

Article

Not peer-reviewed version

M18K: A Multi-Purpose Real-World Dataset for Mushroom Detection, 3D Pose Estimation, and Growth Monitoring

[Abdollah Zakeri](#) , [Mulham Fawakherji](#) , [Jiming Kang](#) , [Bikram Koirala](#) , [Venkatesh Balan](#) , [Weihsang Zhu](#) , [Driss Benhaddou](#) , [Fatima A Merchant](#) *

Posted Date: 25 April 2025

doi: 10.20944/preprints202504.2103.v1

Keywords: public mushroom dataset; automated mushroom harvesting; mushroom detection; computer vision; smart agriculture; growth monitoring



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.

Article

M18K: A Multi-Purpose Real-World Dataset for Mushroom Detection, 3D Pose Estimation, and Growth Monitoring

Abdollah Zakeri ¹, Mulham Fawakherji ², Jiming Kang ², Bikram Koirala ², Venkatesh Balan ², Weihang Zhu ², Driss Benhaddou ² and Fatima A. Merchant ^{1,2,*}

¹ Department of Computer Science, University of Houston, Houston, TX 77004, USA

² Department of Engineering Technology, University of Houston, Houston, TX 77004, USA

* Correspondence: fmerchant@uh.edu

Abstract: Automating agricultural processes holds significant promise for enhancing efficiency and sustainability in various farming practices. This paper contributes to the automation of agricultural processes by providing a dedicated mushroom detection dataset related to automated harvesting, 3D pose estimation, and growth monitoring of the button mushroom produced using *Agaricus Bisporus* fungus. With a total of 2,000 images for object detection, instance segmentation, and 3D pose estimation containing over 100,000 mushroom instances and an additional 3,838 images for yield estimation containing 8 mushroom scenes covering the complete growth period, it fills the gap in mushroom-specific datasets and serves as a benchmark for detection and instance segmentation as well as 3D pose estimation algorithms in smart mushroom agriculture. The dataset, featuring realistic growth environment scenarios with comprehensive 2D and 3D annotations, is assessed using advanced detection and instance segmentation algorithms. The paper details the dataset's characteristics, presents the detailed statistics of mushroom growth and yield, evaluates algorithmic performance, and for broader applicability, we have made all resources publicly available including images, codes, and trained models via our GitHub repository: <https://github.com/abdollahzakeri/m18k>

Keywords: public mushroom dataset; automated mushroom harvesting; mushroom detection; computer vision; smart agriculture;

1. Introduction

The integration of Artificial Intelligence (AI)-driven automation into agricultural processes has gained significant momentum, offering opportunities for increased efficiency and sustainability in various farming practices [1–4]. Among these, the automatic harvesting of edible mushrooms in large farms holds substantial potential for streamlining operations and reducing labor-intensive tasks.

While computer vision and object detection technologies have seen significant advancements, a major challenge remains the lack of publicly available datasets specifically tailored for mushroom detection. This limitation impedes systematic evaluation and comparison of different mushroom detection algorithms. Synthetic mushroom datasets [1,5] have been introduced to aid model pre-training; however, they often fail to capture real-world complexities—such as varied lighting conditions, diverse growth stages, and cluttered backgrounds—necessitating fine-tuning on real images for robust performance. Additionally, in practical agricultural settings, accurate 3D orientations can facilitate automated harvesting by providing robotic arms with precise information about the mushroom's position and tilt [6–8]. Furthermore, reliable 3D pose data are essential for tasks volumetric measurements where a mushroom's overall shape and orientation must be understood in detail. To the best of authors' knowledge, there are currently no other datasets of real mushrooms providing ground truth data for various modalities and purposes such as instance segmentation and object detection, yield estimation, and 3D pose estimation.

To address this gap, we present a comprehensive real-world mushroom dataset featuring 2,000 images dedicated to object detection, instance segmentation, and 3D pose estimation, plus an additional 3,838 images for yield estimation. The 2,000 images contain over 100,000 mushroom instances and the 3,838 images for yield estimation contain 8 different timelapse images of unique scenes covering the entire growth period. Our dataset's growth monitoring and yield estimation subset is an augmented version of the data used in our previous work[9]. All images in the object detection subset are RGB-D pairs captured using an Intel RealSense D405 RGB-D camera, and the yield estimation subset's images were captured by a TLC200 Pro timelapse camera. The dataset focuses on two commonly cultivated varieties—white button mushrooms and baby bella (cremini) mushrooms—and includes detailed 3D pose annotations, generated first by manual labeling of a small subset of 50 images, followed by pseudo-labeling the remaining 3D instances through a trained pose estimation model proposed in[10], and a final round of manual 3D label refinement. Each image captures various lighting conditions and growth stages, covering a broad range of orientations and backgrounds. Additionally, the second subset of 3,838 images includes repeated captures of the 8 different scenes for yield estimation, enabling measurement of growth rates over time and supporting a yield prediction model.

Figure 1 illustrates sample dataset images, corresponding ground truth instance segmentation masks, and depth images for both Baby Bella (BB) and White Button (WB) mushrooms. Figure 2 presents 3D point clouds of brown and white mushroom scenes, showing how our RGB-D data can be leveraged for 3D spatial and volumetric analyses.

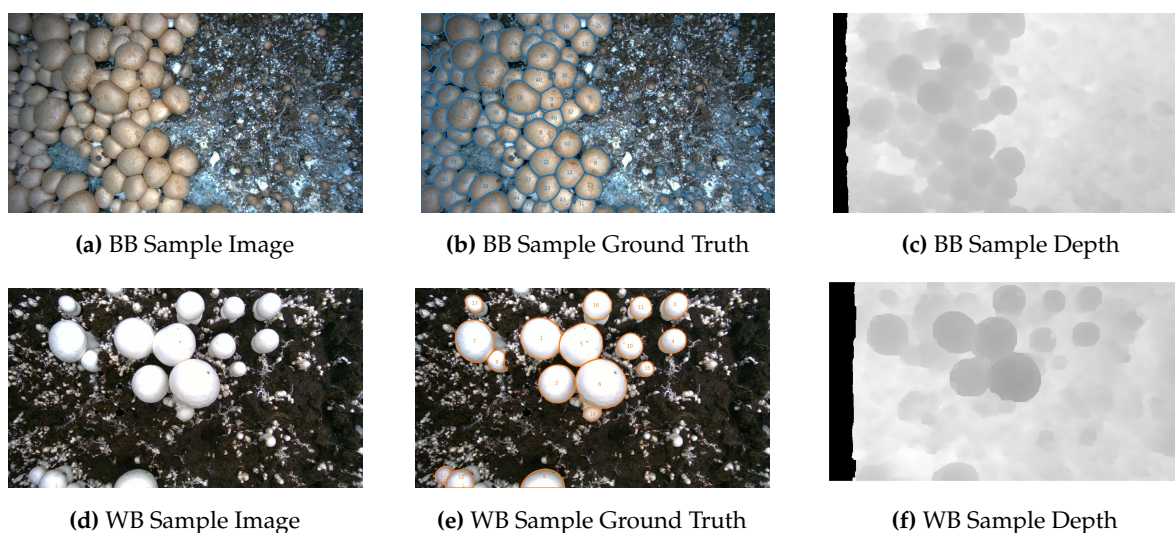
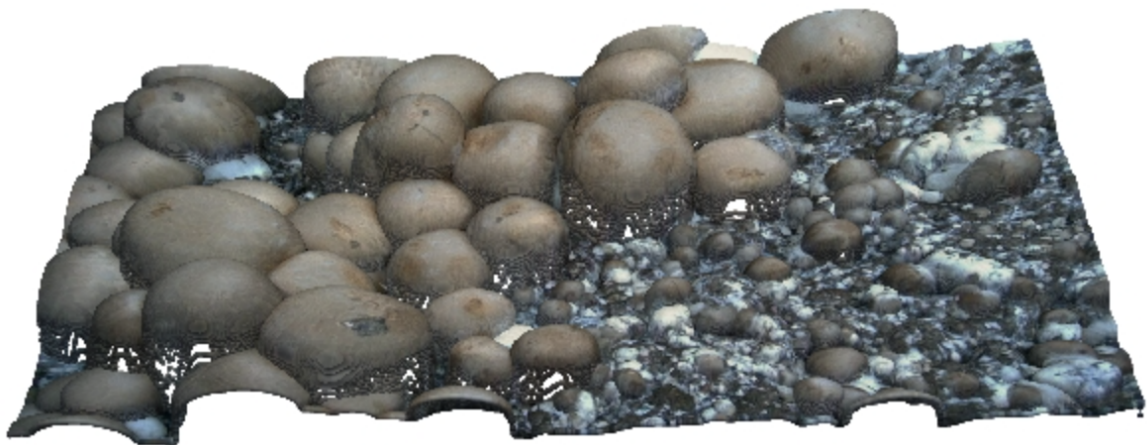


Figure 1. Sample of dataset images, ground truth instance segmentation masks, and depth images for Baby Bella (BB) and White Button (WB) mushroom images.

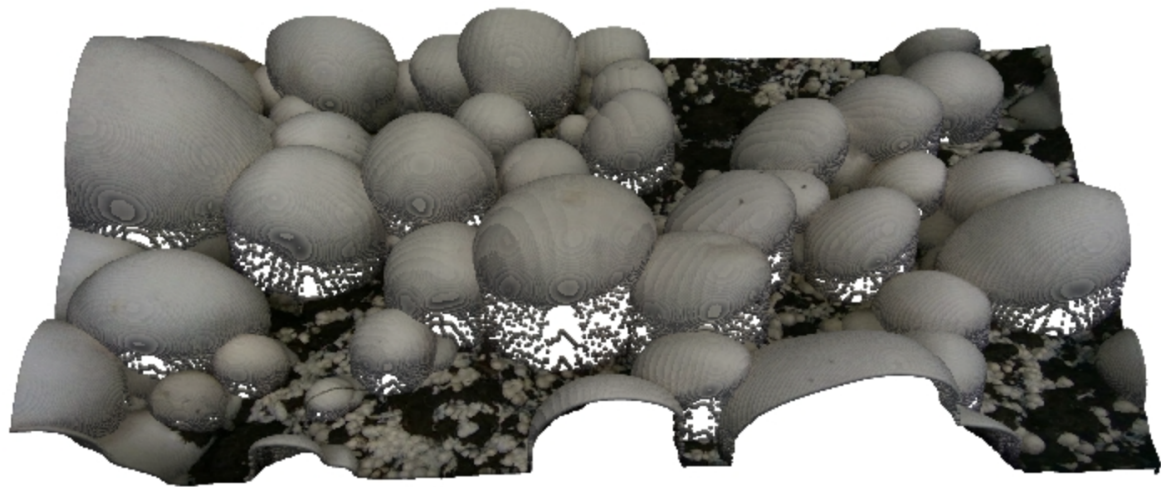
Our main contributions are summarized as follows:

- We introduce a publicly available, real-world mushroom dataset of 2,000 RGB-D images for detection-related tasks, instance segmentation, and 3D pose estimation, as well as 3,838 images for growth monitoring.
- We provide comprehensive ground truth labels for object detection, instance segmentation, 3D pose estimation, and time-series data for yield forecasting.
- We establish a benchmark by evaluating state-of-the-art mushroom detection algorithms, facilitating reliable performance comparison.

In the subsequent sections of this paper, we describe the annotation methodology in detail, present descriptive statistics of the dataset, evaluate benchmark algorithms, and discuss potential applications of our dataset for automated harvesting, growth estimation, and other smart farming tasks.



(a) BB Sample point cloud



(b) WB Sample point cloud

Figure 2. Sample of dataset 3D point clouds for (a) baby bella mushrooms and (b) white button mushrooms.

2. Literature Review

Several articles and papers have been published regarding the application of image processing and machine learning algorithms for mushroom detection, quality assessment, disease detection, growth monitoring, and automatic harvesting. These works can be grouped into three main categories as follows.

Table 1. Summary of task, method, and results of similar studies leveraging mushroom detection.

Year	Author	Task	Method	Dataset Size	Results
2022	Baisa and Al-Diri [2]	Detection, localization, and 3D pose estimation	Segmentation using AC, Detection using CHT, 3D localization using depth information	-	98.99% Precision 99.29% Recall
2024	Jareanpon et al.[11]	Fungal disease detection	DenseNet201, ResNet50, Inception V3, VG-GNet19	2000 images	94.35% Precision 89.47% F1-score
2019	Lee et al.[12]	Detection, and maturity classification	Faster R-CNN for detection, SVM for maturity classification	920 time-lapse image sets	42.00% Precision 82.00% Recall 56.00% F1-score 70.93% Maturity classification accuracy
2023	Moysiadis et al.[13]	Mushroom growth monitoring	YOLOv5 and Detec-tron2	1128 images, 4271 Mush-rooms	76.50% F-1 Score 70.00% Accuracy
2019	Nadim et al.[14]	Mushroom quality control	Neural network and fuzzy logic	250 images	95.60% Accuracy

Table 1. Cont.

Year	Author	Task	Method	Dataset Size	Results
2018	Olpin <i>et al.</i> [15]	Detection	RCNN and RFCN	310 images	92.16% Accuracy
2023	Retsinas <i>et al.</i> [16]	Detection and 3D pose estimation	segmentation using a k-Medoids approach based on FPFH and FCGF	Synthetic 3D dataset	99.80% MAP at 25.00% IOU
2000	Vizhanyo and Felfoldi [17]	Disease Detection	LDA	-	85.00% True Classification Rate
2018	Wang <i>et al.</i> [18]	Automatic sorting	Watershed, Canny, Morphology	-	97.42% Accuracy
2022	Wang <i>et al.</i> [19]	Detection and growth monitoring	YOLOv5 + CBAM + BiFPN	-	99.24% MAP
2022	Wei <i>et al.</i> [20]	Detection and growth monitoring	YOLOv5 + ASPP + CIOU	-	98.00% Accuracy
2022	Yang <i>et al.</i> [21]	Detection	MaskRCNN	-	95.06% AP at 50.00% IOU
2022	Zahan <i>et al.</i> [22]	Disease Classification	AlexNet, GoogleNet, ResNet15	2536 images	90.00% Accuracy 89.00% Precision 90.00% Recall 90.00% F1-score

2.1. Mushroom Detection and Localization

There is a growing body of literature dedicated to mushroom detection and localization, each focusing on different algorithms for mushroom detection and localization, using techniques from classical image processing to advanced deep learning and computer vision.

The YOLO (You Only Look Once) algorithm, a significant advancement introduced in 2016, is a cornerstone in this field. It simplified object detection into a single regression problem, directly predicting bounding boxes and class probabilities, thus enhancing efficiency in detection tasks. Several studies adapted the YOLO algorithm for mushroom detection, yielding notable results.

Wei *et al.*[20] developed Recursive-YOLOv5, significantly surpassing the original YOLOv5 with 98% accuracy, though at the cost of almost doubling the network’s parameters. Wang *et al.*[19] introduced Mushroom-YOLO, an improved version of YOLOv5, achieving a mean average precision of 99.24% by integrating Convolutional Block Attention Module (CBAM) and Bidirectional Feature Pyramid Network (BiFPN) into the architecture. Olpin *et al.*[15] used Region-Based Convolutional Networks for mushroom detection, finding that the Region-based Convolutional Neural Network(RCNN) model was more accurate (92.162%) than the Region-based Fully Convolutional Network (RFCN) (87.631%), albeit slower. Yang *et al.*[21] enhanced *Agaricus Bisporus* recognition using Mask RCNN, achieving an Average Precision (AP50) of 95.061%, though the operation was slow. Retsinas *et al.*[16] created a vision module for 3D pose estimation of mushrooms, resulting in high retrieval scores (up to 99.80% Mean Average Precision at 25% Intersection over Union) and accurate pose estimation (mean angle error as low as 8.70°). Lee *et al.*[12] automated the selective harvesting process, using the Faster R-CNN model for identification and a 3D point cloud for segmentation, achieving 70.93% accuracy in maturity identification. Baisa and Al-Diri [2] focused on robotic picking, detecting, and estimating the 3D pose of mushrooms with high precision (98.99%) and recall (99.29%), using RGB-D data. These studies demonstrate advancements in mushroom detection and localization, showcasing various algorithmic improvements. While they present significant progress, challenges such as computational efficiency, dataset quality, and the need for further accuracy and speed enhancements remain.

2.2. Mushroom Growth Monitoring And Quality Assessment

Lu *et al.*[23] developed a system using YOLOv3 for monitoring mushroom growth in greenhouses. The system could estimate mushroom cap size, count, growth rate, and predict harvest times. It demonstrated effectiveness in identifying growth patterns and provided real-time updates via a mobile interface. Lu and Liaw [24] introduced an image measurement algorithm utilizing YOLOv3, coupled with a Score-Punishment algorithm for accurate mushroom cap diameter measurement. The method excelled in accuracy over traditional methods but faced challenges with unclear mushroom contours or soil particles on caps. Nadim *et al.*’s image processing system [14] assessed mushroom quality based on color, area, weight, and volume, using data mining, neural networks, and fuzzy logic. It achieved

a 95.6% correct detection rate but required image pre-processing to counteract quality issues from normal imaging conditions.

Moysiadis *et al.*[13] automated the monitoring and classification of oyster mushrooms in greenhouses. They used YOLOv5 for detection and classification, and Detectron2 for tracking growth, showing potential improvements in harvesting times but faced difficulties in detecting small mushrooms. Wang *et al.*[18] developed an automatic sorting system for white button mushrooms, using image processing to measure the pileus diameter. The system achieved high grading speed and accuracy, significantly improving over manual grading methods.

Benhaddou *et al.*[9] implemented a computer vision system to estimate mushroom yield and quality in a commercial farm setting, utilizing the Circular Hough Transform algorithm. They achieved detailed tracking of mushroom growth across 1960 frames over 20 days, demonstrating the system's ability to monitor size distribution and development trends. This technological application offers a promising tool for optimizing cultivation and harvesting strategies by providing precise data on mushroom growth patterns and potential yield estimations.

2.3. Mushroom Disease Detection

Zahan *et al.*[22] employed deep learning models to classify mushroom diseases, with ResNet15 showing the best performance in accuracy and precision. The study emphasized the potential of deep learning in agricultural management. Vizhanyo and Felfoldi [17] used advanced image analysis to distinguish diseased spots on mushroom caps. Their method showed an 85% correct classification ratio, effectively differentiating between brown blotch and ginger blotch diseases.

Jareanpon *et al.*[11] developed an intelligent farm system for real-time detection of fungal diseases in Lentinus. The system integrated environment control, an imaging robot, and a deep learning-based prognosis system, achieving high precision in maintaining optimal conditions and disease detection.

In conclusion, the literature review section of this paper highlights a range of innovative approaches to mushroom detection, quality assessment, growth monitoring, and disease detection using image processing and machine learning algorithms. Table 1 lists these studies and their results. The studies vary in focus, ranging from detection and localization to quality and disease detection, each contributing valuable insights and advancements in the field. Notably, while these works demonstrate significant progress, there remains a gap in the standardization of methodologies and results comparison, compounded with challenge of data drift [25]. This gap is primarily due to the lack of a publicly available benchmark dataset. Based on the current literature [26], and to the best of the authors' knowledge, there are no publicly available annotated edible mushroom datasets. A standardized dataset would provide a common ground for evaluating different algorithms and techniques, ensuring comparability and consistency across studies. The availability of such a dataset would benefit the field, allowing for more robust and reliable comparisons of methods and results, driving forward the development of more efficient and effective solutions for mushroom cultivation and processing in the agricultural sector.

3. Labeling Process

3.1. Detection and Segmentation

The images were initially labeled using the Segment Anything Model (SAM) automatic mask generation feature [27]. This feature provides several configurations and post-processing steps to ensure the best quality of automatically generated masks for specific purposes. For instance, a grid of $n \times n$ (with $n = 32$ by default) positive points is placed on the original image and randomly cropped versions of it, and the generated masks are filtered by their confidence scores and stability within different confidence thresholds. The automatic mask generation algorithm considers a mask to be stable if thresholding the probability map at $(0.5 - \delta)$ and $(0.5 + \delta)$ results in similar masks. Additionally, masks can be filtered by their minimum and maximum area, and a certain maximum intersection over union (IOU) threshold can be set to avoid generation of overlapping masks. Nevertheless, these

configurations do not guarantee the generation of perfect ground truth masks and can only be used as a starting point.

These generated masks were meticulously edited manually during several steps to guarantee the quality of the dataset ground truth masks. Initially, all the false positive masks were removed and the missing masks were added. Later, the intersection of all pairs of masks in the same image was calculated for all the images in the dataset, and the overlapping masks were removed manually.

3.2. 3D Pose Estimation

In addition to 2D instance segmentation, a subset of the images was labeled with 3D bounding boxes and rotation annotations. We began by manually labeling an initial subset of 50 images with precise 3D boxes; we then applied the pose-estimation pipeline proposed by [10] to generate pseudo-labels for the remaining images. A final manual refinement phase ensured the accuracy of these 3D labels, particularly in cases where overlapping mushrooms or complex poses challenged the automated predictions.

Each 3D annotation contains

- the object centre (x, y, z) in pixel coordinates of the depth image,
- the bounding-box dimensions (w, h, d) in pixels (width, height, depth), and
- the object orientation in *Euler angles* (ϕ, θ, ψ) —roll, pitch, and yaw in degrees.

Following [28], we convert the Euler angles to the 6D Gram–Schmidt (GS) representation before feeding them to the network and again after inference. Given roll ϕ , pitch θ , and yaw ψ (in radians), we first form the rotation matrix

$$R = R_z(\psi) R_y(\theta) R_x(\phi),$$

where

$$R_x(\phi) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \phi & -\sin \phi \\ 0 & \sin \phi & \cos \phi \end{bmatrix}, R_y(\theta) = \begin{bmatrix} \cos \theta & 0 & \sin \theta \\ 0 & 1 & 0 \\ -\sin \theta & 0 & \cos \theta \end{bmatrix}, R_z(\psi) = \begin{bmatrix} \cos \psi & -\sin \psi & 0 \\ \sin \psi & \cos \psi & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

The GS representation simply concatenates the first two columns of R :

$$\mathbf{g} = [R_{:,1}^\top R_{:,2}^\top]^\top \in \mathbb{R}^6$$

At inference time, we recover the rotation by orthonormalising \mathbf{g} :

$$\hat{\mathbf{r}}_1 = \frac{\mathbf{g}_{1:3}}{\|\mathbf{g}_{1:3}\|_2}, \quad \hat{\mathbf{r}}_2 = \frac{\mathbf{g}_{4:6} - (\hat{\mathbf{r}}_1^\top \mathbf{g}_{4:6}) \hat{\mathbf{r}}_1}{\|\mathbf{g}_{4:6} - (\hat{\mathbf{r}}_1^\top \mathbf{g}_{4:6}) \hat{\mathbf{r}}_1\|_2}, \quad \hat{\mathbf{r}}_3 = \hat{\mathbf{r}}_1 \times \hat{\mathbf{r}}_2,$$

yielding $\hat{R} = [\hat{\mathbf{r}}_1 \ \hat{\mathbf{r}}_2 \ \hat{\mathbf{r}}_3] \in \text{SO}(3)$ and, if needed, the corresponding Euler angles.

This 6D representation avoids the discontinuities and singularities of Euler angles, leading to smoother optimisation during training and more stable pose predictions at inference.

3.3. Growth Monitoring

To facilitate the study of mushroom development over time, consecutive images of each mushroom were captured at multiple stages of growth. Each labeled instance retains a consistent ID across frames, allowing precise tracking and comparison of changes in shape and size throughout different growth phases. During annotation, we used manual inspections to refine the growth stage boundaries, ensuring accurate and reliable labeling for each mushroom across multiple developmental stages.

For the subset of images dedicated to yield estimation, we augmented and labeled the data from a similar work[9] and performed additional labeling to track mushroom growth over time. Specifically, the same scene was captured repeatedly, and each mushroom in the scene was assigned a consistent

identifier across all time points to facilitate growth rate calculations. This subset of our dataset can be utilized to support the development of temporal models for yield forecasting and growth monitoring.

4. Data Description

Our dataset includes 2,000 RGB-D images for object detection, instance segmentation, and 3D pose estimation, plus an additional 3,838 RGB-D images for yield estimation and growth monitoring. In the main subset, we have both baby bella (BB) and white button (WB) mushroom images, whereas the yield estimation subset only consists of repeated captures of WB mushrooms over time. A sample of the main subset's images, depth maps, and label masks is plotted in Fig. 1. Both the RGB images and the depth images in the main subset have a resolution of 1280*720 pixels, and hence, no alignment will be required for using the RGB and depth images as a single RGB-D input. Sample 3D point cloud images of WB and BB mushrooms are plotted in Fig. 2. The camera was placed at vertical distances of 27 cm (about 10.63 in) and 15 cm (about 5.91 in) above the cultivation beds during image acquisition of baby bella and white button mushrooms, respectively.

The total number of mushroom instances in the 2,000-image subset is approximately 70,000 for WB and 30,000 for BB mushrooms. The additional 3,838 images for yield estimation contain 8 different timelapse scenes observed repeatedly with time-steps of 15 minutes to track growth over time. These instances include a wide range of mushroom sizes for both categories, making our dataset suitable for training detection models for various tasks such as mushroom growth monitoring and automatic mushroom harvesting as well as 3D pose estimation. Histograms of mask areas and bounding box diagonal length of mushroom instances are plotted in Figure 4, and scatter plots of instance mask area versus bounding box diagonal length are shown in Figure 3. As illustrated in the scatter plots, some instances lie above the overall trend line, indicating mushrooms that are partially occluded by neighboring mushrooms. These occlusions lead to crescent-shaped masks rather than fully convex shapes. Both Figure 3 and Figure 4 show comparatively larger mask areas for white button mushrooms, explained by the shorter distance (15 cm) between the camera and the cultivation bed during WB image acquisition.

To study the separability of the two mushroom classes in different color spaces, we have plotted 3D scatter graphs (not shown here in detail) for RGB, HSV, and LAB in Figure 5, using the mean values of each instance's mask in all color channels. The two clusters appear more intertwined in the RGB color space, whereas HSV and LAB yield more distinct separations of WB and BB mushrooms.

Furthermore, our main subset including the 2,000 images has been annotated with 3D bounding boxes and rotation information. These 3D labels extend the utility of the dataset for pose estimation tasks and can be used to study volumetric aspects of mushroom growth. For the additional yield-estimation subset, repeated captures of the same scene allow researchers to investigate the temporal evolution of mushroom size and count, thus enabling more accurate yield forecasts.

Figure 6 illustrates two scatter plots showing the spread of rotation and dimension values for a randomly selected subset of our 3D labels having 2500 instances.

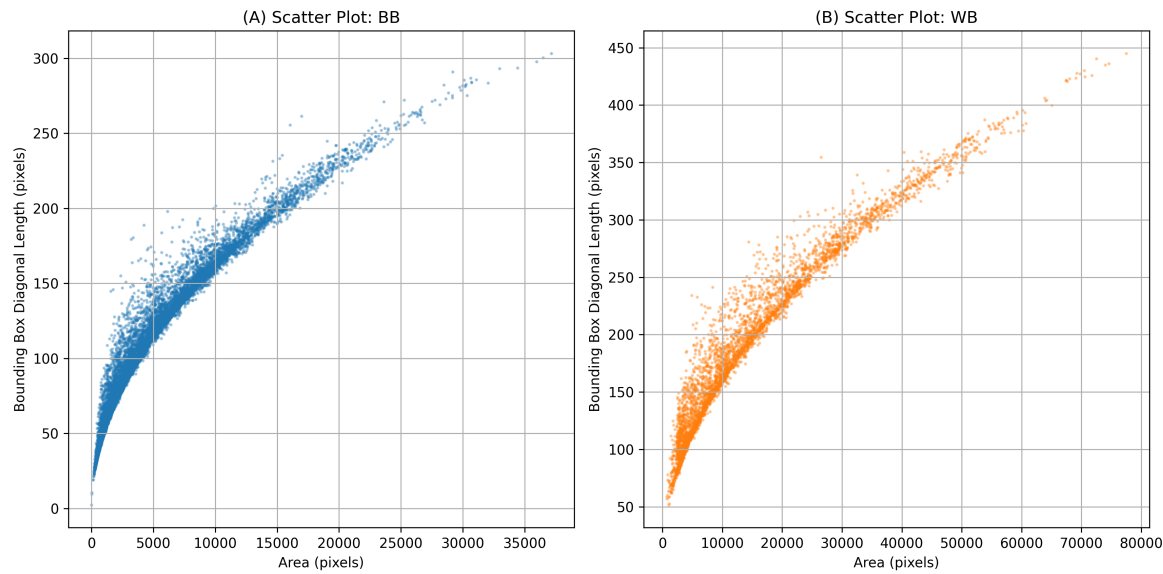


Figure 3. Scatter plots of instance mask area vs. bounding box diagonal length for (A) baby bella mushrooms and (B) white button mushrooms.

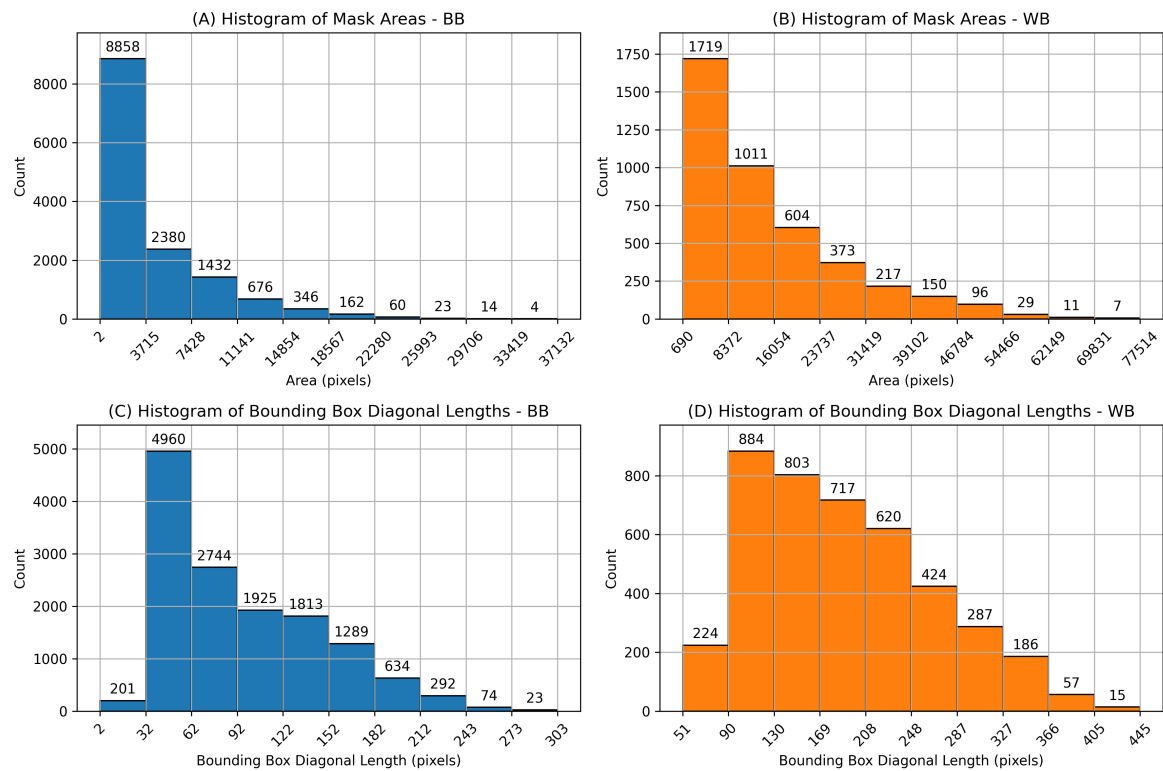


Figure 4. Histograms of mask areas and bounding box diagonal lengths. (A) and (B) show pixel mask areas for Baby Bella (BB) and White Button (WB) mushrooms, respectively, while (C) and (D) show bounding box diagonal lengths for BB and WB mushroom instances.

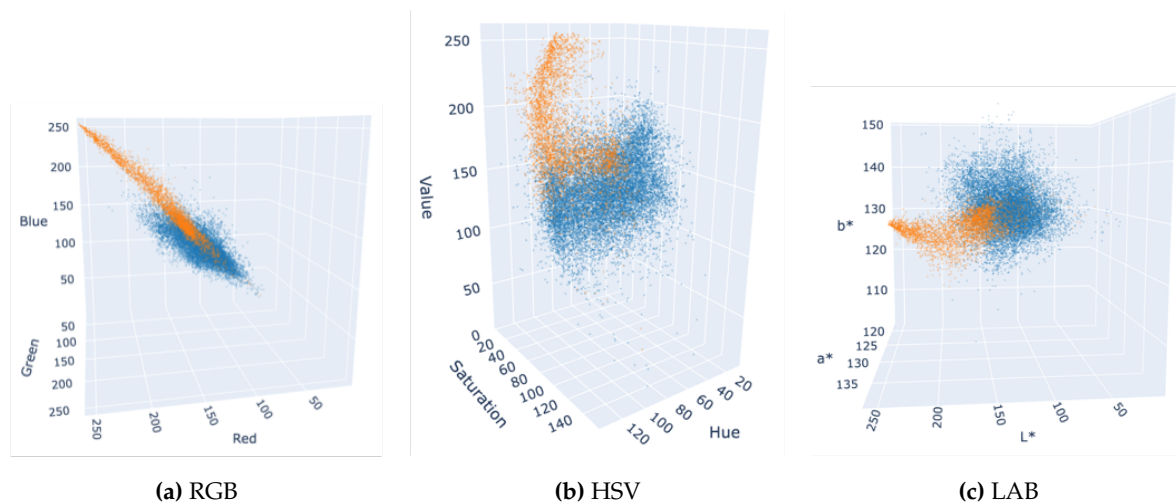


Figure 5. Color distribution of mushroom instances in various color spaces; judging solely by the color distributions, LAB and HSV color spaces appear more suitable than RGB for distinguishing WB and BB instances. WB mushrooms are shown in orange and BB mushrooms are shown in blue.

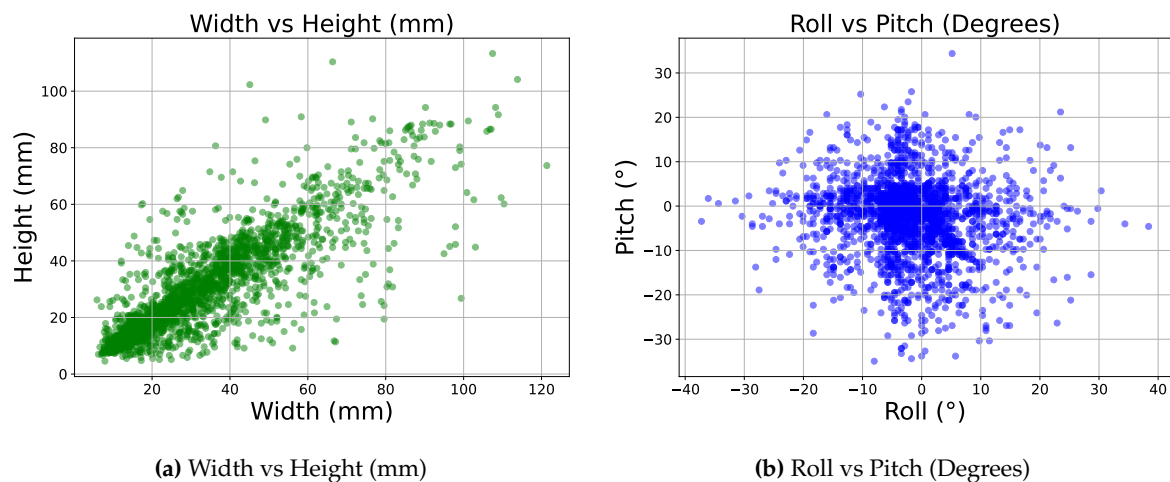


Figure 6. Subfigure (a) shows the distribution of width vs height in our 3D labels in millimeters, and subfigure (b) shows the distribution of Roll vs Pitch in degrees.

The yield estimation subset of our dataset includes time-lapse images of 8 different scenes. For each scene, the images start at the stage where the mushrooms have not appeared yet, until the time when there are no mushrooms left in the scene, taking images every 15 minutes. These images were then labeled using a trained YOLO-V8 model and each mushroom instance was assigned a unique identifier throughout the entire growth period. Figure 7 illustrates a time-series plot where the x-axis represents time and the y-axis represents the mushroom size in millimeters; The average size for each of the 8 scenes is shown, each with a different color. It can be observed that the average time required for a mushroom to reach full maturity is approximately five days. Furthermore, the abrupt decline in the size curve observed on the fifth day can be attributed to the partial harvesting of larger mushrooms. This intervention exposed smaller mushrooms that were previously occluded, thereby altering the recorded size distribution. A sample of the labeled growth monitoring data from a single scene is illustrated in figure 8.

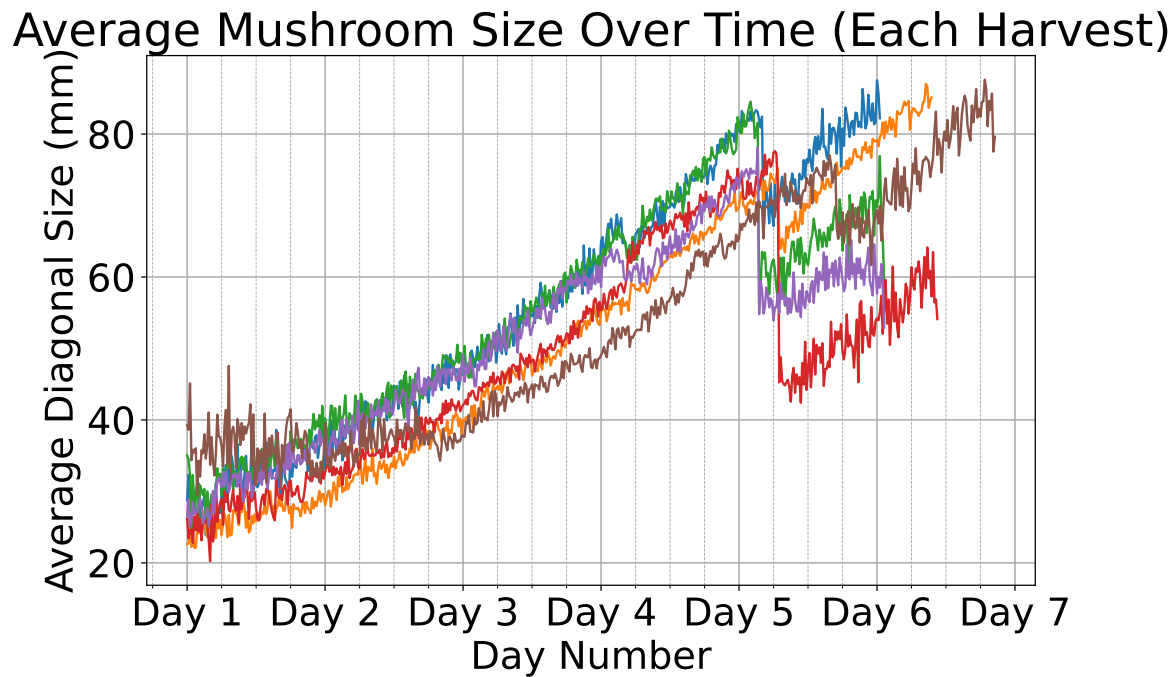


Figure 7. Average diagonal size (in mm) of mushrooms over time across multiple harvest cycles. Each different scene is plotted using a different color.

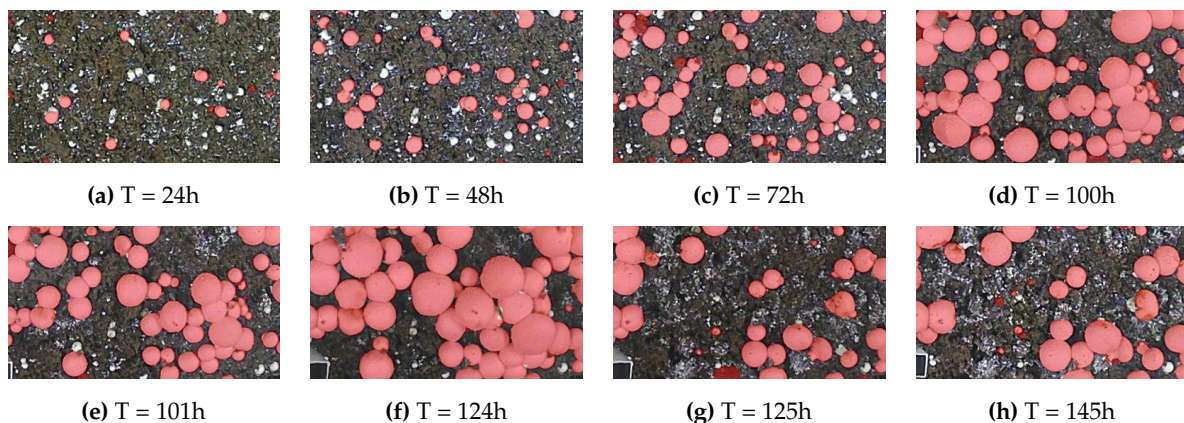


Figure 8. Sample of labeled growth monitoring images from one of the 8 time-lapse scenes. Partial harvest of larger mushrooms is visible at hour 101 and hour 125.

5. Benchmarking and Results

In this section, we provide a detailed analysis of several object detection and instance segmentation algorithms tested on our mushroom dataset, along with the results of the pose-estimation pipeline and analysis of the yield-estimation data. The models were trained on a DGX-2 cluster with sixteen NVIDIA V100 GPUs (32 GB each), providing ample compute for large-scale experiments.

For both detection and 3D pose-estimation tasks, the data were split **70 %/10 %/20 %** into train, validation, and *unseen* test sets, respectively; all metrics reported below refer exclusively to the test split to ensure unbiased generalisation estimates and guard against overfitting.

Every network (detection, segmentation, and pose) was optimised with AdamW, an initial learning rate of 1×10^{-4} , and a cosine-annealing scheduler that decays the learning rate to zero over the full training horizon.

- **Detection / segmentation.** Instead of explicit weight regularisation, we apply online data augmentation—random crops (up to 20 % of area), in-plane rotations of $\pm 15^\circ$, and random translations of $\pm 10\%$ along both axes—to diversify the training distribution and reduce overfitting.

- **3D pose estimation.** Following the protocol in [10], each mushroom instance's point cloud is randomly down-sampled to 1,024 points at every iteration; this stochastic subsampling encourages robustness to point-density variation and further mitigates overfitting.

5.1. Object Detection and Instance Segmentation

All object detection and instance segmentation models underwent thorough evaluations, with their performance presented in Table 2. Our assessment utilized metrics such as F1 score, average precision (AP), and average recall (AR) for all models regardless of the task being detection or instance segmentation.

Table 2. Performance of different object detection and instance segmentation models on our dataset.

Model Name	Backbone	Params(M)	Channels	Task	Image Size	F1	AP	AR
FasterRCNN [29]	efficientnet_b1	86.6	3	Det	1280	0.824	0.807	0.842
	MobileNet-V3 [30]	18.9	3	Det	1280	0.825	0.808	0.843
	ResNet-50 [31]	43.3	3	Det	1280	0.894	0.880	0.909
MaskRCNN [32]	EfficientNet-B1 [33]	91.6	3	Det	1280	0.819	0.802	0.836
				Seg	1280	0.812	0.797	0.828
	MobileNet-V3	64.8	3	Det	1280	0.816	0.800	0.834
				Seg	1280	0.813	0.798	0.828
	ResNet-50	45.9	3	Det	1280	0.891	0.876	0.905
				Seg	1280	0.881	0.866	0.896
	ResNet-50	171	4	Det	1280	0.863	0.846	0.881
				Seg	1280	0.861	0.846	0.876
RT-DETR-L [34]	ResNet-50	31.9	3	Det	640	0.245	0.210	0.295
					1280	0.907	0.891	0.923
RT-DETR-X	ResNet-50	65.4	3	Det	640	0.120	0.099	0.153
					1280	0.906	0.891	0.922
YOLOV8-L	CSPDarknet [35]	43.7	3	Det	640	0.735	0.693	0.782
					1280	0.880	0.869	0.892
				Seg	640	0.700	0.659	0.745
					1280	0.857	0.846	0.868
YOLOV8-M	CSPDarknet	25.9	3	Det	640	0.732	0.690	0.779
					1280	0.868	0.857	0.880
				Seg	640	0.711	0.669	0.760
					1280	0.844	0.833	0.856
YOLOV8-N	CSPDarknet	3.2	3	Det	640	0.775	0.739	0.815
					1280	0.887	0.873	0.901
				Seg	640	0.748	0.712	0.788
					1280	0.859	0.846	0.874
YOLOV8-S	CSPDarknet	11.2	3	Det	640	0.775	0.743	0.809
					1280	0.880	0.868	0.892
				Seg	640	0.753	0.720	0.789
					1280	0.857	0.845	0.869
YOLOV8-X	CSPDarknet	68.3	3	Det	640	0.724	0.667	0.790
					1280	0.887	0.875	0.898
				Seg	640	0.695	0.638	0.763
					1280	0.858	0.847	0.869
YOLOV9-C [36]	GELAN-C	25.5	3	Det	640	0.760	0.726	0.797
					1280	0.884	0.872	0.897
YOLOV9-E	GELAN-E	58.1	3	Det	640	0.736	0.702	0.774
					1280	0.877	0.867	0.887

The data in Table 2 provides details of the performance of a variety of models with distinct backbone architectures, benchmarked on object detection and instance segmentation tasks. These models were trained on image resolutions of 640 * 360 and 1280 * 720. The MaskRCNN with a ResNet-50 backbone surpassed the other networks in instance segmentation, achieving a mean Average Precision (mAP) of 0.866 and an Average Recall (AR) of 0.896 when processing 1280-sized RGB images. When adapted for RGB-D input, which increased the network's parameters over three times the original, there was no enhancement in performance. This lack of improvement may stem from the mandatory re-initialization of weights following the modifications to accommodate the added depth channel, which precluded the use of existing pre-trained weights that typically contribute to superior performance via fine-tuning on a specific dataset. Furthermore, the expanded parameter count makes the network less suitable for real-time use where hardware resources are constrained. Due to the

lack of improvement in performance while adding the depth channel, other networks were not tested with RGB-D images. Furthermore, The RT-DETR Large model exhibited superior object detection performance on 1280-sized images, with a mAP of 0.907 and an AR of 0.923. In contrast, both RT-DETR variants performed poorly on 640-sized images, potentially due to the network not being as effective at capturing the necessary details at a reduced resolution given the small number of dataset images. Figure 9 provides a visual comparison of the prediction results between BB and WB images, generated by Mask R-CNN with a ResNet-50 backbone trained on RGB images of size 1280. Fig. 9(a) and Fig. 9(b) show the prediction results on RGB images for BB (Baby Bella) and WB (White Buttons) mushroom instances, respectively.

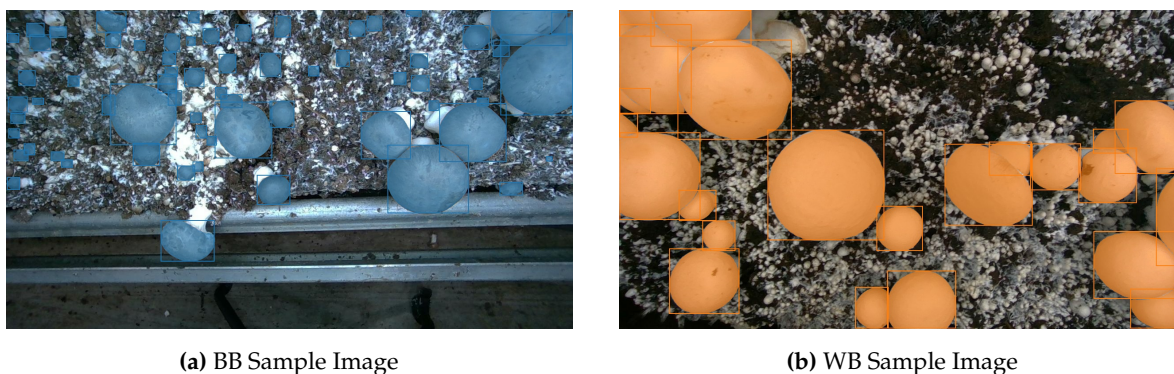


Figure 9. Prediction results of MaskRCNN-ResNet-50 architecture on (a) baby bella images, and (b) white button mushroom images

Overall, the results indicate that while all models performed well across various metrics, the choice of the backbone and the balance between the number of parameters and input channels play a significant role in the trade-off between accuracy and computational efficiency. These insights can inform the development and optimization of mushroom detection algorithms tailored for agricultural automation.

5.2. 3D Pose Estimation

To assure the usability of our 3D data, we trained and evaluated a 3D pose estimation pipeline on the bounding boxes and rotation annotations. Our 3D pose pipeline (based on our previously published approach in [10]) predicts roll, pitch, and yaw angles for each mushroom instance. To measure the accuracy of these predictions, we used a mean $\theta_{difference}^{\circ}$ metric, which averages the absolute degree differences between predicted and ground-truth angles across all mushrooms. Based on the findings of our previous study, we down-sampled each mushroom instance's point cloud to include 1,024 points and used Geodesic loss function [37] along with a 6D Gram-Schmidt rotation representation [28] and achieved a mean $\theta_{difference}^{\circ} = 6.61^{\circ}$ demonstrating the usability of our data and 3D annotations in tasks such as pose estimation. Figure 10 provides a visual comparison of ground-truth versus predicted oriented bounding boxes on a random scene. The predictions closely match the ground-truth orientations, suggesting that the model successfully captures the poses of mushrooms in real-world scenes.

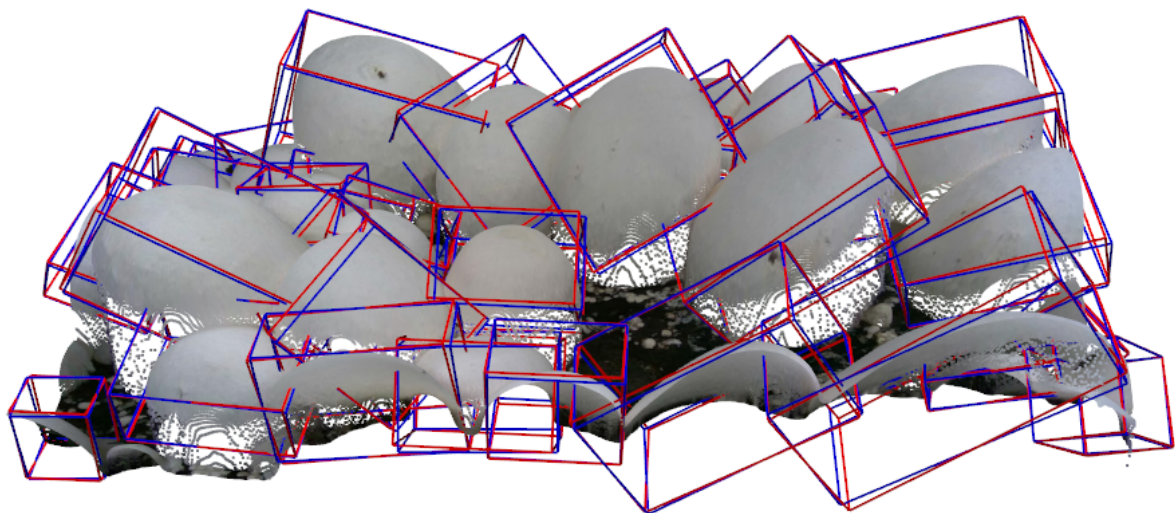


Figure 10. Visualization of ground truth (blue) vs. predicted (red) oriented bounding boxes on point clouds.

5.3. Yield Estimation

To facilitate the development of yield estimation models, we provide a curated subset of 3,838 images encompassing eight distinct scenes, spanning the full growth cycle of mushrooms—from initial emergence from the soil to full maturity and harvest. Figure 11 presents the average mushroom size in millimeters over time, where the solid line represents the mean size and the shaded region denotes the minimum and maximum size bounds. We can observe an abrupt drop in the sizes approximately at day five due to the partial harvesting of larger mushrooms, which enables the appearance of smaller, previously occluded mushrooms. Owing to the large volume of collected data, time series models can be effectively employed to accurately predict future mushroom growth and to estimate the optimal harvesting time.

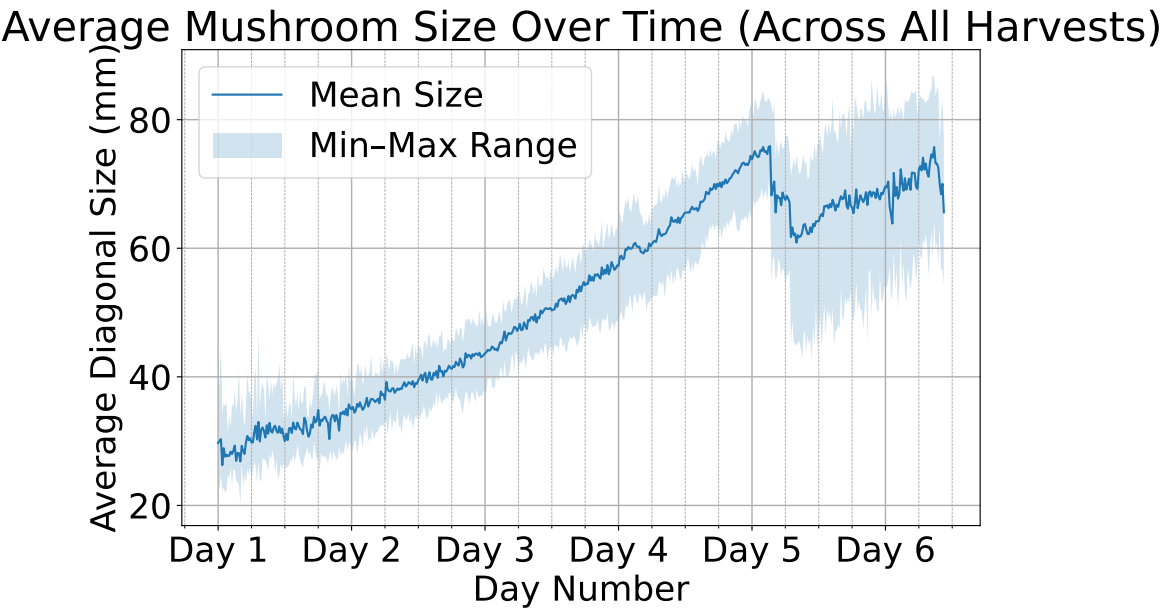


Figure 11. Average mushroom size over time across all harvests, with the solid line indicating the mean size and the shaded region representing the minimum–maximum range.

6. Conclusion

In this work, we have presented a dedicated mushroom detection dataset tailored to train and evaluate object detection and instance segmentation methods for various purposes that require localization of mushrooms in images such as automated mushroom harvesting, mushroom quality assessment,

and disease detection. Through comprehensive data collection and meticulous labeling processes, the dataset encompasses a wide spectrum of scenarios and challenges inherent in real-world mushroom growth environments including clusters of mushrooms with a variety of densities, images with different lighting conditions, occasional occlusion of mushroom instances by soil particles, and mushrooms growing at the edge of cultivation beds. The diverse nature of the dataset, including instances of both white button and baby bella mushrooms, allows for the robust training and benchmarking of detection and instance segmentation algorithms.

Beyond 2D annotations, our dataset also provides 3D bounding box and rotation information for the main subset of mushrooms, enabling pose estimation tasks. Additionally, we introduced a separate subset dedicated to yield estimation, capturing the same scenes over time to facilitate growth forecasting and harvesting schedule optimization.

Our evaluation of 9 different object detection and instance segmentation models highlights their respective strengths and weaknesses, shedding light on their suitability for mushroom detection and instance segmentation tasks. Furthermore, our exploration of color spaces provides insights into potential improvements in classification accuracy. By sharing this dataset and evaluation results, we hope to catalyze research and innovation in the domain of smart edible mushroom farming, fostering collaborations and accelerating progress in detection technologies. We anticipate that the integration of 3D pose estimation and yield forecasting components will further expand the dataset's utility, supporting a broader range of applications in automated mushroom cultivation. The availability of this benchmark dataset, our code, and trained models contributes to the development of advanced mushroom detection techniques, pose estimation methods, and yield prediction models, paving the way for increased efficiency and sustainability in automated mushroom harvesting systems.

Author Contributions: Conceptualization, A.Z., M.F., D.B. and F.A.M.; methodology, A.Z., M.F.; software, A.Z., M.F., J.K.; validation, A.Z., M.F., B.K. and J.K.; formal analysis, A.Z., M.F., F.A.M. and D.B.; investigation, A.Z.; resources, W.Z., V.B. and F.A.M.; data curation, A.Z., M.F.; writing—original draft preparation, A.Z.; writing—review and editing, M.F., B.K., J.K., V.B., W.Z., D.B. and F.A.M.; visualization, A.Z.; supervision, F.A.M., D.B., V.B. and W.Z.; project administration, F.A.M. and W.Z.; funding acquisition, W.Z.; All authors have read and agreed to the published version of the manuscript.

Funding: The work was partially supported by the United States Department of Agriculture grants #2021-67022-34889, 2022-67022-37867, and 2023-51300-40853, as well as the University of Houston Infrastructure Grant.

Data Availability Statement: The dataset and code supporting the findings of this study are publicly available at our GitHub repository: <https://github.com/abdollahzakeri/m18k>

Acknowledgments: We would like to acknowledge Kenneth Wood, Armando Juarez, and Bruce Knobloch from Monterey Mushroom Inc. for allowing us to visit and obtain the necessary information from the mushroom farm in Madisonville, TX, USA.

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

References

1. Károlyi, A.I.; Galambos, P. Automated Dataset Generation with Blender for Deep Learning-based Object Segmentation. In Proceedings of the 2022 IEEE 20th Jubilee World Symposium on Applied Machine Intelligence and Informatics (SAMI), 2022, pp. 000329–000334. <https://doi.org/10.1109/SAMI54271.2022.9780790>.
2. Baisa, N.L.; Al-Diri, B. Mushrooms Detection, Localization and 3D Pose Estimation using RGB-D Sensor for Robotic-picking Applications. *ArXiv* **2022**.
3. Arjun, A.D.; Chakraborty, S.K.; Mahanti, N.K.; Kotwaliwale, N. Non-destructive assessment of quality parameters of white button mushrooms (*Agaricus bisporus*) using image processing techniques. *Journal of Food Science and Technology* **2022**, *59*, 2047–2059. <https://doi.org/10.1007/s13197-021-05219-w>.
4. Nguyen, V.; Ho, T.A.; Vu, D.A.; Anh, N.T.N.; Thang, T.N. Building Footprint Extraction in Dense Areas using Super Resolution and Frame Field Learning. In Proceedings of the 2023 12th International Conference

- on Awareness Science and Technology (iCAST), 2023, pp. 112–117. <https://doi.org/10.1109/iCAST57874.2023.10359290>.
5. Anagnostopoulou, D.; Retsinas, G.; Efthymiou, N.; Filntisis, P.; Maragos, P. A Realistic Synthetic Mushroom Scenes Dataset. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2023, pp. 6282–6289.
 6. Koirala, B.; Shen, G.; Nguyen, H.C.; Kang, J.; Zakeri, A.; Balan, V.; Merchant, F.; Benhaddou, D.; Zhu, W. Development of a Compact Hybrid Gripper for Automated Harvesting of White Button Mushroom. In Proceedings of the International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, 08 2024, Vol. Volume 7: 48th Mechanisms and Robotics Conference (MR), p. V007T07A036. <https://doi.org/10.1115/DETC2024-143056>.
 7. Koirala, B.; Kafle, A.; Nguyen, H.C.; Kang, J.; Zakeri, A.; Balan, V.; Merchant, F.; Benhaddou, D.; Zhu, W. A Hybrid Three-Finger Gripper for Automated Harvesting of Button Mushrooms. *Actuators* **2024**, *13*, 287. Number: 8 Publisher: Multidisciplinary Digital Publishing Institute, <https://doi.org/10.3390/act13080287>.
 8. Koirala, B.; Zakeri, A.; Kang, J.; Kafle, A.; Balan, V.; Merchant, F.A.; Benhaddou, D.; Zhu, W. Robotic Button Mushroom Harvesting Systems: A Review of Design, Mechanism, and Future Directions. *Applied Sciences* **2024**, *14*, 9229. Number: 20 Publisher: Multidisciplinary Digital Publishing Institute, <https://doi.org/10.3390/app14209229>.
 9. Benhaddou, D.; Balan, V.; Garza, A.D.L.; Merchant, F. Estimating Mushroom Yield and Quality Using Computer Vision. In Proceedings of the 2023 International Wireless Communications and Mobile Computing (IWCMC), 2023, pp. 562–567. <https://doi.org/10.1109/IWCMC58020.2023.10182999>.
 10. Zakeri, A.; Koirala, B.; Kang, J.; Balan, V.; Zhu, W.; Benhaddou, D.; Merchant, F.A. SMS3D: 3D Synthetic Mushroom Scenes Dataset for 3D Object Detection and Pose Estimation. *Computers* **2025**, *14*, 128. Number: 4 Publisher: Multidisciplinary Digital Publishing Institute, <https://doi.org/10.3390/computers14040128>.
 11. Jareanpon, C.; Khummanee, S.; Sriputta, P.; Scully, P. Developing an Intelligent Farm System to Automate Real-time Detection of Fungal Diseases in Mushrooms. *CURRENT APPLIED SCIENCE AND TECHNOLOGY* **2024**, pp. e0255708–e0255708. <https://doi.org/10.55003/cast.2023.255708>.
 12. Lee, C.H.; Choi, D.; Pecchia, J.; He, L.; Heinemann, P. Development of A Mushroom Harvesting Assistance System using Computer Vision.
 13. Moysiadis, V.; Kokkonis, G.; Bibi, S.; Moscholios, I.; Maropoulos, N.; Sarigiannidis, P. Monitoring Mushroom Growth with Machine Learning. *Agriculture* **2023**, *13*, 223. Number: 1 Publisher: Multidisciplinary Digital Publishing Institute, <https://doi.org/10.3390/agriculture13010223>.
 14. Nadim, M.; School of Engineering, Deylaman Institute for High Education, Lahijan, Iran.; Ahmadifar, H.; Department of Computer Engineering. University of Guilan. Rasht, Iran.; Mashkinmojeh, M.; School of Engineering, Deylaman Institute for High Education, Lahijan, Iran.; Yamaghani, M.R.; Department of Computer Engineering***** Lahijan Azad University, Lahijan, Iran. Application of Image Processing Techniques for Quality Control of Mushroom. *Caspian Journal of Health Research* **2019**, *4*, 72–75. <https://doi.org/10.29252/cjhr.4.3.72>.
 15. Olpin, A.J.; Dara, R.; Stacey, D.; Kashkoush, M. Region-Based Convolutional Networks for End-to-End Detection of Agricultural Mushrooms. In Proceedings of the Image and Signal Processing; Mansouri, A.; El Moataz, A.; Nouboud, F.; Mammass, D., Eds., Cham, 2018; Lecture Notes in Computer Science, pp. 319–328. https://doi.org/10.1007/978-3-319-94211-7_35.
 16. Retsinas, G.; Efthymiou, N.; Anagnostopoulou, D.; Maragos, P. Mushroom Detection and Three Dimensional Pose Estimation from Multi-View Point Clouds. *Sensors (Basel, Switzerland)* **2023**, *23*, 3576. <https://doi.org/10.3390/s23073576>.
 17. Viézhányó, T.; Felföldi, J. Enhancing colour differences in images of diseased mushrooms. *Computers and Electronics in Agriculture* **2000**, *26*, 187–198. [https://doi.org/10.1016/S0168-1699\(00\)00071-5](https://doi.org/10.1016/S0168-1699(00)00071-5).
 18. Wang, F.; Zheng, J.; Tian, X.; Wang, J.; Niu, L.; Feng, W. An automatic sorting system for fresh white button mushrooms based on image processing. *Computers and Electronics in Agriculture* **2018**, *151*, 416–425. <https://doi.org/10.1016/j.compag.2018.06.022>.
 19. Wang, Y.; Yang, L.; Chen, H.; Hussain, A.; Ma, C.; Al-gabri, M. Mushroom-YOLO: A deep learning algorithm for mushroom growth recognition based on improved YOLOv5 in agriculture 4.0. In Proceedings of the 2022 IEEE 20th International Conference on Industrial Informatics (INDIN), Jul 2022, pp. 239–244. <https://doi.org/10.1109/INDIN51773.2022.9976155>.

20. Wei, B.; Zhang, Y.; Pu, Y.; Sun, Y.; Zhang, S.; Lin, H.; Zeng, C.; Zhao, Y.; Wang, K.; Chen, Z. Recursive-YOLOv5 Network for Edible Mushroom Detection in Scenes With Vertical Stick Placement. *IEEE Access* **2022**, *10*, 40093–40108. Conference Name: IEEE Access, <https://doi.org/10.1109/ACCESS.2022.3165160>.
21. Yang, S.; Huang, J.; Yu, X.; Yu, T. Research on a Segmentation and Location Algorithm Based on Mask RCNN for Agaricus Bisporus. In Proceedings of the 2022 2nd International Conference on Computer Science, Electronic Information Engineering and Intelligent Control Technology (CEI), Sep 2022, pp. 717–721. <https://doi.org/10.1109/CEI57409.2022.9950157>.
22. Zahan, N.; Hasan, M.Z.; Uddin, M.S.; Hossain, S.; Islam, S.F. Chapter 10 - A deep learning-based approach for mushroom diseases classification. In *Application of Machine Learning in Agriculture*; Khan, M.A.; Khan, R.; Ansari, M.A., Eds.; Academic Press, 2022; pp. 191–212. <https://doi.org/10.1016/B978-0-323-90550-3.00005-9>.
23. Lu, C.P.; Liaw, J.J.; Wu, T.C.; Hung, T.F. Development of a Mushroom Growth Measurement System Applying Deep Learning for Image Recognition. *Agronomy* **2019**, *9*, 32. Number: 1 Publisher: Multidisciplinary Digital Publishing Institute, <https://doi.org/10.3390/agronomy9010032>.
24. Lu, C.P.; Liaw, J.J. A novel image measurement algorithm for common mushroom caps based on convolutional neural network. *Computers and Electronics in Agriculture* **2020**, *171*, 105336. <https://doi.org/10.1016/j.compag.2020.105336>.
25. Mirza, S.; Nguyen, V.D.; Mantini, P.; Shah, S.K. Data Quality Aware Approaches for Addressing Model Drift of Semantic Segmentation Models. In Proceedings of the VISIGRAPP (3: VISAPP), 2024, pp. 333–341.
26. Yin, H.; Yi, W.; Hu, D. Computer vision and machine learning applied in the mushroom industry: A critical review. *Computers and Electronics in Agriculture* **2022**, *198*, 107015. <https://doi.org/10.1016/j.compag.2022.107015>.
27. Kirillov, A.; Mintun, E.; Ravi, N.; Mao, H.; Rolland, C.; Gustafson, L.; Xiao, T.; Whitehead, S.; Berg, A.C.; Lo, W.Y.; et al. Segment Anything, 2023, [[arXiv:cs.CV/2304.02643](https://arxiv.org/abs/2304.02643)].
28. Zhou, Y.; Barnes, C.; Lu, J.; Yang, J.; Li, H. On the Continuity of Rotation Representations in Neural Networks. In Proceedings of the 2019 Conference on Computer Vision and Pattern Recognition (CVPR), 2019.
29. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In Proceedings of the Advances in Neural Information Processing Systems; Cortes, C.; Lawrence, N.; Lee, D.; Sugiyama, M.; Garnett, R., Eds. Curran Associates, Inc., 2015, Vol. 28.
30. Howard, A.; Sandler, M.; Chu, G.; Chen, L.C.; Chen, B.; Tan, M.; Wang, W.; Zhu, Y.; Pang, R.; Vasudevan, V.; et al. Searching for MobileNetV3. In Proceedings of the Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), October 2019.
31. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2016.
32. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. In Proceedings of the Proceedings of the IEEE International Conference on Computer Vision (ICCV), Oct 2017.
33. Tan, M.; Le, Q.V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks, 2020, [[arXiv:cs.LG/1905.11946](https://arxiv.org/abs/1905.11946)].
34. Lv, W.; Xu, S.; Zhao, Y.; Wang, G.; Wei, J.; Cui, C.; Du, Y.; Dang, Q.; Liu, Y. DETRs Beat YOLOs on Real-time Object Detection, 2023, [[arXiv:cs.CV/2304.08069](https://arxiv.org/abs/2304.08069)].
35. Wang, C.Y.; Liao, H.Y.M.; Wu, Y.H.; Chen, P.Y.; Hsieh, J.W.; Yeh, I.H. CSPNet: A New Backbone That Can Enhance Learning Capability of CNN. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, June 2020.
36. Wang, C.Y.; Yeh, I.H.; Liao, H.Y.M. YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information, 2024, [[arXiv:cs.CV/2402.13616](https://arxiv.org/abs/2402.13616)].
37. Huynh, D.Q. Metrics for 3D Rotations: Comparison and Analysis. *Journal of Mathematical Imaging and Vision* **2009**, *35*, 155–164. <https://doi.org/10.1007/s10851-009-0161-2>.

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.