Signal, Image and Video Processing (2025) 19:178
https://doi.org/10.1007/s11760-024-03691-3

ORIGINAL PAPER

Deep learning based radish and leaf segmentation for phenotype trait measurement

Nur Alam[1] · A. S. M. Sharifuzzaman Sagar[2] · L. Minh Dang[3] · Wenqi Zhang[1] · Han Yong Park[4] · Moon Hyeonjoon[1]

Received: 22 April 2024 / Revised: 12 August 2024 / Accepted: 16 September 2024

Abstract

The quantification of crop phenotyping traits is essential to understand growth patterns and increase production. Traditional methods of monitoring such properties, particularly in white radishes, are notably labor-intensive and inefficient, necessitating the development of more effective techniques. Moreover, there is a lack of radish dataset to enable monitoring of radish plant growth that combines radish roots and leaves. Addressing these challenges, the current study proposes a radish dataset combining both radish roots and leaves. In addition, we propose an automated approach through the implementation of deep learning and mathematical model that leverages high-resolution imagery for the measurement of white radish phenotype traits. The study utilized a modified Mask Region-based Convolutional Neural Networks (R-CNN) algorithm to accurately segment radish components, facilitating the measurement of leaf and root dimensions. The traditional backbone was improved by introducing a local–global attention mechanism in the feature extraction block. The feature pyramid network (FPN) is also improved by integrating a self-attention mechanism in the top layer. Moreover, we utilize Geometrical Morphological Analysis and the medial axis transform method to measure the height and width of the white radish phenotype traits. Extensive experiments revealed that our proposed modified Mask R-CNN model acquired a mean average precision of 96.3% for segmentation and a mean absolute error (MAE) of 0.51mm for phenotype traits measurement. Our proposed framework demonstrates a significant advancement in the measurement process in agricultural studies, offering a reliable alternative to traditional, time-consuming methods.

Keywords Deep learning · Segmentation · Phenotype trait · Measurement

1 Introduction

Advancements in Artificial Intelligence (AI) and Computer Vision provide opportunities to improve and maximize livestock and agricultural management, leading to increased production [1]. To increase productivity, monitoring plant biophysical traits and taking timely preventive measures to avoid problems are essential for increasing productivity. Monitoring plant growth is the process of identifying the biophysical properties of the plant and measuring the growth of the biophysical properties. Using computer vision methods, researchers can apply advanced image analysis methods

✉ Moon Hyeonjoon
hmoon@sejong.ac.kr

Nur Alam
nur0756@sju.ac.kr

A. S. M. Sharifuzzaman Sagar
sharifsagar80@sejong.ac.kr

L. Minh Dang
minhdl@sejong.ac.kr

Wenqi Zhang
zwqzpq@sju.ac.kr

Han Yong Park
hypark@sejong.ac.kr

[1] Department of Computer Science and Engineering, Sejong University, Seoul 05006, Republic of Korea

[2] Department of Intelligent Mechatronics Engineering, Sejong University, Seoul 05006, Republic of Korea

[3] Department of Information and Communication Engineering and Convergence Engineering for Intelligent Drone, Sejong University, Seoul 05006, Republic of Korea

[4] Department of Bioresource Engineering, Sejong University, Seoul 05006, Republic of Korea

Published online: 03 January 2025

to capture and handle complex growth patterns that support accurate identification of plant's biophysical properties.

Computer vision has emerged as essential methods for the agricultural industry in recent times, providing precise responses to a variety of plant analysis related tasks [2]. These methods are widely applied in plant disease detection, plant growth stage classification, segmentation, plant phenotyping, plant growth monitoring, crop estimations, and so on. Additionally, the attention based convolutional neural networks (ACNN) model is used for accurate crop estimation, resolving issues related to insufficient field samples [3]. Segmentation techniques, such as Mask R-CNN, allow for the detailed outline of radish structures, facilitating tasks such as radish and leaf segmentation and phenotyping [4].

Radishes (Raphanus sativus L.) are widely cultivated root vegetables in Asia, especially in Korea, where they are considered a national vegetable. Radishes, especially the green-shouldered white variety favored for their taste and texture, are a significant ingredient in Korean cuisine, often eaten raw, pickled, or as a fundamental part of kimchi. Although existing research in Korea has focused on cultivation systems and disease resistance [5], thorough growth monitoring is essential to increase radish production, yield, marketability, and cultivation decisions. Traditionally, plant monitoring has been done using manual field measurements and UAV-based technologies [6], latter, despite certain difficulties with sensor calibration and weather, providing benefits in disease detection and prediction of crop growth using high-resolution imaging [7]. Smartphones have developed as sensible alternatives to capture agricultural data, supporting technologies such as light detection and ranging (LiDAR). Recent studies have utilized these methods to investigate crop disease challenge, genetic diversity, and various agronomic traits, underscoring the importance of in-depth phenotypic analysis for agricultural insight and genetic research. However, when it comes to crops such as radishes, current methods often face challenges in terms of precision and accuracy for detecting and segmenting various parts of radishes due to their complex backgrounds and abstract shapes, which later can cause incorrect measurements. Additionally, traditional measurement methods often face difficulties when measuring irregular shapes of radishes.

To overcome the shortcomings of current methods and improve the efficiency of radish analysis, as well as address this gap in underground crops, we first propose a challenging dataset of radish roots and leaves. We then propose a framework combining the deep learning-based segmentation method and the mathematical model for radish root and leaf with measurement of biophysical properties. Specifically, we introduce a modified Mask Region-based Convolutional Neural Network (R-CNN) model with a local–global attention mechanism. The architecture of the Feature Pyramid Network (FPN) has been further refined by incorporating a self-attention mechanism at the top level of its structure. This addition is designed to improve the ability of the model to process and refine feature representations through context-aware focused calculations. Moreover, the mask head module of the Mask R-CNN was replaced with the PointRend algorithm to improve the performance of the model. The median axis analysis and geometrical morphological analysis algorithms are introduced to accurately measure the biophysical properties of the plant extracted from the segmentation model.

The rest of the paper is structured as follows: In Sect. 2, we conduct a comprehensive review of important literature in several crops. Section 3 describes the methodology assigned for this study, along with a detailed description of the dataset used. The experimental results are presented and analyzed in Sect. 4. In Sect. 5, we compare our proposed system with the current state-of-the-art method. Finally, we conclude with a discussion of our findings, and the prospects for future research are in Sect. 6.

2 Related work

Computer vision and deep learning models are used to predict plant growth by analyzing images. For our objective, segmentation models are particularly used, as they can accurately describe the different components of the radish, such as leaves and roots, by predicting object masks. The real-life pixel density in the image is used to accurately measure the traits of the radish by referencing a ruler placed beside it for scale, as shown in Fig. 2b. The segmentation output and pixel density are then used to calculate traits like width and length. Using this data, regression forecasts future growth characteristics such as root width, root length, and other growth parameters.

2.1 Segmentation

In the process of intelligent plant management, Plant identification and growth prediction have drawn an enormous amount of attention. Plant growth information could be directly obtained by optical sensors [8]. Computer vision is frequently applied to detection of plants, predict growth, yield estimation, and other information [9]. Traditional machine vision algorithms are mainly based on color and threshold for plant identification, classification, and segmentation. Traditional machine vision algorithms primarily rely on color and threshold for crop identification, classification, and segmentation. A threshold-based image processing method to examine the effect of the density of the blossoms on the yield of apples. Mizushima and Lu [10] segmented the apples in the image using a combination of the maximum variance between classes method (Otsu) [11] and the support vector machine (SVM) [12]. However, these algorithms were efficient at identifying and segmenting the target, but they were ineffective in environments with complex backgrounds and variable lighting.

As big data processing technologies and GPU computing power continue to advance, an increasing number of deep learning algorithms are being used in agriculture, particularly in plant segmentation [13]. The method used by Dias et al. [14] is classified as semantic segmentation, which is limited to segmenting the area surrounding an apple flower rather than the individual apple flowers. To achieve plant segmentation in a changing light environment, Yu, Zhang, Yang, and Zhang [15] proposed Mask R-CNN [16] approach, which extended Faster R-CNN by combining the bounding box recognition branch with an object mask prediction branch. Using 100 images as a test dataset, the approach offered cutting-edge crop segmentation results from the research. Huang, Huang, Gong, Huang and Wang [17] improved the scoring criteria between the mask of the instance and ground truth and developed the Mask Scoring R-CNN model based on the network architecture of Mask R-CNN. The Mask Scoring R-CNN model boosted the Mask R-CNN's segmentation accuracy and reached a state-of-the-art level in target instance segmentation. However, because these algorithms use a set grid size 28×28 for mask prediction, they often tend to produce overly smoothed outputs for large objects. Although bottom-up approaches can provide more comprehensive results, in benchmark tests, they usually perform worse than region-based methods [18]. To address this, TensorMask [19], a sliding window method, predicts high-resolution masks for big objects. But compared to region-based methods, its accuracy trails significantly.

2.2 Phenotyping

The accurate monitoring and measurement of the physical and phenotyping characteristics of a plant is called phenotyping. Phenotyping is fundamentally vital in various aspects of crop development, including identifying ideal traits, validating genetic predictions, and improving the selection procedure. By analyzing and understanding the phenotypic characteristics of crops, researchers can make informed decisions about breeding programs, genetic modifications, and agronomic practices to improve crop productivity, resilience, and quality. In-field measurements involve manually collecting data directly from field crops or soil [20]. This includes measures such as soil moisture, leaf area index, and plant height. However, collecting data manually and measuring biophysical properties is time consuming, costly, and requires expertise in agriculture. Previous studies in high-throughput phenotyping research have focused mainly on the use of high-tech sensors such as LiDAR [21] and multi-view stereo cameras [22]. These sensors overcome the challenges of in-field measurements to capture detailed three-dimensional data of plants. Although these methods are effective in precisely measuring plant size, they are often expensive and require advanced phenotyping platforms. Liu et al. [23] mention that earlier research examined the manual measuring of phenotypic features and genetic diversity in a variety of plants and crops. But scalability and efficiency issues are often present in this research, especially when dealing with complex characteristics that require repeated, labor-intensive measurements or observations.

3 Materials and methods

3.1 Overall architecture of proposed framework

The overall structure of the proposed system is shown in Fig. 1, which can effectively segment and measure various phenotypic traits of a radish image. Radish data were used as the raw data set to train and test the proposed system. The train data set is then prepossessed and fed into the Modified Mask R-CNN model to learn the characteristics of the radishes of radish for efficient segmentation of their various components. Subsequently, the testing images were used to evaluate the model's ability to segment radish components. Additionally, exact measurements of the phenotypic traits of radishes are gained by determining the real-life pixel density, a process that includes detecting a ruler placed beside the radish. Then, real-life measurements for several radish phenotype features, such as width and length, can be acquired by utilizing the output masks produced by the radish segmentation model and the computed real-life pixel density. The system output is done by embedding segmentation and accurate measurement prediction at the same time.

3.2 Data collection and preprocessing

In the development of our radish dataset, the primary objective was to support the automatic monitoring of radish growth. The data collection captured place in a radish field situated in Kyonggi-do, Korea, over the period of September 2022 to February 2023. To make sure there is consistent lighting, we captured images within a one-hour frame targeted around solar noon, from 11:30 AM to 12:30 PM.

For the purpose of image calibration, we integrated an X-rite 4×6 color checkerboard with 24 colors in every image. The existence of this checkerboard in each image capture period provided a standard reference for accurate color calibration, which is essential for the accurate analysis of the images. The device used to collect the data set was a Samsung Galaxy S22 smartphone, understanding the advantage of its 50 megapixel rear camera feature that includes an f/1.8 aperture and improved autofocus capabilities, producing images

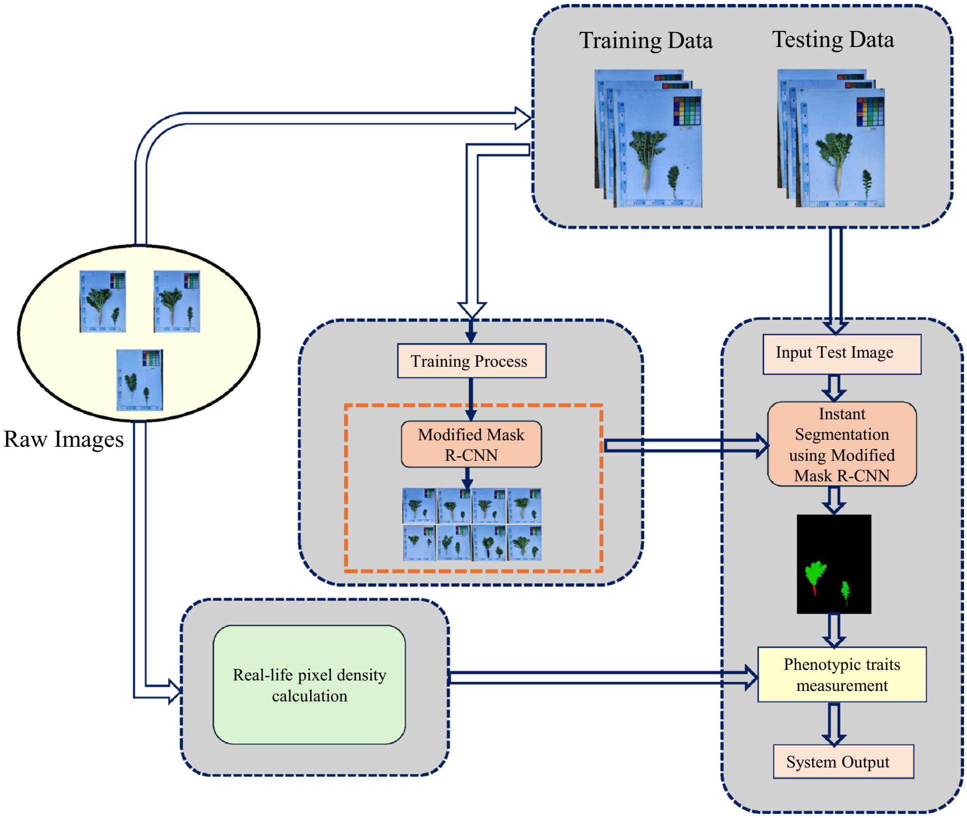

Training Data
Testing Data
Raw Images
Training Process
Modified Mask R-CNN
Input Test Image
Instant Segmentation using Modified Mask R-CNN
Real-life pixel density calculation
Phenotypic traits measurement
System Output

Fig. 1 Overview of the Training and Testing Process Using the Modified Mask R-CNN. The training data and testing data are first prepared and input into the training process. The Modified Mask R-CNN is then utilized for instant segmentation of the input test image, leading to the system output. The real-life pixel density calculation is done to acquire detailed measurements of phenotypic traits including root length, root width, leaf width, and leaf length

at 3000 × 4000 pixels. A total of 1100 high-resolution images were carefully compiled.

Each image received manual annotations with the assist of the LabelMe tool [24], with representations of the color checkerboard, radish fields, and carefully placed rulers to provide a precise assessment of the phenotypic characteristics. The resulting samples of our radish dataset, including the train images and numerical tests obtained from it, are shown in Fig. 2.

3.3 Radish and leaf segmentation

The Mask-RCNN, a deep learning model known for its capabilities in object detection and instance segmentation, represents a significant advancement in the field of computer vision [16]. Adjusting from the Faster-RCNN architecture, it introduces an extra branch granted to generating pixel-level object masks in combining with the existing bounding box recognition branch. Furthermore, Mask-RCNN stands out for its ease of training and adaptability across various computer vision applications. In this study, we utilize a modified Mask-RCNN network with six main components, as illustrated in Fig. 3.

3.3.1 Backbone

The ResNet-50 model in the Residual Networks family has significantly influenced deep learning by enabling the training of much deeper networks than previously possible due to its innovative use of residual connections. These connections help alleviate the problem of vanishing gradients, allowing information to flow through the network more effectively. Despite its successes, ResNet-50 has some limitations, particularly in terms of its convolutional blocks' ability to process complex patterns and understand long-range dependencies within the data. This is mainly because traditional convolutional layers focus on extracting local features, potentially overlooking the global context that is crucial for certain tasks.

To address these limitations, we introduce the Local Global Attention Module (LGAM) to enable ResNet-50's architecture to encompass a broader understanding of the input data. LGAM achieves this by integrating attention mechanisms that enable the network to weigh the importance of different features at both local and global scales. Our proposed LGAM not only enhances the model's ability to capture intricate patterns but also improves its performance in tasks requiring a nuanced comprehension of the entire scene. Therefore, LGAM effectively overcomes the inherent shortcomings of ResNet-50's convolutional blocks, paving

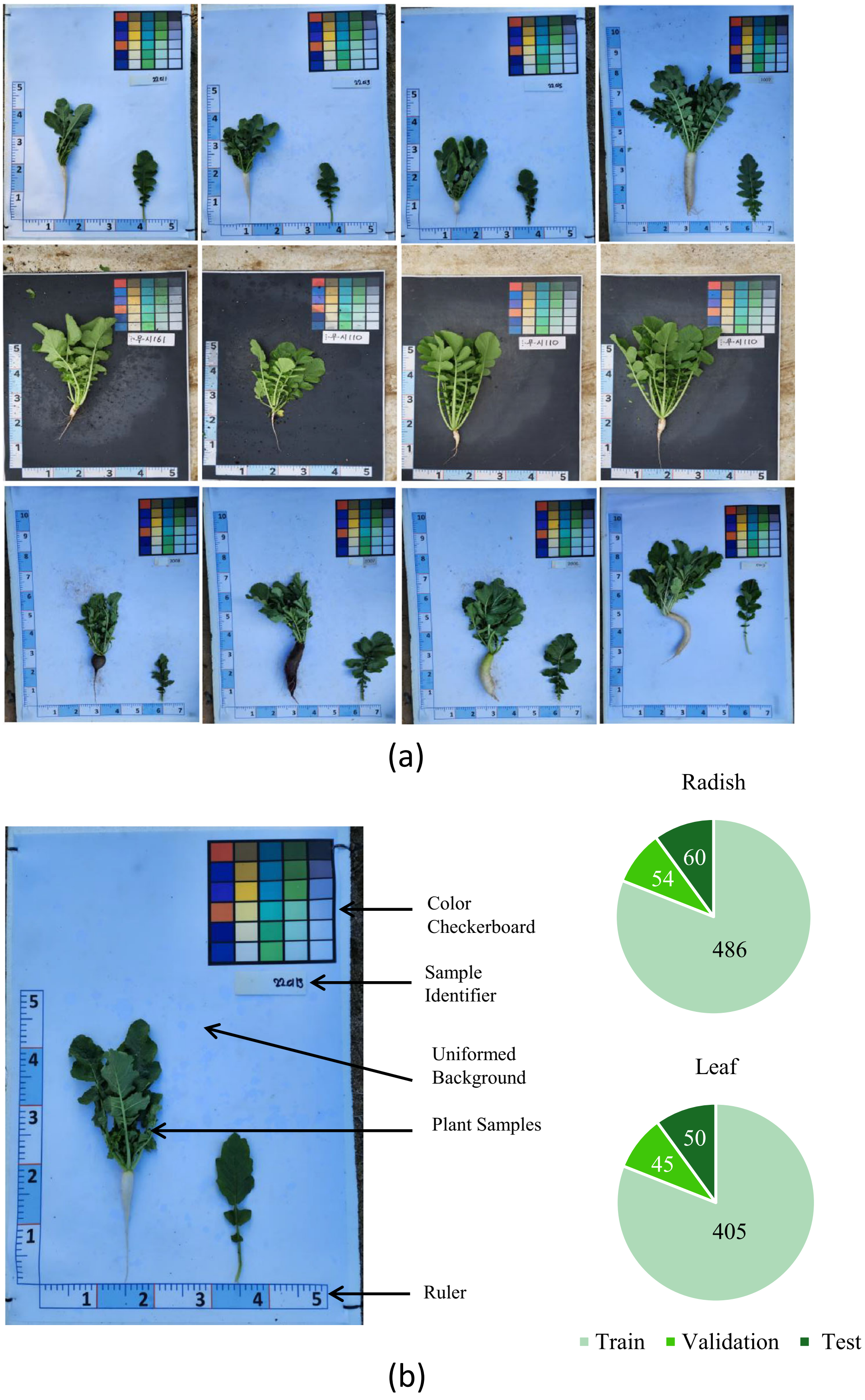


Fig. 2 Illustrating the data collection process, accompanied by details regarding the quantity of training and testing images used in the study. **a** sample images from our proposed dataset; **b** data preprocessing procedure to preserve uniformity over whole dataset. Images alignment relative to the camera was done using the same methodology used in our previous study [4]

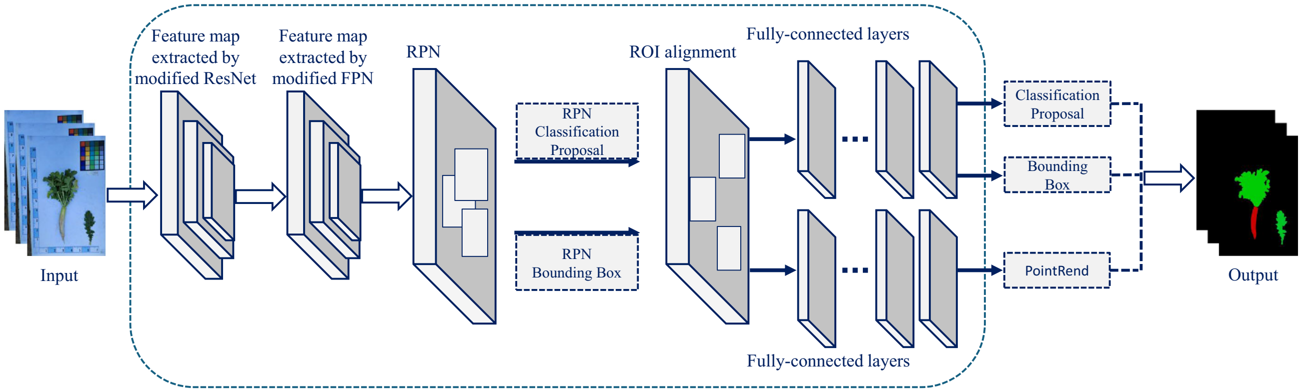


Fig. 3 Radish root and leaf segmentation using the modified Mask R-CNN method, highlighting a balance between efficiency and accuracy

the way for more sophisticated and capable deep learning models.

Figure 4 shows the architecture of the convolution block of the ResNet-50 model with our proposed LGAM. Figure 4(a) shows the modified convolutional block of ResNet-50 model. The convolutional block starts with the two fully convolutional layers with 1*1 and 3*3 kernel size, respectively. LGAM is then introduced to extract complex global and local feature information from input data followed by a 1*1 convolutional layer and batch normalization.

Figure 4(b) shows the proposed LGAM which includes two parallel local and global branches. The local branch begins with an Average Pooling layer that condenses the input feature map. Then we added two Deformable Convolutional layers [25], which adaptively adjust the spatial sampling locations and weights within the convolutional operation, allowing for dynamic receptive field that can capture complex spatial relationships. The deformable convolution can be represented as follows,

$$Y(p_0) = \sum_{p_n \in R} w(p_n) \cdot x(p_0 + p_n + \Delta p_n) \tag{1}$$

where Y is the output, X is the input, W is the weight, R defines the size of the convolutional kernel, p_0 is the position of the central pixel, p_n are the positions of the neighboring pixels, and Δp_n are the learnable offsets. The output of the local attention branch can be defined as follows,

$$Z_1 = \text{DConv}(\text{DConv}(\text{AvgPool}(X))) \tag{2}$$

The global branch employs a Multi-Layer Perceptron (MLP) which captures channel wise dependencies by processing each channel descriptor independently. The MLP, composed of fully connected layers, learns high-level feature representations and inter-channel relationships, which is particularly useful for identifying and emphasizing global, semantic information present across the channels. The MLP can be expressed as

$$Y = (W_2 \cdot \delta(W_1 \cdot A)) \tag{3}$$

where Y is the output, A is the pooled feature map, W_1 and W_2 are the weights of the MLP layers, δ is a non-linear activation function like ReLU. The output of the global branch can be defined as follows,

$$Z_2 = \text{MLP}(\text{AvgPool}(X)) \tag{4}$$

The outputs of both branches are combined through an element-wise multiplication followed by a sigmoid function, which integrates the learned local and global attentions, thereby allowing the model to focus on both local and global features effectively. The output of the LGAM can be defined as follows,

$$Z_2 = \text{MLP}(\text{AvgPool}(X)) \tag{5}$$

3.3.2 FPN

The Feature Pyramid Network (FPN) [26] is an architecture that incorporates a multi-scale approach by leveraging features from different levels of a backbone network, specifically from stages $\{C2, \ldots, C5\}$. These stages provide a diverse set of feature maps that vary in scale and complexity, which the FPN harnesses to construct its own series of feature maps {P2, . . ., P5}. This process is achieved through a top-down pathway that goes from {P5, . . ., P2}, effectively allowing the integration of higher-level, semantically rich feature representations into the lower levels of the pyramid.

A self-attention module is introduced between the C5 and P5 stages. This module is instrumental in capturing global contextual information from the C5 feature maps, which are then infused into the P5 maps. This global information is subsequently disseminated downward along the top-down pathway of the FPN, enriching each subsequent

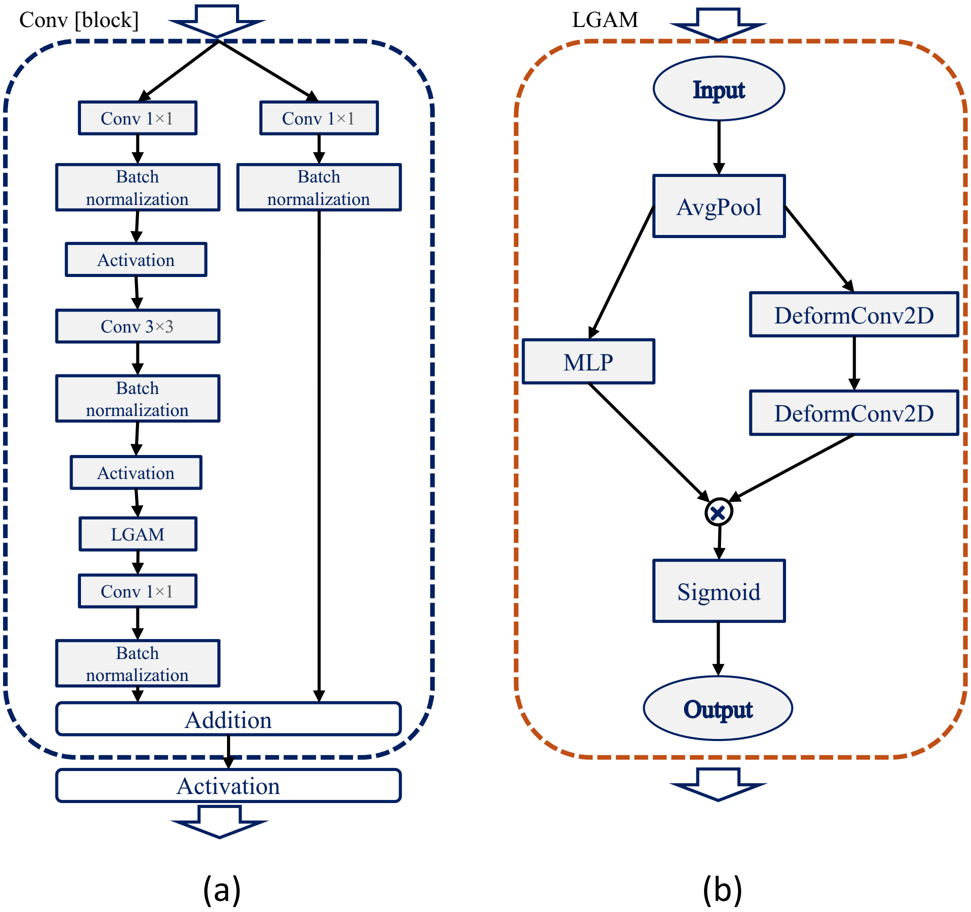


Fig. 4 The structure of the modified backbone from Mask R-CNN, where LGAM is integrated into ResNet-50 convolutional block to improve feature information

level with more contextually aware features. While additional self-attention modules could be inserted between the lower stages of {C2, …, C4} and {P2, …, P4}, empirical evidence from our experiments suggests that placing the self-attention mechanism between C5 and P5 yields the most significant enhancement in performance. This strategic positioning allows the P5 level to serve as an effective distributor of global contextual information, optimizing the feature pyramid for subsequent processing tasks (Fig. 5).

3.3.3 RPN

Region proposal of the objects is generated by RPN using the features that were extracted from the previous model. Three area-scale anchors (64, 128, and 256) and aspect ratios of 1:1, 1:2, and 2:1 were chosen based on the average dimensions of radish leaves and roots found in the dataset. This choice was selected with the 3000 × 4000-pixel picture size in mind from the collected dataset. The RPN proposes regions of interest with bounding boxes and scores using an end-to-end method. By assessing each region proposal's confidence scores, the RPN can determine if a proposal corresponds to any significant features in the keyframes. Subsequently, anchor box regression is used to find the bounding boxes of the radishes within a frame.

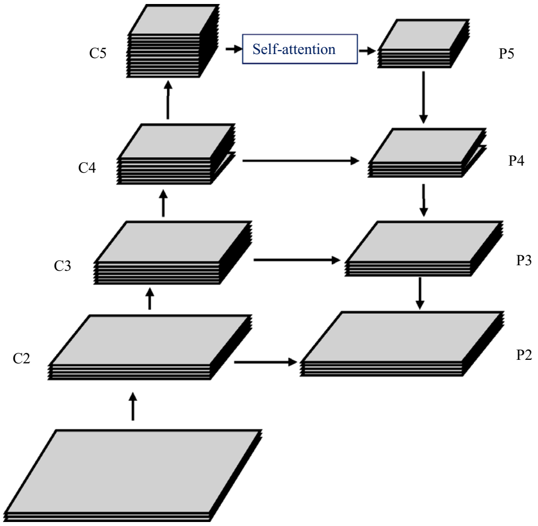


Fig. 5 Feature pyramid networks (FPN) with the self-attention module

3.3.4 ROI alignment

The original ROI Align method uses a single feature map to extract object-related features based on the object proposal. The size of the proposal determines the selection of a feature map from various scales. The larger proposals are sent to larger feature maps, while the smaller proposals are paired with smaller feature maps (5).

3.3.5 Fully connected layer

Once similar dimension features are obtained through the ROI Align layer, the fully connected network uses these features to predict the label of classes and object mask. The PointRend algorithm is used to generate an object mask, which predicts a binary mask representing the pixels that belong to the object and those that don't.

3.3.6 PointRend

PointRend is designed to improve the accuracy of region-based segmentation models by improving the quality of segmentation masks, especially in capturing fine-level details [27]. Using a small MLP (Multi-Layer Perceptron), pointRend is used to enhance coarse instance segmentation masks by selecting a set of points within each detected object and independently predicting the mask value for each point. This MLP makes use of features from the coarse prediction mask as well as a fine-grained feature map of the CNN backbone. The subdivision mask rendering algorithm iteratively refines unclear regions of the predicted mask in a coarse-to-fine manner, drawing inspiration from adaptive subdivision in computer graphics. pointRend computes point-wise feature representations, selects the most uncertain points, up samples the previous prediction, and predicts the labels of those points at each iteration. Until the segmentation reaches the required resolution, this process is repeated. This method efficiently extracts fine-grained information from instance segmentation masks.

3.4 Pixel density conversion

This section explains the physical dimensions of different components of a radish using real-world units. To accurately measure the pixel density in actual dimensions, moreover to determinate, a ruler is identified in the dataset alongside the radish. Figure 6 shows the image processing methods used to apply the Hough line transform [28] operation to locate the ruler line in the input image and then convert it to pixel density. To make edge feature recognition easier, the images are first converted from RGB to grayscale. The grayscale pictures are then subjected to a Gaussian blur to smooth the image and remove unnecessary features that can interfere with line recognition. After that, edges are successfully extracted from the fuzzy grayscale image using the Canny edge method [?]. Finally, the ruler inside the edge-detected imagine is positioned and represented using the Hough line transform method.

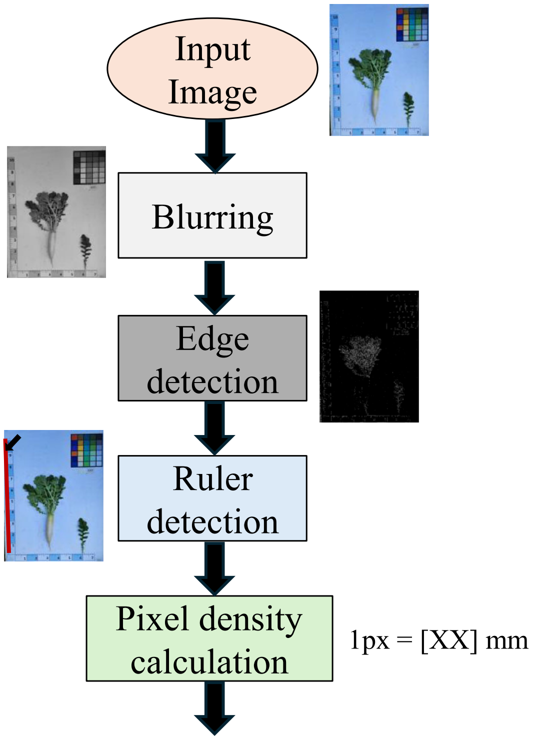


Fig. 6 The key stages of the conversion process determine real-life pixel density, focusing on ruler detection

3.5 Radish phenotypic traits: length and width analysis

This study extensively investigated a comprehensive set of traits, comprising qualitative and quantitative attributes. The quantitative traits measured included root length (RL), root width (RW), leaf length (LL), and leaf width (LW), as shown in Fig. 7. The examination of these traits involved the utilization of customized descriptors, which were derived from the reliable source of the International Union for the Protection of New Varieties of Plants (UPOV, 2021)(https://www.upov.int/portal/index.html.en/). The detailed description for width and length measurement is given below,

3.5.1 Width measurement

we utilize the Geometrical Morphological Analysis (GMA) method which utilizes geometric transformations with morphological operations to estimate width of the object. GMA mitigates the limitations of traditional pixel-based methods,

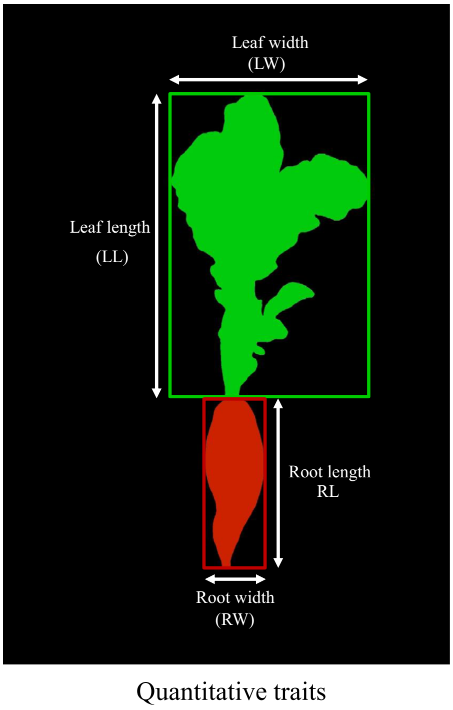


Fig. 7 Illustration of the four phenotypic traits of radishes under investigation in this study

particularly in handling objects with arbitrary orientations and complex shapes. GMA first computes the object's principal orientation using image moment theory [29]. The orientation angle, θ, is derived as follows:

$$\theta = \frac{1}{2} \arctan \left(\frac{2\mu_{11}}{\mu_{20} - \mu_{02}} \right) \tag{6}$$

where μ_{pq} denotes the central moments, capturing the object's intensity distribution. This orientation is fundamental to the alignment process, wherein a rotation matrix,$R(\theta)$, is applied to reorient the object along the image axes as follows,

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \tag{7}$$

After the image alignment, we employ a sequence of morphological operations - specifically dilation ($\oplus$) and erosion ($\ominus$), which can be formulated as follows,

$$A \oplus B, \quad A \ominus B \tag{8}$$

where A represents the binary image, and B is the structuring element. These operations are instrumental in refining the object's representation by smoothing edges and bridging discontinuities.

Finally, the largest width is measured along the axis orthogonal to the object's major axis by analyzing the distribution of pixels in the aligned and processed image. The width W is the count of pixels in the row (or column, depending on orientation) with the maximum sum of pixel values, indicating the broadest part of the object:

$$W = \max_i \left(\sum_j I(i, j) \right) \tag{9}$$

where $I(i, j)$ signifies the intensity of the pixel at position (i, j).

3.5.2 Root length measurement

we utilize the medial axis transform based root length measurement method. It is a commonly employed method within the skeletonization process, aiming to compute the centerline of an object and generate a skeleton of one pixel wide, making the representation simpler without sacrificing structural integrity and important characteristics. Figure 8 displays the efficacy of the medial axis skeletonization process by clearly displaying the abstract-shaped radish root. Using this technique, a binary image was generated with the skeleton's pixels assigned a value of 1 and all other pixels assigned a value of 0. Based on earlier studies [4], the measurement of root length (RL) is made easier once the root skeleton is extracted. When root (RL) is skeletonized by single-pixel wide representations, the length of root can be calculated by:

$$RL = \int_c pdl \approx \sum pdl \tag{10}$$

The root finite length RL is represented by the variable dl, while the geometric calibration factor is denoted by p. Initially, p was representing as a pixel displacement calibration parameter in the outputs mask. However, since the dataset utilized in this study did not exhibit any geometric distortion, the value of p was set to 1. The number of pixels along the skeleton and the root length can be correctly determined without the need for complex calibration factors.

4 Experiment and result analysis

4.1 Experimental platforms and parameters

For this study, Ubuntu was selected to operate as an experimental platform. An Intel(R) Core (TM) i7-6700 CPU running at 3.40 GHz and an Nvidia Tesla V100 GPU with 32 GB of RAM made up the hardware configuration. PyTorch 1.9 was chosen as the preferred deep learning framework,

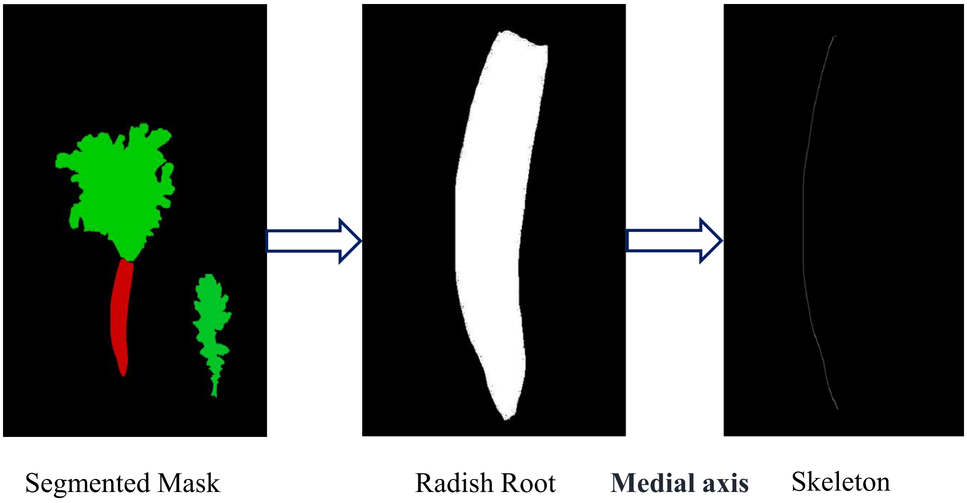


Fig. 8 The process for measuring root length using the medial axis skeletonization algorithm

while Python 3.7 was used as the programming language in this study.

We trained our modified Mask R-CNN model with the stochastic gradient descent (SGD) [30] optimizer on the radish dataset. The optimizer was configured using a 0.9 momentum, 0.0001 weight decay, 0.02 learning rate, and 0.0001 momentum. Furthermore, we use the PyTorch-based open source object detection framework MMdetection [31] to develop our method. We utilized a pre-trained ResNet-50 model using ImageNet as the model's base in order to facilitate reliable testing.

4.2 Comparison of the different segmentation models

The main goal of this section is to showcase the performance of the proposed model over the previous model. To evaluate the performance of the segmentation model, we Utilize the mean average precision (mAP). The mAP is calculated by averaging the average precision (AP) values across all object classes. AP is derived from the precision-recall (PR) curve, which plots the trade-off between precision and recall for each class. Based on values ranging from 0 to 1, the area under the PR curve denotes the AP. Therefore, mAP offers a comprehensive metric for evaluating the performance of the model across all object classes. The equations for mAP are defined as follows.

$$mAP = \frac{1}{K}\sum_{i=1}^{K}\left(\text{precision}_i \times \text{recall}_i\right) \tag{11}$$

where k represents the total number of classes.

Initially, we compared the proposed segmentation model with other state-of-the-art segmentation models to evaluate their performance on our radish dataset. We made use of seven popular segmentation models, to get mAP and inference time. Table 1 represents a performance comparison of the different model mAP and inference time (s) for segmentation using our radish dataset. We observed that QueryInst [32] performed poorly on our dataset, reaching a maximum mAP of 66.9%. However, our proposed (modified Mask R-CNN) model and SOLOv2 [33] demonstrated superior performance among the models evaluated in our data set, where the proposed model achieved the highest mAP of 96.3%.

In terms of processing speed, QueryInst [32] had the longest inference time on our dataset, taking 25.47 s per image. By comparison, our proposed (modified Mask R-CNN) model demonstrated the fastest inference time among the evaluated models, achieving an inference time of 0.53 s per image. This confirms that our modifications not only improve the segmentation accuracy, achieving a segm mAP of 96.3%, but also maintain efficient inference times, comparable to other state-of-the-art methods.

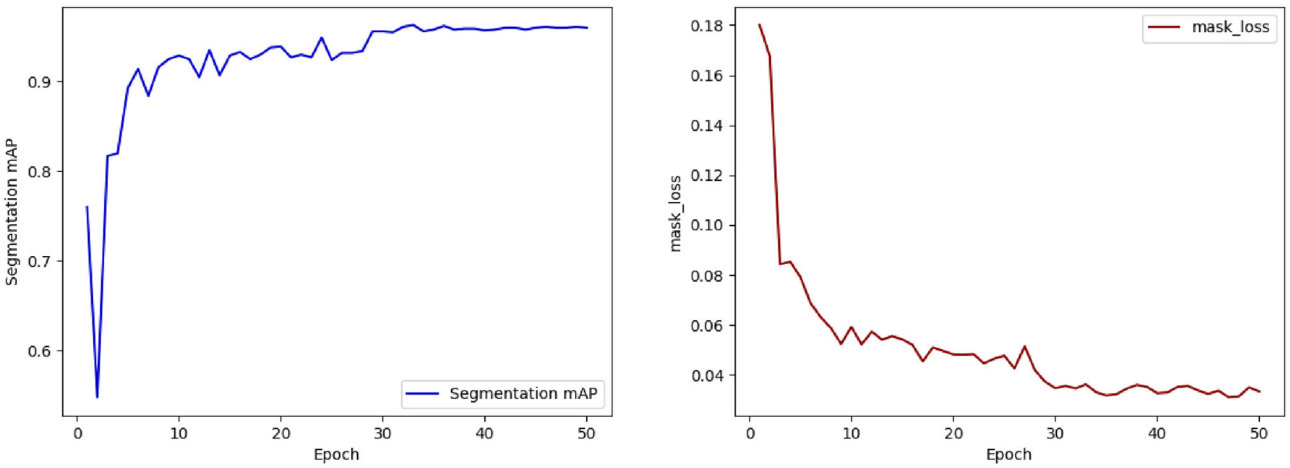
The evaluation of the performance of the proposed method, including the mAP segmentation and the training loss mask, is visually represented in Fig. 9 through those plots. Regarding the model's performance, the validation mAP provides as a key indicator. It demonstrates promising results by rapidly increasing to more than 0.91 in just six epochs and gradually refining to a maximum value of 0.96 at epoch of 33. This highlights our model's proficiency in precisely segmenting radish images. During the training process, the mask loss gradually reduces and achieves a notable decrease to nearly 0.08 after 3 epoch. It continues to converge progressively, obtaining a loss at the conclusion of the training procedure of less than 0.03, which ends at epoch of 50.

4.3 Ablation study

This section is dedicated to the Ablation study of the modified Mask R-CNN with PointRend, exploring introduced

Table 1 Performance comparison of mAP scores and inference time of different models for segmentation using our radish dataset

Method	Backbone	Segm_mAP	Inference time (s)
ConvNeXt [34]	ResNet50	89.5	0.61
ConvNeXt-V2 [35]	ResNet50	89.6	1.28
QueryInst [32]	ResNet50	66.9	25.47
Mask R-CNN + mask head [16]	ResNet50	89.4	0.57
Mask R-CNN + PointRend [27]	ResNet50	92.1	0.51
SOLO [36]	ResNet50	87.4	1.05
SOLOv2 [33]	ResNet50	93.7	0.99
Cascade Mask R-CNN [37]	ResNet50	88.6	0.59
Ours	ResNet50	96.3	0.53



Validation mAP

Training loss mask

Fig. 9 The validation mAP and training mask loss of modified Mask R-CNN over different number of epochs

mechanisms. We selected four ablation study configurations to evaluate the performance of the model. First, we performed hyperparameter optimization for the vanilla Mask R-CNN with PointRend, with training and evaluation conducted on the training and validation sets respectively to select the best hyperparameters. Second, we introduced the Self-Attention Mechanism in the FPN and repeated the hyperparameter optimization process. Third, we introduced the Local–Global Attention Mechanism in isolation, with similar hyperparameter optimization. Finally, we combined both the Self-Attention Mechanism and the Local–Global Attention Mechanism, again performing hyperparameter optimization. Throughout these experiments, we kept the PointRend module fixed. A detailed description of each study is discussed in this section.

4.3.1 Ablation study of vanilla Mask R-CNN hyperparameter optimization

Table 2 presents the accuracy results of different learning rates (0.02, 0.002, 0.0002) across various values of gamma (γ) (0.1, 0.3, 0.06, 0.09). For $\gamma = 0.1$, the highest accuracy (92.7%) is achieved with a learning rate of 0.02, while lower learning rates yield slightly reduced accuracies. At $\gamma = 0.3$, the accuracy remains relatively stable across learning rates, with minor fluctuations. When $\gamma = 0.06$, all learning rates show a decrease in accuracy, with the highest being 91.7% at a learning rate of 0.02. For $\gamma = 0.09$, the highest accuracy (92.6%) occurs at a learning rate of 0.002, indicating an optimal combination at this parameter setting. Overall, the learning rate of 0.02 with $\gamma = 0.1$ provides the best performance in this analysis. After determining the best hyperparameters (learning rate of 0.02 and $\gamma = 0.1$), we conducted a final test on the test-set using this best combination. The results of this final evaluation confirmed the robustness and effectiveness of our hyperparameter selection, with an optimal accuracy of 92.1%.

4.3.2 Ablation study of self-attention mechanism in FPN

Table 3 illustrates the performance outcomes for different learning rates (0.02, 0.002, 0.0002) with various values of

Table 2 Accuracy results of different learning rates and γ values for vanilla Mask R-CNN with PointRend hyperparameter optimization

γ	Larning rate (0.02)	Larning rate (0.002)	Larning rate (0.0002)	Optimal hyperparameter
0.1	92.7	92.4	92.0	92.1
0.3	92.5	92.2	92.3	
0.06	91.7	91.5	91.1	
0.09	92.3	92.6	92.2	

γ (0.1, 0.3, 0.06, 0.09). When γ is set to 0.1, the highest accuracy (94.9%) is achieved with a learning rate of 0.02, with slightly lower accuracies observed at reduced learning rates. For $\gamma = 0.3$, the accuracies decline across all learning rates, with the maximum accuracy (93.7%) at the highest learning rate. At $\gamma = 0.06$, the learning rate of 0.002 yields the highest accuracy (93.3%), showing a slight improvement over the other rates. Finally, for $\gamma = 0.09$, the highest accuracy (93.4%) is achieved at a learning rate of 0.02, while other rates show moderately lower values. Overall, the learning rate of 0.02 combined with $\gamma = 0.1$ results in the best performance in this set of experiments. Subsequently, we proceeded with a final test on the test-set using this optimal combination. The results of this final evaluation validated the robustness and effectiveness of our chosen hyperparameters, achieving an optimal accuracy of 94.3%.

4.3.3 Ablation study of local–global attention mechanism in the feature extractor

Table 4 shows accuracy results for various learning rates (0.02, 0.002, 0.0002) across different γ values (0.1, 0.3, 0.06, 0.09). For $\gamma = 0.1$, the highest accuracy (94.6%) is obtained with a learning rate of 0.02, with accuracies decreasing slightly as the learning rate decreases. At $\gamma = 0.3$, the performance remains relatively stable, with the highest accuracy (93.7%) also occurring at the learning rate of 0.02. For $\gamma = 0.06$, the learning rate of 0.002 achieves the highest accuracy (93.5%), outperforming the other rates. Lastly, for $\gamma = 0.09$, the learning rate of 0.02 again shows the highest accuracy (93.3%), though with less variation among the rates. In summary, the learning rate of 0.02 paired with $\gamma = 0.1$ yields the best accuracy in this analysis. Then, we proceeded with a final test on the test-set using this best hyperparameters combination. The results of this final evaluation validated the robustness and effectiveness of our chosen hyperparameters, achieving an optimal accuracy of 93.8%.

4.3.4 Ablation study of combined self-attention and local–global attention mechanisms in FPN

Table 5 details the accuracy results for various learning rates (0.02, 0.002, 0.0002) across different γ values (0.1, 0.3, 0.06, 0.09). With γ set to 0.1, the highest accuracy (96.3%) is achieved with a learning rate of 0.02, with a slight decrease in accuracy as the learning rate lowers. For $\gamma = 0.3$, while the accuracy remains relatively high, the highest value (95.7%) is observed with a learning rate of 0.0002. When $\gamma = 0.06$, the learning rate of 0.002 results in the highest accuracy (94.8%), showing an improvement over the other rates. For $\gamma = 0.09$, the learning rate of 0.02 again provides the highest accuracy (94.9%), although the differences among the rates are minimal. Finally, we performed a final test on the test-set using this ideal combination of hyperparameters (learning rate of 0.02 and $\gamma = 0.1$). The results of this final evaluation confirmed the robustness and effectiveness of our hyperparameter selection, achieving an optimal accuracy of 96.3%.

4.4 Cross-validation

We conducted a series of cross-validation experiments to evaluate the performance of proposed model enhanced with Self-Attention and Local–Global Attention Mechanisms within the FPN. The goal was to determine the model's accuracy and robustness across multiple folds. We performed a 5-fold cross-validation, where the dataset was randomly split into five sized subsets. Each fold was used once as a test set while the remaining four folds composed of the training set. The primary metric for evaluation was the mAP.

As shown in Table 6, the mAP values across the five folds ranged from 95.0% to 95.5, with an average mAP of 95.2%. These results demonstrate the consistent performance of the modified Mask R-CNN model with the incorporated attention mechanisms across different subsets of the data.

4.5 Robust radish segmentation analysis

Figure 10 illustrates the effective segmentation capabilities of our proposed model when applied to segment radish roots and leaves. Significantly, the precision and accuracy observed in complex scenarios are highlighted in Fig. 10(a) and Fig. 10(b). The model accurately segments the unique characteristics of a slender, reverse C-shaped root in Fig. 10(a), and clearly separates the segments of the root and the leaf, even in the case where Fig. 10(a) contains an item, such a glove, whose color is like that of radish peel. This confirms

Table 3 Performance outcomes of different learning rates with various γ values for the incorporation of self-attention mechanism in FPN

γ	Larning rate (0.02)	Larning rate (0.002)	Larning rate (0.0002)	Optimal hyperparameter
0.1	94.9	94.3	94.5	94.3
0.3	93.7	92.9	92.7	
0.06	93.1	93.3	92.1	
0.09	93.4	92.7	93.2	

Table 4 Performance outcomes of different learning rates with various γ values for the incorporation of local–global attention mechanism in the feature extractor

γ	Larning rate (0.02)	Larning rate (0.002)	Larning rate (0.0002)	Optimal hyperparameter
0.1	94.6	94.0	93.6	93.8
0.3	93.7	93.3	93.5	
0.06	93.2	93.5	92.3	
0.09	93.3	93.1	92.6	

Table 5 Accuracy results for different learning rates with various γ values for the combined self-attention and local–global attention mechanisms in FPN

γ	Larning rate (0.02)	Larning rate (0.002)	Larning rate (0.0002)	Optimal hyperparameter
0.1	96.8	95.9	95.3	96.3
0.3	95.3	95.2	95.7	
0.06	94.5	94.8	94.4	
0.09	94.9	94.7	94.2	

the model's robustness and its capability to handle various challenging conditions.

Similarly, in Fig. 10(c), the model continues to highlight by completely segmenting the small and detailed parts of the radish, indicating its refined discrimination between various parts. Despite the similar coloration, the root and leaves are clearly separated in Fig. 10(d), further demonstrates the strong segmentation performance under testing situation. These instances show the efficiency of the proposed model in handling constraint scenarios which can be found in the real world.

However, as shown in Fig. 11 (a), (b), the model encounters struggle and incorrectly segments certain areas. In Fig. 11(a), where the soil obscuring the radish root causes the model to miss part of the root region. Another scenario in Fig. 11(b), a dark spot between the radish and leaf confuses the model, leading it to incorrectly classify that region as background. These examples point out the model's limitations, showcasing its capabilities in difficult scenarios while also indicating areas requiring improvement.

4.6 Measurement result

Figure 12 shows how our system can be used to efficiently measure various features of the radish phenotyping. The original radish image is shown in Fig. 12(a), and the measurement of various traits, comprising root length (RL), root width (RW), leaf length (LL), and leaf width (LW), are shown in Fig. 12(b).

Table 7 shows the measurements of phenotypic traits for seven radish samples using our proposed algorithm. The table provides Ground Truth (GT) values, measured manually with a tape measure for precision, and compares them with values predicted (PT) by our algorithm. The accuracy of the predictions of our algorithm can be determined by comparing these predicted values (PT) with the GT.

For example, the GT for sample 1 is 34.4 mm for root length and 6.21 mm for root width, and similar GT measurements for other traits are prepared for all samples. Following the GT value, the table records the predicted (PT) measurements. The close correspondence between the ground truth (GT) and the predicted values (PT) denotes a high prediction accuracy, with Mean Absolute Error (MAE) values indicating a near match to the GT. The formula in the following is utilized to determine the MAE:

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{12}$$

where y_i represents the GT value for the trait, $\hat{y}_i$ is the predicted trait value and N is the number of traits. The use of absolute value $|.|$ guarantees that the magnitude errors are positive. Furthermore, the low MAE indicates that our algorithm measures phenotypic traits with high precision and accuracy.

Table 6 Illustrates the mAP results for each fold as well as the average mAP across all folds

Fold_1 (mAP)	Fold_2 (mAP)	Fold_3 (mAP)	Fold_4 (mAP)	Fold_5 (mAP)	Average (mAP)
95.4	95.0	95.0	95.5	95.1	95.2

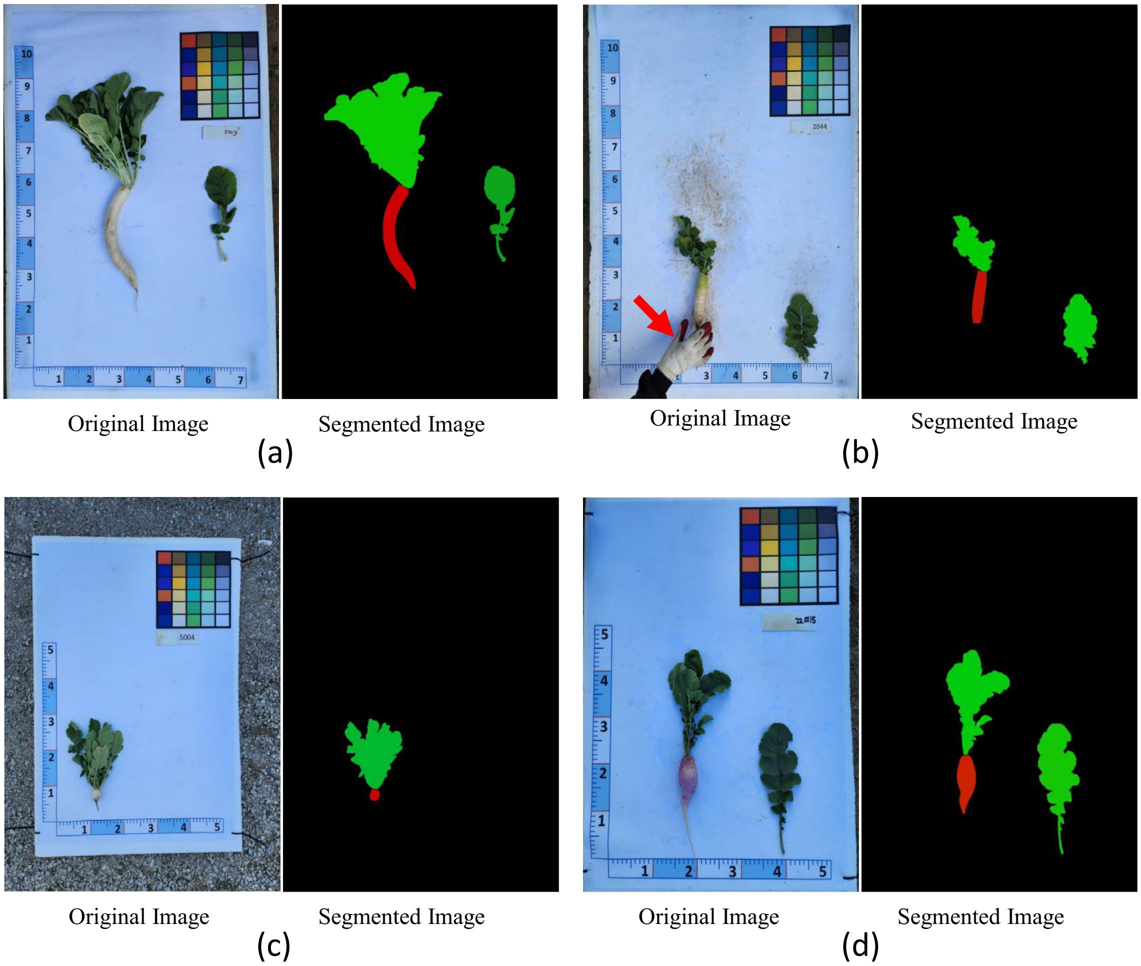


Fig. 10 The prediction results of modified Mask -RCNN on our testing dataset. It can be seen that our proposed model can accurately segment radish roots and leaves in **a** and **b**. The model can segment small details, separating components effectively even in similar coloration situations in **c** and **d**

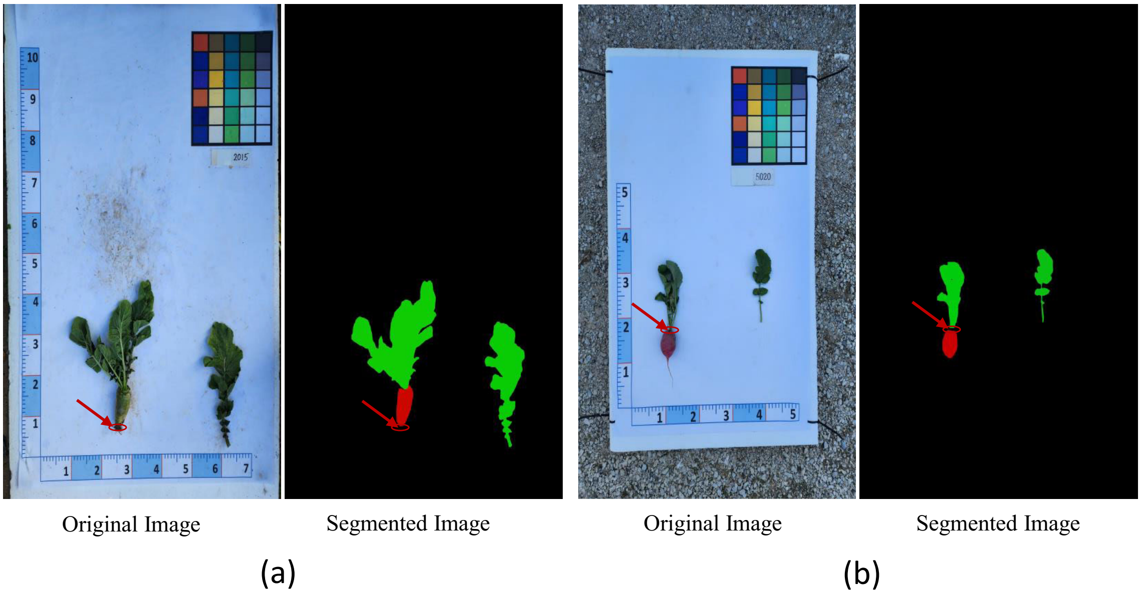
5 Discussion

5.1 Comparison with existing methodology

To compare the performance of our proposed system, we conducted a comparative analysis of the proposed segmentation model along with other state-of-the-art segmentation models. Table 8 presents the comparisons of the data set between the proposed model and recent radish and leaf segmentation approaches considering accuracy. The total of class types and sample sizes that were investigated in this study was lower than the earlier studies except Zhang et. ai. [38]. This study achieved the highest mAP of 96.3% for the segmentation of the roots and leaves of the radish.

5.2 Advantages, limitations and future directions

The existing study on radish segmentation and trait measurement has been limited in providing an efficient framework that includes both precise segmentation techniques along with faster inference time and robust measurement methods. This study introduces a robust and efficient method for the segmentation and measurement of radish phenotypic traits using modified Mask R-CNN model. The results demonstrate a significant advancement over traditional methods, achieving a high mAP of 96.3% for segmentation and a low MAE of 0.51mm for trait measurement. These improvements are attributed to the combination of an improved local–global attention mechanism and an enhanced feature



Original Image Segmented Image Original Image Segmented Image

(a) (b)

Fig. 11 Demonstrating the model's segmentation challenges. **a** Soil obstruction causes the model to miss parts of the radish root region. **b** A dark spot between the radish and leaf leads to incorrect background classification

Table 7 Comparison of phenotypic trait measurements between ground truth (GT) and model predictions (PT) for seven radish samples

	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6	Sample 7
Ground truth (GT)							
RL	34.40	27.09	7	15.20	34.5	25.70	11
RW	6.21	6.17	4.75	6.70	6.1	5.3	7.91
LL	36.56	45.6	15.6	28.9	41.89	43.9	40.49
LW	41.15	34.9	8.1	28.1	42.5	29.5	34
Predicted (PT)							
RL	34.14	27.55	7.05	15.20	34.42	25.71	11
RW	6.33	5.92	4.98	6.68	6.24	5.49	7.99
LL	35.96	44.55	15.94	27.31	43.23	44.13	40.31
LW	40.31	34.15	8.61	27.08	43.84	28.30	33.17
MAE	0.45	0.62	0.28	0.65	0.72	0.40	0.27

Table 8 Dataset comparisons between the proposed model and recent state-of-the-art radish segmentation approaches

Papers	Segmentation method	Dataset size	Class type	Accuracy
Dang et al. [4]	Mask R-CNN	1100	Radish root, leaf, and background (3 classes)	87%
Singh et al. [39]	UNet	6404	–	86.43%
Dang et al. [7]	Inception-V3	4811	Radish, soil, and plastic mulch (3 classes)	95.7%
Zhang et al. [38]	Solo	450	Radish, leaf, and background (3 classes)	87.6%
This study	Modified Mask-RCNN	1100	Radish root and leaf (2 classes)	96.3%

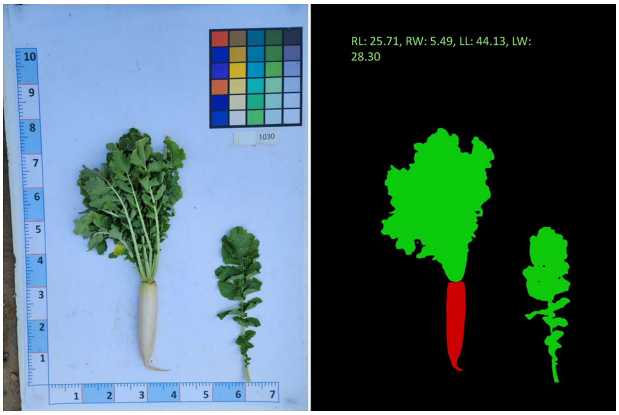


Fig. 12 The measurement calculation of the sample image using our proposed measurement technique

pyramid network (FPN), which allow precise extraction of radish components even under challenging scenarios, such as various obstruction and similar coloration between radish and background elements.

One notable strength of the proposed method is its ability to maintain high accuracy across various scenarios, as shown in Fig. 10. Even in challenging scenarios, the model effectively handles complex shapes and similar colorations, ensuring accurate segmentation. However, the study also points limitations, such as the model's struggle with soil obstruction and dark spots that can lead to incorrect background classification. These limitations denote areas for potential improvement, such as improving the model's robustness in handling varied and complex backgrounds. Future work could focus on further refining the model to address its current limitations and exploring its application to other types of crops and phenotypic traits.

The comparative analysis in Table 8 confirms that the proposed model outperforms existing state-of-the-art approaches, achieving the highest mAP among similar studies. This demonstrates the model's efficacy in radish segmentation and its potential application in broader agricultural settings. By providing a detailed comparison with other segmentation approaches, the study validates the performance of the proposed method and sets a new benchmark for future research in this domain.

6 Conclusion

This research presents a comprehensive framework for the measurement of radish phenotypic traits, designed to automate the growth monitoring of white radish production. We introduced a dataset of 1100 high-resolution images spanning three distinct growth stages of radish, the study ensures precise analysis of phenotypic traits. A critical initial step in this framework was segmentation of the phenotypic traits, which is done by proposed Mask R-CNN to further perform measurement for growth monitoring. To evaluate the effectiveness of segmentation model, different models were trained and assessed. The evaluation highlighted the superiority of the proposed modified Mask-RCNN model, which demonstrated an mAP of 96.3% in the efficient segmentation of the radish components. Furthermore, the implementation of GMA and medial axis analysis allowed the precise measurement of radish phenotypic traits under real-world conditions, acquiring an MAE of 0.51mm.

Although the framework was developed with the measurement of radish traits in mind, it possesses the flexibility to be adapted for other crops, such as cucumbers and pumpkins, with appropriate adjustments and sufficient data for segmentation. There are some limitations in implementing our proposed frameworks on real-time phenotypic trait measurement, primarily due to its complex nature. Future enhancements should focus on refining the framework to enhance robustness and efficiency, thereby facilitating real-time measurement applications. Furthermore, alternative methods for more accurate estimation of radish root width, especially for irregularly shaped roots, is needed to further improve the measurement accuracy.

Acknowledgements This work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2020R1A6A1A03038540) and by the Korean Institute of Planning and Evaluation for Technology in Food, Agriculture, Forestry and Fisheries(IPET) through Digital Breeding Transformation Technology Development Program, funded by Ministry of Agriculture, Food and Rural Affairs (MAFRA) (322063-03-1-SB010) and by the Institute of Information and Communications Technology Planning and Evaluation (IITP) grant funded by the Korea government (MSIT) (No.2022-0-00106, Development of explainable AI-based diagnosis and analysis frame work using energy demand big data in multiple domains.

Author Contributions Nur Alam: Conceptualization of this study, Methodology, implementation, Software, Writing - original draft, Writing - review and editing. A S M Sharifuzzaman Sagar: Conceptualization of this study, methodology, Writing - review and editing. Wenqi Zhang: Data curation, implementation, Writing - Original draft preparation. L. Minh Dang: Conceptualization of this study, implementation, Software, reviewed, edited. Han Yong Park: Data curation, implementation, Writing - Original draft preparation. Moon Hyeonjoon: Conceptualization of this study, Methodology, Proofreading, Software.

Data availability The data underlying this article will be shared on reasonable request to the corresponding author.

References

1. Malhotra, M., Jaiswar, A., Shukla, A., Rai, N., Bedi, A., Iquebal, M.A., et al.: Application of AI/ML approaches for livestock improvement and management. In: Biotechnological Interventions Augmenting Livestock Health and Production. pp. 377–394. Springer, Springer Nature (2023)

2. Karunathilake, E., Le, A.T., Heo, S., Chung, Y.S., Mansoor, S.: The path to smart farming: Innovations and opportunities in precision agriculture. Agriculture **13**(8), 1593 (2023)
3. Xiao, G., Zhang, X., Niu, Q., Li, X., Li, X., Zhong, L., et al.: Winter wheat yield estimation at the field scale using Sentinel-2 data and deep learning. Computers Electron. Agric. **216**, 108555 (2024)
4. Dang, L.M., Min, K., Nguyen, T.N., Park, H.Y., Lee, O.N., Song, H.K., et al.: Vision-based white radish phenotypic trait measurement with smartphone imagery. Agronomy **13**(6), 1630 (2023)
5. Park, C.H., Ki, W., Kim, N.S., Park, S.Y., Kim, J.K., Park, S.U.: Metabolic profiling of white and green radish cultivars (Raphanus sativus). Horticulturae **8**(4), 310 (2022)
6. Xie, C., Yang, C.: A review on plant high-throughput phenotyping traits using UAV-based sensors. Computers Electron. Agric. **178**, 105731 (2020)
7. Dang, L.M., Hassan, S.I., Suhyeon, I., kumar Sangaiah, A., Mehmood, I., Rho, S., et al.: UAV based wilt detection system via convolutional neural networks. Sustain. Comput.: Inform. Syst. **28**, 100250 (2020)
8. Barbedo, J.G.A.: A review on the use of unmanned aerial vehicles and imaging sensors for monitoring and assessing plant stresses. Drones **3**(2), 40 (2019)
9. Patrício, D.I., Rieder, R.: Computer vision and artificial intelligence in precision agriculture for grain crops: A systematic review. Comput. Electron. Agric. **153**, 69–81 (2018)
10. Mizushima, A., Lu, R.: An image segmentation method for apple sorting and grading using support vector machine and Otsu's method. Computers Electron. Agric. **94**, 29–37 (2013)
11. Otsu, N., et al.: A threshold selection method from gray-level histograms. Automatica **11**(285–296), 23–27 (1975)
12. Hearst, M.A., Dumais, S.T., Osuna, E., Platt, J., Scholkopf, B.: Support vector machines. IEEE Intell. Syst. Appl. **13**(4), 18–28 (1998)
13. Kamilaris, A., Prenafeta-Boldú, F.X.: Deep learning in agriculture: A survey. Computers Electron. Agric. **147**, 70–90 (2018)
14. Dias, P.A., Tabb, A., Medeiros, H.: Apple flower detection using deep convolutional networks. Computers Ind. **99**, 17–28 (2018)
15. Yu, Y., Zhang, K., Yang, L., Zhang, D.: Fruit detection for strawberry harvesting robot in non-structural environment based on Mask-RCNN. Computers Electron. Agric. **163**, 104846 (2019)
16. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask r-cnn. In: Proceedings of the IEEE International Conference on Computer Vision; pp. 2961–2969 (2017)
17. Huang, Z., Huang, L., Gong, Y., Huang, C., Wang, X.: Mask scoring r-cnn. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; pp. 6409–6418 (2019)
18. Liu, S., Jia, J., Fidler, S., Urtasun, R.: Sgn: Sequential grouping networks for instance segmentation. In: Proceedings of the IEEE International Conference on Computer Vision; pp. 3496–3504 (2017)
19. Chen, X., Girshick, R., He, K., Dollár, P.: Tensormask: A foundation for dense object segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision; pp. 2061–2069 (2019)
20. Kolhar, S., Jagtap, J.: Plant trait estimation and classification studies in plant phenotyping using machine vision-a review. Inform. Process. Agric. **10**(1), 114–135 (2023)
21. Wu, S., Wen, W., Xiao, B., Guo, X.: An accurate skeleton extraction approach from 3D point clouds of maize plants. Front. Plant Sci. **10**, 426822 (2019)
22. Wang, Y., Wen, W., Wu, S., Wang, C., Yu, Z., Guo, X., et al.: Maize plant phenotyping: Comparing 3D laser scanning, multi-view stereo reconstruction, and 3D digitizing estimates. Remote Sens. **11**(1), 63 (2018)
23. Liu, G., Nouaze, J.C., Touko Mbouembe, P.L., Kim, J.H.: YOLO-tomato: A robust algorithm for tomato detection based on YOLOv3. Sensors **20**(7), 2145 (2020)
24. Russell, B.C., Torralba, A., Murphy, K.P., Freeman, W.T.: LabelMe: A database and web-based tool for image annotation. Int J Computer Vision **77**, 157–173 (2008)
25. Dai, J., Qi, H., Xiong, Y., Li, Y., Zhang, G., Hu, H., et al.: Deformable convolutional networks. In: Proceedings of the IEEE International Conference on Computer Vision; pp. 764–773 (2017)
26. Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature pyramid networks for object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; pp. 2117–2125 (2017)
27. Kirillov, A., Wu, Y., He, K., Girshick, R.: Pointrend: Image segmentation as rendering. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; pp. 9799–9808 (2020)
28. Duda, R.O., Hart, P.E.: Use of the Hough transformation to detect lines and curves in pictures. Commun. ACM **15**(1), 11–15 (1972)
29. Teague, M.R.: Image analysis via the general theory of moments. Josa **70**(8), 920–930 (1980)
30. Ruder, S.: An overview of gradient descent optimization algorithms. arXiv preprint arXiv:1609.04747. (2016)
31. Chen, K., Wang, J., Pang, J., Cao, Y., Xiong, Y., Li, X., et al.: MMDetection: Open mmlab detection toolbox and benchmark. arXiv preprint arXiv:1906.07155. (2019)
32. Fang, Y., Yang, S., Wang, X., Li, Y., Fang, C., Shan, Y., et al.: Instances as queries. Proceedings of the IEEE/CVF International Conference on Computer Vision, pp.6910–6919 (2021)
33. Wang, X., Zhang, R., Kong, T., Li, L., Shen, C.: Solov2: Dynamic and fast instance segmentation. Adv. Neural Inform. Process. Syst. **33**, 17721–17732 (2020)
34. Liu, Z., Mao, H., Wu, C. Y., Feichtenhofer, C., Darrell, T., Xie, S.: A convnet for the 2020s. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; pp. 11976–11986 (2022)
35. Woo, S., Debnath, S., Hu, R., Chen, X., Liu, Z., Kweon, I.S., et al.: Convnext v2: Co-designing and scaling convnets with masked autoencoders. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 16133–16142 (2023)
36. Wang, X., Kong, T., Shen, C., Jiang, Y., Li, L.: Solo: Segmenting objects by locations. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVIII 16. Springer; pp. 649–665 (2020)
37. Cai, Z., Vasconcelos, N.: Cascade R-CNN: High quality object detection and instance segmentation. IEEE Trans. Pattern Anal. Mach Intell. **43**(5), 1483–1498 (2019)
38. Zhang, W., Dang, L.M., Li, Y., Wang, H., Lee, S., Moon, H.: Enhanced solo-based instance segmentation algorithm for efficient plant growth assessment: A radish case study. Korean Society of Broadcasting and Media Engineering Conference. pp. 274–277 (2023)
39. Singh, S., Singh, D., Agarwal, S., Saini, M.: IRPD: In-field radish plant dataset. International Conference on Agriculture-Centric Computation. Springer, pp. 87–98 (2023)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.