

Article

An Improved 2D Pose Estimation Algorithm for Extracting Phenotypic Parameters of Tomato Plants in Complex Backgrounds

Yawen Cheng ^{1,2} , Ni Ren ^{1,2}, Anqi Hu ^{1,2}, Lingli Zhou ^{1,2}, Chao Qi ^{1,2}, Shuo Zhang ³  and Qian Wu ^{1,2,*} 

¹ Institute of Agricultural Information, Jiangsu Academy of Agricultural Sciences, Nanjing 210031, China; 20210102@jaas.ac.cn (Y.C.); rn@jaas.ac.cn (N.R.); haq93@163.com (A.H.); 20210054@jaas.ac.cn (L.Z.); chaoqi@jaas.ac.cn (C.Q.)

² Key Laboratory of Intelligent Agricultural Technology (Changjiang Delta), Ministry of Agriculture and Rural Affairs, Nanjing 210014, China

³ School of Information Management, Nanjing University, Nanjing 210023, China; zhanshuo@smail.nju.edu.cn

* Correspondence: wuqian@jaas.ac.cn; Tel.: +86-025-84391622

Abstract: Phenotypic traits, such as plant height, internode length, and node count, are essential indicators of the growth status of tomato plants, carrying significant implications for research on genetic breeding and cultivation management. Deep learning algorithms such as object detection and segmentation have been widely utilized to extract plant phenotypic parameters. However, segmentation-based methods are labor-intensive due to their requirement for extensive annotation during training, while object detection approaches exhibit limitations in capturing intricate structural features. To achieve real-time, efficient, and precise extraction of phenotypic traits of seedling tomatoes, a novel plant phenotyping approach based on 2D pose estimation was proposed. We enhanced a novel heatmap-free method, YOLOv8s-pose, by integrating the Convolutional Block Attention Module (CBAM) and Content-Aware ReAssembly of FEatures (CARAFE), to develop an improved YOLOv8s-pose (IYOLOv8s-pose) model, which efficiently focuses on salient image features with minimal parameter overhead while achieving a superior recognition performance in complex backgrounds. IYOLOv8s-pose manifested a considerable enhancement in detecting bending points and stem nodes. Particularly for internode detection, IYOLOv8s-pose attained a *Precision* of 99.8%, exhibiting a significant improvement over RTMPose-s, YOLOv5s6-pose, YOLOv7s-pose, and YOLOv8s-pose by 2.9%, 5.4%, 3.5%, and 5.4%, respectively. Regarding plant height estimation, IYOLOv8s-pose achieved an *RMSE* of 0.48 cm and an *rRMSE* of 2%, and manifested a 65.1%, 68.1%, 65.6%, and 51.1% reduction in the *rRMSE* compared to RTMPose-s, YOLOv5s6-pose, YOLOv7s-pose, and YOLOv8s-pose, respectively. When confronted with the more intricate extraction of internode length, IYOLOv8s-pose also exhibited a 15.5%, 23.9%, 27.2%, and 12.5% reduction in the *rRMSE* compared to RTMPose-s, YOLOv5s6-pose, YOLOv7s-pose, and YOLOv8s-pose. IYOLOv8s-pose achieves high precision while simultaneously enhancing efficiency and convenience, rendering it particularly well suited for extracting phenotypic parameters of tomato plants grown naturally within greenhouse environments. This innovative approach provides a new means for the rapid, intelligent, and real-time acquisition of plant phenotypic parameters in complex backgrounds.

Keywords: plant height; internode length; 2D pose estimation; YOLOv8-pose; CBAM; CARAFE; plant phenotyping; tomato



Citation: Cheng, Y.; Ren, N.; Hu, A.; Zhou, L.; Qi, C.; Zhang, S.; Wu, Q. An Improved 2D Pose Estimation Algorithm for Extracting Phenotypic Parameters of Tomato Plants in Complex Backgrounds. *Remote Sens.* **2024**, *16*, 4385. <https://doi.org/10.3390/rs16234385>

Academic Editor: Jose Moreno

Received: 24 September 2024

Revised: 13 November 2024

Accepted: 18 November 2024

Published: 24 November 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The tomato is widely recognized as one of the most important vegetable crops, owing to its rich nutritional composition and distinctive flavor [1,2]. Monitoring tomato growth and capturing real-time phenotypic data play a significant role in advancing genetic breeding, optimizing greenhouse management, and increasing yields. Plant height, internode

length, and node count are pivotal metrics for monitoring the growth and development of the tomato [3]. These morphological parameters exert a substantial influence on various physiological processes, such as nutrient transportation, photosynthesis, the establishment of plant architecture, as well as the occurrence and spread of pests and diseases [4,5]. These traits are frequently utilized to elucidate the intricacies of tomato growth patterns, evaluate plant performance, and investigate the interplay between genetic and environmental determinants that shape tomato growth and development [6–9].

High-throughput plant phenotyping technology enables non-destructive monitoring and quantitative analysis of plant morphological and physiological traits through periodic image acquisition [10–13]. Non-destructive measurements of plant morphology generally rely on two-dimensional (2D) or three-dimensional (3D) imaging techniques [14]. Three-dimensional imaging provides accurate spatial information for quantifying phenotypic parameters [15–19] by the digital capture of plant structural characteristics. While 3D digitization techniques facilitate the extraction of precise phenotypic parameters through 3D plant modeling [20–24], these methods are labor-intensive, time-consuming, and costly. Furthermore, 3D data acquisition is susceptible to environmental influences that can compromise the accuracy of the generated 3D models [25]. Consequently, the instability of point cloud data acquisition in complex environments poses challenges for accurate, efficient, and automated acquisition of high-throughput plant phenotypic parameters.

Plant phenotypic traits extraction based on 2D images remains a prominent approach for high-throughput plant phenotyping [26–29]. Machine vision techniques have been widely utilized in plant phenotyping based on 2D images, including object detection, fruit recognition and localization, yield estimation, and disease identification, achieving promising results [30,31]. Previous studies have developed methods for extracting parameters such as plant height, stem diameter, and internode length from 2D images. Pound M P et al. [32] employed CNNs to extract plant leaf tip and root tip information. Fu et al. [33] employed the Mask R-CNN segmentation algorithm to extract leaf length and plant height from images of potted wheat. Guo et al. [34] utilized Re-YOLOv5 and a detection region search algorithm to obtain the soybean's plant height along with the number of stem nodes and branches. Fan et al. [35] deployed the YOLOv4 model on a field phenotyping robot to measure maize stem diameter.

The aforementioned approaches for extracting parameters such as plant height and internode length predominantly rely on image segmentation or object detection techniques. However, segmentation-based methods are time-consuming and labor-intensive, requiring extensive annotation for training. Moreover, the accuracy of the geometric features extracted is highly dependent on the integrity of the instance masks. Object detection algorithms, while faster, primarily provide information based on bounding boxes, limiting their ability to simultaneously learn their intricate structural features. Furthermore, most existing studies are conducted in a controlled laboratory with strict requirements for data acquisition backgrounds and environments. In contrast, high-throughput image acquisition in greenhouse environments often yields images with complex backgrounds containing multiple plants. This necessitates a significant effort in pre-processing and struggles to accurately extract phenotypic information from images with complex backgrounds, and there are other challenging factors [36]. Extracting phenotypic information from such images requires not only accurate identification of individual plants but also the detection of keypoints, including stem bending points and nodes, within each identified plant. This calls for a method capable of simultaneously considering and learning both global and local relationships within the image.

To overcome the above limitations, this paper endeavors to develop a model capable of learning plant structural characteristics from 2D images with complex backgrounds. This approach strives to reduce time and labor costs while ensuring accurate extraction of plant phenotypic parameters. Drawing inspiration from 2D human pose estimation, which treats human joints (e.g., head, wrist, elbow) as keypoints to estimate their spatial relationships, we extend this concept to plant phenotyping. By treating bending points

and nodes within the plant structure as keypoints for learning, this approach aims to achieve accurate extraction of phenotypic parameters even in complex backgrounds, while simultaneously acquiring more comprehensive phenotypic information.

Pose estimation stands as a pivotal area within deep learning applications. In 2014, the successful introduction of CNN for single-person attitude estimation [37] marked a significant milestone, leading to the development of numerous subsequent algorithms [38,39]. Most of the current mainstream human pose estimation algorithms are based on heatmaps, but existing heatmap-based two-stage methods are not optimal because they often lack end-to-end trainability, and training relies on substituting L1 losses. At the same time, they require additional upsampling operations and a higher processing cost, which affects the reasoning speed, making them limited in practical applications. With the emergence of object detection algorithms such as SSD [40] and YOLO [41], the object detection capability has been significantly improved. In 2022, Debapriya Maji et al. [42] explored the correlation between human pose estimation and object detection by introducing YOLO-pose, a novel heatmap-free approach for 2D pose estimation. By adopting the end-to-end training method, YOLO-pose has optimized the OKS index itself, which can jointly detect the human frame and corresponding 2D pose in a single forward propagation, avoiding the post-processing process in the bottom-up method, and improving the overall efficiency and accuracy. Building upon advancements in object detection algorithms, this method surpassed the performance of all previous bottom-up approaches. Later, with the continuous improvement of YOLO series algorithm performance in target detection tasks, YOLOv7 [43] and YOLOv8 successively achieved breakthroughs in pose estimation tasks. YOLOv7-pose adopted a series of advanced technologies, including multi-scale training, data enhancement, etc., which could improve detection accuracy while ensuring speed. Compared with YOLOv7-pose, YOLOv8-pose introduced the new EfficientRep backbone network, which can improve the inference speed and also optimizes the anchor-free mechanism, which improves the generalization ability of detecting targets of different scales, thus improving the positioning accuracy and detection efficiency. Nowadays, pose estimation has also found applications in agriculture [44,45]. Russello et al. [46] employed it for monitoring cow behavior, while Taehyeong Kim et al. [47] utilized OpenPose to identify tomato pedicle cutting points, thereby enhancing the success rate of robotic harvesting.

This paper analogized the morphological structure of plants to the human skeletal system, and employed 2D human pose estimation techniques to investigate the detection of plant structural keypoints and the extraction of phenotypic parameters. Building upon YOLOv8s-pose, we replaced its original nearest-neighbor interpolation upsampling method with Content-Aware ReAssembly of FEatures (CARAFE) [48] and introduced the Convolutional Block Attention Module (CBAM) [49]. This approach enabled the identification of plants and their keypoints in complex backgrounds. By detecting these keypoints, the model reconstructed the plant's skeletal features, facilitating the efficient and accurate extraction of multiple parameters, including the plant height, the internode length, and number of internodes. The primary contributions of this study are as follows:

- (1) utilizing 2D pose estimation to extract the main stem skeleton structure of tomato plants, enabling the efficient, accurate, and real-time extraction of phenotypic parameters under realistic greenhouse conditions with complex backgrounds;
- (2) enhancing the YOLOv8s-pose model by incorporating the CBAM and replacing the original upsampling method with CARAFE to develop an improved YOLOv8s-pose (IYOLOv8s-pose) algorithm, which efficiently focuses on salient image features with minimal parameter overhead while achieving a superior recognition performance in complex backgrounds.

2. Materials and Methods

2.1. The Development of the Plant Phenotyping Platform

To efficiently capture image data of tomato plants, a plant phenotyping platform (Figure 1) was designed and constructed in our glass greenhouse. This customized platform

integrated advanced imaging technology and precise motion control mechanisms. The core components included an XYZ-axis 3D rail guidance system and an imaging unit which featured a detachable multi-functional mount equipped with three industrial cameras (Teledyne FLIR, BFS-PGE-50S5, 2448×2048 pixels, 5MP), enabling the comprehensive acquisition of RGB image data from various perspectives, including top, squint, and side-view (Figure 1a,b). These cameras were connected to a local Jetson device via Ethernet connection, allowing the seamless transmission and storage of acquired data in a pre-defined format both locally as well as on cloud servers (Figure 1c). The imaging unit was mounted at the end of the Z-axis of the rail guidance system, enabling automated movement to pre-defined locations for comprehensive imagery capture across all tomato plants. A dedicated cloud-based software system was developed for this plant phenotyping platform (Figure 1d). The platform's software was meticulously designed using the lightweight Django, featuring a user-friendly graphical user interface (GUI) and comprehensive user management. It provided users with Python scripting for algorithm deployment and seamlessly integrated a robust database system. The system enables automated data collection through a programmable framework and remote, unsupervised operation. This configuration facilitates high-throughput, continuous, and rapid measurements of plants, significantly enhancing research efficiency.

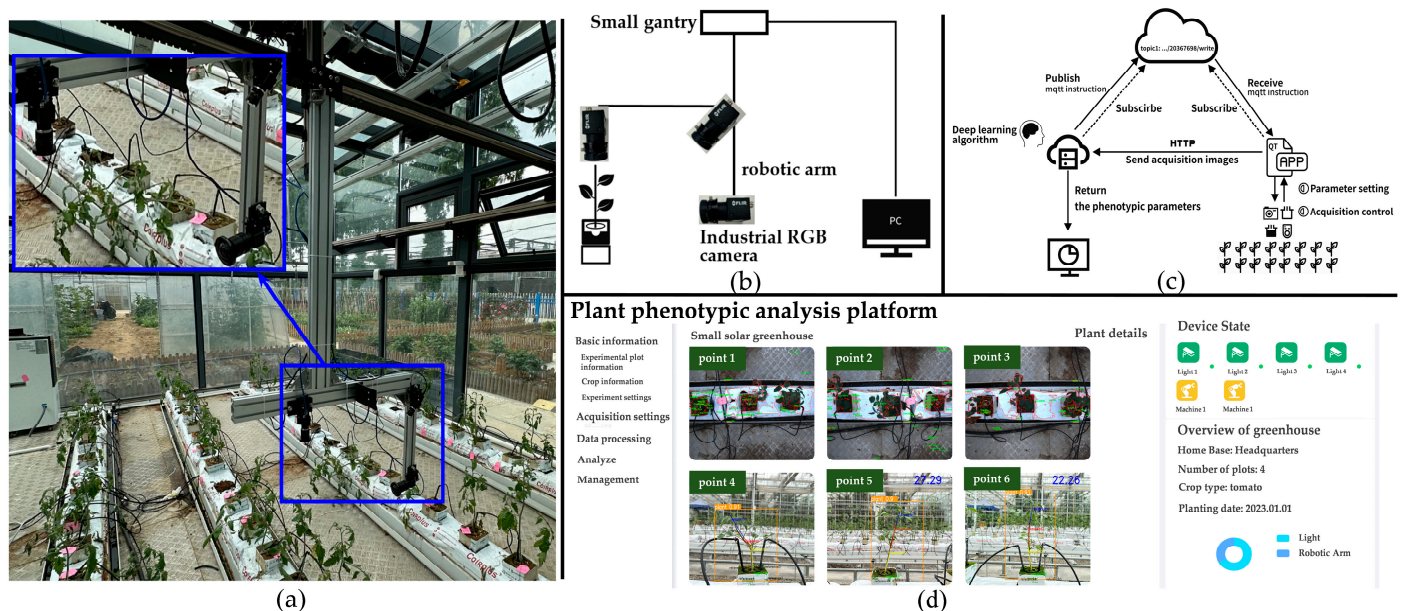


Figure 1. Plant phenotyping platform. (a) Physical diagram of the plant phenotyping platform; (b) Schematic diagram of the plant phenotyping platform; (c) The working flow of the cloud-based software system of the plant phenotyping platform; (d) A cloud-based plant phenotyping software system.

2.2. Image Pre-Processing and Dataset Preparation

2.2.1. Data Collection

The data collection experiments were conducted at two greenhouses of the Jiangsu Academy of Agricultural Sciences, Nanjing, China (32.03°N , 118.87°E), in 2022–2023. In order to study the extraction algorithms applicable to the extraction of phenotypic traits such as plant height and internode length of tomato seedlings in greenhouse environments with realistic and complex backgrounds, our dataset should be diverse and needs to fully consider the different situations in greenhouse environments, and for this reason we collected the image data of tomato seedlings at different times and under different light conditions. At the same time, considering that different tomato cultivars will have some

differences in growth structure, we selected five tomato cultivars for experimental research to improve the generalization performance of the model.

Then, we acquired 686 profile images of tomato seedlings at different time points and under different lighting conditions by the custom-developed high-throughput plant phenotyping platform. These images were captured at a resolution of 2448×2048 pixels. To enhance the diversity of the dataset, an additional 303 side-view images of multiple tomato varieties with a resolution of 4032×3024 pixels were collected in another greenhouse using a mobile phone. After meticulously removing duplicates and blurry images, a final set of 800 side-view images representing a range of tomato cultivars was selected (Table 1). At the same time, in order to more intuitively show the data differences between different tomato varieties and under lighting conditions, we provide the presentation of part of the dataset (Figure 2).

Table 1. Basic information of the tomato image dataset.

Tomato Varieties	The Number of Plants
FengZhu	300
Ruby	260
Cherry Tomato	140
White Tomato	80
Pink Brandywine Tomato	20
Total	800

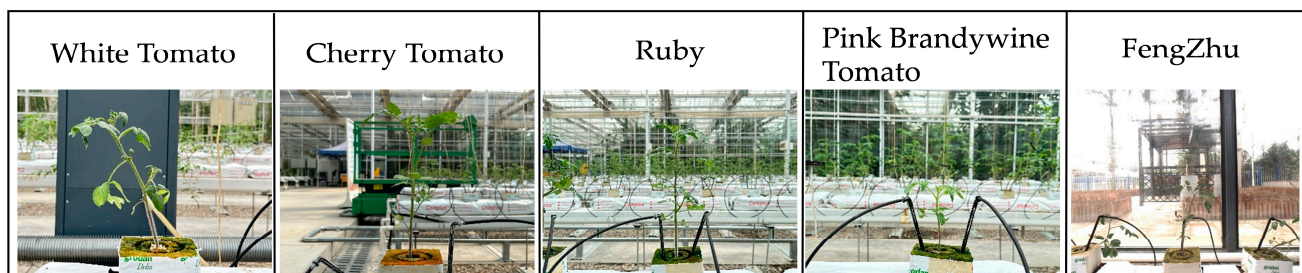


Figure 2. Tomato seedling side-view dataset.

In Figure 2 we can see that the White Tomato has more branches, the Cherry Tomato and Ruby are less curved than other tomato seedlings, and FengZhu and the White Tomato have similar structures with larger leaves at the seedling stage.

2.2.2. Truth Measurement Method

During the image data acquisition process, we measured the real value of each tomato plant (tomato plant height, internode length) manually and synchronously with a tape measure. Measurement standards follow agricultural norms, and the height of tomato plants is determined by measuring a tape along the main stem from the base to the tip (not the relative height of the tip to the ground). The rule for measuring pitch is given by the diagram in Figure 3, i.e., the length between each node defines the distance between two adjacent nodes on the main stem.

2.2.3. Data Annotation

Image annotation, encompassing both image class labels and keypoint annotations, was conducted using the open-source tool LabelMe [50]. The annotated data were initially formatted as a VOC dataset and subsequently converted to the YOLO format to facilitate model training. The dataset underwent meticulous annotation to delineate bounding boxes and keypoints, with particular emphasis on accurately reflecting plant height and internode length keypoints. The annotation process involved precise definition of a single bounding

box, multiple keypoints, and four reference points for each image (Figure 4). The keypoints for extracting plant height were primarily selected based on the inflection points along the curvature of plant's main stem to enable accurate height estimation for tomato seedlings (Figure 4a). For the internode length/number keypoints, each plant branching node was annotated with a single keypoint (Figure 4b).

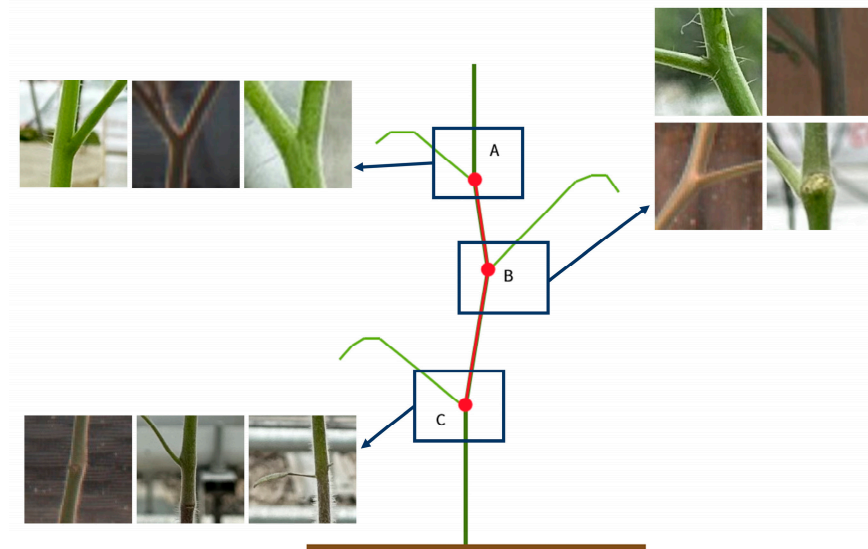


Figure 3. Schematic diagram of the internode length measurement. A, B and C represent the stem nodes on the plant.

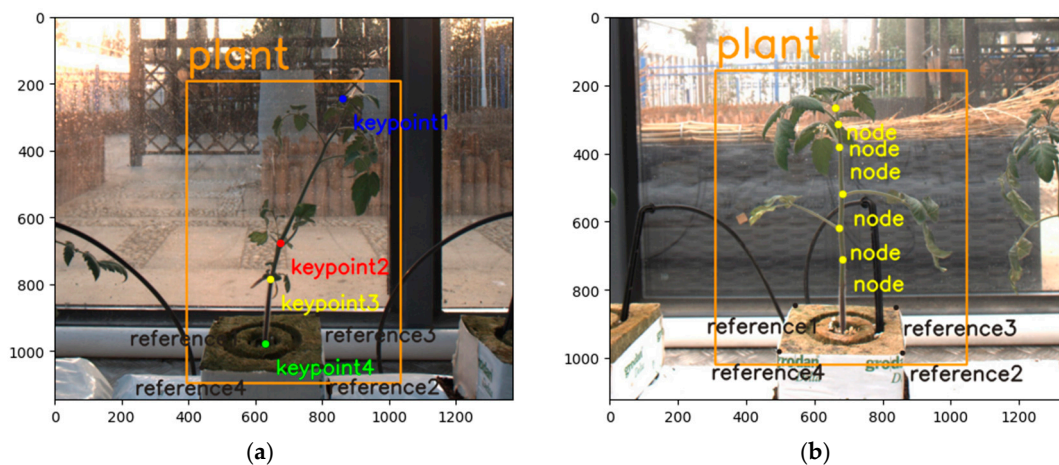


Figure 4. Annotation visualization. (a) Keypoints annotation for plant height; (b) Keypoints annotation for internode length.

2.2.4. Phenotypic Parameters Estimation

The complete dataset was divided into training, validation, and test sets in an 8:1:1 ratio. The annotated dataset was used to train the model for estimating plant height and internode length. Subsequently, the trained model weights were applied to the test dataset. Uniform square rockwool blocks were employed as reference objects. Plant height was then calculated using Equation (1) based on the detected plant and its corresponding keypoints, and the i -th internode length d_i was determined by Equation (2):

$$H_{plant} = \frac{d_{real} \times (l_{K1-2} + l_{K2-3} + l_{K3-4})}{(l_{R1-2} + l_{R3-4}) \times \frac{1}{2}} = \frac{2d_{real} \times (l_{K1-2} + l_{K2-3} + l_{K3-4})}{(l_{R1-2} + l_{R3-4})} \quad (1)$$

$$d_i = \frac{2d_{real} \times l_i}{(l_{R1-2} + l_{R3-4})} \quad (2)$$

where d_{real} represents the actual distance between reference 1 and reference 2 (reference 3 and reference 4) on the rockwool block. $l_{K1-2} + l_{K2-3} + l_{K3-4}$ denotes the total distance between keypoints on the tomato main stem, l_{R1-2} , l_{R3-4} denotes the distance between the reference points of the square rockwool blocks, and l_i represents the distance between two nodes.

2.3. The IYOLOv8s-Pose Module

As a lightweight YOLOv8-pose architecture, YOLOv8s-pose features a streamlined parameter structure comprising three primary components: a backbone network, a neck network, and a prediction head. The backbone network employs convolutional operations to extract multi-scale features from RGB images. Subsequently, the neck network fuses these extracted features, typically employing a Feature Pyramid Network (FPN) architecture to aggregate low-level features into higher-level representations. Finally, the prediction head layer is responsible for utilizing three detector groups of varying sizes to selectively detect image content.

Considering the spatial characteristics and structural complexities of tomato plants, we proposed an improved YOLOv8s-pose (IYOLOv8s-pose) model for extracting plant traits from 2D images captured within an intricate greenhouse background. Despite the low computational requirements and high accuracy of YOLOv8s-pose, it encounters challenges in effectively capturing crucial feature information when confronted with the complex backgrounds and plant self-occlusion. To address these limitations, this study incorporates CBAM (Figure 5) between the c2f layers and convolutional layers within the model's network. This integration serves to focus the network's attention to critical feature details while accounting for both spatial and channel dimensions. By refining feature maps and mitigating information loss, CBAM significantly enhances detection performance.

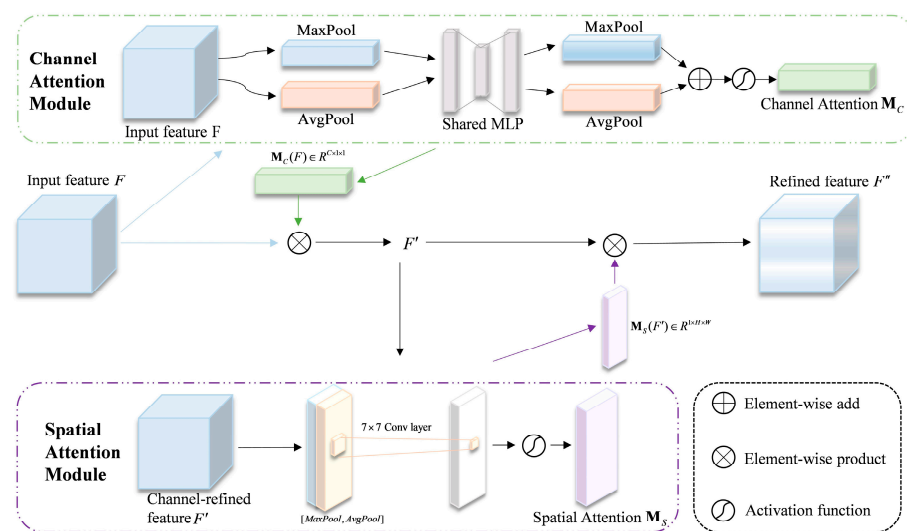


Figure 5. Schematic diagram of the CBAM model.

YOLOv8s-pose utilizes nearest-neighbor interpolation for upsampling. Although this method is computationally efficient and simple to implement, it only considers the grayscale value of the nearest pixel when determining the grayscale value of the upsampled point. It overlooks the influence of neighboring pixels and fails to leverage semantic information from the feature map, resulting in noticeable discontinuities in grayscale values after resampling and a significant degradation in image quality. To address this issue, we substituted the upsampling module with the CARAFE module. CARAFE aggregates contextual information from a broader receptive field, enhancing the utilization of seman-

tic information while maintaining lightweight properties. CARAFE's Kernel Prediction Module predicts a resampling kernel for each target location based on its content (Figure 6). Subsequently, the Content-Aware ReAssembly of FEatures Module employs this predicted kernel for feature reassembly.

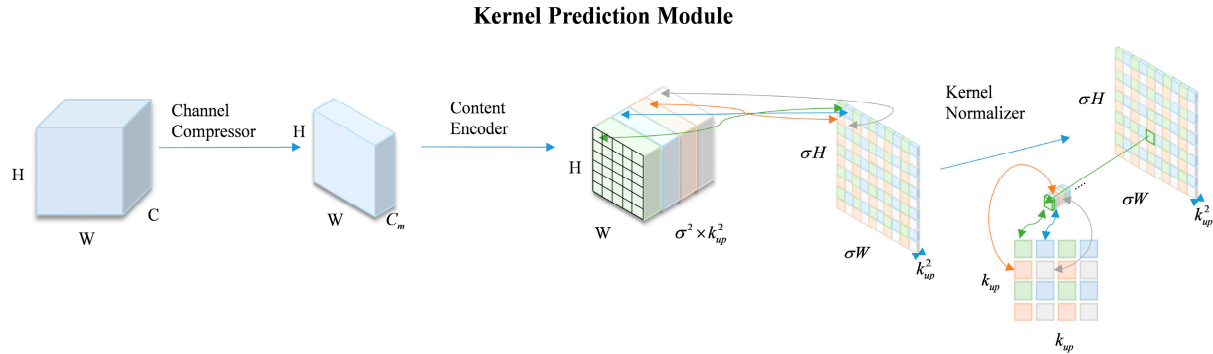


Figure 6. Kernel Prediction Module of CARAFE.

The improved IYOLOv8s-pose model demonstrates enhanced performance in terms of both computational complexity and model accuracy, effectively addressing the challenges associated with plant keypoint detection in complex backgrounds. The overall IYOLOv8s-pose model was shown in Figure 7. To ensure the consistency of the network input and to reduce the computational burden of the model, the input standard of the model is 640×640 , and the original image must be resized to the standard input network. To avoid the problem of unbalanced (distorted) target proportions caused by direct stretching of the image, the model adopts the method of the YOLOv8 base model. This means that the scale is equal (width/height up to 640; the rest is filled with background).

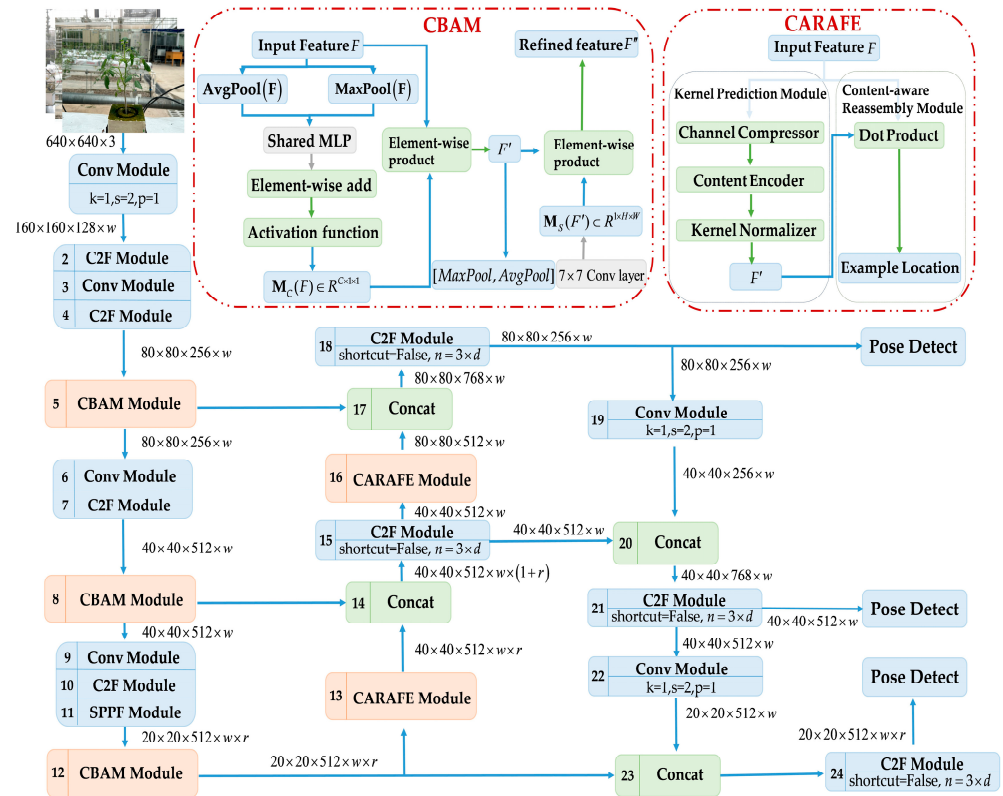


Figure 7. Structure of the IYOLOv8s-pose model.

2.3.1. Model Training

Network training was conducted on the Ubuntu operating system using the PyTorch framework. Our server was equipped with an Intel Core i9-9900k CPU and an NVIDIA A100-PCIE-40GB GPU. We utilized CUDA 11.0, CUDNN 8.0.1, Python 3.8.5, and the torch-2.0.1+cu117 environment. Recognizing the critical impact of hyperparameters on model performance [51], we employed specific settings: an initial learning rate and a cyclic learning rate of 0.01, 200 epochs, a batch size of 16, and the SGD optimizer with a momentum of 0.937 to accelerate convergence.

2.3.2. Evaluation Indicators

Model performance was evaluated based on *Precision*, *Recall*, *Accuracy*, and mean average precision (*mAP*) at two Intersection over Union (*IoU*) thresholds: 0.5 and 0.95 (*mAP*₅₀, *mAP*₅₀₋₉₅). The accuracy of our method in extracting plant phenotypic parameters was assessed using the root mean squared error (*RMSE*), relative root mean squared error (*rRMSE*), and the coefficient of determination (R^2), calculated between the ground truth measurements and the predicted values.

(1) Precision, Recall, and Accuracy

$$precision = \frac{TP}{TP + FP} \quad (3)$$

$$recall = \frac{TP}{TP + FN} \quad (4)$$

$$accuracy = \frac{TN + TP}{TN + FP + TP + FN} \quad (5)$$

TP = True positive, TN = True negative, FP = False positive, FN = False negative.

(2) Mean average precision (*mAP*)

The *mAP* metric measures the mean average precision across multiple object classes and various *IoU* thresholds by averaging *AP* values for each class to provide a comprehensive assessment of detection performance considering both object class and localization accuracy (*IoU*).

$$AP = \int_0^1 P(R) dR$$

$$mAP = \frac{1}{C} \sum_{i=1}^C AP_i \quad (6)$$

where $P(R)$ represents the precision P as a function of recall R , commonly referred to as the P - R curve; AP denotes the area under the P - R curve; and C represents the number of object detection categories.

(3) *RMSE*, *rRMSE*, and R^2

The *RMSE* was used to determine the accuracy of our proposed method in extracting plant parameters. The formula for calculating the *RMSE* was provided in Equation (7). Furthermore, the *rRMSE* (Equation (8)), which quantifies the deviation between predicted and actual values, was employed as a complementary accuracy measure. R^2 was used to measure the fit of the model to the data and was calculated as Equation (9).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_{pi} - x_{ri})^2} \quad (7)$$

$$rRMSE = RMSE / \bar{x}_{ri} \quad (8)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (x_{ri} - x_{pi})^2}{\sum_{i=1}^n (x_{ri} - \bar{x}_{ri})^2} = 1 - \frac{MSE}{Var} \quad (9)$$

where n represents the number of samples, x_{pi} denotes the predicted values from the model, x_{ri} represents the manually measured true values, and $\bar{x}_{ri}$ represents the mean value.

3. Results

3.1. Ablation Experiment

To assess the efficacy of the enhanced IYOLOv8s-pose network model in detecting keypoints of tomato plants, ablation experiments were conducted on each module, and the results were summarized in Table 2. The results indicated that incorporating the CBAM module and substituting the upsampling operator with CARAFE in the baseline YOLOv8s-pose model led to a minor increment in the number of floating-point operations (GFLOPs) during training and inference, while only marginally increasing the model's parameter. This demonstrated that the computational efficiency of the improved model was maintained and potentially enhanced. In contrast to the original YOLOv8s-pose, the addition of CBAM effectively improved the ability of keypoints detection. In particular, it is helpful to improve the ability to extract complex features. For the detection of main stem bending points, the *Precision* rose from 99.8% to 100%. For the detection of stem nodes, the *Accuracy*, *Precision*, *Recall*, *mAP50*, and *mAP50-95* improved significantly by 9.2%, 0.9%, 5.0%, 2.9%, and 8.0%, respectively. Replacing the upsampling operator with CARAFE significantly enhanced both *Recall* and *mAP50*. For the detection of main stem bending points, *Recall* increased from 99% to 100%, and *mAP50-95* rose from 93.6% to 94.1%. Notably, for the detection of stem nodes, *Precision*, *Recall*, and *mAP50* improved substantially, by 4.3%, 12.9%, and 5.2%, respectively. Compared with the baseline model YOLOv8s-pose, our optimized IYOLOv8s-pose model demonstrated comprehensive improvements in *Accuracy*, *Precision*, *Recall*, *mAP50*, and *mAP50-95*, particularly for the detection of main stem bending points, which improved by 11.0%, 5.4%, 13%, 5.2%, and 3.9%, respectively. After adding the attention mechanism and improving the upsampling method, our model has significantly improved the two indices of accuracy and recall, which means that the model reduces the cases of missed detection and improves the integrity of detection. These results demonstrated that our proposed method, IYOLOv8s-pose, effectively extracts crucial information from tomato seedlings in complex greenhouse environments, resulting in more accurate measurements of the plant height, internode length, and node count.

Table 2. Results of ablation experiments.

Detection Object	CBAM	CARAFE	Parameters	GFLOPs	Accuracy (%)	Precision (%)	Recall (%)	mAP50 (%)	mAP@50-95 (%)
Bending points of main stem	✓	✓	11,424,443	29.6	95.1	99.8	99.0	99.4	93.6
			11,769,697	29.9	95.1 (+0.0)	100 (+0.2)	97.9 (−1.1)	99.4 (+0.0)	93.8 (+0.2)
			11,589,451	30.0	97.0 (+1.9)	99.8 (+0.0)	100 (+1.0)	99.5 (+0.1)	94.1 (+0.5)
			11,934,705	30.2	98.0 (+2.9)	99.9 (+0.1)	100 (+1.0)	99.5 (+0.1)	95.3 (+1.7)
Nodes of main stem	✓	✓	11,626,046	30.4	87.0	94.4	87.0	94.3	48.7
			11,971,300	30.7	96.2 (+9.2)	95.3 (+0.9)	92.0 (+5.0)	97.2 (+2.9)	56.7 (+8.0)
			11,791,054	30.7	91.0 (+4.0)	98.7 (+4.3)	99.9 (+12.9)	99.5 (+5.2)	48.4 (−0.3)
			12,136,308	31	98.0 (+11.0)	99.8 (+5.4)	100 (+13.0)	99.5 (+5.2)	52.6 (+3.9)

Note: “✓” indicates that the corresponding improvement strategy was included in the network.

To further validate the accuracy of IYOLOv8s-pose in extracting stem-related phenotypic parameters, we compared the model's output with manually measured values for plant height, internode length, and node count. Correlation analyses were conducted to assess the relationship between estimated and ground truth values (Figures 8 and 9). For plant height estimation (Figure 8), IYOLOv8s-pose achieved an *RMSE* of 0.48 cm, an *rRMSE* of 2%, and an *R²* of 0.98, demonstrating high precision. Notably, the integration of both the CBAM and CARAFE modules contributed to these improvements. Compared with the baseline YOLOv8s-pose, IYOLOv8s-pose exhibited a 6% increase in *R²*, a 50% reduction in the *RMSE*, and a 60% decrease in the *rRMSE*. When tasked with the more intricate extraction of internode length, IYOLOv8s-pose achieved an *RMSE* of 0.63 cm and an *R²* of 0.78 compared to the actual internode length values (Figure 9). This indicates a 10%

improvement in R^2 , a 13.8% reduction in the $RMSE$, and a 12.5% decrease in the $rRMSE$ compared to YOLOv8s-pose. Furthermore, IYOLOv8s-pose demonstrated an accuracy of 78% in detecting the number of stem nodes (Figure 9).

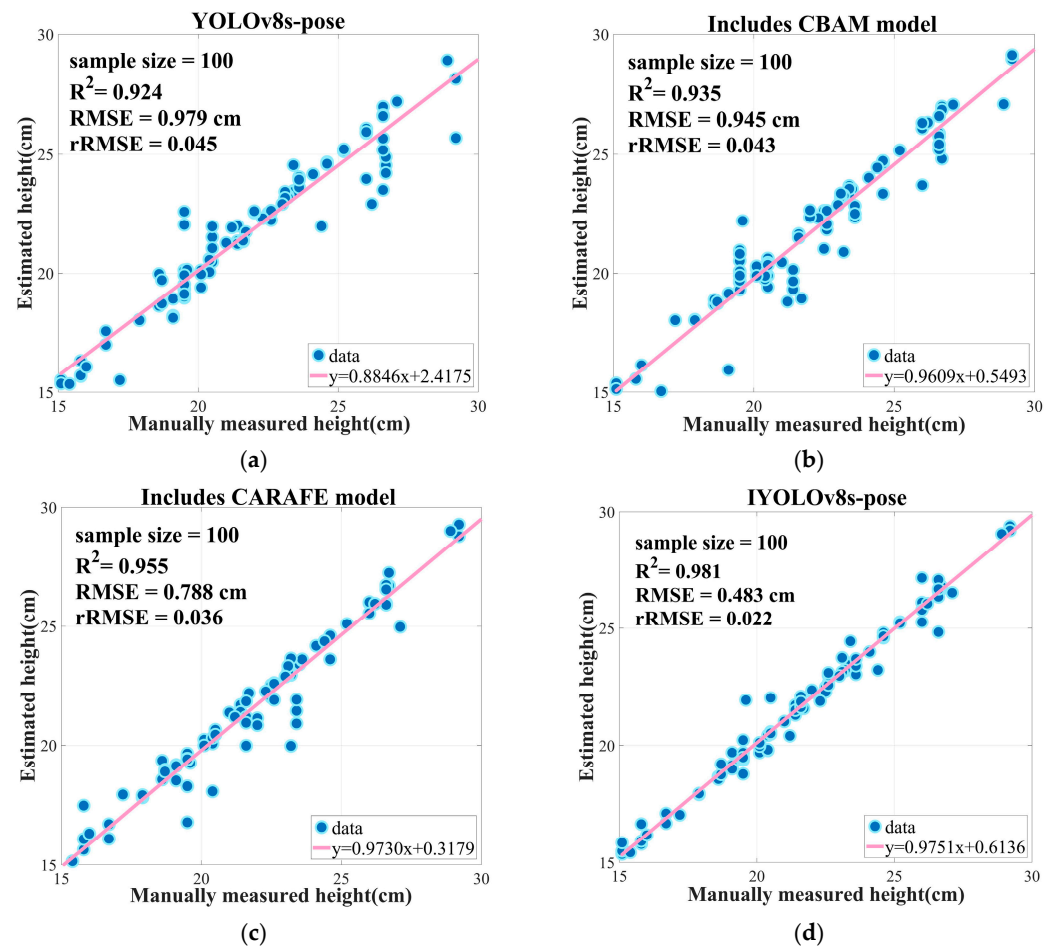


Figure 8. (a–d) Correlation analyses of the relationship between the measured and estimated values of tomato height.

To more intuitively demonstrate the effectiveness of the human pose estimation algorithm for the acquisition of plant stem-related parameters, we visualized and analyzed the results of stem keypoints detection. As depicted in Figure 10, ideal detection keypoints show the target detection points of the plant, and the true value in the image is obtained by manual measurement with a tape measure. In the plant height estimation task, we expect to detect the bending and inflection points of the plant, and the connection between the keypoints should extend the main stem as far as possible to restore the actual measurement effect, as shown in Figure 10a. The more the detected keypoints match the main trunk, the more accurate the plant height output will be. Similarly, in the internode length estimation task, the keypoints we hope to detect are the nodes defined during the actual measurement. The closer the detected keypoints in Figure 10b are to the actual trunk nodes, the more accurate the internode length output will be.

The results are shown in Figure 10. IYOLOv8s-pose detected keypoints that exhibited only minor deviations from the target locations. The combined analysis of the estimation results of plant height (Figure 10a) and internode length and number (Figure 10b,c) confirmed the feasibility of adapting human pose estimation algorithms for accurately acquiring stem-related phenotypic parameters in tomato seedlings. The ablation study emphasized the superior performance of our proposed method in effectively handling complex stem information extraction, outperforming both the baseline model and models

incorporating only a single improvement module. These results collectively demonstrated that IYOLOv8s-pose enables the non-destructive and rapid extraction of stem-related phenotypic parameters of tomato seedlings.

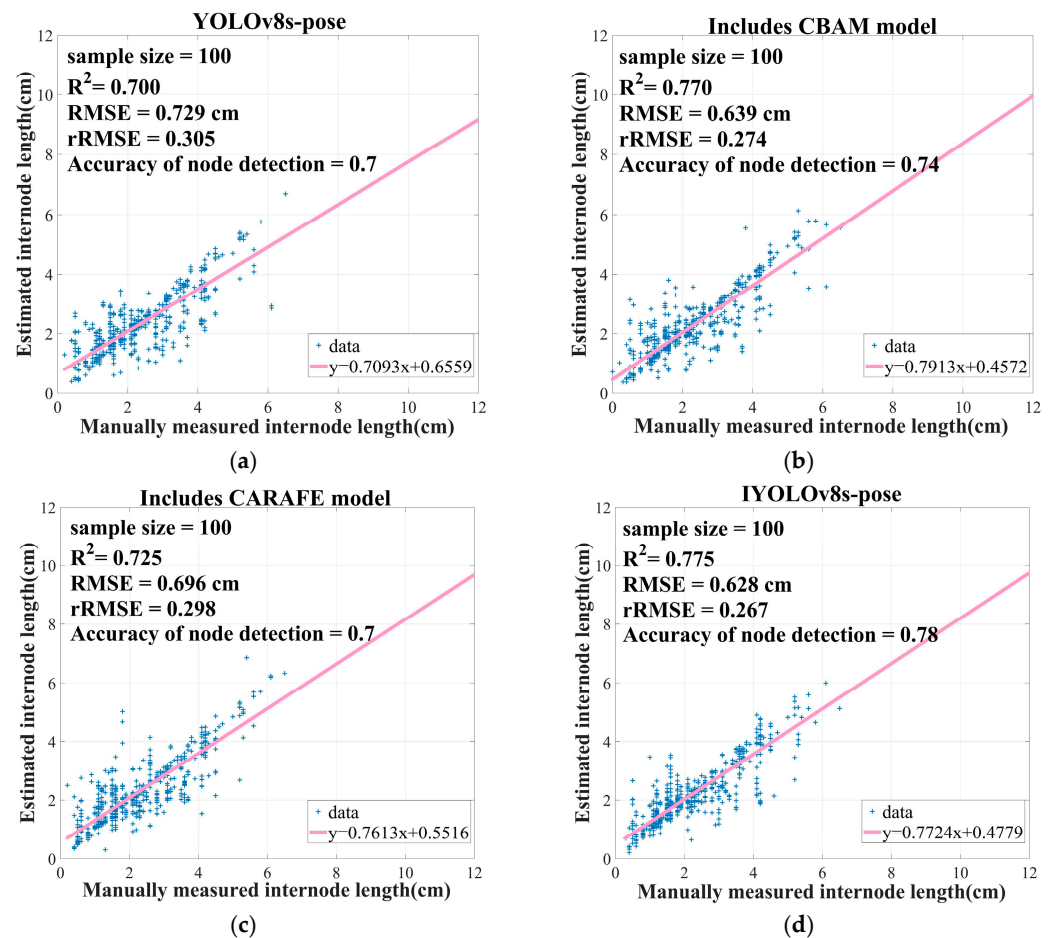


Figure 9. (a–d) Correlation analyses of the relationship between the measured and estimated values of tomato internode length.

3.2. Mainstream Algorithm Comparison

To evaluate the effectiveness of the enhanced IYOLOv8s-pose network model in detecting keypoints of tomato plants, we compared it with YOLOv5s6-pose and YOLOv7-tiny-pose. YOLOv5s6-pose is a model with a similarly lightweight parameter structure from the YOLO-pose family. Additionally, our improved model, IYOLOv8s-pose, was contrasted with a YOLOv7-tiny-pose model based on the lightweight YOLOv7-tiny architecture (Table 3). In addition to novel heatmap-free methods, in order to better prove its performance superiority in keypoint estimation tasks, we selected the model RTMPose (Real-Time Multi-Person Pose Estimation) [52], which is also suitable for real-time analysis of application scenarios, for experimental comparison and analysis. RTMPose, proposed in 2023, is a top-down method based on a thermal map. It uses an unsupervised learning method called MoCo v2 to improve the accuracy and speed of pose estimation, and it also has a strong anti-interference capability to accurately detect and track human poses in complex background environments. For the experimental comparison, we again chose the lightweight version RTMPose-s.

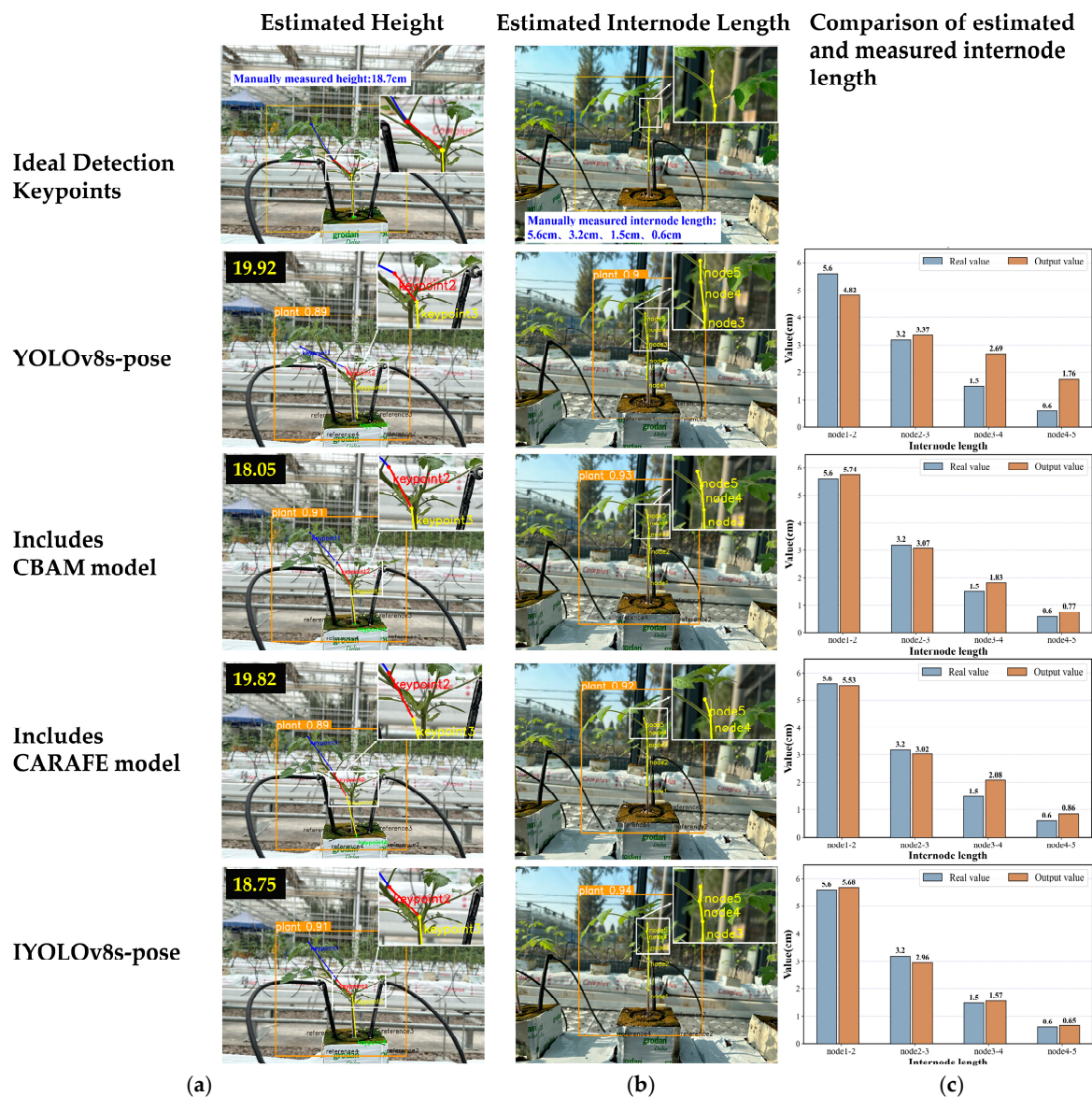


Figure 10. Comparison of the estimation results of different algorithms. (a) Visual comparison of the detection results for plant height estimation. (b) Visual comparison of the detection results for internode length and number estimation. (c) Comparison of the estimated and measured values of internode length; the x-axis represents the internode length: for example, node1–2 represents the internode length between stem nodes 1 and 2.

The results demonstrated that our proposed model has fewer parameters than YOLOv5s6-pose and demonstrated superior computational efficiency compared to RTMPose-s, YOLOv5s6-pose, and YOLOv7-tiny-pose. Furthermore, IYOLOv8s-pose exhibited notable improvements in *Accuracy*, *Precision*, *mAP*, and *Recall*. For detecting plant bending points, the *Precision* of IYOLOv8s-pose reached 99.8%, surpassing that of RTMPose-s, YOLOv5s6-pose, and YOLOv7-tiny-pose by 0.1%, 0.9%, and 0.8%, respectively; the *mAP50-95* of IYOLOv8s-pose reached 95.3%, surpassing that of RTMPose-s, YOLOv5s6-pose, and YOLOv7-tiny-pose by 2.6%, 21%, and 20.4%, respectively. In addition, the *Accuracy* has also been significantly improved: the *Accuracy* of IYOLOv8s-pose reached 98.0%, surpassing that of RTMPose-s, YOLOv5s6-pose, and YOLOv7-tiny-pose by 1.5%, 5%, and 6.5%, respectively. For the identification of tomato seedling stem nodes, the *Precision*, *Recall*, *mAP50*, and *mAP50-95* of our model surpassed those of both YOLOv5s6-

pose and YOLOv7-tiny-pose by 5.4%, 13%, 5.2%, and 3.9%, and 3.5%, 23.1%, 8.1%, and 2.8%, respectively.

Table 3. Results of comparative experiments.

Detection Object	Model	Parameters	GFLOPs	Accuracy (%)	Precision (%)	Recall (%)	mAP50 (%)	mAP50-95 (%)
Bending points of main stem	RTMPose-s	5,295,809	8.5	96.5	99.7	100	98.0	92.7
	YOLOv5s6-pose	15,010,856	20.3	93.0	98.9	99.9	99.4	74.3
	YOLOv7-tiny-pose	7,868,132	16.8	91.5	99.0	100	99.5	74.9
	IYOLOv8s-pose	11,934,705	30.0	98.0	99.8	100	99.5	95.3
Nodes of main stem	RTMPose-s	5,521,610	9.6	99.6	96.9	93.8	97.0	51.7
	YOLOv5s6-pose	15,114,860	20.5	99.5	94.4	87.0	94.3	48.7
	YOLOv7-tiny-pose	7,940,951	17.0	99.5	96.3	76.9	91.4	49.8
	IYOLOv8s-pose	12,136,308	31.0	100	99.8	100	99.5	52.6

It is worth noting that YOLOv8s-pose belongs to a one-stage model, i.e., YOLOv8s-pose can simultaneously complete target detection and pose estimation in one step, while RTMPose requires two steps: first, target detection is performed and then pose estimation is performed. Therefore, although it performs well in terms of experimental accuracy, it increases the complexity and time cost of training. We only report the metrics for the pose estimation models and omit the object detection models in Table 3.

To further validate the accuracy of IYOLOv8s-pose in predicting stem-related phenotypic parameters, we conducted a correlation analysis by comparing the model's output with manually measured values for plant height, internode length, and node count (Figures 11 and 12). Compared to RTMPose-s, YOLOv5s6-pose, and YOLOv7-tiny-pose, our proposed IYOLOv8s-pose exhibited a 64.7% (0.89 cm), 67.9% (1.03 cm), and 65.7% (0.93 cm) reduction in the RMSE, a 65.1%, 68.1%, and 65.6% decrease in the rRMSE, and an 14%, 15%, and 14% increase in R^2 , respectively, for plant height estimation. For internode length estimation (Figure 12), IYOLOv8s-pose exhibited an improvement of 12.9% (0.09 cm), 27.9% (0.24 cm), and 26.5% (0.23 cm) in the RMSE; 15.5%, 23.9%, and 27.2% in the rRMSE; 5.6%, 5.2%, and 10% in R^2 , compared to RTMPose-s, YOLOv5s6-pose, and YOLOv7-tiny-pose, respectively. Furthermore, our method exhibited an improvement of 14%, 24%, and 22% in detection accuracy for the number of stem nodes when compared to RTMPose-s, YOLOv5s6-pose, and YOLOv7-tiny-pose, respectively, showcasing its capability to handle more complex keypoint detection tasks compared to the other algorithms. According to the linear correlation analysis, the correlation coefficients of different algorithms are relatively small, while the RMSE has a large difference, with the highest surface phase difference of 67.9%. It can be seen that different pose estimation algorithms show good correlation and can be used to extract some phenotypic parameters of tomato plants at seedling stage. However, the effect of the different models on the test set is very different, reflecting the difference in their generalization performance.

To visually illustrate the superiority of our proposed algorithm in extracting phenotypic parameters, we conducted a visual analysis of estimation results (Figure 13). As shown in Figure 13, IYOLOv8s-pose exhibited minor deviations between predicted keypoints and target locations. The comprehensive analysis of plant height (Figure 13a) and internode length and number detection results (Figure 13b,c) demonstrated the superior performance of our enhanced IYOLOv8s-pose algorithm in extracting stem-related phenotypic parameters from tomato seedlings.

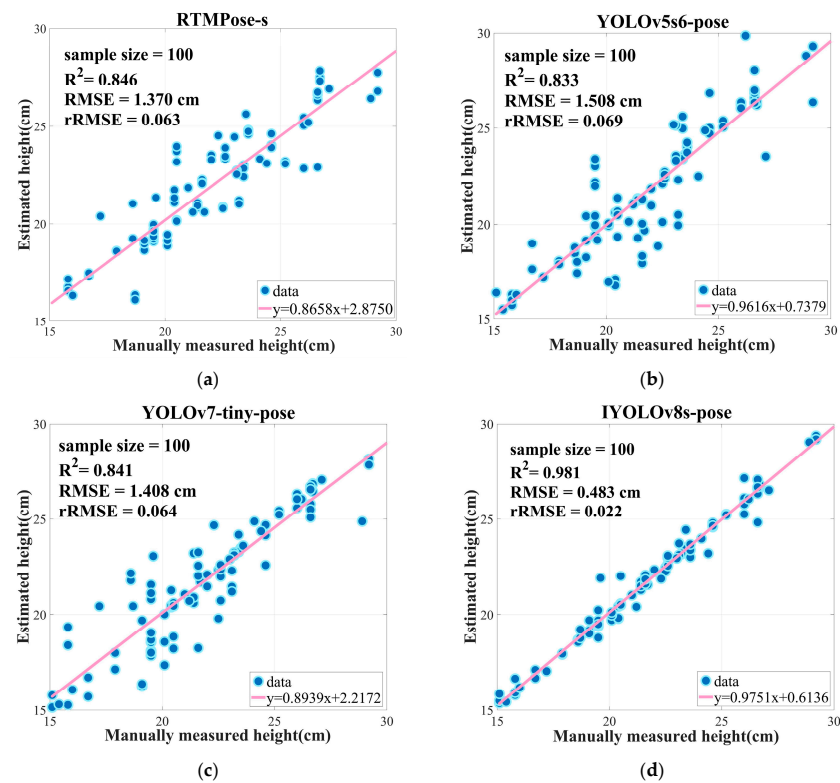


Figure 11. (a–d) Correlation analyses of the relationship between the measured and estimated values of tomato height using different algorithms.

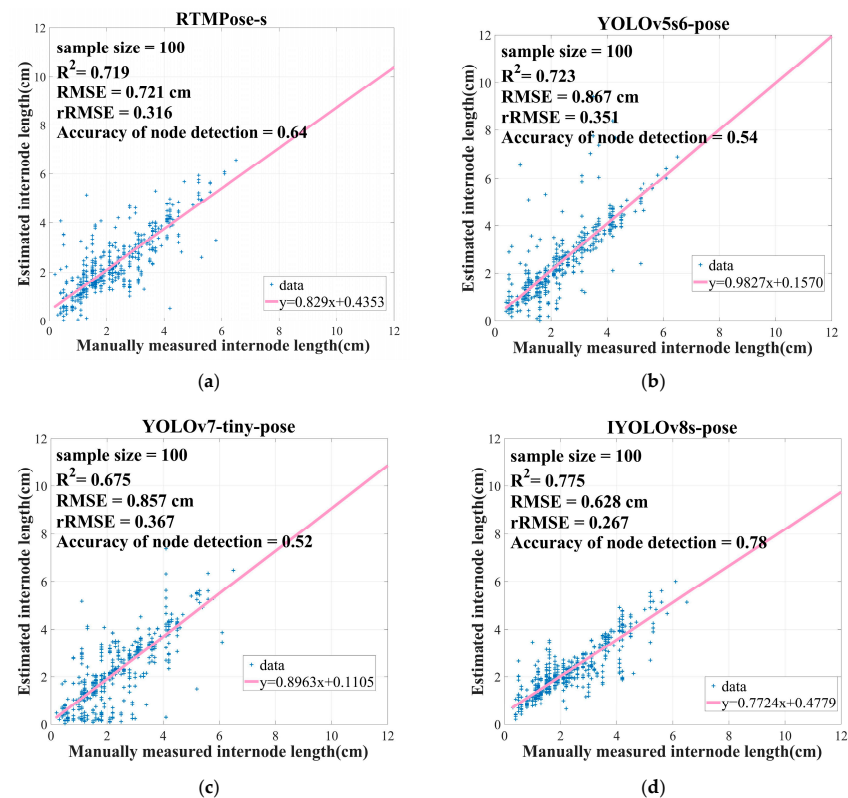


Figure 12. (a–d) Correlation analyses of the relationship between the measured and estimated values of tomato internode length using different algorithms.

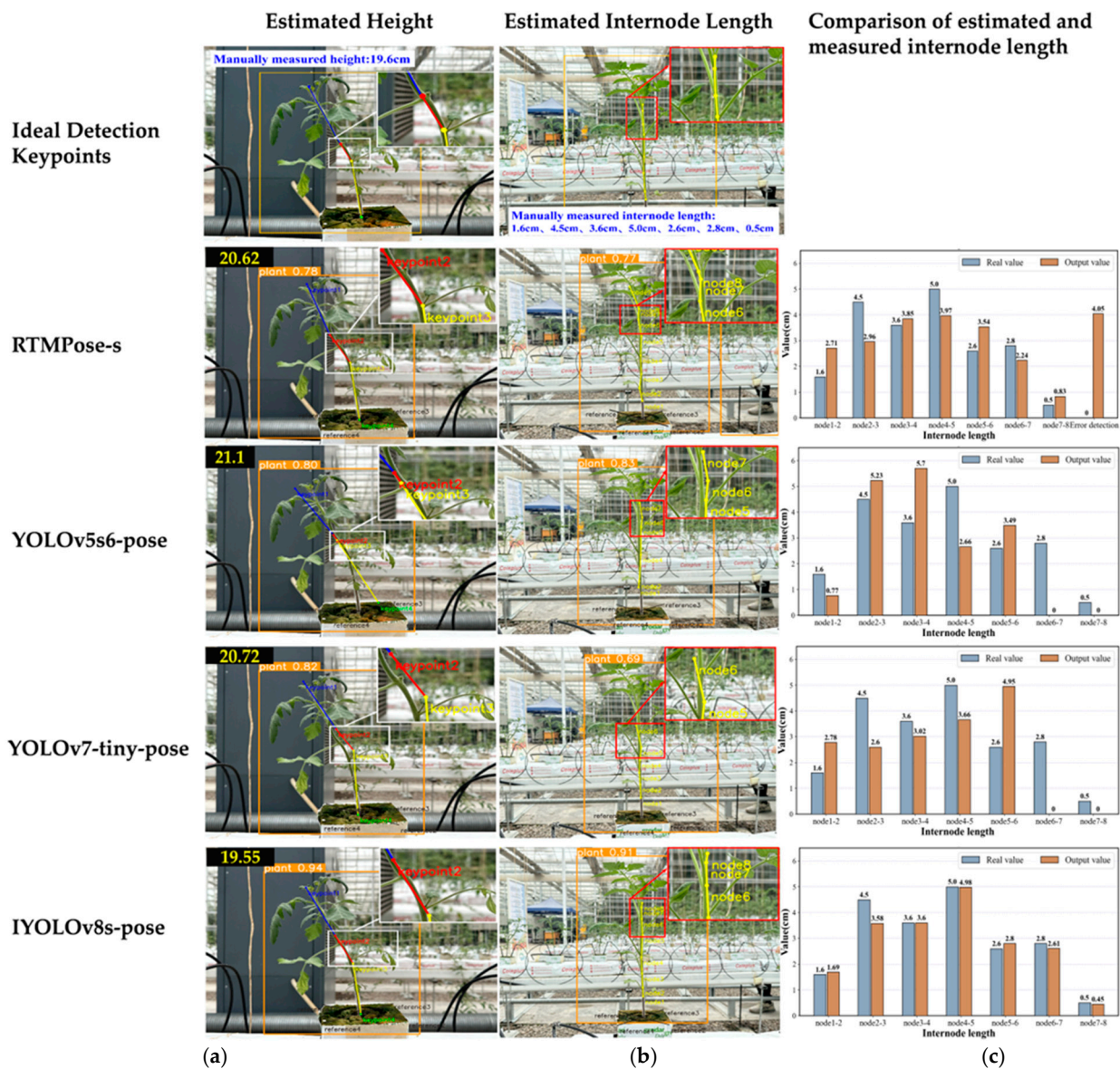


Figure 13. Comparison of the estimation results of different algorithms. (a) Visual comparison of the detection results for plant height estimation. (b) Visual comparison of the detection results for internode length and number estimation. (c) Comparison of the estimated and measured values of internode length; the x-axis represents the internode length: for example, node1–2 represents the internode length between stem nodes 1 and 2.

4. Discussion

The application of utilizing 2D pose estimation techniques in phenotypic analysis allows for the extraction of keypoints from plants, thereby facilitating the reconstruction of structural characteristics and extraction of architectural phenotypic parameters. This approach opens avenues for extensive research and practical applications. The real-time extraction of plant architecture traits can be utilized to develop growth models that guide decision making in greenhouse management. In addition, future endeavors could be made to extend 2D keypoint detection deep into 3D reconstruction to achieve a more comprehensive characterization of plant structure.

The plant height extracted using IYOLOv8s-pose is of great value for assessing plant growth dynamics. By continuously monitoring a single plant and capturing side-view images at different time points, we can monitor changes in plant height over time, enabling the generation of growth curves (Figure 14). This analysis can provide insights for precise

and standardized cultivation practices in greenhouse tomato production. Furthermore, analyzing seedling growth patterns can assist in selecting tomato varieties with desirable traits, such as rapid growth, a high yield, and strong stress resistance, thereby contributing to efficient tomato breeding and improvement programs.

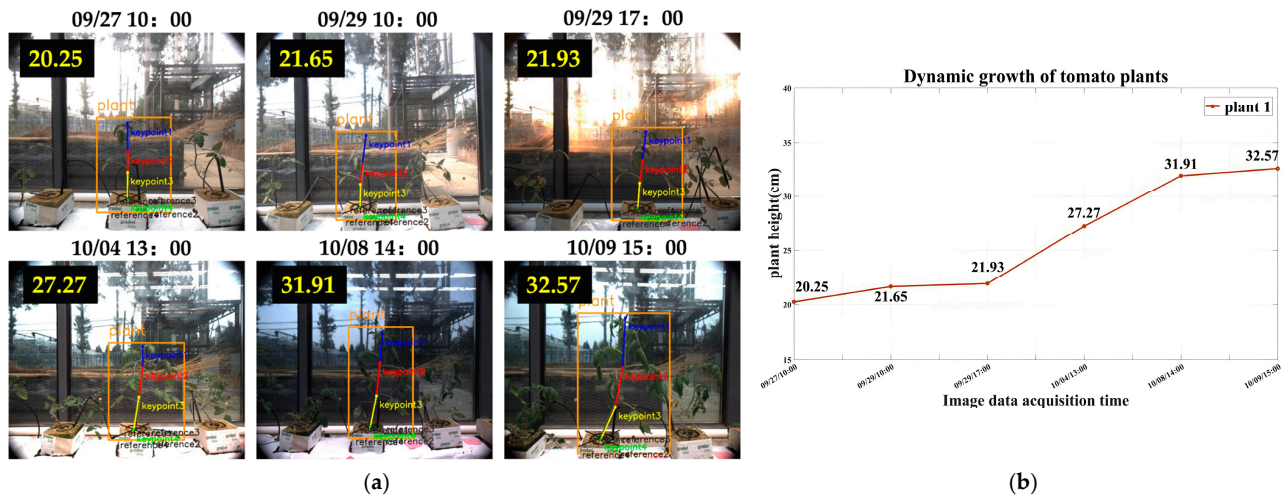


Figure 14. The plant height extraction of the same tomato in different time periods (a) and dynamic growth analysis diagram (b).

Considerable advancements have been achieved in the field of 3D human pose estimation from 2D images [53–55]. Martinez et al. [56] utilized an hourglass network for 2D skeleton pose estimation, followed by two fully connected layers to predict 3D joint coordinates. Tian et al. [57] proposed a method for 3D human reconstruction from monocular images by estimating human pose and shape. Depth information can be indirectly estimated by utilizing cues from grayscale images, such as shading, texture variations, and motion patterns. Moreover, depth information could be integrated alongside the extracted plant keypoints to facilitate rudimentary 3D reconstructions. In this study, we utilized grayscale information from 2D images to obtain an initial depth estimation. The obtained depth information, combined with our 2D keypoint detection technique, was used to reconstruct the 3D skeleton of the plant structure (Figure 15).

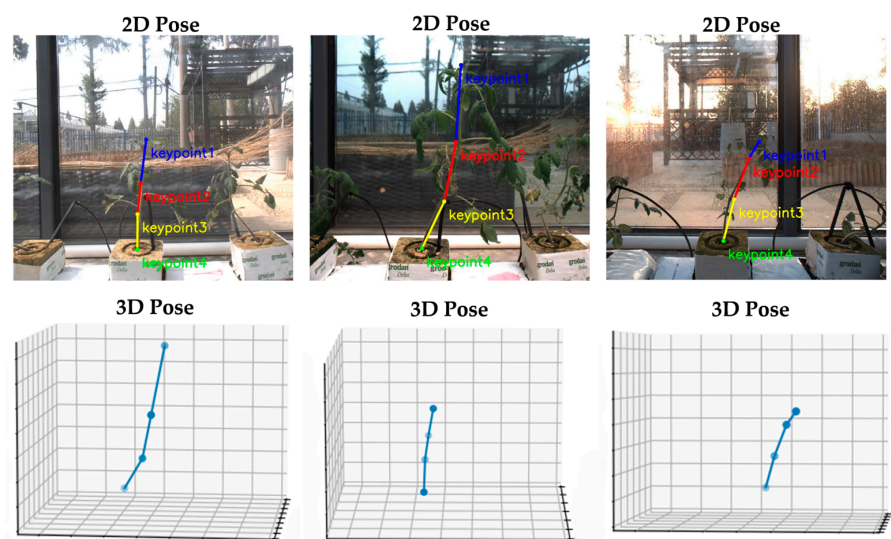


Figure 15. Three-dimensional skeleton reconstruction of the main stems of three randomly selected tomato plants.

5. Conclusions

A novel plant phenotyping approach based on 2D pose estimation was proposed in this study. The proposed model, IYOLOv8s-pose, was utilized to extract the main stem skeleton structure of tomato plants, enabling the efficient, accurate, and real-time extraction of plant height and internode length of tomato seedlings under realistic greenhouse conditions with complex backgrounds. Considering the spatial characteristics and structural features of plant growth, our methodology prioritizes keypoint detection, providing a novel perspective on 2D image-based plant phenotyping. In comparison to segmentation-based models, which often necessitate time-consuming and labor-intensive data annotation, our proposed method efficiently extracts phenotypic parameters for diverse tomato cultivars with significantly reduced annotation requirements. Moreover, in contrast to existing 2D image-based phenotyping methods reliant on object detection, our algorithm demonstrates superior accuracy in identifying and monitoring plants within complex backgrounds while capturing specific structural characteristics of plants. In comparison with 3D image-based phenotyping models, our approach offers a cost-effective and time-efficient alternative without compromising accuracy. This study introduces an innovative direction for machine vision-based phenotyping algorithms, demonstrating its potential for broader application in extracting diverse phenotypic parameters across various plant species. Looking ahead, several aspects of this method necessitate further refinement. Improving the accuracy of internode length calculations represents a substantial opportunity for enhancement. Additionally, exploring strategies to reduce computational time complexity while expanding the types of extracted phenotypic parameters will be a central focus of future research.

Author Contributions: Conceptualization, Y.C. and Q.W.; methodology, Y.C.; supervision, N.R. and Q.W.; software, Y.C.; validation, A.H., L.Z. and C.Q.; formal analysis, Y.C. and S.Z.; investigation, L.Z. and C.Q.; data curation, Y.C., Q.W. and N.R.; visualization, Y.C. and A.H.; writing—original draft preparation, Y.C.; writing—review and editing, Y.C., N.R. and Q.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (32201664), the Key R&D Plan of Jiangsu Province (BE2021379), and the Agricultural Independent Innovation of Jiangsu Province (CX(22)5008).

Data Availability Statement: The raw data will be made available on the request.

Conflicts of Interest: The authors declare that we have no conflicts of interest.

References

1. Tian, Y.; Gao, L. Theory and Technology for Facility Cultivation of High-quality Tomato. *China Veg.* **2021**, *2*, 30–40.
2. Kehili, M.; Choura, S.; Zammel, A.; Allouche, N.; Sayadi, S. Oxidative stability of refined olive and sunflower oils supplemented with lycopene-rich oleoresin from tomato peels industrial by-product, during accelerated shelf-life storage. *Food Chem.* **2018**, *246*, 295–304. [[CrossRef](#)] [[PubMed](#)]
3. Li, J.; Shen, L.-X.; Zhang, K.-L.; Li, B.-N. Effects of drip irrigation soluble organic fertilizer on tomato growth and quality. *Water Sav. Irrig.* **2022**, *12*, 37–41.
4. Yamamoto, K.; Guo, W.; Ninomiya, S. Node detection and internode length estimation of tomato seedlings based on image analysis and machine learning. *Sensors* **2016**, *16*, 1044. [[CrossRef](#)]
5. Sabina, S.; Jithesh, M.N. Mechanical rubbing of tomato internode influence stem growth, improve tensile strength but negatively impact flavonoid levels. *Adv. Hortic. Sci.* **2020**, *34*, 373–380.
6. Boogaard, F.P.; Rongen, K.S.; Kootstra, G.W. Robust node detection and tracking in fruit-vegetable crops using deep learning and multi-view imaging. *Biosyst. Eng.* **2020**, *192*, 117–132. [[CrossRef](#)]
7. Hu, J.; Li, G.; Mo, H.; Lv, Y.; Qian, T.; Chen, M.; Lu, S. Crop node detection and internode length estimation using an improved YOLOv5 model. *Agriculture* **2023**, *13*, 473. [[CrossRef](#)]
8. Kim, E.-Y.; Park, S.-A.; Park, B.-J.; Lee, Y.; Oh, M.-M. Growth and antioxidant phenolic compounds in cherry tomato seedlings grown under monochromatic light-emitting diodes. *Hortic. Environ. Biotechnol.* **2014**, *55*, 506–513. [[CrossRef](#)]
9. Sibomana, I.; Aguyoh, J.; Opiyo, A. Water stress affects growth and yield of container grown tomato (*Lycopersicon esculentum* Mill) plants. *Gjbb* **2013**, *2*, 461–466.
10. Huang, C.; Li, Y.; Luo, S.; Yang, W.; Zhu, L. Cotton seedling leaf traits extraction method from 3d point cloudbased on structured light imaging. *Trans. Chin. Soc. Agric. Mach.* **2019**, *50*, 243–248.

11. Tardieu, F.; Cabrera-Bosquet, L.; Pridmore, T.; Bennett, M. Plant phenomics, from sensors to knowledge. *Curr. Biol.* **2017**, *27*, R770–R783. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Das Choudhury, S.; Samal, A.; Awada, T. Leveraging image analysis for high-throughput plant phenotyping. *Front. Plant Sci.* **2019**, *10*, 508. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Wang, Q.; Fan, X.; Zhuang, Z.; Tjahjadi, T.; Jin, S.; Huan, H.; Ye, Q. One to All: Towards a Unified Model for Counting Cereal Crop Heads Based on Few-Shot Learning. *Plant Phenomics* **2024**. [\[CrossRef\]](#)
14. Qiu, R.; Wei, S.; Zhang, M.; Li, H.; Sun, H.; Liu, G.; Li, M. Sensors for measuring plant phenotyping: A review. *Int. J. Agric. Biol. Eng.* **2018**, *11*, 1–17. [\[CrossRef\]](#)
15. Yano, K.; Morinaka, Y.; Wang, F.; Huang, P.; Takehara, S.; Hirai, T.; Ito, A.; Koketsu, E.; Kawamura, M.; Kotake, K. GWAS with principal component analysis identifies a gene comprehensively controlling rice architecture. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 21262–21267. [\[CrossRef\]](#)
16. Zhong, H.; Liu, S.; Meng, X.; Sun, T.; Deng, Y.; Kong, W.; Peng, Z.; Li, Y. Uncovering the genetic mechanisms regulating panicle architecture in rice with GPWAS and GWAS. *BMC Genom.* **2021**, *22*, 86. [\[CrossRef\]](#)
17. Lu, H.; Tang, L.; Whitham, S.A.; Mei, Y. A robotic platform for corn seedling morphological traits characterization. *Sensors* **2017**, *17*, 2082. [\[CrossRef\]](#)
18. Huang, X.; Zheng, S.; Zhu, N. High-throughput legume seed phenotyping using a handheld 3D laser scanner. *Remote Sens.* **2022**, *14*, 431. [\[CrossRef\]](#)
19. Wang, Y.; Hu, S.; Ren, H.; Yang, W.; Zhai, R. 3DPhenoMVS: A low-cost 3D tomato phenotyping pipeline using 3D reconstruction point cloud based on multiview images. *Agronomy* **2022**, *12*, 1865. [\[CrossRef\]](#)
20. Zhu, T.; Ma, X.; Guan, H.; Wu, X.; Wang, F.; Yang, C.; Jiang, Q. A method for detecting tomato canopies' phenotypic traits based on improved skeleton extraction algorithm. *Comput. Electron. Agric.* **2023**, *214*, 108285. [\[CrossRef\]](#)
21. Wei, B.; Ma, X.; Guan, H.; Yu, M.; Yang, C.; He, H.; Wang, F.; Shen, P. Dynamic simulation of leaf area index for the soybean canopy based on 3D reconstruction. *Ecol. Inform.* **2023**, *75*, 102070. [\[CrossRef\]](#)
22. He, W.; Ye, Z.; Li, M.; Yan, Y.; Lu, W.; Xing, G. Extraction of soybean plant trait parameters based on SfM-MVS algorithm combined with GRNN. *Front. Plant Sci.* **2023**, *14*, 1181322. [\[CrossRef\]](#)
23. Qiu, R.; Zhang, M.; He, Y. Field estimation of maize plant height at jointing stage using an RGB-D camera. *Crop J.* **2022**, *10*, 1274–1283. [\[CrossRef\]](#)
24. Wang, Y.; Liu, Q.; Yang, J.; Ren, G.; Wang, W.; Zhang, W.; Li, F. A Method for Tomato Plant Stem and Leaf Segmentation and Phenotypic Extraction Based on Skeleton Extraction and Supervoxel Clustering. *Agronomy* **2024**, *14*, 198. [\[CrossRef\]](#)
25. Zhang, W.; Li, X.; Wan, H.; Li, C.; Zhang, W.; Jin, B.; Liu, Y. Stem and leaf segmentation of maize plant image based on skeleton extraction and binary tree analysis. *J. Henan Agric. Sci.* **2020**, *49*, 166.
26. Arefi, A.; Motlagh, A.M.; Mollazade, K.; Teimourlou, R.F. Recognition and localization of ripen tomato based on machine vision. *Aust. J. Crop Sci.* **2011**, *5*, 1144–1149.
27. Chen, S.; Liao, Y.; Lin, F.; Huang, B. An improved lightweight YOLOv5 algorithm for detecting strawberry diseases. *IEEE Access* **2023**, *11*, 54080–54092. [\[CrossRef\]](#)
28. Das Choudhury, S.; Goswami, S.; Bashyam, S.; Samal, A.; Awada, T. Automated stem angle determination for temporal plant phenotyping analysis. In Proceedings of the 2017 IEEE International Conference on Computer Vision Workshops, Venice, Italy, 22–29 October 2017; pp. 2022–2029.
29. ZHANG, R.; Zongwei, J.; Ruibin, W.; Simin, Y.; Zhang, J. Phenotypic parameter extraction for wheat ears based on an improved Mask-RCNN algorithm. *INMATEH-Agric. Eng.* **2022**, *66*, 267–278. [\[CrossRef\]](#)
30. Gehan, M.A.; Fahlgren, N.; Abbasi, A.; Berry, J.C.; Callen, S.T.; Chavez, L.; Doust, A.N.; Feldman, M.J.; Gilbert, K.B.; Hodge, J.G. PlantCV v2: Image analysis software for high-throughput plant phenotyping. *PeerJ* **2017**, *5*, e4088. [\[CrossRef\]](#)
31. Guan, H.; Fu, C.; Zhang, G.; Li, K.; Wang, P.; Zhu, Z. A lightweight model for efficient identification of plant diseases and pests based on deep learning. *Front. Plant Sci.* **2023**, *14*, 1227011. [\[CrossRef\]](#)
32. Pound, M.P.; Atkinson, J.A.; Townsend, A.J.; Wilson, M.H.; Griffiths, M.; Jackson, A.S.; Bulat, A.; Tzimiropoulos, G.; Wells, D.M.; Murchie, E.H. Deep machine learning provides state-of-the-art performance in image-based plant phenotyping. *Gigascience* **2017**, *6*, gix083. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Fu, J.; Shi, J.; Zhang, J.; Yang, W. Extraction of single leaf length and plant height of potted wheat based on Mask R-CNN. *J. Jilin Agric. Univ.* **2021**, *43*, 163–170.
34. Guo, X.; Li, J.; Zheng, L.; Zhang, M.; Wang, M. Soybean plant phenotypic parameters were obtained by Re-YOLOv5 and region search algorithm. *Trans. Chin. Soc. Agric. Eng.* **2022**, *38*, 186–194.
35. Fan, Z.; Sun, N.; Qiu, Q.; Li, T.; Zhao, C. A high-throughput phenotyping robot for measuring stalk diameters of maize crops. In Proceedings of the 2021 IEEE 11th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER), Jiaying, China, 27–31 July 2021; pp. 128–133.
36. Minervini, M.; Scharr, H.; Tsafaris, S.A. Image analysis: The new bottleneck in plant phenotyping [applications corner]. *IEEE Signal Process. Mag.* **2015**, *32*, 126–131. [\[CrossRef\]](#)
37. Jain, A.; Tompson, J.; Andriluka, M.; Taylor, G.W.; Bregler, C. Learning human pose estimation features with convolutional networks. *arXiv* **2013**, arXiv:1312.7302 2013.

38. Chen, Y.; Wang, Z.; Peng, Y.; Zhang, Z.; Yu, G.; Sun, J. Cascaded pyramid network for multi-person pose estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18 June 2018; pp. 7103–7112.
39. Sun, K.; Xiao, B.; Liu, D.; Wang, J. Deep high-resolution representation learning for human pose estimation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 11–15 June 2019; pp. 5693–5703.
40. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. Ssd: Single shot multibox detector. In Proceedings of the Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; pp. 21–37.
41. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
42. Maji, D.; Nagori, S.; Mathew, M.; Poddar, D. Yolo-pose: Enhancing yolo for multi person pose estimation using object keypoint similarity loss. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 2637–2646.
43. Wang, C.Y.; Bochkovskiy, A.; Liao, H.Y.M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Vancouver, BC, Canada, 17–24 June 2023; pp. 7464–7475.
44. Berrigan, E.M.; Wang, L.; Carrillo, H.; Echegoyen, K.; Kappes, M.; Torres, J.; Ai-Perreira, A.; McCoy, E.; Shane, E.; Copeland, C.D. Fast and efficient root phenotyping via pose estimation. *Plant Phenomics* **2024**, *6*, 0175. [\[CrossRef\]](#)
45. Li, X.; Sun, K.; Fan, H.; He, Z. Real-Time Cattle Pose Estimation Based on Improved RTMPose. *Agriculture* **2023**, *13*, 1938. [\[CrossRef\]](#)
46. Russello, H.; van der Tol, R.; Kootstra, G. T-LEAP: Occlusion-robust pose estimation of walking cows using temporal information. *Comput. Electron. Agric.* **2022**, *192*, 106559. [\[CrossRef\]](#)
47. Kim, T.; Lee, D.-H.; Kim, K.-C.; Kim, Y.-J. 2D pose estimation of multiple tomato fruit-bearing systems for robotic harvesting. *Comput. Electron. Agric.* **2023**, *211*, 108004. [\[CrossRef\]](#)
48. Wang, J.; Chen, K.; Xu, R.; Liu, Z.; Loy, C.C.; Lin, D. Carafe: Content-aware reassembly of features. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 3007–3016.
49. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
50. Russell, B.C.; Torralba, A.; Murphy, K.P.; Freeman, W.T. LabelMe: A database and web-based tool for image annotation. *Int. J. Comput. Vis.* **2008**, *77*, 157–173. [\[CrossRef\]](#)
51. Nguyen, H.-C.; Nguyen, T.-H.; Scherer, R.; Le, V.-H. YOLO series for human hand action detection and classification from egocentric videos. *Sensors* **2023**, *23*, 3255. [\[CrossRef\]](#)
52. Jiang, T.; Lu, P.; Zhang, L.; Ma, N.; Han, R.; Lyu, C.; Chen, K. RtmPose: Real-time multi-person pose estimation based on mmpose. *arXiv* **2023**, arXiv:2303.07399.
53. Zhao, L.; Peng, X.; Tian, Y.; Kapadia, M.; Metaxas, D.N. Semantic graph convolutional networks for 3d human pose regression. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 3425–3435.
54. Cai, Y.; Ge, L.; Liu, J.; Cai, J.; Cham, T.-J.; Yuan, J.; Thalmann, N.M. Exploiting spatial-temporal relationships for 3d pose estimation via graph convolutional networks. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 2272–2281.
55. Wandt, B.; Rosenhahn, B. RepNet: Weakly supervised training of an adversarial reprojection network for 3d human pose estimation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 7782–7791.
56. Martinez, J.; Hossain, R.; Romero, J.; Little, J.J. A simple yet effective baseline for 3d human pose estimation. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2640–2649.
57. Tian, Y.; Zhang, H.; Liu, Y.; Wang, L. Recovering 3d human mesh from monocular images: A survey. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*; IEEE: New York, NY, USA, 2023; pp. 15406–15425.

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.