[90]	Fishery	Faster R-CNN	-	-	0.99%-12.19%	0.715-0.905	-
[34]	Agriculture	Mask R-CNN, ORCNN, ResNeXt-101	3.2 mm-41.5 mm	4.1 mm-88.5 mm	-	70.2%- 97.9%	90.1%- 98.4%
[34]	Agriculture	Mask R-CNN, ResNeXt-101	6.4 mm- 41.5 mm	4.1 mm- 4.1 mm	-	70.2%- 97.9%	93.0% - 98.4%
[8]	Horticulture	CNN, Mask R-CNN	2.05 mm- 5.23 mm	2.80 mm-6.27 mm	3.79% - 9.55%	0.86	0.90-0.94
[91]	Agriculture	YOLOv8	0.10 mm-0.35 mm	-	-	0.86- 0.995	0.984- 0.987
[46]	Fishery	DGCNN	0.014-0.817	0.022- 1.25	15.59%-28.28%	-	-
[54]	Robotic	YOLO, GAN, Pixel2Graph	0.33 m- 0.89 m	0.45 m- 1.99 m	3.72%- 4.85%	90.82%	90%
[50]	Robotic	Faster RCNN, MonoLite3D Network	-	-	-	82.27 %	-
[39]	Food	GAN	40.05 kCal - 76.74 kCal	-	22.0%- 33.0%	-	-
[92]	Manufacture	CNN, Fast R-CNN	-	-	1.3%- 3.5%	99.5%- 100%	-
[43]	Metrology	Metric3D v2 models, Convnets,	0.024- 0.127	0.015- 3.258	-	0.925 - 0.998	-

[38]	Food and health	GAN	-	-	-	-	-
[3]	Many domains	Deep Learning Network, Adam optimizer	-	-	0.95%- 3.40%	97.63%	-
[41]	Many domains	CNN, ResNet-101	9.5%- 30.26%	-	9.5%- 30.26%	50.93%- 99.86%	-
[93]	Many domains	Mask R-CNN	-	-	0-10%	83.5%	-
[52]	Manufature	CAD-ClassNet	-	-	0.05%- 20.37%	81.85%-99.70%	-
[48]	Many domains	CNN	0.03 pixels- 0.08%	-	-	-	-
[55]	Agriculture	ANN regression model, RF regression model	8.56 g/m²- 28.94 g/m²	9.13 g- 730.90 g	-	0.90- 1.0	0.89 - 0.90
[61]	Printing	YOLO	-	-	-	0.98894 - 0.99496	0.995- 0.99496
[68]	Agriculture	Neural Radiance Fields	-	-	0.089- 0.135	0.96	0.953
[40]	Agriculture	PVSegNet	-	0.0311 cm - 0.5988	-	0.9182- 0.9489	96.38%
[7]	Construction	YOLO	-	-	22% - 41%	-	-
[44]	Ocean observation	YOLO	-	3.6 mm- 5.1 mm	-	81%- 97%	93% - 94.8%
[76]	Robotic	ProtoCar Network	-	-	-	-	-
[45]	Construction	YOLO	-	2-3 cm	-	92.03%- 90.89%	-
[56]	Autonomus vehicle	Spiking Neural Network, CNN, VGG-11 dan VGG-13	-	-	-	14%- 84.10%	52%- 65%
[70]	*Augmented reality*	CNN	-	-	-	-	-
[85]	Construction	VGGNet-13	-	0.003 mm- 3.0195 mm	-	7.91%- 14.57%	0.932- 0.978
[37]	Robotic	CNN3D, PointNet++, SegNet	-	-	-	0.72 AUC- 0.89 AUC	-
[75]	Agriculture	YOLOv8	2.99 unit	2.08 mm- 3.95 unit	-	0.5 - 0.95	0.739 - 0.888