

Article

Not peer-reviewed version

Comparative Evaluation of YOLOv8 and YOLOv11 Models for Efficient Phenotypic Segmentation of Edible Mushrooms

Doo-Ho Choi , [Youn-Lee Oh](#) , [Minji Oh](#) , [Eun-Ji Lee](#) , Sung-I Woo , Minseek Kim , [Ji-Hoon Im](#) *

Posted Date: 3 February 2026

doi: 10.20944/preprints202602.0065.v1

Keywords: YOLOv8; YOLOv11; phenotypic analysis; *Pleurotus ostreatus*; *Flammulina velutipes*



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

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

Article

Comparative Evaluation of YOLOv8 and YOLOv11 Models for Efficient Phenotypic Segmentation of Edible Mushrooms

Doo-Ho Choi, Youn-Lee Oh, Minji Oh, Eun-Ji Lee, Sung-I Woo, Minseek Kim and Ji-Hoon Im *

Mushroom Research Division, National Institute of Horticultural and Herbal Science, RDA, Eumseong, Chungbuk 27709, Republic of Korea

* Correspondence: gamchoduh@korea.kr; jihooni24@korea.kr

Abstract

Digital phenotyping is increasingly recognized as a critical tool for quantitative analysis of fungal morphology, particularly in controlled indoor cultivation systems where large numbers of fruiting bodies must be assessed consistently and non-destructively. While YOLOv8-based deep learning approaches have been previously applied to phenotypic analysis of edible mushrooms, the applicability of newer YOLO architectures to fungal phenotyping remains unexplored. In this study, we present a controlled-environment digital phenotyping framework for indoor mushroom cultivation and evaluate the feasibility of YOLOv11 for phenotypic analysis through a direct comparison with YOLOv8. Using bottle-cultivated *Pleurotus ostreatus* and *Flammulina velutipes* as representative edible basidiomycetes, we conducted a systematic comparison of YOLOv8-seg and YOLOv11-seg under identical datasets, preprocessing pipelines, and hyperparameter configurations. The results demonstrate that YOLOv11 achieves segmentation performance comparable to YOLOv8 across all evaluated metrics ($\Delta mAP_{50-95} < 0.01$), while substantially reducing computational complexity, including fewer trainable parameters, lower FLOPs, and decreased gradient load. Validation against caliper-based physical measurements revealed moderate, trait-dependent agreement, whereas inter-model consistency between YOLOv8 and YOLOv11 was consistently high across diverse morphological and segmentation scenarios. These findings indicate that architectural refinements in YOLOv11 primarily enhance computational efficiency and training behavior without altering phenotypic interpretation. Collectively, this study provides the first validation of YOLOv11 for mushroom phenotyping and highlights its potential as an efficient analytical tool for future high-throughput fungal morphology studies.

Keywords: YOLOv8; YOLOv11; phenotypic analysis; *Pleurotus ostreatus*; *Flammulina velutipes*

1. Introduction

Since ancient times, mushrooms have been valued for their distinctive flavor, economic importance, and medicinal properties [1,2]. To date, more than 15,000 mushroom species have been described, of which approximately 2,000 are considered edible [3]. Owing to their saprophytic lifestyle, mushrooms can grow on diverse substrates, including soil as well as food and agricultural residues. Taxonomically, mushrooms are sporulating fungi belonging to the phylum Basidiomycota and the class Agaricomycetes. Structurally, the fruiting body is composed of a stipe (stem), pileus (cap), and lamellae (gills), and variation in these external morphological components is directly linked to commercial quality, developmental stage, and cultivation performance [4].

Phenotype refers to the observable characteristics arising from interactions between genetic background and environmental conditions [5]. In cultivated mushrooms, phenotypic traits such as fruiting body size, shape, and proportional morphology are essential indicators for assessing market quality, breeding outcomes, and cultivation stability. Throughout the history of mushroom

domestication, selective breeding has been used to enhance desirable phenotypic traits for edible purposes [6]. However, the intensification and automation of modern mushroom cultivation systems have increased production efficiency while simultaneously challenging conventional phenotypic evaluation. Traditional phenotyping approaches remain labor-intensive, time-consuming, and prone to operator-dependent variability, particularly under high-throughput indoor cultivation conditions [7,8]. As a result, there is a growing demand for efficient, objective, and automated phenotyping strategies. In parallel with advances in digital agriculture, artificial intelligence (AI)-based image analysis has emerged as a promising tool for quantitative phenotyping, with increasing relevance for fungal morphology assessment [9,10].

Instance segmentation represents a key methodological advance by integrating object detection and semantic segmentation to enable pixel-level delineation of biological structures within convolutional neural network (CNN)-based frameworks [11,12]. In plant phenotyping, instance segmentation has been successfully applied to tasks such as disease assessment and yield-related trait extraction [13–15]. In fungal research, computer vision approaches based on the You Only Look Once (YOLO) family have been applied to toxic mushroom classification, fruiting body detection, and phenotypic analysis in cultivated systems [16–18]. Owing to its favorable balance between inference speed and accuracy, YOLO has become a widely adopted architecture for real-time object detection. Since its initial introduction in 2016 [19], the YOLO framework has undergone continuous architectural refinement. Among the Ultralytics YOLO series, only selected versions, including YOLOv5, YOLOv8, and the recently released YOLOv11, support instance segmentation, thereby enabling pixel-level phenotypic analysis of complex biological structures [16].

Beyond accuracy-oriented performance, recent studies in digital phenotyping have emphasized the importance of system-level requirements in AI model selection, particularly under large-scale, automated, and resource-constrained analysis environments. In indoor mushroom cultivation, phenotypic monitoring often requires continuous image acquisition and processing under constrained computational resources [20]. In such contexts, computational efficiency, architectural compactness, and training stability are critical factors that directly affect the scalability and sustainability of phenotyping pipelines. Consequently, the evaluation of phenotyping models must account for architectural suitability and long-term deploy ability, especially for high-throughput fungal morphology analysis. Recent studies in automated cultivation and digital phenotyping further indicate that lightweight deep learning models play a critical role in enabling scalable and context-aware farming systems, especially where real-time perception and continuous operation are required [21].

Although a growing number of studies have explored AI-based phenotypic analysis, systematic investigations focusing on instance segmentation-based phenotyping of cultivated mushrooms remain limited. Existing studies have predominantly relied on YOLOv8 for the detection and phenotypic characterization of mushroom fruiting bodies [9,17,22]. In contrast, despite recent advances in model architecture and computational efficiency, applications of YOLOv11 to mushroom phenotypic analysis have rarely been reported [23]. To address this gap, we establish a controlled-environment digital phenotyping benchmark using cultivated mushrooms grown in standardized polypropylene bottle systems. Within this framework, YOLOv8 and YOLOv11 are independently applied under identical training conditions to evaluate segmentation performance, agreement with physical ground-truth measurements, and computational efficiency. Through this systematic comparison, we aim to assess whether YOLOv11 can maintain phenotypic measurement fidelity comparable to YOLOv8 while offering improved computational efficiency, thereby supporting its practical applicability in scalable indoor mushroom phenotyping pipelines.

2. Materials and Methods

2.1. Preparation of Mushroom Material and Image Acquisition

For phenotypic analysis under controlled-environment conditions, *Pleurotus ostreatus* and *Flammulina velutipes* were cultivated in 2024 based on established bottle cultivation protocols reported in previous studies [6,7,24]. Both edible mushroom species were grown in 800 mL polypropylene bottles under standardized and controlled conditions. For the cultivation of *P. ostreatus*, a substrate composed of poplar sawdust, beet pulp, and cottonseed meal was prepared in a 50:30:20 (v/v) ratio. In the case of *F. velutipes*, the cultivation medium consisted of corn cob, rice bran, beet pulp, soybean hull, wheat bran, crushed oyster shell, and waste limestone mixed in a 35:33:10:6:6:6:4 (v/v) ratio. All media were sterilized at 121 °C for 90 min prior to inoculation. Fruiting body development was conducted at 23 °C with a relative humidity of 80–90%, following conventional controlled bottle cultivation practices. All images used in this study were acquired from mature fruiting bodies at the commercial harvest stage to ensure consistency in morphological traits and to focus the analysis on phenotypic features directly relevant to market-oriented evaluation.

Side-view images of mushroom fruiting bodies were acquired under strictly controlled illumination and spatial conditions to construct a standardized dataset suitable for deep learning-based digital phenotyping. To minimize background noise, shadow interference, and illumination heterogeneity, a custom-built imaging chamber (PODO Co., Ltd., Korea) was constructed in 2024. The chamber was designed to provide reproducible imaging conditions across all samples. Within the chamber, a high-resolution RGB camera (EOS M50; Canon Inc., Tokyo, Japan) equipped with a Fujinon CF8ZA-1S industrial lens (8 mm focal length, f/1.8) was mounted at a fixed position. Illumination was provided by dual diffused LED panels symmetrically installed at 45° angles relative to the sample plane (Figure 1). During image acquisition, mushroom bottles were placed sequentially at four predefined positions on the chamber floor to account for minor spatial variability while maintaining consistent imaging geometry. These controlled imaging conditions were intentionally adopted to minimize environmental variability and to ensure a fair and reproducible comparison between deep learning models, rather than to replicate all complexities of commercial production environments.

All acquired images were stored in RGB color mode using lossless compression (Figure 2). Image preprocessing was performed using Python-based routines implemented with the OpenCV library (version 4.x). The preprocessing pipeline included background normalization to reduce residual illumination heterogeneity, global color correction based on channel-wise intensity normalization, and resizing of all images to 640 × 640 pixels using bilinear interpolation to comply with YOLO input requirements [25]. No additional data augmentation or manual color manipulation was applied during preprocessing. The identical preprocessing workflow was consistently applied to the training, validation, and test datasets to ensure reproducibility and fair comparison between deep learning models. Image resizing was performed solely to meet standardized model input requirements and was consistently applied across all datasets, ensuring that relative performance comparisons between models were not biased by preprocessing differences.

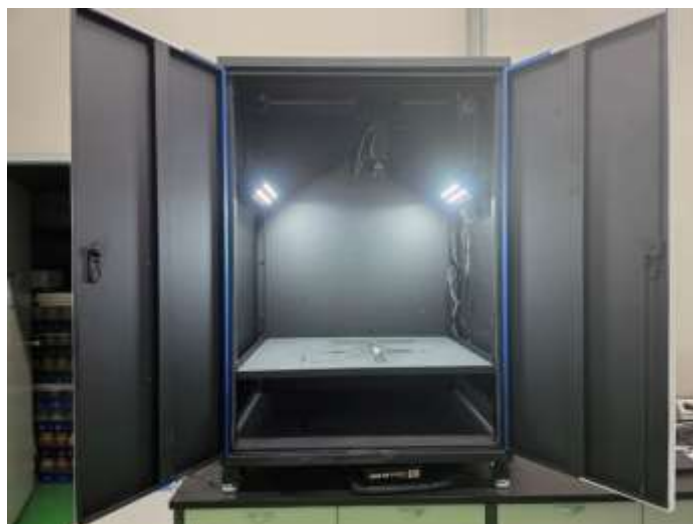


Figure 1. Custom-built imaging chamber for standardized acquisition of mushroom phenotypic data. The chamber was designed to minimize background noise, shadow artifacts, and illumination variability during image capture. Captured images were saved in RGB mode with lossless compression, followed by preprocessing steps including background normalization, color correction, and resizing to 640×640 pixels for YOLO-based phenotypic analysis.

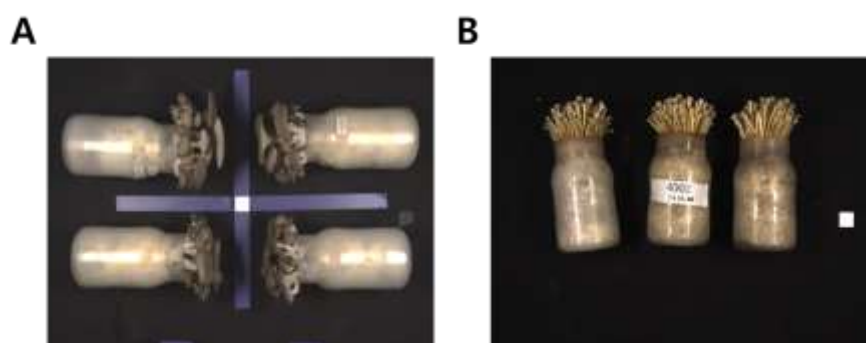


Figure 2. Raw images of *P. ostreatus* (A) and *F. velutipes* (B). Under controlled conditions in the imaging chamber, images illustrate variability in growth stage, individual size, and clustering patterns included in the dataset.

2.2. Dataset Construction

Following image acquisition, all valid images were organized and preprocessed to construct a reliable dataset for AI model training and evaluation under controlled-environment conditions. The dataset consisted exclusively of mature mushroom fruiting bodies cultivated in bottle-based systems, as mature stages exhibit clearly distinguishable pileus and stipe structures that are essential for instance segmentation-based phenotypic analysis. Although the dataset focused on mature fruiting bodies, substantial morphological variability was preserved, including variations in individual size, stipe length, pileus shape, and degrees of fruiting body overlap, reflecting realistic heterogeneity within controlled indoor cultivation systems.

Image annotation was performed on mushroom images (403 side-view images of *P. ostreatus* and 201 images of *F. velutipes*) using polygon-based instance segmentation with Label Studio (v1.13; Heartex, San Francisco, USA) in 2025. Two phenotypic categories—pileus and stipe—were manually delineated for each fruiting body (Figure 3). To ensure annotation reliability, all labels were cross-validated by two independent annotators, and discrepancies were resolved through consensus-based review to maintain inter-annotator consistency. The unequal number of images between species

reflects inherent differences in cultivation yield and image availability and was consistently preserved across all subsets to avoid introducing artificial sampling bias.

The annotated dataset was partitioned into training (60%), validation (20%), and test (20%) subsets using stratified random sampling based on predefined morphological categories to ensure balanced representation across subsets [26]. All annotation management and dataset processing steps were conducted using Visual Studio Code (VS Code; Microsoft Corp., Redmond, WA, USA) and integrated with the Ultralytics YOLO framework, ensuring seamless compatibility between the annotation schema and the model input pipeline [16,27]. The final dataset conformed to the COCO-format instance segmentation structure, enabling reproducible and fair comparative analysis across different YOLO architectures. Stratification was performed based on coarse morphological attributes, including relative fruiting body size and the presence of overlapping individuals, rather than species identity, to maintain phenotypic diversity across subsets.

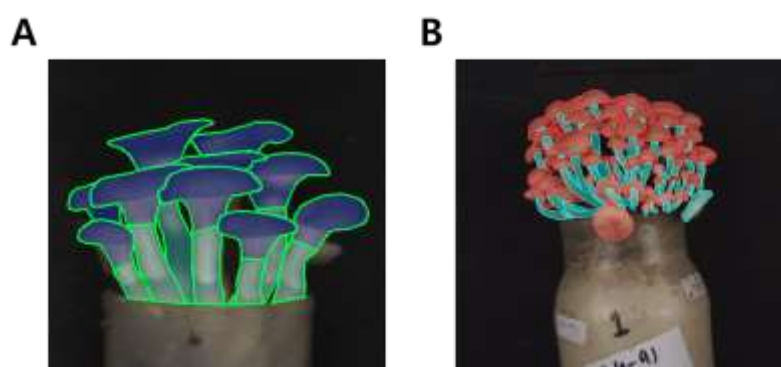


Figure 3. Images of polygon-based annotation used for dataset construction in *P. ostreatus* (A) and *F. velutipes* (B). Pileus and stipe regions were manually annotated using label studio to support instance segmentation tasks.

2.3. Model Configuration and Training Procedure

Comparative analyses of analytical performance across different YOLO architectures have been reported in various agricultural and phenotyping studies [16,28,29]. In this study, two YOLO-based instance segmentation models, YOLOv8-seg and YOLOv11-seg, were employed for a systematic comparison of mushroom phenotypic segmentation performance. Both models were implemented using the Ultralytics YOLO framework (v8.2.0) within Python 3.12 and PyTorch 2.3.1 environments. All computations were conducted on a Windows 11 workstation equipped with an NVIDIA RTX 4080 GPU (16 GB VRAM) and 32 GB RAM, providing sufficient computational capacity for high-resolution instance segmentation.

To ensure controlled and reproducible comparison between the two architectures, both YOLOv8-seg and YOLOv11-seg were trained under identical hyperparameter settings, following configuration guidelines reported in previous benchmarking studies [8]. Prior to training, all input images were resized to 640×640 pixels. Batch size was dynamically adjusted based on available GPU memory. The learning rate was set to 0.01, momentum to 0.937, and weight decay to 0.0005. Stochastic gradient descent (SGD) was used as the optimizer, and training was performed for 120 epochs, including three warm-up epochs to stabilize gradient updates.

The adoption of identical hyperparameters was intended to isolate the effects of architectural differences between YOLOv8 and YOLOv11, rather than to maximize the absolute performance of each model through architecture-specific tuning. Under this controlled setup, observed differences in segmentation accuracy and computational efficiency can be attributed primarily to variations in network design and information flow.

YOLOv8 employs a conventional C2f backbone combined with a spatial pyramid pooling-fast (SPPF) neck, which supports hierarchical multi-scale feature extraction [30]. In contrast, YOLOv11 introduces C2f-Fusion and RepNCSPELAN modules, which enhance feature reuse and contextual

representation while reducing redundant computation [31]. These architectural modifications primarily alter feature aggregation and propagation mechanisms. To maintain fairness in performance evaluation, both YOLOv8 and YOLOv11 were trained using the same datasets, preprocessing pipeline, and hyperparameter configurations, enabling direct comparison of their suitability for instance segmentation-based mushroom phenotypic analysis [16].

2.4. Evaluation Metrics

2.4.1. Detection and Segmentation Accuracy

Following the training procedure, quantitative evaluation of the YOLO models was performed using four standard COCO-based metrics: Precision (P), Recall (R), mAP₅₀, and mAP₅₀₋₉₅. Precision measures the proportion of correctly identified positive detections among all detections, while Recall represents the proportion of correctly detected targets among all ground-truth instances. The mean Average Precision (mAP) was computed using the COCO evaluation protocol, where mAP₅₀ corresponds to an intersection-over-union (IoU) threshold of 0.50 and mAP₅₀₋₉₅ represents the mean value across thresholds from 0.50 to 0.95 at 0.05 increments [32].

$$P = \frac{TP}{TP+FP} \quad R = \frac{TP}{TP+FN} \quad mAP = \frac{1}{n} \sum_{i=0}^n AP_i \quad AP_i = \int_0^1 p(r) dr$$

Where TP, FP, and FN denote true positives, false positives, and false negatives, respectively, and AP_i is the area under the precision–recall curve for class i . For mask-based segmentation, these metrics were computed on a pixel-wise IoU basis between predicted and ground-truth masks.

2.4.2. Computational Efficiency and Model Complexity

To evaluate a computational efficiency and scalability of the YOLO models, the following model descriptors were quantified using the THOP (PyTorch) library:

- FLOPs (B) : total floating-point operations per forward pass, reflecting computational demand;
- Params (M) : number of trainable parameters, representing model size and memory usage;
- Gradients (G) : number of gradient tensors updated during backpropagation, indicating optimization cost;
- Layers (L) : total number of computational blocks.

These metrics were used to quantitatively assess segmentation accuracy and computational efficiency of the YOLO models. Evaluation was designed to verify whether the architectural refinement of YOLOv11 could achieve a lower computational burden and parameter count compared with YOLOv8 while maintaining comparable accuracy across segmentation metrics [33].

2.4.3. Validation Against Physical Measurements

To verify the biological validity of YOLO-derived measurements, the predicted dimensional traits (pileus diameter, pileus thickness, stipe length, and stipe thickness) were compared against physically measured ground-truth data obtained using a digital caliper. All pixel-based outputs were converted into metric units (mm) using a calibration target captured under identical imaging conditions. Based on previous studies, Agreement between predicted and observed measurements was quantified using three statistical indices [34,35]:

1. Pearson's correlation coefficient (r) to evaluate linear association;
2. Coefficient of determination (R^2) from simple linear regression to assess explanatory power;
3. Mean Absolute Error (MAE) to describe the mean magnitude of deviation between predicted and actual values.

These indices were computed according to the following formulations:

$$r = \frac{\sum_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2} \sqrt{\sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})^2}}, \quad R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

Where y_i and \hat{y}_i represent the observed and predicted measurements, respectively. All statistical analyses were performed using Python 3.12, NumPy 1.26, and SciPy 1.11, and graphical summaries were visualized using Matplotlib 3.8 and Pandas 2.2.

2.4.4. Statistical Analysis and Visualization

For all performance values, including both accuracy and computational indicators, were analyzed separately for *P. ostreatus* and *F. velutipes*. Mean and standard deviation were calculated for each metric, and independent-sample t-tests were conducted at a 95% confidence level ($p < 0.05$) to evaluate the statistical significance of differences between YOLOv8 and YOLOv11. Correlation analyses were further performed between computational complexity (FLOPs, Params, Gradients) and accuracy (mAP₅₀₋₉₅) to determine the balance between efficiency and precision. Visualization of FLOPs, Params, and Gradient values was expressed as bar graphs, while the learning loss and mAP convergence trends were depicted using epoch-wise line charts to assess stability during training.

3. Results

3.1. Comparative Performance of YOLOv8 and YOLOv11

Both YOLOv8-seg and YOLOv11-seg successfully detected and segmented pileus and stipe regions across all test images of *Pleurotus ostreatus* and *Flammulina velutipes* under controlled imaging conditions (Table 1). Quantitative evaluation indicated that YOLOv11 achieved segmentation performance comparable to YOLOv8 across all accuracy metrics, with computational efficiency examined separately in subsequent sections.

For *P. ostreatus*, YOLOv8 achieved a precision (P) of 0.80 and recall (R) of 0.79 for bounding-box detection, which served as a preliminary indicator of object localization performance, with a mean average precision (mAP₅₀) of 0.85 and mAP₅₀₋₉₅ of 0.55. Corresponding mask-based segmentation results yielded P = 0.81, R = 0.75, mAP₅₀ = 0.82, and mAP₅₀₋₉₅ = 0.45. Under identical training conditions, YOLOv11 produced comparable bounding-box detection results (P = 0.82, R = 0.78) and identical mask-based segmentation performance (P = 0.81, R = 0.75, mAP₅₀ = 0.82, mAP₅₀₋₉₅ = 0.45), indicating that segmentation accuracy for *P. ostreatus* was maintained despite architectural differences between the two models.

For *F. velutipes*, both models exhibited relatively lower precision and recall values compared with *P. ostreatus*, which can be attributed to the smaller and more slender morphology of the fruiting bodies, resulting in reduced pixel occupancy and increased sensitivity to partial occlusion in side-view images. YOLOv8 achieved box-level P = 0.73, R = 0.75, mAP₅₀ = 0.77, and mAP₅₀₋₉₅ = 0.46, whereas YOLOv11 achieved P = 0.72, R = 0.76, mAP₅₀ = 0.77, and mAP₅₀₋₉₅ = 0.48. For mask-based segmentation, YOLOv11 showed a slightly higher recall (R = 0.68) than YOLOv8 (R = 0.65), with a marginal increase in mAP₅₀₋₉₅ from 0.23 to 0.24. However, these differences were minor and did not indicate a substantial divergence in segmentation accuracy between the two architectures.

Qualitative inspection of segmentation outputs further demonstrated that both YOLOv8 and YOLOv11 generated consistent mask delineation of pileus and stipe regions across diverse morphological conditions. As shown in Figure 4, both models accurately segmented fruiting bodies with varying sizes and degrees of overlap, producing visually comparable mask boundaries. Although YOLOv11 occasionally produced visually smoother boundary delineation along pileus margins, these differences were qualitative in nature and did not correspond to statistically significant differences in quantitative accuracy metrics.

Overall, the quantitative and qualitative results collectively confirm that YOLOv11 preserves segmentation-based phenotypic measurement accuracy at a level comparable to YOLOv8 under controlled-environment conditions, providing a robust phenotypic performance baseline for subsequent evaluation of computational efficiency and model scalability.

3.2. Computational Efficiency and Model Complexity

As illustrated in Figure 5, YOLOv11 demonstrated a clear reduction in computational requirements compared with YOLOv8. The total number of floating-point operations (FLOPs) decreased from 39.9 G to 32.8 G, corresponding to an approximate reduction of 17.8%. Similarly, the number of trainable parameters was reduced from 11.8 M to 10.01 M (15.2% reduction), and the gradient load decreased from 11.79 M to 10.08 M, suggesting reduced optimization overhead during training. These results indicate that YOLOv11 adopts a more compact network configuration with reduced computational and memory demands. Although the total number of network layers increased slightly (from 85 to 113), this increase reflects architectural restructuring rather than added computational burden, enabled by modular components such as C2f-Fusion and RepNCSPeLAN.

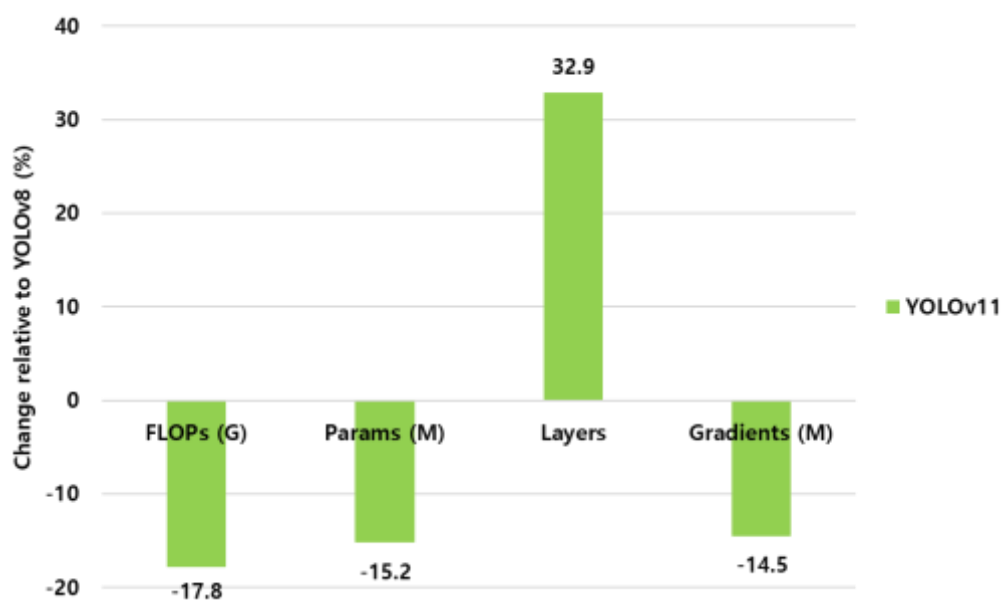


Figure 5. Relative changes in FLOPs, parameters, layers, and gradients of YOLOv11 compared with YOLOv8. YOLOv11 achieves a 17.8% reduction in FLOPs and approximately 15% fewer parameters and gradients, despite an increase in network depth (113 vs. 85 layers).

Training and validation loss curves are presented in Figure 6. Across both *P. ostreatus* and *F. velutipes* datasets, YOLOv11 exhibited smoother and more stable convergence behavior compared with YOLOv8, particularly for segmentation loss. Validation loss curves for YOLOv11 reached stable plateaus approximately 15–20% earlier in terms of training epochs than those of YOLOv8. After approximately 60 training epochs, both box and segmentation losses of YOLOv11 remained consistently 10–18% lower than those observed for YOLOv8. These trends indicate improved training efficiency and convergence stability under identical training conditions, rather than direct gains in segmentation accuracy.

Despite the reduced computational complexity, YOLOv11 maintained segmentation accuracy comparable to that of YOLOv8 ($\Delta mAP_{50-95} < 0.01$), demonstrating that the observed efficiency gains were not accompanied by a loss of representational capability. Collectively, these results suggest that YOLOv11 providing a favorable balance between segmentation performance and computational efficiency for high-throughput phenotyping under controlled-environment conditions.

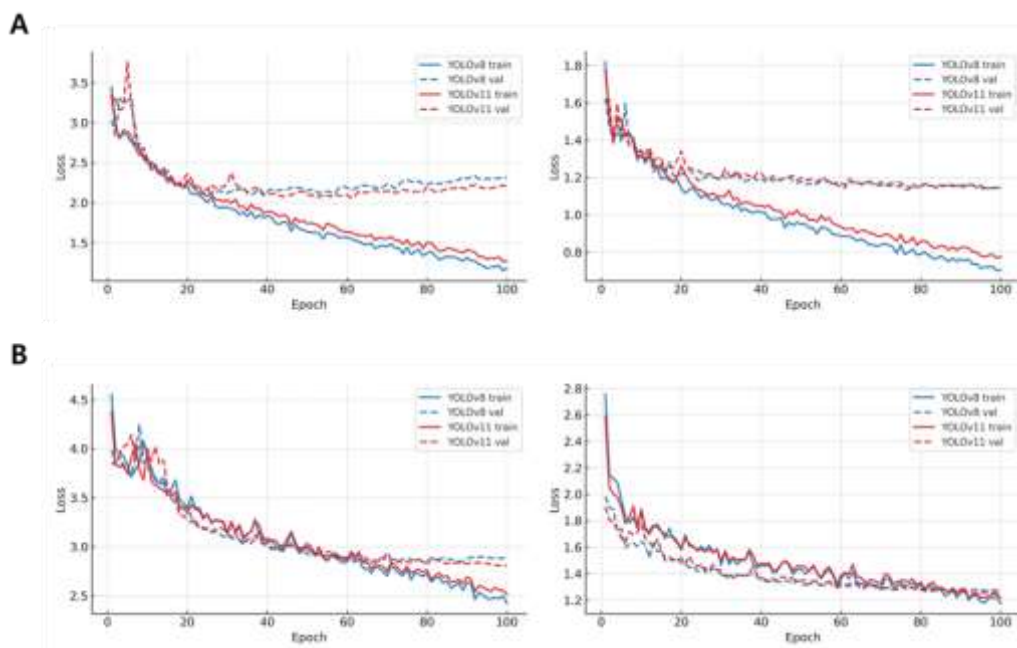


Figure 6. Training and validation loss trends for YOLOv8 and YOLOv11 models in (A) *Pleurotus ostreatus* and (B) *Flammulina velutipes*. The graphs illustrate the epoch-wise decrease in segmentation (left) and bounding-box (right) losses for both species, showing stable convergence patterns. YOLOv11 exhibited slightly faster convergence and lower validation loss, indicating improved learning efficiency and robustness during phenotypic segmentation.

3.3. Validation against Physical Measurements

To assess the biological consistency of YOLO-derived phenotypic measurements, model outputs were compared with digital caliper-based physical measurements for four key traits: pileus diameter, pileus thickness, stipe length, and stipe thickness (Tables 2 and 3). Agreement between predicted and observed values was evaluated using Pearson's correlation coefficient (r), the coefficient of determination (R^2), and mean absolute error (MAE), providing complementary perspectives on association strength, explanatory power, and absolute deviation.

Table 2. Validation of measurements on predicted phenotypic traits by YOLO in *Pleurotus ostreatus*.

		Pearson (r)	R-score (R^2)	MAE (mm)
Pileus (diameter)	YOLOv08/ Physical Measurements	0.20	0.04	5.06
	YOLOv11/ Physical Measurements	0.23	0.05	5.02
	YOLOv08/ YOLOv11	0.95	0.91	1.16
	YOLOv08/ Physical Measurements	0.16	0.03	5.68
Pileus (thickness)	YOLOv11/ Physical Measurements	0.17	0.03	5.54
	YOLOv08/ YOLOv11	0.94	0.89	0.86

Stipe (thickness)	YOLOv08/ Physical Measurements	0.43	0.18	4.60
	YOLOv11/ Physical Measurements	0.39	0.15	4.51
Stipe (length)	YOLOv08/ YOLOv11	0.93	0.86	0.63
	YOLOv08/ Physical Measurements	0.42	0.17	9.03
	YOLOv11/ Physical Measurements	0.43	0.19	8.88
	YOLOv08/ YOLOv11	0.98	0.96	1.44

Table 3. Validation of measurements on predicted phenotypic traits by YOLO in *Flammulina velutipes*.

		Pearson (r)	R-score (R ²)	MAE (mm)
Pileus (diameter)	YOLOv08/ Physical Measurements	0.42	0.18	2.01
	YOLOv11/ Physical Measurements	0.41	0.17	1.99
Pileus (thickness)	YOLOv08/ YOLOv11	0.98	0.96	0.22
	YOLOv08/ Physical Measurements	0.22	0.05	3.23
Stipe (thickness)	YOLOv11/ Physical Measurements	0.19	0.04	3.23
	YOLOv08/ YOLOv11	0.98	0.95	0.18
Stipe (length)	YOLOv08/ Physical Measurements	0.39	0.15	3.71
	YOLOv11/ Physical Measurements	0.39	0.15	3.90
Stipe (length)	YOLOv08/ YOLOv11	0.94	0.88	0.28
	YOLOv08/ Physical Measurements	-0.27	0.07	11.19
	YOLOv11/ Physical Measurements	-0.27	0.07	11.09
	YOLOv08/ YOLOv11	0.95	0.90	1.38

For *P. ostreatus* (Table 2), image-derived measurements exhibited trait-dependent levels of agreement with physical measurements. Pileus diameter showed limited correspondence, with YOLOv8 achieving $r = 0.20$, $R^2 = 0.04$, and $MAE = 5.06$ mm, while YOLOv11 produced similar values ($r = 0.23$, $R^2 = 0.05$, $MAE = 5.02$ mm). Pileus thickness displayed consistently low correlation across both models ($r \approx 0.16$ – 0.17 , $R^2 \approx 0.03$), whereas stipe-related traits demonstrated relatively stronger associations, with r values ranging from 0.39 to 0.43 and R^2 exceeding 0.15. Despite these differences in absolute agreement with physical measurements, inter-model consistency between YOLOv8 and YOLOv11 was exceptionally high across all traits ($r \geq 0.93$, $R^2 \geq 0.86$, $MAE \leq 1.44$ mm), indicating that both models produced nearly identical dimensional predictions under identical training conditions, thereby supporting analytical equivalence between the two architectures.

A comparable trend was observed for *F. velutipes* (Table 3), reflecting the slender morphology and frequent overlap of fruiting bodies. For pileus diameter, moderate agreement with physical measurements was obtained for both YOLOv8 ($r = 0.42$, $R^2 = 0.18$, $MAE = 2.01$ mm) and YOLOv11 ($r = 0.41$, $R^2 = 0.17$, $MAE = 1.99$ mm). Pileus thickness exhibited limited correspondence across models, while stipe thickness and length showed low to moderate agreement with physical values. Notably, inter-model reproducibility remained extremely high for all traits ($r \geq 0.94$, $R^2 \geq 0.88$, $MAE \leq 1.38$ mm), demonstrating consistent phenotypic interpretation between the two architectures.

Overall, these results indicate that YOLO-derived phenotypic measurements capture systematic morphological variation in a consistent manner across models. Although absolute agreement with physical measurements was moderate and strongly trait dependent, likely influenced by species-specific morphology, occlusion, and structural complexity, YOLOv11 reproduced the analytical behavior of YOLOv8 with comparable R^2 values and similar MAE. When considered together with the observed reductions in computational complexity, these findings support the applicability of YOLOv11 as a non-destructive analytical tool for controlled-environment phenotyping workflows rather than as a direct replacement for physical measurement tools.

4. Discussion

This study systematically compared the analytical performance and computational efficiency of two YOLO segmentation models, YOLOv8 and YOLOv11, for the efficient and automated phenotypic analysis of *Pleurotus ostreatus* and *Flammulina velutipes* [36]. Both models successfully identified and segmented pileus and stipe regions in side-view mushroom images, demonstrating stable performance across diverse morphological conditions. Importantly, the results indicate that YOLOv11 achieves segmentation accuracy equivalent to YOLOv8 while achieving clear improvements in computational efficiency, suggesting that recent architectural refinements in YOLOv11 primarily influence processing efficiency and training behavior rather than yielding statistically significant gains in accuracy [37,38].

Although YOLOv11 produced smoother and more visually consistent segmentation boundaries, particularly along the pileus margins, these improvements were qualitative rather than statistically significant. Quantitative indices, including precision, recall, and mAP_{50-95} , showed only marginal differences ($\Delta mAP_{50-95} < 0.01$) between the two models, indicating that their predictive performance was essentially comparable. The slightly enhanced visual coherence of YOLOv11 may stem from its use of the C2f-Fusion and RepNCSPPELAN modules, which improve feature reuse and stabilize training dynamics [39]. These architectural refinements appear to influence convergence behavior rather than altering final detection precision. Accordingly, the observed differences between YOLOv8 and YOLOv11 should be interpreted as efficiency-oriented refinements rather than true performance divergence.

Computational analysis further confirmed that YOLOv11 operates with reduced FLOPs, fewer trainable parameters, and lower gradient load compared with YOLOv8, despite incorporating a greater number of network layers. Together with earlier stabilization of training loss curves, these results demonstrate that YOLOv11 converges more efficiently under identical data and hyperparameter conditions. Such efficiency gains are consistent with prior studies emphasizing

lightweight and modular architectures for scalable deployment under hardware constraints [40,41]. In phenotyping workflows where computational sustainability and throughput are critical, these characteristics provide a practical advantage.

Validation against physical measurements revealed inherent limitations of image-based phenotyping, particularly for thickness-related traits and highly occluded structures. Pearson's correlation and R^2 values indicated moderate to low correspondence between YOLO-derived and physically measured traits, reflecting geometric ambiguity and scale sensitivity associated with two-dimensional imaging. Notably, however, exceptionally high inter-model consistency between YOLOv8 and YOLOv11 was observed across all evaluated traits ($r \geq 0.94$, $R^2 \geq 0.86$) [42,43]. This consistency indicates that YOLOv11 preserves the same phenotypic information content as YOLOv8 despite its reduced computational footprint.

Several limitations of this study should be acknowledged. First, the YOLO models were trained and evaluated using images of *P. ostreatus* and *F. velutipes* cultivated under bottle-based systems, and application to other mushroom species with distinct morphologies may require additional data and model adaptation. Second, image acquisition was conducted under controlled illumination and spatial conditions using a custom-built imaging chamber. While this ensured data consistency and minimized background interference, model performance under uncontrolled or field-scale conditions remains to be evaluated. In addition, although polygon-based annotation was cross-validated by independent annotators, manual labeling inevitably involves a degree of subjectivity, particularly for thickness-related traits and overlapping structures where clear boundaries are difficult to define.

Despite these limitations, the proposed framework demonstrates clear practical relevance for mushroom phenotyping and breeding systems. The ability to non-destructively extract phenotypic traits from mature fruiting bodies directly within cultivation bottles enables efficient monitoring of size distribution, morphological uniformity, and growth consistency, which are closely associated with yield stability and quality assessment. Combined with its reduced computational burden, YOLOv11 provides a practical foundation for scalable indoor phenotyping pipelines and digital phenotyping pipelines supporting phenotype-informed breeding and morphological assessment.

In summary, these findings demonstrate that the primary contribution of YOLOv11 lies not in improving absolute measurement accuracy, but in delivering phenotypic interpretation equivalent to YOLOv8 with substantially enhanced computational efficiency. By maintaining analytical equivalence while reducing processing cost and improving training stability, YOLOv11 represents a practical and scalable option for high-throughput, non-destructive phenotyping under controlled-environment conditions.

5. Conclusions

This study compared the performance of YOLOv8 and YOLOv11 segmentation models for automated phenotypic analysis of *Pleurotus ostreatus* and *Flammulina velutipes* cultivated in bottle-based systems. In contrast to conventional phenotyping approaches that require destructive sampling, the proposed framework enables non-destructive extraction of key morphological traits directly from side-view images acquired during cultivation. Both YOLO models reliably detected and segmented pileus and stipe structures, exhibiting stable performance across two morphologically distinct mushroom species.

Quantitative evaluation demonstrated that YOLOv11 achieved segmentation accuracy comparable to that of YOLOv8 ($\Delta mAP_{50-95} < 0.01$), while reducing computational complexity as evidenced by lower FLOPs, parameter counts, and gradient loads. These efficiency gains, achieved through architectural refinements such as the C2f-Fusion and RepNCSPPELAN modules, contributed to improved training efficiency and convergence stability without compromising segmentation accuracy.

Comparison with caliper-based physical measurements indicated that YOLO-derived phenotypic traits captured systematic morphological variation, although absolute correspondence with physical measurements remained moderate and trait-dependent due to occlusion effects and

structural complexity inherent to mushroom fruiting bodies. Importantly, inter-model consistency between YOLOv8 and YOLOv11 was exceptionally high across all evaluated traits (inter-model $r \geq 0.94$, $R^2 \geq 0.86$, $MAE \leq 1.44$ mm), confirming that the computational efficiency improvements of YOLOv11 did not alter phenotypic interpretation or analytical reliability.

Overall, these findings demonstrate that YOLOv11 provides an efficient and analytically equivalent alternative to YOLOv8 for non-destructive phenotyping under controlled-environment cultivation conditions. By maintaining phenotypic consistency while reducing computational demand, YOLOv11 offers a practical foundation for scalable indoor phenotyping pipelines. Future work should explore the extension of this approach to more diverse cultivation environments and larger-scale phenotyping scenarios, thereby supporting the continued development of AI-based phenotyping methodologies for fungal morphology assessment and controlled-environment cultivation systems.

Supplementary Materials: The supplementary data supporting the findings of this study are available in Zenodo at <https://doi.org/10.5281/zenodo.18372009>. Table S1: Quantitative comparison of predicted and measured phenotypic traits in *Pleurotus ostreatus*. Table S2: Quantitative comparison of predicted and measured phenotypic traits in *Flammulina velutipes*. Figure S1: Instance segmentation results of *Pleurotus ostreatus* obtained using YOLOv8. Figure S2: Instance segmentation results of *Pleurotus ostreatus* obtained using YOLOv11. Figure S3: Instance segmentation results of *Flammulina velutipes* obtained using YOLOv8. Figure S4: Instance segmentation results of *Flammulina velutipes* obtained using YOLOv11.

Author Contributions: Conceptualization, D.H., Y.L. and J.H.; Data curation, M.J. and S.I.; Formal analysis, Y.L. and J.H.; Funding acquisition, J.H.; Investigation, D.H., E.J. and M.S.; Methodology, D.H. and J.H.; Project administration, J.H.; Resources, Y.L., M.J., S.I. and M.S.; Supervision, Y.L. and J.H.; Validation, D.H. and J.H.; Visualization, D.H.; Writing—original draft, D.H.; Writing—review and editing, D.H., Y.L. and J.H. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Development of core mushroom resources and safety preservation technology, the postdoctoral research program support project (PJ01716801) as part of the results conducted by the Rural Development Administration (PJ01733103).

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The data supporting the findings of this study are available in Zenodo at <https://doi.org/10.5281/zenodo.18372009>. Additional data and materials are available from the corresponding authors upon reasonable request.

Acknowledgments: D.H. gratefully acknowledges the support of the Mushroom Research Division, Ginseng and Specialty Crops Department, Rural Development Administration (RDA), Republic of Korea. This study was carried out as part of a professional researcher support program within the Mushroom Research Division, under which the experimental work and manuscript preparation were conducted.

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

AI-Assisted Editing Statement: During revision, an AI-assisted tool was used solely for English language editing (grammar and readability). It was not used for data analysis, interpretation, or generating scientific content. The authors take full responsibility for the manuscript.

References

1. Morais, M.H.; Ramos, A.C.; Matos, N.; Santos-Oliveira, E.J. Production of shiitake mushroom (*Lentinus edodes*) on ligninocellulosic residues. *Food Sci. Technol. Int.* 2000, 6, 123–128.
2. Sánchez, C. Modern aspects of mushroom culture technology. *Appl. Microbiol. Biotechnol.* 2004, 64, 756–762.
3. Chang, S.-T.; Miles, P.G. *Mushrooms: Cultivation, Nutritional Value, Medicinal Effect, and Environmental Impact*, 2nd ed.; CRC Press: Boca Raton, FL, USA, 2004.

4. Chang, S.-T.; Hayes, W.A. *The Biology and Cultivation of Edible Mushrooms*; Academic Press: Massachusetts, USA, 2013; pp. 3–9.
5. Suarez, E.; Blaser, M.; Sutton, M. Automating leaf area measurement in citrus: Development and validation of a Python-based tool. *Appl. Sci.* 2025, 15, 9750.
6. Cheong, J.C.; Lee, C.J.; Suh, J.S.; Moon, Y.H. Comparison of physico-chemical and nutritional characteristics of pre-inoculation and post-harvest *Flammulina velutipes* media. *J. Mushroom Sci. Prod.* 2012, 10, 174–178.
7. Cheong, J.C.; Lee, C.J.; Moon, J.W. Comprehensive model for medium composition for mushroom bottle cultivation. *J. Mushrooms* 2016, 14, 111–118.
8. Sapkota, R.; Karkee, M. Object detection with multimodal large vision-language models: An in-depth review. *Inf. Fusion* 2025, 126, 103575.
9. Wei, Z.; Wang, J.; You, H.; Ji, R.; Wang, F.; Shi, L.; Yu, H. A lightweight context-aware framework for toxic mushroom detection in complex ecological environments. *Ecol. Inform.* 2025, 90, 103256.
10. Dhanya, V.G.; Subeesh, A.; Kushwaha, N.L.; Vishwakarma, D.K.; Kumar, T.N.; Ritika, G.; Singh, A.N. Deep learning-based computer vision approaches for smart agricultural applications. *Artif. Intell. Agric.* 2022, 6, 211–229.
11. Hafiz, A.M.; Bhat, G.M. A survey on instance segmentation: State of the art. *Int. J. Multimed. Inf. Retr.* 2020, 9, 171–189. <https://doi.org/10.1007/s13735-020-00195-x>.
12. Coulibaly, S.; Kamsu-Foguem, B.; Kamissoko, D.; Traore, D. Deep learning for precision agriculture: A bibliometric analysis. *Intell. Syst. Appl.* 2022, 16, 200102. <https://doi.org/10.1016/j.iswa.2022.200102>.
13. Sapkota, R.; Ahmed, D.; Karkee, M. Comparing YOLOv8 and Mask R-CNN for instance segmentation in complex orchard environments. *Artif. Intell. Agric.* 2024, 13, 84–99.
14. Rashid, J.; Khan, I.; Ali, G.; Alturise, F.; Alkhalifah, T. Real-time multiple guava leaf disease detection from a single leaf using a hybrid deep learning technique. *Comput. Mater. Continua* 2023, 74, 1–15. <https://doi.org/10.32604/cmc.2023.032005>.
15. Maji, A.K.; Marwaha, S.; Kumar, S.; Arora, A.; Chinnusamy, V.; Islam, S. SlyphNet: Spikelet-based yield prediction of wheat using advanced plant phenotyping and computer vision techniques. *Front. Plant Sci.* 2022, 13, 889853. <https://doi.org/10.3389/fpls.2022.889853>.
16. Sapkota, R.; Karkee, M. Comparing YOLOv11 and YOLOv8 for instance segmentation of occluded and non-occluded immature green fruits in complex orchard environment. *arXiv* 2025, arXiv:2410.19869.
17. Xie, L.; Jing, J.; Wu, H.; Kang, Q.; Zhao, Y.; Ye, D. MPG-YOLO: Enoki mushroom precision grasping with segmentation and pulse mapping. *Agronomy* 2025, 15, 432. <https://doi.org/10.3390/agronomy15020432>.
18. Qi, W.; Chen, H.; Zheng, X.; Zhang, T.; Liu, Y. Detection and classification of shiitake mushroom fruiting bodies based on Mamba YOLO. *Sci. Rep.* 2025, 15, 133. <https://doi.org/10.1038/s41598-025-00133-z>.
19. Redmon, J. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; IEEE: Las Vegas, NV, USA, 2016.
20. Gong, R.; Zhang, H.; Li, G.; He, J. Edge Computing-Enabled Smart Agriculture: Technical Architectures, Practical Evolution, and Bottleneck Breakthroughs. *Sensors* 2025, 25(17), 5302. <https://doi.org/10.3390/s25175302>.
21. Tariq, M.U.; Saqib, S.M.; Mazhar, T.; Khan, M.A.; Shahzad, T.; Hamam, H. Edge-enabled smart agriculture framework: Integrating IoT, lightweight deep learning, and agentic AI for context-aware farming. *Results in Engineering* 2025, 28, 107342. <https://doi.org/10.1016/j.rineng.2025.107342>.
22. Xu, X.; Li, J.; Zhou, J.; Feng, P.; Yu, H.; Ma, Y. Three-dimensional reconstruction, phenotypic traits extraction, and yield estimation of shiitake mushrooms based on structure from motion and multi-view stereo. *Agriculture* 2025, 15, 298.
23. Khan, A.T.; Jensen, S.M. LEAF-Net: A unified framework for leaf extraction and analysis in multi-crop phenotyping using YOLOv11. *Agriculture* 2025, 15, 196. <https://doi.org/10.3390/agriculture15020196>.
24. Ho Bao Thuy, Q.; Suzuki, A. Technology of mushroom cultivation. *Viet. J. Sci. Technol.* 2019, 57, 265–286.
25. Badgujar, C.M.; Poulouse, A.; Gan, H. Agricultural object detection with YOLO: A bibliometric and systematic review. *Comput. Electron. Agric.* 2024, 223, 109090. <https://doi.org/10.1016/j.compag.2024.109090>.
26. Zakeri, R.; Zamani, A.; Taghizadeh, A.; Abbaszadeh, M.; Saadatfar, B. M18K: A comprehensive RGB-D dataset and benchmark for mushroom detection and instance segmentation. *arXiv* 2024, arXiv:2407.11275.

27. Abdullah, A.; Amran, G.A.; Tahmid, S.M.A.; Alabrah, A.; Al-Bakhrani, A.A.; Ali, A. Deep-learning-based detection of diseased tomato leaves. *Agronomy* 2024, *14*, 1593. <https://doi.org/10.3390/agronomy14071593>.
28. Wang, C.; Li, H.; Deng, X.; Liu, Y.; Wu, T.; Liu, W.; Xiao, R.; Wang, Z.; Wang, B. Improved YOLOv8 model for precision detection of tea leaves. *Agriculture* 2024, *14*, 2324. <https://doi.org/10.3390/agriculture14122324>.
29. Wang, N.; Liu, H.; Li, Y.; Zhou, W.; Ding, M. Segmentation and phenotype calculation of rapeseed pods using YOLOv8 and Mask R-CNN. *Plants* 2023, *12*, 3328. <https://doi.org/10.3390/plants12183328>.
30. Solimani, F.; Cardellicchio, A.; Dimauro, G.; Petrozza, A.; Summerer, S.; Cellini, F.; Renò, A. Optimizing tomato plant phenotyping using an enhanced YOLOv8 architecture. *Comput. Electron. Agric.* 2024, *218*, 108728.
31. Wu, C.; Zhang, S.; Wang, W.; Wu, Z.; Yang, S.; Chen, W. Phenotypic parameter computation using YOLOv11-DYFP keypoint detection. *Aquac. Eng.* 2025, *111*, 102571. <https://doi.org/10.1016/j.aquaeng.2025.102571>.
32. Sanchez, S.A.; Romero, H.J.; Morales, A.D. Comparison of performance metrics of pretrained object detection models. *IOP Conf. Ser. Mater. Sci. Eng.* 2020, *844*, 012024. <https://doi.org/10.1088/1757-899X/844/1/012024>.
33. Murat, A.A.; Kiran, M.S. A comprehensive review on YOLO versions for object detection. *Eng. Sci. Technol. Int. J.* 2025, *70*, 102161. <https://doi.org/10.1016/j.jestch.2025.102161>.
34. Lu, C.P.; Liaw, J.J.; Wu, T.C.; Hung, T.F. Development of a mushroom growth measurement system applying deep learning. *Agronomy* 2019, *9*, 32. <https://doi.org/10.3390/agronomy9010032>.
35. Frossard, E.; Liebis, F.; Hgaza, V.K.; Kiba, D.I.; Kirchgessner, N.; Müller, L.; Müller, P.; Pouya, N.; Ringger, C.; Walter, A. Image-based phenotyping of water yam growth and nitrogen status. *Agronomy* 2021, *11*, 249. <https://doi.org/10.3390/agronomy11020249>.
36. Shi, Y.; Zhang, C.; Sun, Z.; Liu, J.; Li, B. OMC-YOLO: A lightweight grading detection method for oyster mushrooms. *Horticulturae* 2024, *10*, 742. <https://doi.org/10.3390/horticulturae10070742>.
37. He, L.H.; Zhou, Y.Z.; Liu, L.; Cao, W.; Ma, J.H. Research on object detection and recognition in remote sensing images based on YOLOv11. *Sci. Rep.* 2025, *15*, 14032. <https://doi.org/10.1038/s41598-025-96314-x>.
38. Mihajlovic, M.; Stojanovic, A.; Petrovic, S. Enhancing instance segmentation in high-resolution aerial imagery with YOLOv11s-Seg. *Mathematics* 2025, *13*, 3079.
39. Su, C.; Lin, H.; Wang, D. Nav-YOLO: A lightweight and efficient object detection method for edge devices. *ISPRS Int. J. Geo-Inf.* 2025, *14*, 364.
40. Padilla, R.; Netto, S.; Da Silva, E. Performance metrics for object detection algorithms. *Electronics* 2020, *9*, 279.
41. Long, X.; Deng, K.; Wang, G.; Zhang, Y. PP-YOLO: An effective and efficient implementation of object detector. *arXiv* 2020, arXiv:2007.12099.
42. Lu, C.P.; Cheng, S.H.; Hsiao, Y.T. Development of a mushroom growth measurement system using image processing. *Agronomy* 2019, *9*, 32.
43. Kiba, D.I.; Ofori, E.; Bationo, A. Image-based phenotyping methods for measuring water yam growth and nitrogen nutritional status. *Agronomy* 2021, *11*, 1529.

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