agriculture

MDPI

Article

A Multi-Phenotype Acquisition System for *Pleurotus eryngii* Based on RGB and Depth Imaging

Yueyue Cai [1,2,†], Zhijun Wang [1,3,†], Ziqin Liao [1], Yujie Li [1], Weijie Shi [1], Peijie Huang [1], Bingzhi Chen [4], Jie Pang [2], Xiangzeng Kong [3,*] and Xuan Wei [1,2,3*]

[1] College of Mechanical and Electrical Engineering, Fujian Agriculture and Forestry University, Fuzhou 350002, China
[2] College of Food Science, Fujian Agriculture and Forestry University, Fuzhou 350002, China
[3] Institute of Artificial Intelligence in Agriculture and Forestry, Fujian Agriculture and Forestry University, Fuzhou 350002, China
[4] College of Life Science, Fujian Agriculture and Forestry University, Fuzhou 350002, China
* Correspondence: xzkong@fafu.edu.cn (X.K.); xuanwei@fafu.edu.cn (X.W.)
† These authors contributed equally to this work.

Abstract

High-throughput phenotypic acquisition and analysis allow us to accurately quantify trait expressions, which is essential for developing intelligent breeding strategies. However, there is still much potential to explore in the field of high-throughput phenotyping for edible fungi. In this study, we developed a portable multi-phenotypic acquisition system for *Pleurotus eryngii* using RGB and RGB-D cameras. We developed an innovative Unet-based semantic segmentation model by integrating the ASPP structure with the VGG16 architecture. This allows for precise segmentation of the cap, gills and stem of the fruiting body. By leveraging depth images from RGB-D cameras, we can effectively collect phenotypic information about *Pleurotus eryngii*. By combining K-means clustering with Lab color space thresholds, we are able to achieve more precise automatic classification of *Pleurotus eryngii* cap colors. Moreover, AlexNet is utilized to classify the shapes of the fruiting bodies. The Aspp-VGGUnet network demonstrates remarkable performance with a mean Intersection over Union (mIoU) of 96.47% and a mean pixel accuracy (mPA) of 98.53%. These results reflect respective improvements of 3.03% and 2.23% compared to the standard Unet model, respectively. The average error in size phenotype measurement is just 0.15 ± 0.03 cm. The accuracy for cap color classification reaches 91.04%, while fruiting body shape classification achieves 97.90%. The proposed multi-phenotype acquisition system reduces the measurement time per sample from an average of 76 s (manual method) to about 2 s, substantially increasing data acquisition throughput and providing robust support for scalable phenotyping workflows in breeding research.

Keywords: *Pleurotus eryngii*; assisted breeding; phenotypic analysis; deep learning; depth camera

Academic Editor: Maciej Zaborowicz

Received: 28 October 2025
Revised: 5 December 2025
Accepted: 8 December 2025
Published: 11 December 2025

Citation: Cai, Y.; Wang, Z.; Liao, Z.; Li, Y.; Shi, W.; Huang, P.; Chen, B.; Pang, J.; Kong, X.; Wei, X. A Multi-Phenotype Acquisition System for *Pleurotus eryngii* Based on RGB and Depth Imaging. *Agriculture* **2025**, *15*, 2566. https://doi.org/10.3390/agriculture15242566

1. Introduction

Pleurotus eryngii is highly valued not only for its rich nutritional content but also for its health-promoting properties, such as its hypotensive and hypolipidemic effects (Zhang et al., 2020) [1]. It is characterized by a thick flesh texture, a unique almond-like aroma and flavor profiles reminiscent of abalone. Its high oligosaccharide content supports

prebiotic activity through beneficial interactions with *Bifidobacterium*, offering potential dual advantages for gastrointestinal well-being and cosmetic applications (Wang et al., 2024) [2]. In modern cultivation techniques, accurately measuring phenotypic traits like stipe length, pileus diameter, color characteristics and morphological features is essential for breeding programs. However, traditional manual measurement methods can be inefficient and inconsistent due to the risk of human error and subjective judgment.

Thanks to the rapid advancement of intelligent sensing, computer vision and deep learning technologies, agricultural phenotypic analysis is gradually moving toward automation, high-throughput processing and improved precision. In response to the increasing need for comprehensive data acquisition, several automated platforms have been developed, such as track-based systems (Li et al., 2023) [3], pipeline and imaging chamber systems (Dengyu et al., 2016) [4] and mobile platforms (Wang et al., 2023) [5]. Zhou et al. (2021) [6] created a portable and affordable plant phenotyping system that incorporates image sensors to assess soybean salt tolerance, thus supporting breeding programs. Zhu et al. (2023) [7] developed a mushroom measurement system using resistance strain gauges to accurately determine position, size and weight. An et al. (2016) [8] created a high-throughput phenotyping system using dual digital cameras to measure leaf length and rosette leaf area. These crop-specific systems greatly boost the efficiency of data collection, minimize human error and improve measurement accuracy.

Sensors like RGB imaging, hyperspectral imaging, RGB-D sensor and LiDAR are commonly used in plant phenotyping platforms to measure and analyze crop traits. These sensors play a valuable role in monitoring crop growth, conducting trait analysis and enabling precise breeding (Dengyu et al., 2016 [4]; Li et al., 2023 [3]; Wang Y. et al., 2023 [5]). RGB-D camera and LiDAR are used to capture spatial information such as crop structure, volume and biomass (Raj et al., 2021 [9]; Rong et al., 2023) [10]. Compared to LiDAR, RGB-D sensor have lower hardware and computational costs (Ravaglia et al., 2019) [11], which makes them a great choice for medium and small-scale phenotypic acquisition and automated systems. Baisa and Al-Diri (2022) [12] investigated the potential of using affordable consumer-grade RGB-D sensor to create an innovative algorithm. This algorithm focuses on detecting, locating and estimating the 3D pose of mushrooms to enable efficient robotic picking capabilities. Zhou et al. (2024) [13] introduced a method to align RGB and depth images and convert 2D coordinates into 3D coordinates via reverse prediction, allowing for precise measurement of corn stem diameters. These studies enhance RGB images through the introduction of depth information, reducing the requirement for external references (such as checkerboards or calibration objects) and thereby enhancing the accuracy of measurement and the practicality of the system.

In phenotypic analysis research, image segmentation is a crucial step in extracting morphological parameters (Narisetti et al., 2022) [14], particularly when it comes to tasks such as size measurement and area calculation, where precise semantic segmentation becomes very important for achieving accurate results. The segmentation results can be used to determine the contour area and length-width dimensions of crops through geometric calculation methods (Raj et al., 2021) [9]. For instance, Yang et al. (2022) [15] introduced a segmentation and localization method for *Agaricus bisporus* using Mask R-CNN, applying elliptical fitting techniques to determine the centroid position and stem size of the mushrooms. Meanwhile, Kolhar and Jagtap (2023) [16] adopted the DeepLabV3 Plus to perform contour segmentation of Komatsuna leaves, enabling the estimation of leaf count, projected area and growth stages. The segmented regions of interest not only help to reduce the influence of environmental and other factors but also allow us to integrate the color quantization and grading algorithm within the chromaticity space (Ahmad et al., 2021) [17] with the shape classification algorithm based on contour features (Vazquez et al., 2024)

[18]. The integration allows for the precise classification of various phenotypic parameters, such as color and shape.

Recent advancements have been made in non-destructive geometric measurement of *Pleurotus eryngii*. Luo et al. (2024) [19] used RGB-D imaging to measure *Pleurotus eryngii* volume, filling an early gap in mushroom phenotyping and greatly reducing detection time. However, their method focuses solely on volume and does not consider gill-related phenotypic information. Similarly, Yin et al. (2025) [20] used mobile multi-view imaging combined with structured light to reconstruct the 3D geometry of *Pleurotus eryngii*, enabling detailed modeling of surface features. Nevertheless, their method still relies on external structured-light modules, involves a relatively complex and time-consuming acquisition process, and is not well suited for high-throughput, and real-time phenotyping in breeding or production environments.

One of the key challenges in acquiring multiple phenotypes of *Pleurotus eryngii* is to develop an efficient, portable and precise method for phenotype collection and analysis. To this end, the main contributions of this study encompass the following three aspects:

(1) We introduce a portable multi-phenotype acquisition system for *Pleurotus eryngii* using RGB and RGB-D cameras. This system addresses common inefficiencies and errors in traditional manual measurements, providing a fast, real-time, and precise approach for phenotyping edible fungi.
(2) We created an enhanced semantic segmentation model that improves the accuracy of mask segmentation for *Pleurotus eryngii* fruiting bodies and enables the automatic extraction of multiple phenotypic parameters.
(3) A set of algorithms was developed to efficiently measure the size, color and shape of *Pleurotus eryngii*, offering strong support for intelligent breeding analysis of edible fungi.

2. Materials and Methods

2.1. Phenotypic Acquisition Device

Figure 1 presents the phenotypic acquisition device, which consists of a dark chamber integrated with a RealSense D405 (Intel Corporation, Santa Clara, CA, USA), an RGB camera (Shenzhen Ruier Weishi Technology Co., Ltd, Shenzhen, China), a pressure sensor (Shenzhen Xintai Microelectronics Technology Co., Ltd, Shenzhen, China), a temperature/humidity sensor (Guangzhou Aosong Electronics Co., Ltd, Guangzhou, China), a light source controller with a 6-watt white LED, and an STM32F103C8T6 microcontroller (STMicroelectronics, Geneva, Switzerland). The microcontroller collects data from the sensors via UART serial communication and transmits them to the application terminal for real-time processing and interaction.

2.2. Data Acquisition

The *Pleurotus eryngii* samples utilized in this study were sourced from Zhongyan Mushroom Industry Co., Ltd., located in Zhangzhou City, China Using the self-developed phenotypic acquisition platform, the samples were precisely positioned so that the caps were directed toward the RGB camera and the fruiting bodies aligned with the RGB-D camera at a working distance of 30 cm, as shown in Figure 1. In total, more then 300 samples were systematically collected for analysis. In this study, 607 images of the fruiting bodies were captured using the RGB-D camera, while 304 images of the caps were acquired with the RGB camera. To ensure data consistency, all images were uniformly resized to 640 × 460 pixels.

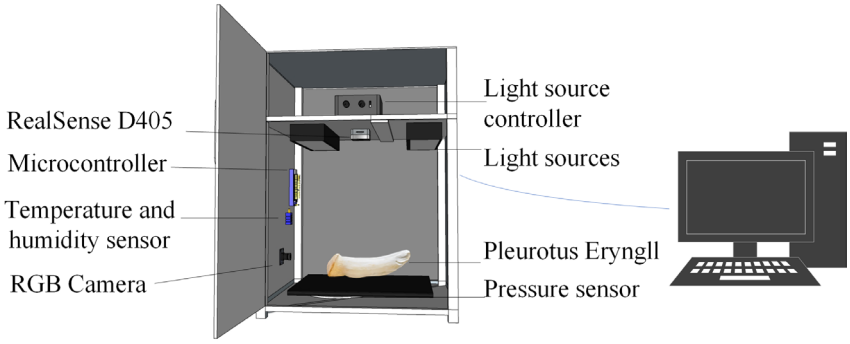

Figure 1. *Pleurotus eryngii* phenotype collection device.

In addition, key morphological parameters of each sample, including cap diameter and stipe length, were measured manually using a ruler. Cap color was quantitatively assessed with a standardized color chart. Based on manual observation and classification, the fruiting body morphology was categorized into three types: bowling-pin shaped (swollen on one side of the stipe), cylindrical (uniform width of cap, gills, and stipe), and standard type (broad cap with slender stipe). Similarly, the cap color was manually classified into three categories: yellowish-brown, dark brown, and light brown, based on reference color cards.

2.3. Data Preprocessing

In the image segmentation task, we assigned labels to the cap, gills and stipe of the *Pleurotus eryngii* fruiting body to help distinguish between different regions. Using Labelme (v4.5.11), each part of the fruiting body was manually annotated and organized into three categories: cap, gills, and stipe. The data were randomly divided into training, validation and test sets in 7:2:1 ratio.

To enhance model robustness and mitigate overfitting, we applied a variety of data augmentation techniques before training. Specifically, we randomly rotated and flipped images to introduce diversity, adjusted brightness, contrast and saturation to mimic real-world lighting conditions, scaled and cropped images to help the model adapt to different sizes and added random deformations to improve its ability to handle shape variations. The dataset was enriched by increasing the number of training images from 607 to 1136.

2.4. Model Structure and Improvement

2.4.1. U-Net Basic Network

Ronneberger et al. (2015) [21] proposed the U-Net architecture, which has demonstrated strong effectiveness in medical image segmentation, particularly under limited training data. As illustrated in Figure 2, the U-Net architecture primarily consists of an encoder, a series of skip connections, and a decoder. The encoder extracts hierarchical features through repeated convolution and downsampling operations, where the feature representations gradually transition from low-level spatial details to high-level semantic information. The skip connections transmit feature maps from the encoder to the corresponding decoder layers, enabling the model to retain fine-grained spatial information that may be lost during downsampling. This mechanism helps preserve localization accuracy and enhances the mode's ability to delineate structural boundaries. The decoder then progressively restores spatial resolution through upsampling and convolution, integrating encoder features via skip connections to reconstruct detailed segmentation maps with high accuracy.

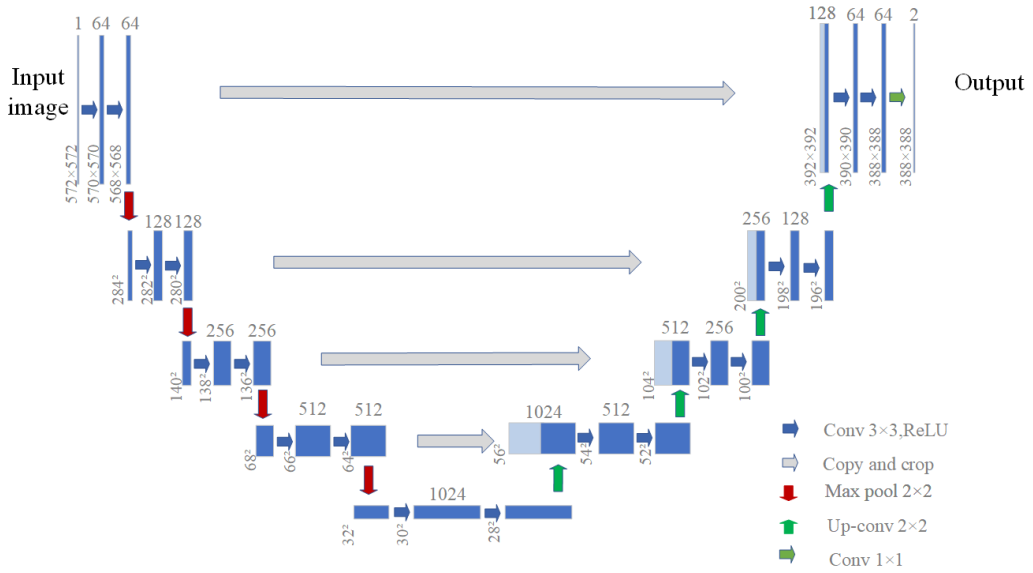

Figure 2. U-net network structure.

2.4.2. VGG16 Network Structure

Simonyan and Zisserman (2015) [22] proposed the VGG16 network, a classic deep convolutional neural network architecture and its network structure is depicted in Figure 3. VGG16 significantly improved model performance by increasing network depth and achieved outstanding results in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC), establishing itself as an important foundation in modern computer vision research.

VGG16 construct a deep network by layering multiple small convolutional kernels (3 × 3 kernels). Compared with earlier architectures relying on large convolutional filters, the use of small kernels allows the network to increase depth while keeping the number of parameters manageable. This thoughtful design enables VGG16 to achieve an excellent balance between depth and computational efficiency. By progressively stacking multiple small convolutional kernels, the network can effectively learn local details while gradually capturing broader image features through deeper-level combinations. This thoughtful design enables the network to excel in both precision and scope of feature extraction.

VGG16 also exhibits strong transfer-learning capability. It delivers competitive performance on ImageNet and performs well in downstream tasks such as image classification, object detection, and semantic segmentation. Thanks to its robust capabilities, it has become a cornerstone for computer vision applications. Researchers can efficiently develop domain-specific models by fine-tuning pre-trained VGG16. Its architecture is easy to work with and excels at feature extraction, which makes it as a favored choice in deep learning research. Although more advanced architectures such as ResNet and EfficientNet have emerged, VGG16 continues to serve as a reliable and influential benchmark due to its simplicity, robustness, and ease of implementation.

2.4.3. The Pyramid Pooling Module

The Atrous Spatial Pyramid Pooling (ASPP) module (He et al., 2014) [23] is widely employed in computer vision tasks, including image segmentation and object detection, to enhance a model's capacity for multi-scale feature extraction As shown in Figure 4, ASPP utilizes atrous convolutions with multiple dilation rates (e.g., 1, 2, 4, and 8) to capture both local details and global contextual information. Its primary advantage lies in expanding the receptive field while maintaining computational efficiency.

To further improve performance, the ASPP module integrates a global average pooling layer, which computes the mean of the entire feature map, providing global context and ensuring that essential semantic information is retained during feature fusion. Subsequently, multi-scale and global features are aggregated along the channel dimension to form a comprehensive feature representation, significantly enhancing the effectiveness of deep learning models in visual tasks. For example, its incorporation into the DeepLab series has demonstrated superior performance in image segmentation, particularly in scenes with complex backgrounds and objects of varying scales, improving both segmentation accuracy and recall.

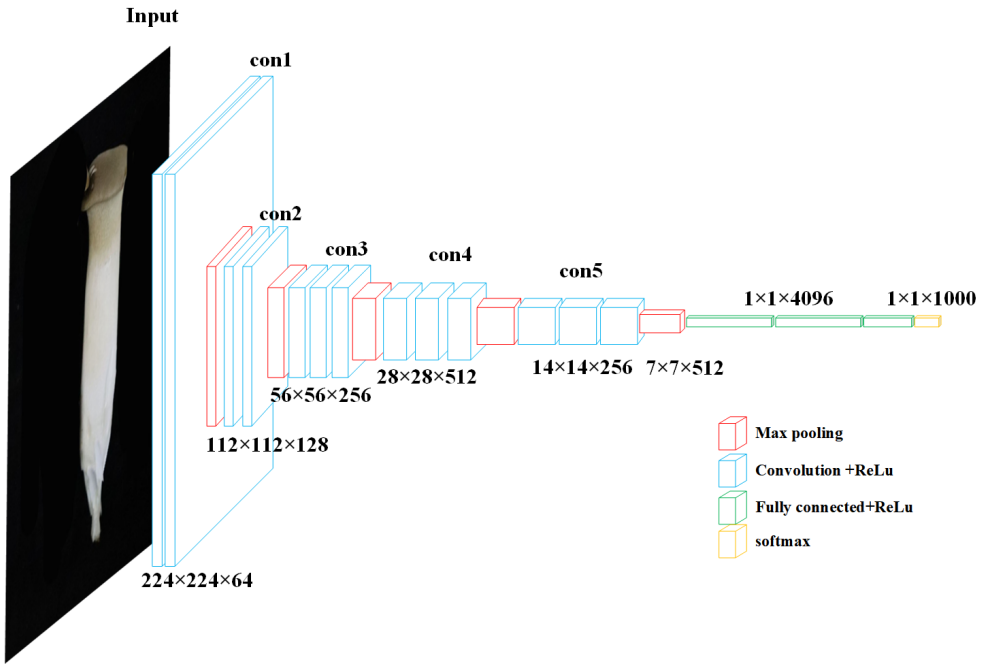

Figure 3. The VGG16 network structure.

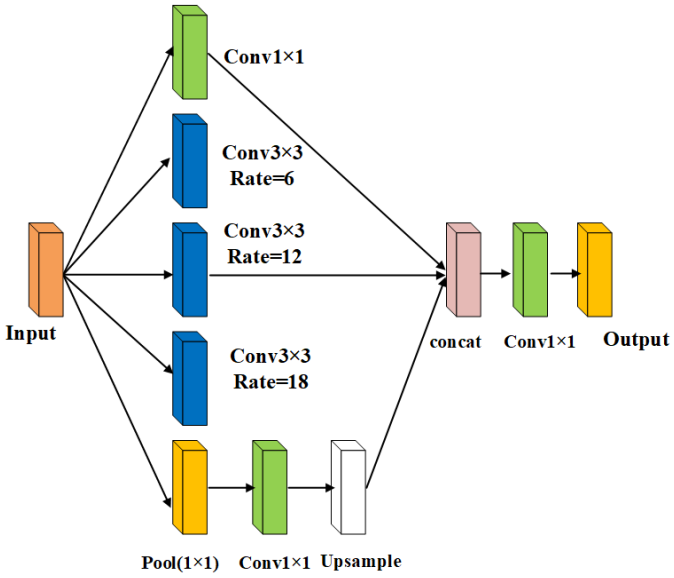

Figure 4. The pyramid pooling module.

2.4.4. Aspp-VGGUnet Network Structure

In this work, we enhanced the encoder of the U-net structure by replacing its original downsampling pathway with the first 13 convolutional layers of VGG16 and incorporating a pyramid structure into the Copy and Crop phase. The improved U-net network is shown in Figure 5, primarily consisting of an enhanced encoder feature extraction network and skip connections.

Specifically, the five convolutional blocks of VGG16 were adopted to substitute the "convolution + max pooling" structure in the standard U-net encoder. The feature maps output from each VGG16 block were used as multi-scale representations for the corresponding encoding stages. Leveraging the continuous small-kernel convolutions and pre-trained weights of VGG16 enables the encoder to extract richer low-level textures, structural patterns, and semantic information while reducing feature loss during downsampling.

Furthermore, the ASPP module was incorporated into the skip connections. For each skip connection, the feature map from the encoder is first processed by ASPP, where parallel atrous convolutions with different dilation rates generate multi-scale contextual responses. These enhanced features are then concatenated with the corresponding decoder features to facilitate more effective feature fusion. By expanding the receptive field without compromising spatial resolution, ASPP provides strong multi-scale context perception and significantly improves boundary delineation, localization accuracy, and the overall robustness of the segmentation network.

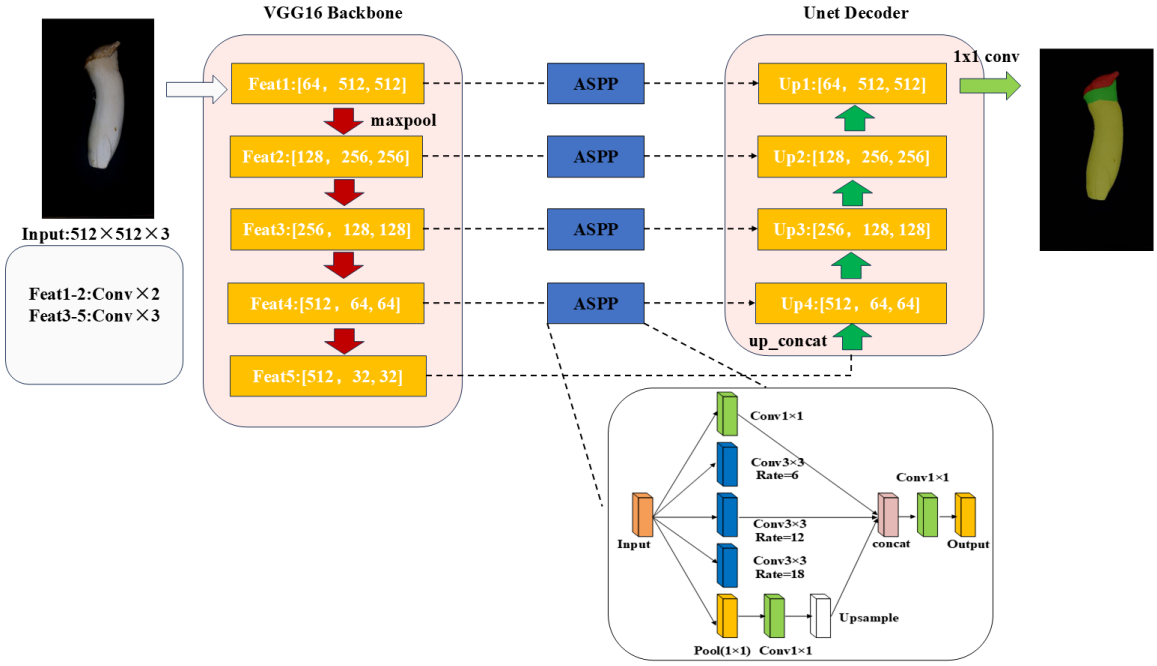

Figure 5. Aspp-VGGUnet network structure.

2.5. Phenotypic Measurement Algorithm

To evaluate the effectiveness of the method introduced in this work for measuring the phenotypic parameters of *Pleurotus eryngii*, we selected eight key characteristics as measurement indicators. These include the color of the cap, the maximum diameter of the cap, the thickness and length of the cap, the length and width of the gills, the length of the stem and the overall shape of the fruiting body.

2.5.1. Algorithm for Measuring Dimensional Phenotypes

This approach employs optical triangulation technology to calculate the spatial position of an object relative to the camera, thereby generating depth images in the process. To estimate the size information of fruiting bodies, caps, gills and stems, their regions of interest are typically extracted first. Subsequently, the length of their minimum bounding rectangles is measured to obtain phenotypic size data. For this study, the measurement method consists of three critical stages, The detailed steps of the process are presented in Figure 6:

(1) The regions of interest for the cap, gills and stem of the mushroom are identified using the Aspp-VggUnet semantic segmentation network.

(2) The mask block filters out irrelevant depth data to identify the ROI and preserve its depth values.

$$D_{ROI}(u,v) = M(u,v) \odot D(u,v) \quad (1)$$

Defined by Equation (1), (u,v) denotes the pixel coordinates (column and row indices) where M is the binary segmentation mask, D is the raw depth map, and $\odot$ denotes element-wise multiplication.

Since the object's placement direction is predefined, its 2D mask contour is rotated by a known angle θ before calculating the minimum bounding rectangle (MBR). This aligns the MBR's sides with the coordinate axes. The 2D rotation transformation is applied:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x - x_c \\ y - y_c \end{bmatrix} + \begin{bmatrix} x_c \\ y_c \end{bmatrix} \quad (2)$$

Defined by Equation (2) (x,y) are the original contour points, (x,y) are the rotated points, and is the centroid. This method standardizes orientation, ensuring measurement consistency.

This method standardizes the rectangle's orientation and ensures consistency with the coordinate system.

(3) The horizontal dimensions of the 3D object's projection are determined by mapping the ROI's 3D contour coordinates to a 2D plane. Traditional calibration methods often produce significant size errors due to difficulty maintaining alignment between the calibration and measured objects during real-time measurements. RGB-D cameras, however, acquire depth information actively or passively, generating 3D point cloud data P = $\{x_i, y_i, z_i\}$, enabling direct Euclidean distance calculation for true dimensions:

$$W_{true} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} \quad (3)$$

This enables more precise measurement of *Pleurotus eryngii*'s true size, reducing errors from height variations and complex shapes common in traditional methods.

2.5.2. Algorithm for Color Classification

In order to accurately capture the color details of *Pleurotus eryngii* caps, we applied the K-means clustering algorithm to identify and group the primary cap colors. K-means is a helpful unsupervised learning technique that organizes image pixels according to their color characteristics.

The specific procedures are presented as follows:

(1) The region of interest (ROI) in the cap is identified using the semantic segmentation algorithm, allowing us to focus on the most relevant areas.

(2) To identify the dominant color of the mushroom cap, the K-means clustering algorithm was first employed with K = 3 to group pixels into three major color clusters.

Each pixel is assigned to the nearest cluster center based on Euclidean distance in RGB space:

$$d(p, c_i) = \sqrt{(\mathrm{R}_p - \mathrm{R}_{c_i})^2 + (\mathrm{G}_p - \mathrm{G}_{c_i})^2 + (\mathrm{B}_p - \mathrm{B}_{c_i})^2} \quad (4)$$

Defined by Equation (4), p denotes a pixel represented as 3D vector in RGB space (R, G, and B intensity channels), c_i is the i-th cluster center, and d (p,c_i) denotes their Euclidean distance.

In this equation, p denotes a pixel point, and$c_i$denotes the i-th cluster center. The cluster centers are iteratively updated based on the average color values of the assigned pixels until the algorithm converges. After convergence, the cluster containing the largest number of pixels is selected. The centroid color of this dominant cluster is then taken as the single representative foreground color of the cap.

(3) For finer perceptual color grading, the obtained cluster centroids (or the dominant centroid) are transformed into the perceptually uniform Lab color space. By analyzing their positions in this space—particularly the lightness (L) and yellow–blue (b) components—and establishing empirically derived thresholds based on sample statistics, the color grades (light brown, yellowish brown, and dark brown) can be accurately classified, ensuring precise and consistent color differentiation.

The classification rules are defined as follows: Light Brown corresponds to L* > 60 and b* > 10; Yellowish Brown corresponds to 45 ≤ L* ≤ 60 and b* > 15; and Dark Brown corresponds to L* < 45 and b* ≤ 15. These thresholds were optimized through statistical analysis of color centroids obtained from 134 labeled samples, ensuring that the grading aligns with human visual perception while maintaining objective measurability and reproducibility.

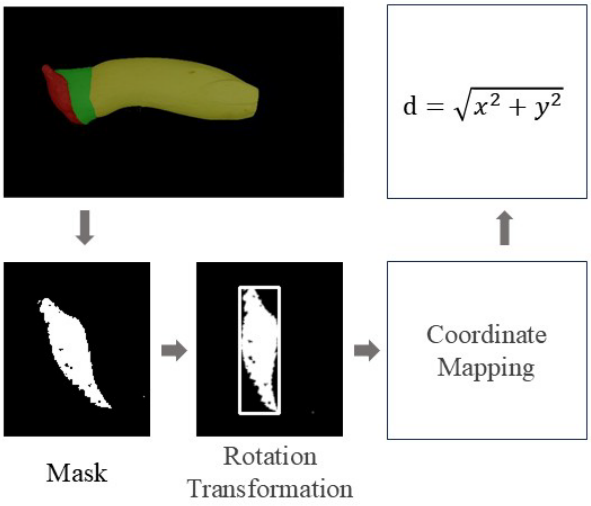

Figure 6. Flowchart for acquiring dimension information.

2.5.3. Algorithm for Shape Classification

As shown in Figure 7, the AlexNet network consists of eight convolutional layers arranged consecutively. Each layer is followed by a max pooling layer and uses the ReLU activation function to enhance non-linear features. Additionally, the network includes three fully connected layers that help achieve the final classification output. During the data preprocessing stage, all input images are resized to 227 × 227 pixels to meet the requirements of the AlexNet model. The number of output channels in the convolutional layers is 96, 256, 384, 384 and 256. Through this step-by-step feature extraction and compression process, the model learns the shape characteristics of the *Pleurotus eryngii* fruiting

body, enabling it to make accurate predictions. Finally, the model makes decisions using a fully connected layer with 1000 neurons and outputs the classification results.

2.6. Experimental Environment

For this experiment, we utilized an Intel® Xeon® Silver 4214R CPU with a clock speed of 2.40 GHz, 90 GB of memory and an RTX 3080 Ti GPU. The deep learning framework was established on the Ubuntu 20.04 system using Python 3.9.0 and PyTorch 2.1.0. In the segmentation task, we employed the Adam optimizer with a momentum value of 0.9, an initial learning rate of 0.0001, a batch size of 8 and a total of 135 training iterations.

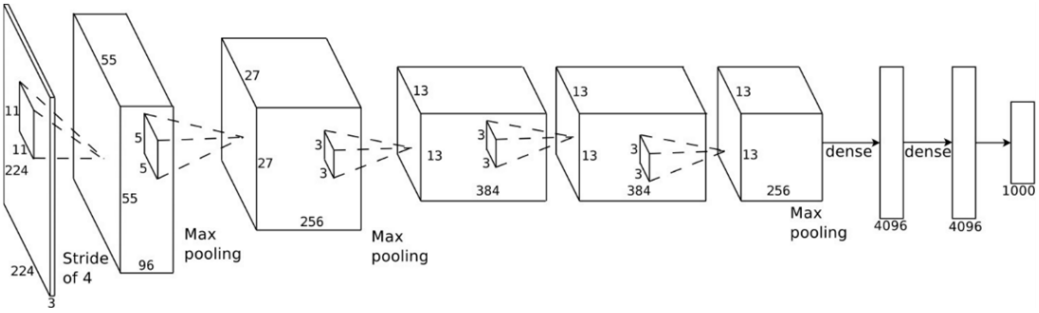

Figure 7. Structural diagram of the AlexNet network.

2.7. Evaluation Indices

In this work, commonly adopted evaluation metrics were employed to comprehensively assess model performance. For the semantic segmentation of *Pleurotus eryngii* fruiting bodies, mPA, mIoUand F1-score (also known as the Dice coefficient) were used. In addition, Mean Squared Error (MSE) and the coefficient of determination (R^2) were utilized to evaluate the accuracy of the size phenotype extraction algorithm, mPA reflects the average pixel accuracy across all categories and evaluates how accurately each category's pixels are classified.

MPA is a core evaluation metric for semantic segmentation. It computes the pixel-wise classification accuracy for each class independently and then averages these accuracies across all classes. Compared to overall PA, mPA treats each class equally, thereby effectively mitigating evaluation bias caused by class imbalance in pixel count.

$$mPA = \frac{1}{k}\sum_{i=1}^{k}\frac{TP_i}{TP_i + FN_i} \tag{5}$$

Defined by Equation (5), K denotes the total number of semantic classes;TP_i (True Positives) denotes the number of pixels correctly predicted as class i; FN_i (False Negatives) denotes the number of pixels that actually belong to class i but are incorrectly predicted as other classes.

MIoU can be used to evaluate the performance of semantic segmentation. It calculates the IoU for each category and then determines the average value.

$$mIoU = \frac{1}{k}\sum_{i=1}^{k}\frac{TP_i}{TP_i + FP_i + FN_i} \tag{6}$$

Defined by Equation (3), FN_i (False Negatives) denotes the number of pixels that belong to class i in the ground truth but are incorrectly predicted as other classes;

The F1 Score (or Dice coefficient) is a metric that thoughtfully balances precision and recall, making it particularly well-suited for scenarios where classes are imbalanced.

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

$$recall = \frac{TP}{TP + FN} \tag{8}$$

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{9}$$

MSE is a commonly used metric for understanding the difference between a model's predicted values and its true values. Calculating the average of the squared prediction errors it provides insight into how accurate the model is. A smaller RMSE value indicates that the model's predictions are closer to the actual values, which is a sign of better performance.

R^2 is used to evaluate how well the size phenotype algorithm explains the variability in the actual size data of *Pleurotus eryngii*. Its value ranges between 0 and 1, and the closer it gets to 1, the better the algorithm fits the data.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y_i})^2 \tag{10}$$

$$R^2 = 1 - \frac{\sum_i^n (y_i - \widehat{y_i})^2}{\sum_i^n (y_i - \bar{y})^2} \tag{11}$$

n is the number of pixels or samples. y_i is the ground truth value. $\widehat{y_i}$ is the predicted value produced by the model.

By combining the MSE and R^2, the model's performance can be evaluated from two complementary angles. R^2 provides insight into how well the model fits the size data, while MSE quantifies the discrepancy between predicted and actual measured values. Together, these metrics help gain a deeper understanding of the algorithm's accuracy and reliability, offering valuable guidance for further refinement.

3. Results and Analysis

3.1. Experimental Outcomes

3.1.1. Model Validation

To demonstrate the effectiveness of the improvements made to the Aspp-VGGUnet algorithm, this study systematically compared the original Unet, PSPNet, DeeplabV3+, This selection encompasses foundational, multi-scale context modeling, and advanced encoder–decoder architectures, providing a comprehensive baseline to assess our model's performance in capturing both global semantics and fine-grained details. The enhanced Aspp-VGGUnet model under the same experimental conditions and dataset. All experiments adhered to a standardized protocol. Models were trained and evaluated on the same proprietary dataset of 1136 annotated *Pleurotus eryngii* images, split 7:2:1 into training, validation, and test sets. All inputs were resized to 512 × 512 pixels and augmented via random horizontal flipping. A consistent training setup was used: the Adam optimizer (initial learning rate: 1×10^{-4}, batch size: 8), 150 epochs on an NVIDIA RTX 3080 Ti GPU. Performance was evaluated using the standard metrics mIoU and mPA.

The experimental results are presented in Table 1—Comparisons of Segmentation Models. The results show that the improved Aspp-VGGUnet achieved impressive scores of 96.47% and 98.53% in mIoU and mPA, significantly outperforming other comparative models. This highlights its outstanding ability in segmentation accuracy and detail detection.

As shown in Figure 8, Figure 8a displays the original image of the king oyster mushroom, while Figure 8b presents the visualization result of overlaying the annotated mask (ground truth) on the original image. Figure 8c, Figure 8d and Figure 8e show the segmentation results generated by the PSPNet, DeepLabV3+, and proposed Aspp-VGGUnet models, respectively. The results highlight how Aspp-VGGUnet excels at extracting contours in the edge areas of *Pleurotus eryngii* images, with greater precision and an enhanced ability to capture fine details. Compared to our proposed model, DeeplabV3+ and PSPNet exhibit slightly lower performance, with mIoU values of 90.94% and 85.02%, respectively. Although these models incorporate advanced encoder–decoder architectures and multi-scale feature fusion techniques, they still encounter challenges in accurately capturing fine-grained details, especially in the precise extraction of edges and contours for *Pleurotus eryngii*, indicating room for further improvement in this regard. PSPNet's multi-scale pooling module demonstrates robust capabilities in capturing global information. However, inaccuracies persist in detailed areas, especially those with complex structures characteristic of *Pleurotus eryngii*. Similarly, DeeplabV3+ leverages atrous convolution to improve its global context modeling, yet it still lags behind Aspp-VGGUnet in terms of detailed segmentation accuracy.

Table 1. Comparisons of Segmentation Models.

Models	mIoU (%)	F1_score (%)	mPA (%)
Unet	93.44	93.30	96.30
PSPNet	85.02	91.07	93.05
DeeplabV3+	90.94	95.32	93.10
Aspp-VGGUnet	**96.47**	**96.00**	**98.53**

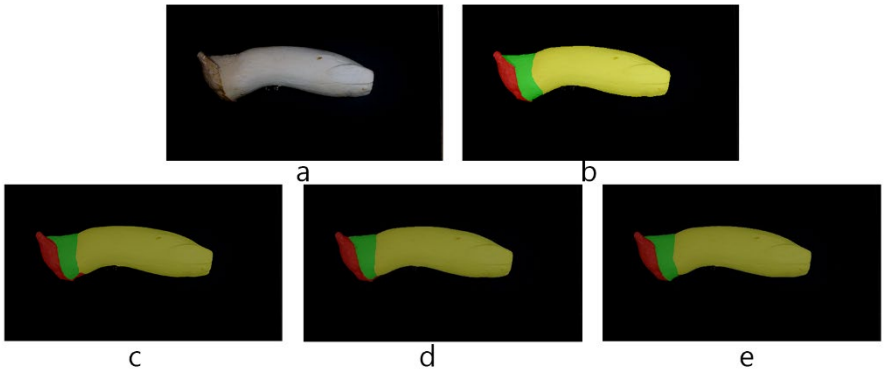

Figure 8. Segmentation effect comparison; (**a**) Original fruiting body image, (**b**) Ground truth, (**c**) PSPNet result, (**d**) DeeplabV3+ result, (**e**) Aspp-VGGUnet result.

3.1.2. Component-Wise Ablation Analysis

To better understand how each module contributes to the model's performance, we carried out a series of ablation experiments. The detailed results can be found in Table 2—Ablation experiments evaluating the impact of components A (VGG16) and B (ASPP). We separately examined how the VGG16 pre-trained backbone network (Component A) and the ASPP module (Component B) enhance the model's feature extraction capabilities.

The U-Net architecture employed VGG16 as its feature extractor, with the network initialized using pre-trained weights to facilitate more efficient and effective training. Compared to the original Unet, the VGG16 backbone significantly improves key metrics such as mIou, F1_score and mPA. This indicates that it effectively enhances the model's ability to extract meaningful features. Following this, we integrated the ASPP module,

which uses multi-scale dilated convolutions to more effectively capture the contextual information across different scales in images. The performance of Aspp-VggUnet is significantly enhanced across all metrics compared to the original VGGUnet, particularly excelling in multi-scale feature extraction.

In summary, the ablation experiments indicate that incorporating the ASPP and VGG16 modules leads to a notable improvement of Unet's capability for feature extraction and multi-scale context comprehension. This is particularly evident when the sample size is limited, as the model's overall performance is significantly strengthened.

Table 2. Ablation experiments evaluating the impact of components A (VGG16) and B (ASPP).

Method	A	B	mIoU (%)	F1_score (%)	mPA (%)
Baseline	✗	✗	93.44	93.30	96.30
Baseline + A	✓	✗	95.00	94.50	96.71
Baseline + A + B	✓	✓	**96.47**	**96.00**	**98.53**

3.2. Phenotypic Analysis

3.2.1. Semantic Segmentation Analysis

Based on the comparative analysis presented in Table 1, it is evident that the Aspp-VGGUnet model, which incorporates VGG16 pre-trained weights and ASPP modules dilated convolution modules, achieved the best performance in the experiment. Consequently, it was selected as the default model for extracting phenotypic parameters of *Pleurotus eryngii*.

The segmentation results for *Pleurotus eryngii* are shown in Figure 9. In this figure, (1), (2) and (3) represent three mature fruiting body samples of the same cultivar of *Pleurotus eryngii*, while Figure 9a, Figure 9b, Figure 9c and Figure 9d correspond to the original frontal images, frontal mask images, original dorsal images and dorsal mask images of the fruiting bodies, respectively. By capturing images from both the front and back and applying semantic segmentation, we observe that in the segmented mask regions, the red areas indicate the caps of the fruiting bodies, the green areas represent the gills and the yellow areas correspond to the stipes.

The enhanced Aspp-VGGUnet algorithm demonstrates exceptional performance in segmenting different parts of the *Pleurotus eryngii* fruiting body, with high segmentation accuracy and rapid processing speed. Moreover, the proposed method eliminates the need for preprocessing steps such as filtering, binarization, and edge trimming, thereby simplifying the entire workflow. Traditional segmentation algorithms are highly sensitive to threshold settings and often struggle with generalization across different scenarios. In contrast, the improved semantic segmentation model developed in this study effectively distinguishes the structural components of *Pleurotus eryngii*, resulting in more accurate, stable, and robust segmentation outcomes.

3.2.2. Dimensional Phenotype Analysis

To verify the accuracy of the size algorithm, this study conducted systematic tests on various phenotypic dimensions of *Pleurotus eryngii*. By analyzing the correlation between the calculated values and the actual measured values, the specific indicators are clearly illustrated in Figure 10.

The evaluated traits included the maximum cap diameter, cap thickness (horizontal length), cap longitudinal width, gill length, stem length, stem width, and the total length of the fruiting body. The sample measurement data were acquired using RGB cameras, RGB-D cameras, and advanced image-processing techniques. Length and width estimates

were computed from the three-dimensional coordinates within the ROI, ensuring accurate and reliable quantification of each phenotypic trait.

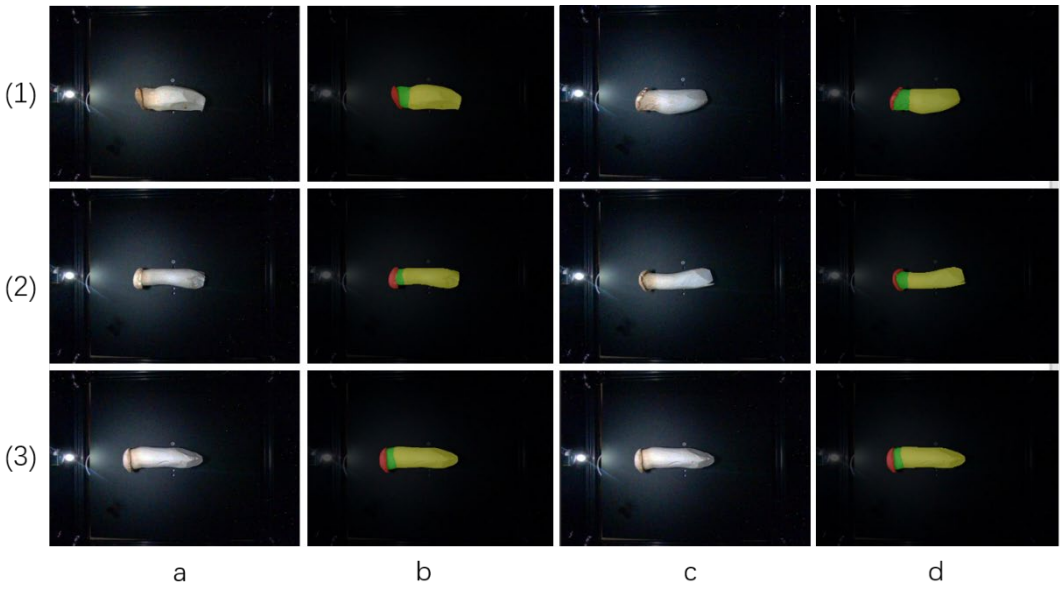

Figure 9. Semantic segmentation result diagram; (**a**) Original front image of the sub-entity, (**b**) Front mask image of the sub-entity, (**c**) Original back image of the sub-entity, (**d**) Back mask image of the sub-entity.

The subsequent figure presents the R^2 and MSE. Data analysis reveals that the correlation between the estimated values and the actual measured values for each dimension reflects the accuracy of the measurement. The results show that the R^2 values for all dimensions fall consistently within the range of 0.91 to 0.95, highlighting the algorithm's strong consistency and reliability in predicting sizes. Additionally, the MSE ranges from 1.3 to 2.1, reflecting a commendably small difference between calculated and measured values, which further confirms the algorithm's precision.

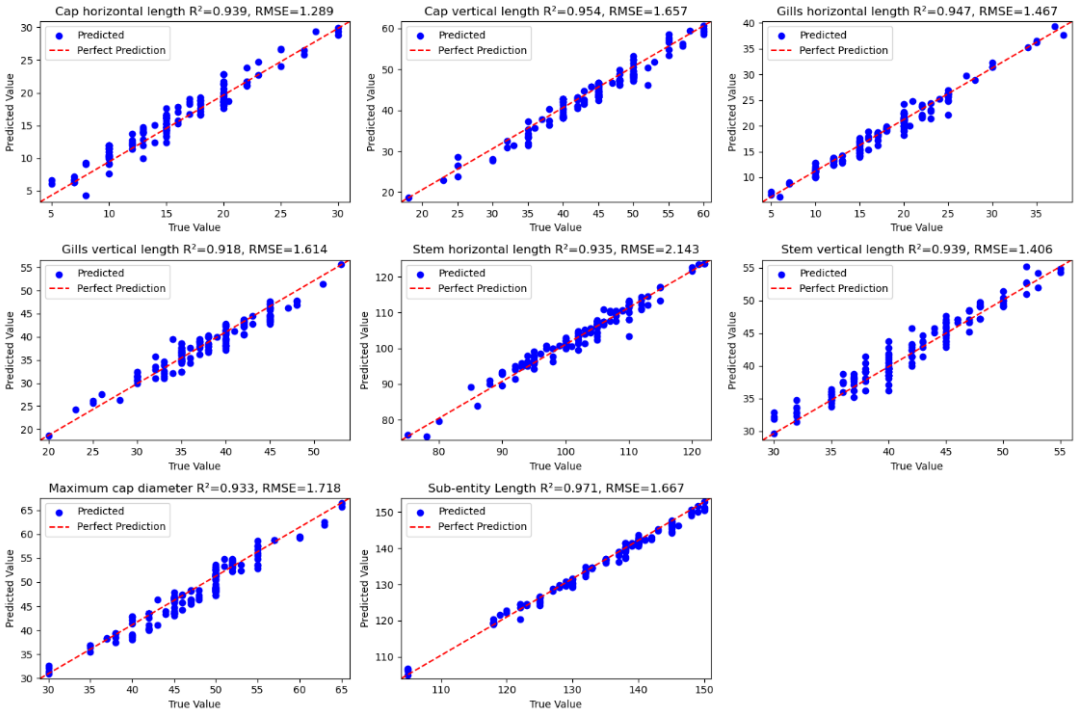

Figure 10. Linear fitting diagram of dimensions.

3.2.3. Color Phenotypic Outcomes

The color clustering algorithm employs an iterative optimization process to classify pixels with similar colors pixels into distinct categories. Through multiple iterations, it continuously refines pixel color assignments, gradually grouping analogous colors and extracting the foreground color of the cap image. Figure 11 shows sub-images (1), (2), and (3) for different caps: Figure 11a is the original image, Figure 11b is the ROI extraction result, and Figure 11c is the main color extraction result. The cap is first segmented from the original image using an Aspp-VGGUnet network to extract the ROI. Subsequently, K-means clustering is applied to obtain the dominant foreground color. Specifically, Specifically, pixels within the ROI are grouped into K = 3 clusters based on their color similarity in RGB space. The cluster containing the largest number of pixels is identified, and the centroid (mean color) of this dominant cluster is taken as the representative foreground color of that individual cap.

During the implementation of the algorithm, we carefully analyzed multiple experimental datasets and established a suitable threshold in the LAB color space. This step helps achieve precise quantitative classification of cap colors while ensuring consistent and reliable outcomes. We measured the cap colors of 134 samples, and the resulting confusion matrix is shown in Figure 12. The findings reveal that our color classification method achieves an impressive accuracy rate of 91.04%, demonstrating strong performance.

Figure 11. Color Clustering Algorithm Results; (**a**) Original Cap Image, (**b**) Cap ROI Extraction, (**c**) Main Color Extraction.

3.2.4. Shape Characterization

In the field of *Pleurotus eryngii* breeding, morphological differences in mushroom shape are closely linked to key agronomic traits, including growth patterns, yield, disease resistance, environmental adaptability, and marketability. Shape-based classification therefore facilitates the selection of superior varieties with high economic value and stable performance, enabling refined cultivation strategies and enhanced production efficiency.

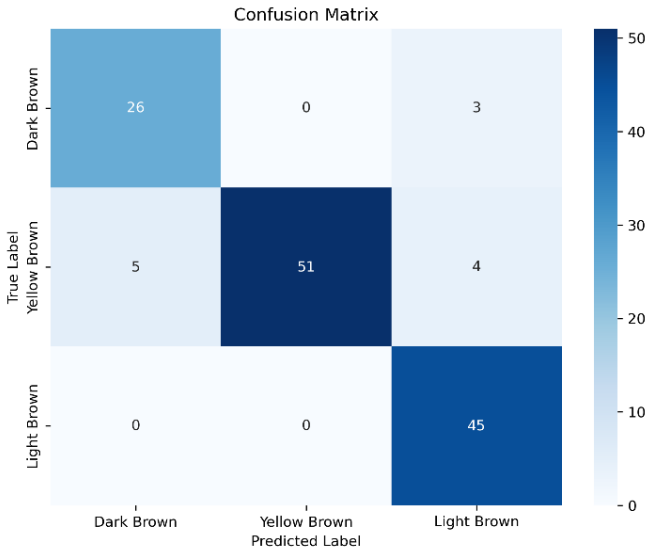

Figure 12. Confusion Matrix Diagram of Color Grading Outcomes.

In this study, we developed an AlexNet-based classification model to distinguish three *Pleurotus eryngii* shapes: bowling ball-shaped, straight-shaped, and standard-shaped. A total of over 1136 annotated samples were used, divided into training and validation sets. During training, images were augmented through random cropping and horizontal flipping, while validation images were resized to 224 × 224 and normalized. The model was optimized with cross-entropy loss and the Adam optimizer at a learning rate of 0.0002, trained over 300 epochs with a batch size of 32 and 8 parallel data-loading threads. The final model achieved an impressive accuracy of 97.90% on the test set. Figure 13 illustrates the experimental results of AlexNet in three different classification tasks, clearly showing how effective the model is in each task.

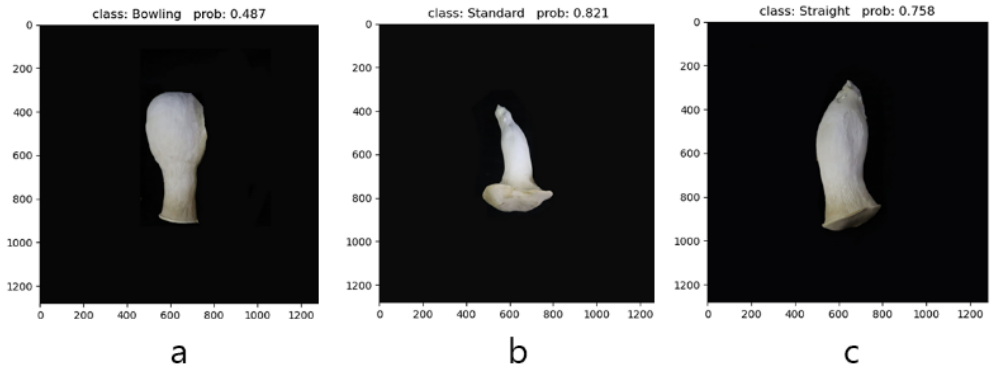

Figure 13. Classification of *pleurotus eryngii* shapes; (**a**) Bowling ball type, (**b**) Standard type, (**c**) Straight type.

3.3. System Interface Display

Figure 14 shows the system interface and the displayed results of the *Pleurotus eryngii* phenotypic acquisition system. In real-time monitoring and data collection, the software platform can display image data from both the RGB-D camera and the RGB camera instantly. Using the previously introduced phenotypic measurement algorithm, breeders can efficiently acquire detailed phenotypic information about the size of various parts of *Pleurotus eryngii* through an intuitive interactive interface. The system also provides real-time outputs for cap color classification (where the label "Light" corresponds to "Light Brown" within the defined categories of light brown, yellowish brown, and dark brown)

and fruiting body shape, with an overall phenotypic response time of approximately 2 s. Environmental parameters, including temperature and humidity inside the acquisition chamber, are monitored in real time via integrated sensors and displayed on the interface to provide continuous feedback on experimental conditions. Simultaneously, the weight of each fruiting body is directly measured by a built-in pressure sensor and automatically recorded within the system. In terms of data management, it automatically saves and exports batch data, while storing multi-dimensional phenotypic data in a structured format. This not only simplifies the process but also greatly facilitates subsequent data analysis and processing for breeding research.

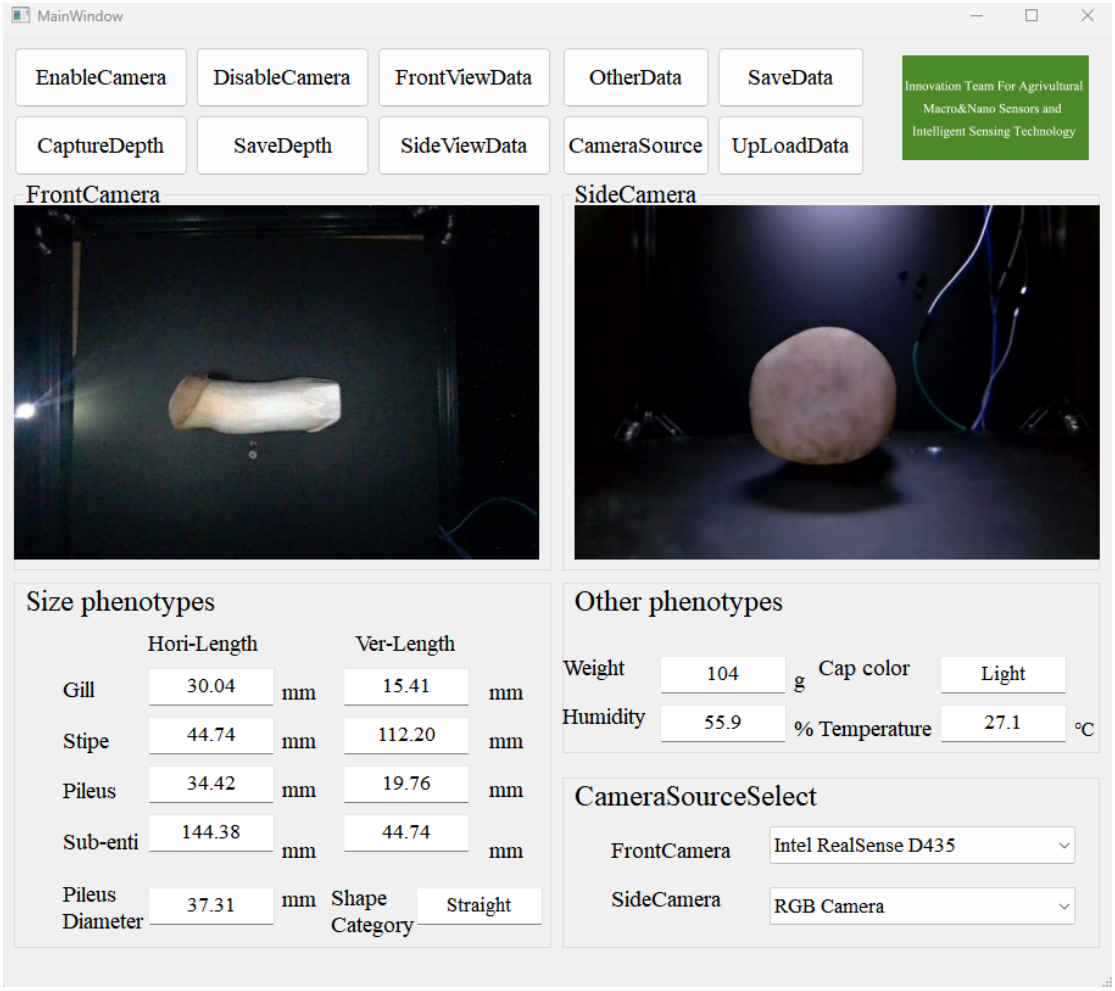

Figure 14. Interface of the *Pleurotus eryngii* Phenotypic Acquisition System.

4. Discussion

A portable phenotyping system for *Pleurotus eryngii* was developed, integrating an RGB camera and the RealSense D405 to simultaneously capture RGB and depth images for 3D phenotypic measurement (Zhou et al., 2021) [24]. Furthermore, the proposed modified U-Net architecture based on ASPP and the VGG16 network achieved remarkable performance enhancements in the segmentation tasks of the cap, gills, and stem, with the mIoU reaching 96.47% and mPA attaining 98.53%. Although excellent results were achieved in the current task, the performance of this network might be constrained by specific datasets, and its generalization ability in other domains (such as medical images, remote sensing images) and complex outdoor scenarios still requires further validation.

In the task of phenotypic acquisition and analysis, this study proposes a method that utilizes depth information from the camera to calculate three-dimensional coordinates

and directly compute size parameters based on spatial data. Unlike traditional 2D imaging or fixed-distance measurement approaches, this method relies on depth-derived 3D coordinates for calculation, thereby fundamentally avoiding the scale errors caused by variations in imaging distance (Guo & Chen, 2018) [25]. The method stands out for its ability to meet size measurement requirements in complex scenarios, providing robust and reliable support for phenotypic analysis. This advantage becomes particularly evident when compared with multi-view neural radiance field (NeRF)-based reconstruction methods, such as that presented by Yin et al. (2025) [20], which require extended data acquisition and processing time. Our system maintains measurement accuracy while significantly improving throughput, making it better suited for high-throughput or near-real-time applications such as on-site breeding screening. In contrast, our method demonstrates superior acquisition efficiency, providing rapid and reliable phenotypic measurements suitable for high-throughput or real-time applications.

For color and shape classification, the perceptually uniform Lab color space was adopted to decouple luminance from chromaticity, establishing an illumination-robust pipeline that enhanced the repeatability and standardization of color assessment. Meanwhile, a standard AlexNet model effectively resolved subtle morphological variations across mushroom phenotypes, attaining a classification accuracy of 97.90%. These results confirm that established computer vision and deep learning architectures, when embedded within a customized hardware-software pipeline, can reliably execute complex phenotypic analysis under controlled imaging environments. Nevertheless, the present performance remains dependent on consistent acquisition conditions and dataset uniformity. Extending the system to scenarios with variable illumination, wider genetic backgrounds, or less structured field settings would require further validation and could be improved through adaptive preprocessing or transfer-learning strategies to enhance generalization capability.

In summary, the method introduced in this study has achieved notable success in tasks such as phenotype acquisition, color classification and shape classification. However, there are still some areas for improvement, including the reliance on depth information accuracy and challenges with color stability under extreme lighting conditions. While this study focuses on the automated extraction of mature phenotypes, investigating phenotypic variations across different maturity levels and growth conditions will be an important direction for future research. Further development, for example, by incorporating time-series models and dynamic phenotypic data at different environmental conditions for a given variety, will help broaden the understanding of mushroom morphological changes.

5. Conclusion

This study established a portable and cost-effective phenotypic acquisition system for *Pleurotus eryngii*, offering a practical solution for high-throughput and precise phenotypic analysis in edible mushroom breeding. By integrating an improved deep learning model with readily deployable sensors, the system successfully enabled the automated extraction of multiple morphological traits, significantly enhancing measurement efficiency and objectivity while considerably lowering the technical barrier to adoption. This work provides reliable technical support for transitioning *Pleurotus eryngii* breeding from experience-based selection to data-driven decision-making and lays a practical foundation for its scalable application within modern precision breeding systems.

Author Contributions: Conceptualization, Conceptualization, Y.C., Z.W. and X.W.; methodology, Z.W. and Z.L.; software, Y.L. and W.S.; validation, P.H., Z.L. and Y.L.; formal analysis, Y.C. and

X.W.; investigation, Z.W., W.S. and P.H.; resources, B.C. and X.K.; data curation, Y.C. and J.P.; writing—original draft preparation, Y.C., Z.W. and X.W.; writing—review and editing, Y.C., Z.W. and X.W.; visualization, Y.L. and P.H.; supervision, X.K. and X.W.; project administration, Y.C. and P.H.; funding acquisition, X.K. and X.W. All authors have read and agreed to the published version of the manuscript.

Funding: The authors gratefully acknowledge the financial support provided by Outstanding Young Research Talents Program of Fujian Agriculture and Forestry University (xjq202116), Special Fund Project for Scientific and Technological Innovation of Fujian Agriculture and Forestry University (KFB22100XA), National Key Laboratory (KFXZ23002) and Natural Science Foundation of Fujian Province (2021J01131).

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Acknowledgments: The authors gratefully acknowledge the support provided by the funders of this work.

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

References

1. Zhang, B.; Li, Y.; Zhang, F.; Linhardt, R.J.; Zeng, G.; Zhang, A. Extraction, structure and bioactivities of the polysaccharides from Pleurotus eryngii: A review. *Int. J. Biol. Macromol.* **2020**, *150*, 1342–1347. https://doi.org/10.1016/j.ijbiomac.2019.10.144.
2. Wang, S.; Qi, J.; Cai, X.; Wu, W.; Wang, Z.A.; Jiao, S.; Dong, C.; Li, Y.; Yang, Z. Advancements and future perspectives in the study of oligosaccharides derived from edible-medicinal mushrooms. *Food Biosci.* **2024**, *61*, 104874. https://doi.org/10.1016/j.fbio.2024.104874.
3. Li, Y.; Wen, W.; Fan, J.; Gou, W.; Gu, S.; Lu, X.; Yu, Z.; Wang, X.; Guo, X. Multi-Source data fusion improves time-series phenotype accuracy in maize under a field high-throughput phenotyping platform. *Plant Phenomics* **2023**, *5*, 0043. https://doi.org/10.34133/plantphenomics.0043.
4. Dengyu, X.; Liang, G.; Chengliang, L.; Yixiang, H. Phenotype-based robotic screening platform for leafy plant breeding. *IFAC PapersOnLine* **2016**, *49*, 237–241. https://doi.org/10.1016/j.ifacol.2016.10.044.
5. Wang, Y.; Fan, J.; Yu, S.; Cai, S.; Guo, X.; Zhao, C. Research advance in phenotype detection robots for agriculture and forestry. *Int. J. Agric. Biol. Eng.* **2023**, *16*, 14–25. https://doi.org/10.25165/j.ijabe.20231601.7945.
6. Zhou, S.; Mou, H.; Zhou, J.; Zhou, J.; Ye, H.; Nguyen, H.T. Development of an automated plant phenotyping system for evaluation of salt tolerance in soybean. *Comput. Electron. Agric.* **2021**, *182*, 106001. https://doi.org/10.1016/j.compag.2021.106001.
7. Zhu, X.; Zhu, K.; Liu, P.; Zhang, Y.; Jiang, H. A special robot for precise grading and metering of mushrooms based on Yolov5. *Appl. Sci.* **2023**, *13*, 10104. https://doi.org/10.3390/app131810104.
8. An, N.; Palmer, C.M.; Baker, R.L.; Markelz, R.J.C.; Ta, J.; Covington, M.F.; Maloof, J.N.; Welch, S.M.; Weinig, C. Plant high-throughput phenotyping using photogrammetry and imaging techniques to measure leaf length and rosette area. *Comput. Electron. Agric.* **2016**, *127*, 376–394. https://doi.org/10.1016/j.compag.2016.04.002.
9. Raj, R.; Walker, J.P.; Pingale, R.; Nandan, R.; Naik, B.; Jagarlapudi, A. Leaf area index estimation using top-of-canopy airborne RGB images. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *96*, 102282. https://doi.org/10.1016/j.jag.2020.102282.
10. Rong, J.; Zhou, H.; Zhang, F.; Yuan, T.; Wang, P. Tomato cluster detection and counting using improved YOLOv5 based on RGB-D fusion. *Comput. Electron. Agric.* **2023**, *207*, 107741. https://doi.org/10.1016/j.compag.2023.107741.
11. Ravaglia, J.; Fournier, R.A.; Bac, A.; Véga, C.; Côté, J.-F.; Piboule, A.; Rémillard, U. Comparison of three algorithms to estimate tree stem diameter from terrestrial laser scanner data. *Forests* **2019**, *10*, 599. https://doi.org/10.3390/f10070599.
12. Baisa, N.L.; Al-Diri, B. Mushrooms detection, localization and 3D pose estimation using RGB-D sensor for robotic-picking application. *arXiv* **2022**, arXiv:2201.02837
13. Zhou, J.; Cui, M.; Wu, Y.; Gao, Y.; Tang, Y.; Jiang, B.; Wu, M.; Zhang, J.; Hou, L. Detection of maize stem diameter by using RGB-D cameras' depth information under selected field condition. *Front. Plant Sci.* **2024**, *15*, 1371252. https://doi.org/10.3389/fpls.2024.1371252.

14. Narisetti, N.; Henke, M.; Neumann, K.; Stolzenburg, F.; Altmann, T.; Gladilin, E. Deep learning based greenhouse image segmentation and shoot phenotyping (DeepShoot). *Front. Plant Sci.* **2022**, *13*, 906410. https://doi.org/10.3389/fpls.2022.906410.
15. 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 2nd International Conference on Computer Science, Electronic Information Engineering and Intelligent Control Technology (CEI), Nanjing, China, 23–25 September 2022; pp. 717–721. https://doi.org/10.1109/CEI57409.2022.9950157.
16. Kolhar, S.; Jagtap, J. Phenomics for komatsuna plant growth tracking using deep learning approach. *Expert Syst. Appl.* **2023**, *215*, 119368. https://doi.org/10.1016/j.eswa.2022.119368.
17. Ahmad, N.; Asif, H.M.S.; Saleem, G.; Younus, M.U.; Anwar, S.; Anjum, M.R. Leaf image-based plant disease identification using color and texture features. *Wirel. Pers. Commun.* **2021**, *121*, 1139–1168. https://doi.org/10.1007/s11277-021-09054-2.
18. Vazquez, D.V.; Spetale, F.E.; Nankar, A.N.; Grozeva, S.; Rodríguez, G.R. Machine learning-based tomato fruit shape classification system. *Plants* **2024**, *13*, 2357. https://doi.org/10.3390/plants13172357.
19. Luo, S.; Tang, J.; Peng, J.; Yin, H. A novel approach for measuring the volume of Pleurotus eryngii based on depth camera and improved circular disk method. *Sci. Hortic.* **2024**, *336*, 113382. https://doi.org/10.1016/j.scienta.2024.113382.
20. Yin, H.; Cheng, W.; Yuan, L.; Chen, M.; Sun, Y.; Xu, Y.; Wang, Y. A pipeline for columnar agricultural products volume measurement based on neural radiance fields and mobile phone. *J. Food Meas. Charact.* **2025**, *29*, 1–21 https://doi.org/10.1007/s11694-025-03704-w.
21. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015, Munich, Germany, 5–9 October 2015; 9351, pp. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28.
22. Simonyan, K.; Zisserman, A. Very Deep convolutional networks for large-scale image recognition. In Proceedings of the 3rd International Conference on Learning Representations, San Diego, CA, USA, 7–9 May 2015; pp. 1–14. https://doi.org/10.48550/arXiv.1409.1556.
23. He, K.; Zhang, X.; Ren, S.; Sun, J. Spatial pyramid pooling in deep convolutional networks for visual recognition. In *European Conference on Computer Vision 2014*; Springer: Berlin/Heidelberg, Germany, 2014; Volume 8691, pp. 346–361. https://doi.org/10.1007/978-3-319-10578-9_23.
24. Zhou, T.; Fan, D.-P.; Cheng, M.-M.; Shen, J.; Shao, L. RGB-D salient object detection: A survey. *Comput. Vis. Media* **2021**, *7*, 37–69. https://doi.org/10.1007/s41095-020-0199-z.
25. Guo, Y.; Chen, T. Semantic segmentation of RGBD images based on deep depth regression. *Pattern Recognit. Lett.* **2018**, *109*, 55–64. https://doi.org/10.1016/j.patrec.2017.08.026.