

RESEARCH

Open Access



Automatic plant phenotyping analysis of Melon (*Cucumis melo* L.) germplasm resources using deep learning methods and computer vision

Shan Xu¹, Jia Shen², Yuzhen Wei³, Yu Li⁴, Yong He¹, Hui Hu⁵ and Xuping Feng^{1,4*}

Abstract

Cucumis melo L., commonly known as melon, is a crucial horticultural crop. The selection and breeding of superior melon germplasm resources play a pivotal role in enhancing its marketability. However, current methods for melon appearance phenotypic analysis rely primarily on expert judgment and intricate manual measurements, which are not only inefficient but also costly. Therefore, to expedite the breeding process of melon, we analyzed the images of 117 melon varieties from two annual years utilizing artificial intelligence (AI) technology. By integrating the semantic segmentation model Dual Attention Network (DANet), the object detection model RTMDet, the keypoint detection model RTMPose, and the Mobile-Friendly Segment Anything Model (MobileSAM), a deep learning algorithm framework was constructed, capable of efficiently and accurately segmenting melon fruit and pedicel. On this basis, a series of feature extraction algorithms were designed, successfully obtaining 11 phenotypic traits of melon. Linear fitting verification results of selected traits demonstrated a high correlation between the algorithm-predicted values and manually measured true values, thereby validating the feasibility and accuracy of the algorithm. Moreover, cluster analysis using all traits revealed a high consistency between the classification results and genotypes. Finally, a user-friendly software was developed to achieve rapid and automatic acquisition of melon phenotypes, providing an efficient and robust tool for melon breeding, as well as facilitating in-depth research into the correlation between melon genotypes and phenotypes.

Keywords Plant phenotyping, Machine learning, Deep learning, Computer vision

*Correspondence:

Xuping Feng
fengxp@zju.edu.cn

¹College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou 310058, China

²Institute of Vegetables, Zhejiang Academy of Agricultural Sciences, Hangzhou 310021, China

³School of Information Engineering, Huzhou University, Huzhou 313000, China

⁴Agricultural Experiment Station & Agricultural Sci-Tech Park Management Committee, Zhejiang University, Hangzhou 310058, China

⁵Sichuan Yuheyan Agricultural Technology Co., Ltd, Chengdu 610066, China



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Introduction

Melon, an important crop of the *Cucurbitaceae* family, is highly favored by consumers due to its sweet taste and nutritional value [1]. There exists a wide variety of melon types, among which *Cucumis melo* L., renowned for its superior quality and attractive appearance, holds strong competitiveness in the consumer market. With the rapid development of modern economic society, the demand for fruit quality continues to escalate. The external appearance of melons directly influences consumer choices, determining their commercial value. Therefore, in the breeding process of melons, breeders focus on phenotypic traits such as fruit length, width, shape index, and color, which reflect melons' appearance. Additionally, the length and width of melons' pedicel determine whether the fruit is prone to dropping and the ease of fruit picking post-maturation, underscoring the significance of obtaining corresponding phenotypic traits.

Crop breeding is the primary approach to improving crop quality [2]. Breeders select superior varieties and preserve their germplasm resources by evaluating the phenotypic traits of fruits as representations of corresponding genotypes [3–5]. Therefore, phenotypic detection is a crucial step in melon breeding [6]. However, currently the acquisition of phenotypic data relies predominantly on expert judgment and complex manual measurements. This high-cost, low-efficiency mode of work severely limits the quantity and quality of phenotypic data, and it is difficult to quantify traits due to individual differences [7, 8]. Therefore, finding a rapid and accurate automated method for plant phenotypic extraction is of great practical significance.

In recent years, the field of AI has continuously achieved crucial technological breakthroughs, and automated phenotypic analysis methods based on computer vision and deep learning have demonstrated powerful application potential [9], providing strong support for crop breeding. Currently, RGB images are the most common data source, which is easy to obtain, cost-effective, and applicable to multiple models [10, 11]. Tu et al. [12] utilized the Fast Region-based Convolutional Neural Network (Faster R-CNN) algorithm to detect passion fruits in natural environments and extract relevant phenotypic information, achieving classification of five different levels of ripeness by inputting it into a classifier. Wu et al. [13] proposed two models, linear regression and deep learning, to calculate the grain number per panicle in rice, assisting in rice variety selection. Ni et al. [14] implemented the segmentation of individual blueberries based on Mask Region-based Convolutional Neural Network (Mask R-CNN), obtaining the cluster compactness and fruit maturity, and estimating berry number per clusters, which are of great significance for blueberry breeding. Li et al. [15] proposed the SPM-IS

algorithm, which can obtain phenotypic traits of mature soybean stems, pods, and seeds. The results showed a high correlation coefficient between predicted and true values, making it a powerful tool for accelerating soybean breeding processes.

However, related research on melons remains notably scarce. Ho et al. [16] utilized the UNet to segment melon fruit peel images, obtaining masks for four parts and extracting 12 phenotypic traits. These traits were then input into a Deep Neural Network (DNN) to achieve sweetness grading. Similarly, Qian et al. [17] used algorithms to extract color and texture features of melon skin to predict weight loss rates, enabling rapid assessment of melon storage conditions. Cho et al. [18] developed two machine learning algorithms based on hyperspectral images to predict melon solids concentration and moisture content as indicators of melon ripeness. Kalantar et al. [19] used drone images to identify melons in images using object detection algorithms and extracted contours. By approximating the contours to standard ellipses, the weight of melons was calculated based on ellipse dimensions and density formulas, combined with fruit quantity to achieve yield prediction. Although deep learning algorithms perform well in phenotypic extraction of melons, its applications are mostly focused on post-harvest grading, ripeness detection, yield prediction, etc., with relatively limited applications in breeding.

Another salient point to note is that, although there has been some progress in the recognition and segmentation of plant stems and fruit pedicels using deep learning methods, there is almost no research on extracting their phenotypes. In the related studies on melon phenotype acquisition, there has been no focus on pedicel, despite the importance of its phenotypic traits for melon breeding. Meanwhile, we have noticed that in recent years, keypoint detection algorithms have been widely applied in fruit picking location research due to their unique advantages in detecting stem-like target objects. And this has important implications for our research [20–22].

Therefore, addressing the current lack of rapid and automatic phenotypic acquisition methods in the field of melon breeding, this study established a comprehensive framework for automatic extraction of melon fruit and pedicel phenotypes based on multiple deep learning algorithms. It can efficiently and accurately obtain multiple important phenotypic traits of melon. The main contributions of this study are as follows: [1] Analyzed and compared the performance of six classical semantic segmentation models for melon fruit. Achieved high-precision segmentation of various image components, including fruit and scale, on the optimal model DANet [2]. Employed the object detection algorithm RTMDet and keypoint detection algorithm RTMPose to achieve pedicel localization. Utilized the predicted bounding box

and keypoint coordinates as cues, inputting them into the MobileSAM model to accomplish pedicel segmentation [3]. Proposed a series of phenotype extraction algorithms based on the obtained masks of segmented components and the regression-derived keypoint coordinates. Through comparison and conversion with the scale, the genuine phenotypic traits of melon were obtained [4]. Integrated the algorithms into a comprehensive framework and developed corresponding software. When fed with specified format melon image, the software can achieve rapid and automatic acquisition of its phenotype, effectively applying our research outcomes to the breeding process of melons.

Materials and methods

This study can be divided into four stages (Fig. 1): data acquisition, model selection and algorithm framework construction, proposal of phenotype extraction algorithms, and software development.

Data

In two annual cycles, we cultivated 117 different varieties of melon to construct our dataset, comprising 35 varieties

cultivated in 2020 and 82 in 2022. Fruit morphological appearances varied significantly among different varieties. For each variety, 1–3 images were collected. During image acquisition, all melon fruits were positioned within a dark enclosure and illuminated from above. Positioned directly in front of the melon was a 6 cm scale for reference. A handheld camera with a resolution of 5184×3456 pixels was employed to capture the images. And care was taken to ensure that the melon and scale did not overlap in the frame and were roughly centered. To facilitate training and ensure uniformity in size measurement, all images underwent preprocessing involving cropping and resizing, ultimately resized to 600×400 pixels before being input into the model.

The ground truth used for training was manually annotated using the data labeling tool LabelMe. For the two distinct tasks of fruit segmentation and pedicel segmentation, two different types of annotations were conducted: [1] For semantic segmentation of melon fruit, annotations were made along the contours of the fruit, intact pedicel, and the scale; [2] For the detection of melon pedicel, we first annotated the required portion of the pedicel area using rectangular bounding

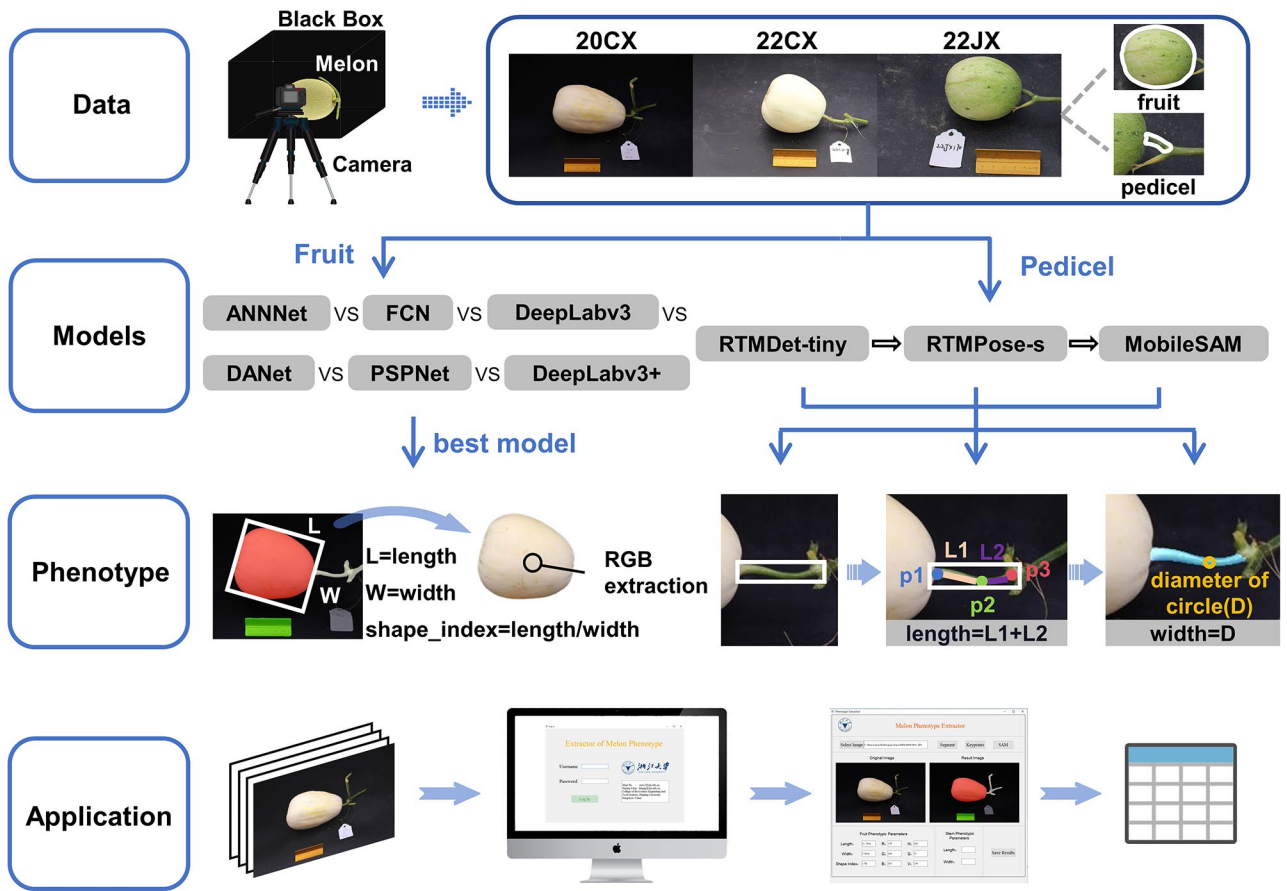


Fig. 1 Deep learning framework for melon phenotypic traits acquisition. Four stages of this study: data acquisition, model selection and algorithm framework construction, proposal of phenotype extraction algorithms, and software development

boxes. Additionally, upon comprehensive evaluation of all images, we found that the bending situations of most pedicels were not overly complex, with three key points being sufficient to describe their shape. Therefore, in this study, we annotated three key points for each pedicel: the intersection point between the pedicel and fruit, the bending point of the pedicel (or the midpoint if the pedicel was not bent), and the intersection point between the required pedicel portion and other parts of the pedicel. These key points were named p1, p2, and p3 based on their proximity to the melon fruit. Upon completion of annotation, each image generated a corresponding JSON file, which was subsequently converted into the required dataset format according to the specific model requirements. Furthermore, for ease of subsequent model training and development, the dataset was randomly divided into training set (137 images), validation set (35 images), and test set (20 images) in a 7:2:1 ratio.

Model

Semantic segmentation

Segmentation of melon fruits is a task involving single-object segmentation of multiple classes against a pure black background. To meet the specific requirements of this task, we selected a total of six classic semantic segmentation models for comparison. These six models are Asymmetric Non-local Neural Network (ANNet) [23], DANet [24], DeepLabV3 [25], DeepLabV3+ [26], Fully Convolutional Network (FCN) [27], and Pyramid Scene Parsing Network (PSPNet) [28]. Under identical experimental conditions, all models were trained on our constructed melon dataset, and their performance was subsequently evaluated on the test set.

Compared to object detection and classification, applying deep learning methods to semantic segmentation faces two major challenges. One is resolution: in the process of convolution and pooling, the feature map size shrinks continuously in deep convolutional neural networks, leading to the loss of spatial information. However, semantic segmentation, as a dense prediction problem, requires not only deep semantic information but also high spatial resolution. The other challenge is the balance between receptive field size and computational complexity. Establishing long-range dependencies for features at different scales can significantly improve segmentation results, but this leads to increased computational costs. The six semantic segmentation models selected in this study employ different strategies to address these challenges from various perspectives. For example, the Atrous Spatial Pyramid Pooling (ASPP) module introduced in the DeepLab series and the Pyramid Pooling Module (PPM) module introduced in PSPNet adapt to the needs of different tasks. By comparing the prediction results, we would select the model that

best suits our study and better extract melon fruit phenotype features.

The backbone of all models is selected as ResNet50. All deep learning code used in the study was executed in a GPU environment on Google Colab, utilizing Tesla V4 GPU resources. The code framework was built using the MMSegmentation and MMPose algorithm libraries.

In evaluation, to compare the performance of different models, F-score is used as the primary evaluation metric. While Precision measures the subjective accuracy of the model and Recall gauges the objective accuracy, F-score achieves a balance between the two, offering a more holistic assessment of model performance.

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

$$F - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (3)$$

Where *TP* (True Positive) refers to the pixels correctly identified as part of the object class, meaning they are labeled as belonging to the object both in the predicted segmentation mask and the ground truth. *FP* (False Positive) occurs when background pixels are mistakenly classified as part of the object, meaning the prediction shows them as object pixels, but the ground truth labels them as background. *FN* (False Negative) refers to object pixels that are incorrectly classified as background, meaning the prediction misses them as part of the object, even though they are labeled as such in the ground truth.

Furthermore, a series of metrics such as Accuracy and Precision, among others not listed, were also used as references in our study.

Three deep learning models used for pedicel segmentation

It is important to clarify that the “pedicel” described in this study does not refer to the entire pedicel as depicted in images, but only to a small segment that is connected to the fruit and has not yet branched. To obtain corresponding phenotypic information, we combined three deep learning models to acquire keypoints and mask.

In general, for cases such as the one in our study where there is only a single pedicel target in the image, top-down keypoint detection algorithms are more commonly employed.

We trained an RTMDet model on the constructed melon dataset to serve as the target detector for keypoint detection (Fig. 2a). RTMDet is a one-stage object detection model proposed only in 2022 [29], which shares similarities in architecture with the traditional You Only

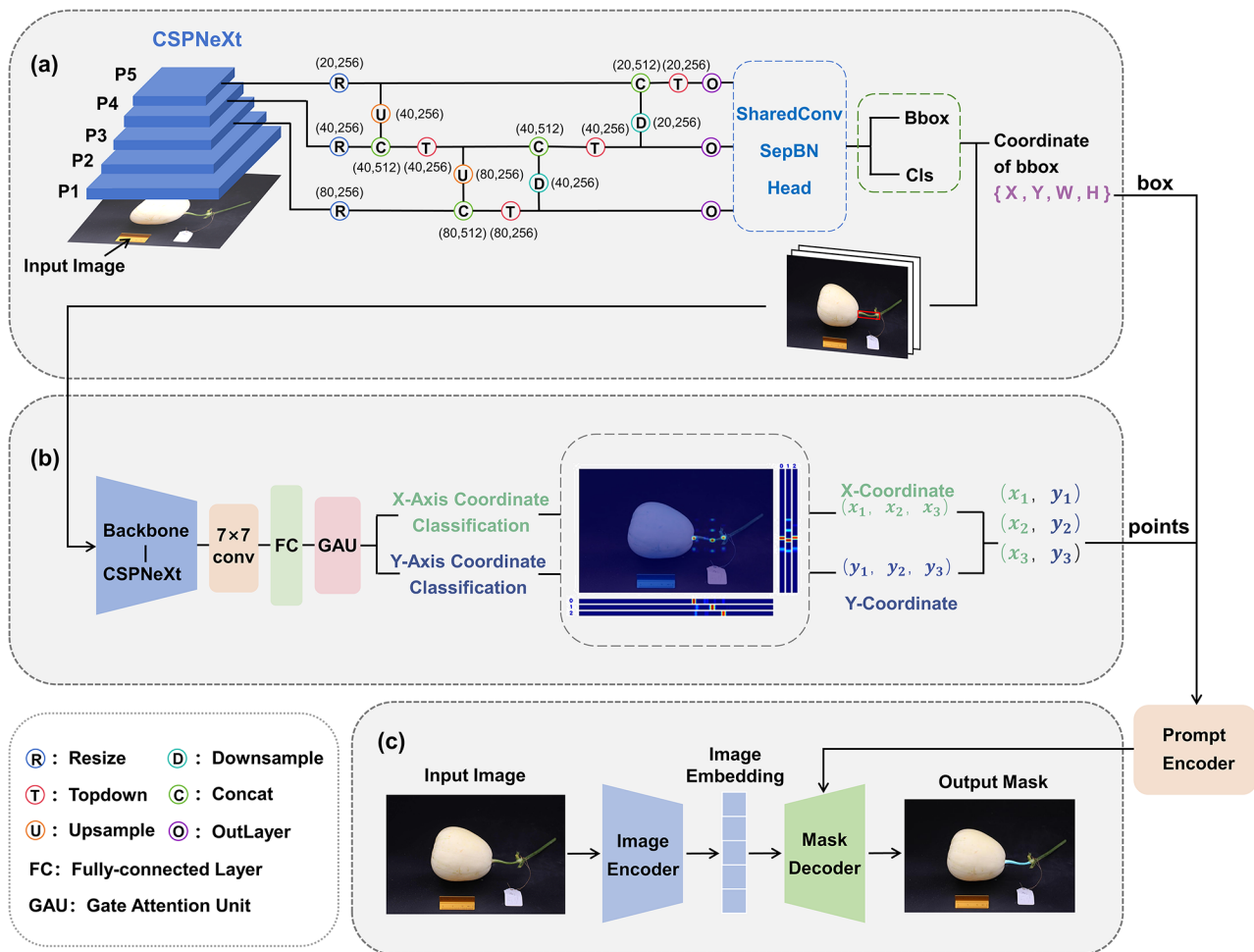


Fig. 2 Deep learning model for melon pedicel segmentation. **(a)** Architecture of the RTMPose-tiny model employed for pedicel object detection. **(b)** Detailed structure of the RTMPose-s model used for melon pedicel keypoint detection, where the RTMDet model serves as its target detector. **(c)** Structure of the zero-shot trained MobileSAM model, where the prediction results from RTMDet-tiny and RTMPose-s serve as prompt inputs

Look Once (YOLO) series, but has not been widely explored in research [30–32]. Since the introduction of the Cross Stage Partial (CSP) structure into DarkNet by YOLOv4, CSPDarkNet has been widely applied in various versions of the YOLO series due to its simplicity and efficiency. The backbone of RTMDet is also built upon CSPDarkNet, with its improvement lying in the incorporation of 5×5 large convolutional kernels to increase the effective receptive field, thereby enhancing the feature extraction capability of the basic building blocks. Subsequently, it utilizes the same building blocks as the neck part for multi-scale feature fusion. Finally, the features are inputted into detection heads with shared convolution weights and separated batch normalization (BN) layers to predict the classification and regression results of the bounding boxes. Considering the relatively small size of the dataset, we opted for its tiny version.

Upon this foundation, RTMPose is employed as the keypoint detection algorithm (Fig. 2b), which is a novel

model introduced by Jiang et al. [33] in March 2023. RTMPose is a high-performance real-time multi-person pose estimation framework based on MMPose following a “top-down” approach. It utilizes the same building blocks as RTMDet to construct the backbone of the main network, achieving a fine balance between speed and accuracy. Notably, in the keypoint prediction part, RTMPose adopts an algorithm based on SimCC [34] to predict keypoints, which treats keypoint localization as a classification task of two pixel points along the x and y coordinate axes. Compared to heatmap-based algorithms, the SimCC-based algorithm achieves competitive accuracy with lower computational cost. Furthermore, SimCC employs a very simple architecture of only two fully connected layers for prediction, making it easy to deploy on various backends. Currently, there is no research applying it in agriculture-related fields. Similarly, for considerations of data volume, RTMPose-s is employed in our study.

Finally, to achieve accurate segmentation of the pedicel, we adopted the latest lightweight SAM model based on prompt-guided input, namely MobileSAM [35]. Compared to the original SAM model [36], it significantly improves the inference speed without sacrificing accuracy, and can be directly transferred for use without training. It consists of an image encoder for extracting image features and a prompt encoder for embedding prompts (Fig. 2c). The coordinates of the bounding boxes and keypoints predicted by the above algorithm are embedded into the MobileSAM model as prompts to reveal the specific locations of the target objects. The keypoints, when used as prompts, can have two different labels: background points or foreground points. In this study, we tested four different prompt input methods, and ultimately, a prompt input consisting of one target box, two foreground points, and one background point was proven to be the most effective (Fig. 3a).

Phenotypic traits extraction

The acquisition of phenotypic traits for melon is based on the results of all the aforementioned models. The specific calculation methods for phenotypic parameters are as follows:

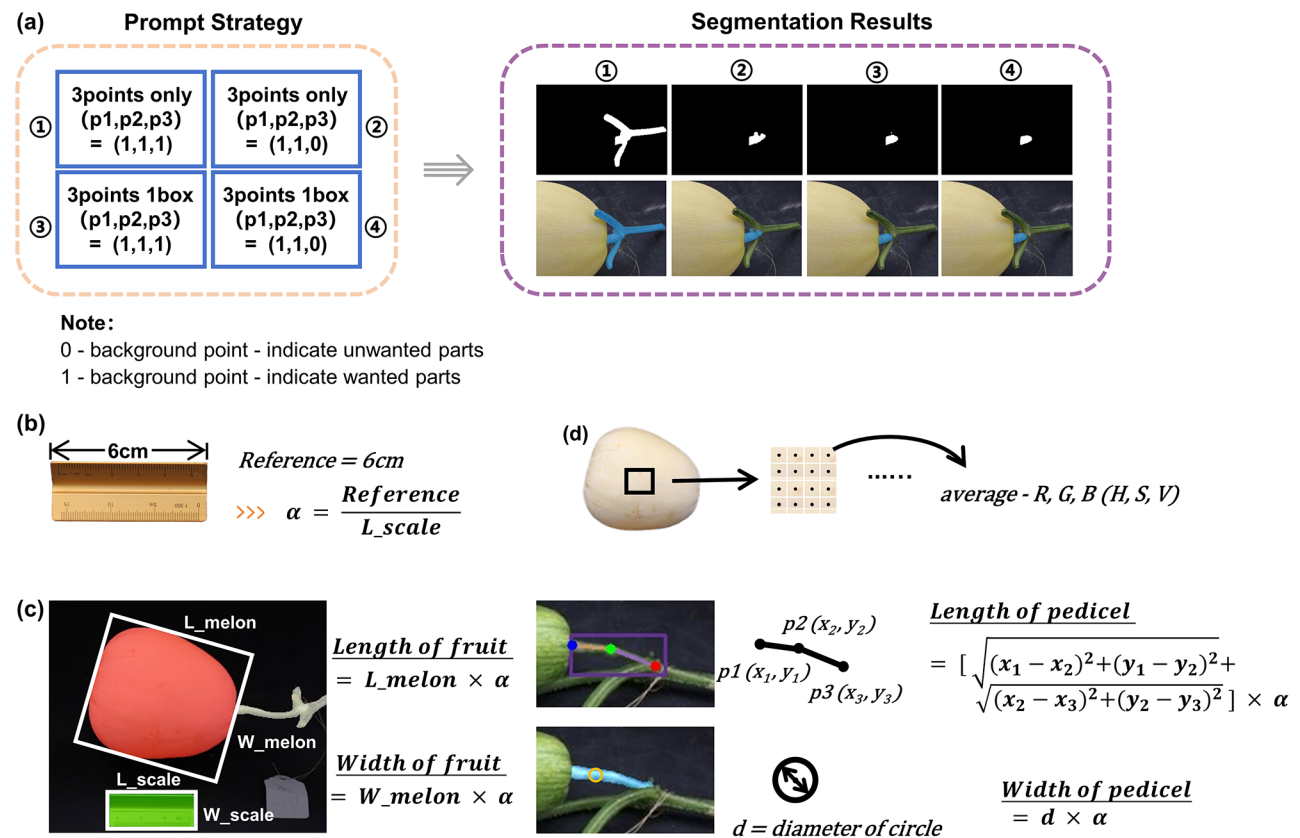


Fig. 3 Prompt input selection for the MobileSAM model and melon phenotypic extraction algorithm. **(a)** Segmentation results of the MobileSAM model under four different prompt inputs. **(b)** The scale used in our study and the computation of the scaling factor. **(c)** Method for obtaining size information of melon fruit and pedicel. **(d)** Method for acquiring color information of melon fruit peel

averaged to represent the color features of the fruit skin (Fig. 3d).

Third, obtain the length and width of the melon pedicel. On one hand, by applying the Pythagorean theorem, the lengths of the lines between p1 and p2, as well as p2 and p3, were calculated based on the coordinates of three keypoints obtained through regression. The sum of these two segments is taken as an approximation of the pedicel length. On the other hand, based on the mask obtained from MobileSAM segmentation, the maximum inscribed circle is extracted. In this process, to avoid the influence of small noise generated by segmentation on the results, the same processing is applied to all independent mask parts on the image, and only the maximum inscribed circle with the largest radius is selected. The diameter of this circle is taken as an approximation of the pedicel width. Finally, using the scaling factor obtained from the first part, the actual values of the pedicel length and width are obtained through conversion (Fig. 3c).

Results and discussion

Melon fruit segmentation

For all six semantic segmentation models, a batch size of 4 and a training epoch of 5000 were configured. The total training time for each model was approximately 2 h, with no significant speed differences observed.

After training, we evaluated the performance of the aforementioned six semantic segmentation models on our test dataset, contrasting the segmentation results of various parts of the image (Fig. 4). Here, mFscore represents the average F-score of these parts. Firstly, it is intuitive and clear that all models achieve satisfactory segmentation results for the fruit, scale, and background, with F-scores all above 90%. However, at the same time, the segmentation of the pedicel by each model are less than ideal, far from optimal compared to other parts. We annotated the complete pedicel contour here, but the final segmentation result is still not up to par. It is conceivable that segmenting part of the pedicel we need is undoubtedly challenging using semantic segmentation models alone, especially when there is interference from other parts of the pedicel nearby. Therefore, in the following study, we employed more complex and novel methods to accomplish this task. Considering that the format of the images to be processed in the actual application of the model to the breeding program is essentially consistent with that of the training data, we therefore do not conduct tests on the model's generalization or robustness in complex scenarios.

Although the segmentation performance of each model is relatively excellent, it can be observed that there are still certain differences between the results of different parts of the images. The melon fruit and the scale, which will be used as reference in subsequent phenotypic traits

extraction, are the two main parts of interest on the image. Therefore, the relevant segmentation metrics of these two parts became the focus of comparison.

After comprehensive consideration of various factors, we selected DANet as the final model for melon fruit segmentation, which has advantages in both accuracy and time efficiency. Although the differences in performance between different models are not significant in terms of numerical indicators, DANet wins by a slight margin. However, from the specific prediction results, DANet demonstrates the best segmentation performance. It achieves high-precision segmentation for melons of different varieties, colors, and placement positions. In contrast, other models tend to exhibit incomplete segmentation in some particularly unique or atypical melon fruits (Fig. 4b).

DANet, proposed by Fu et al. in 2019, is a dual-attention network designed to address the problem of intra-class inconsistency caused by local receptive fields. It is based on self-attention mechanisms to capture feature dependencies between different sizes and channels, thereby enhancing the model's understanding and representation capabilities of image content by establishing associations between features to explore global context information. The Position Attention module and the Channel Attention module are key components in achieving high-precision segmentation.

For the sake of rigor, we conducted experiments to test the effectiveness of semantic segmentation models in segmenting part of the pedicel. We manually annotated a new batch of data, delineating along the contours of the melon fruit, the desired pedicel, and the scale, while defining the remaining pedicel parts as background. Using the same experimental parameters on this new dataset, we trained DANet and observed the prediction performance on the test dataset. Although the model can roughly locate the position of the pedicel, the segmentation accuracy is low, with only a few melons having their pedicels segmented relatively completely (Fig. 4c). Therefore, this experiment also proves that obtaining the mask of the pedicel solely through semantic segmentation is impractical. Moreover, it is challenging to extract the length phenotype of the pedicel based on obtaining the mask only.

Melon pedicel segmentation

As mentioned earlier, for pedicel phenotypic traits extraction, we innovatively combined the keypoint detection model primarily used in fields such as human pose detection, face recognition, and action analysis with the “zero-training” segmentation model MobileSAM. In our study, both RTMDet-tiny and RTMPose-s were trained for an adequate number of epochs on the melon dataset constructed by us. After training, they were tested on

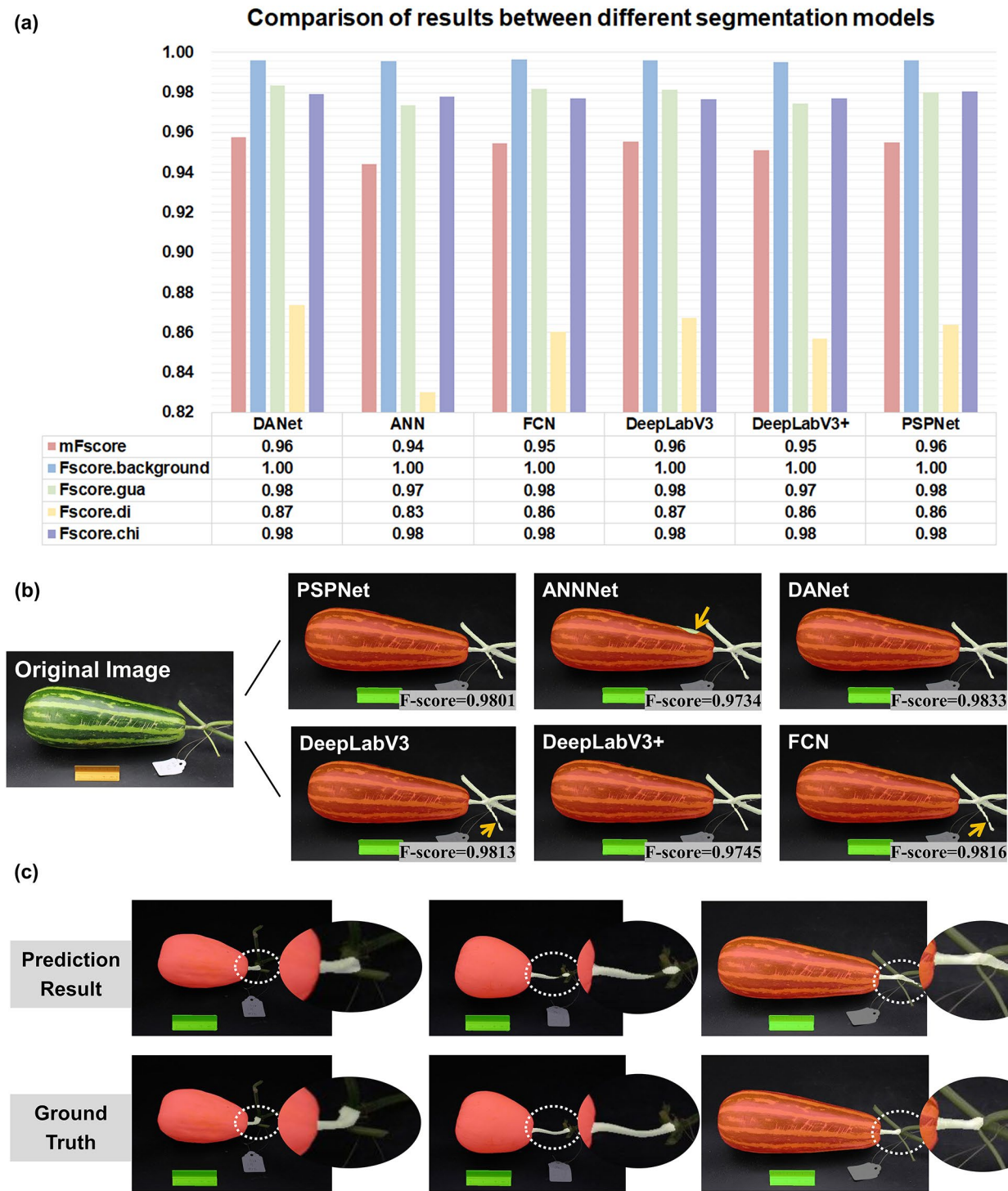


Fig. 4 Semantic segmentation results of melon fruit. **(a)** Comparison of segmentation results among different models. **(b)** Visualization results of predicting a certain variety of melon image using different segmentation models, with yellow arrows indicating areas of inaccurate segmentation. **(c)** Comparison between the segmentation results obtained using the DANet model and the ground truth for the desired pedicel part

the test dataset, and their prediction results were visualized. It was found that both models achieved satisfactory detection performance (Fig. 5). From the images, it can be observed that the pedicels were accurately located in the object detection task, and the detection accuracy of different varieties of melons was ideal. Moreover, on this basis, RTMPose could effectively identify the three key-points on the pedicel. Especially for pedicels with bends,

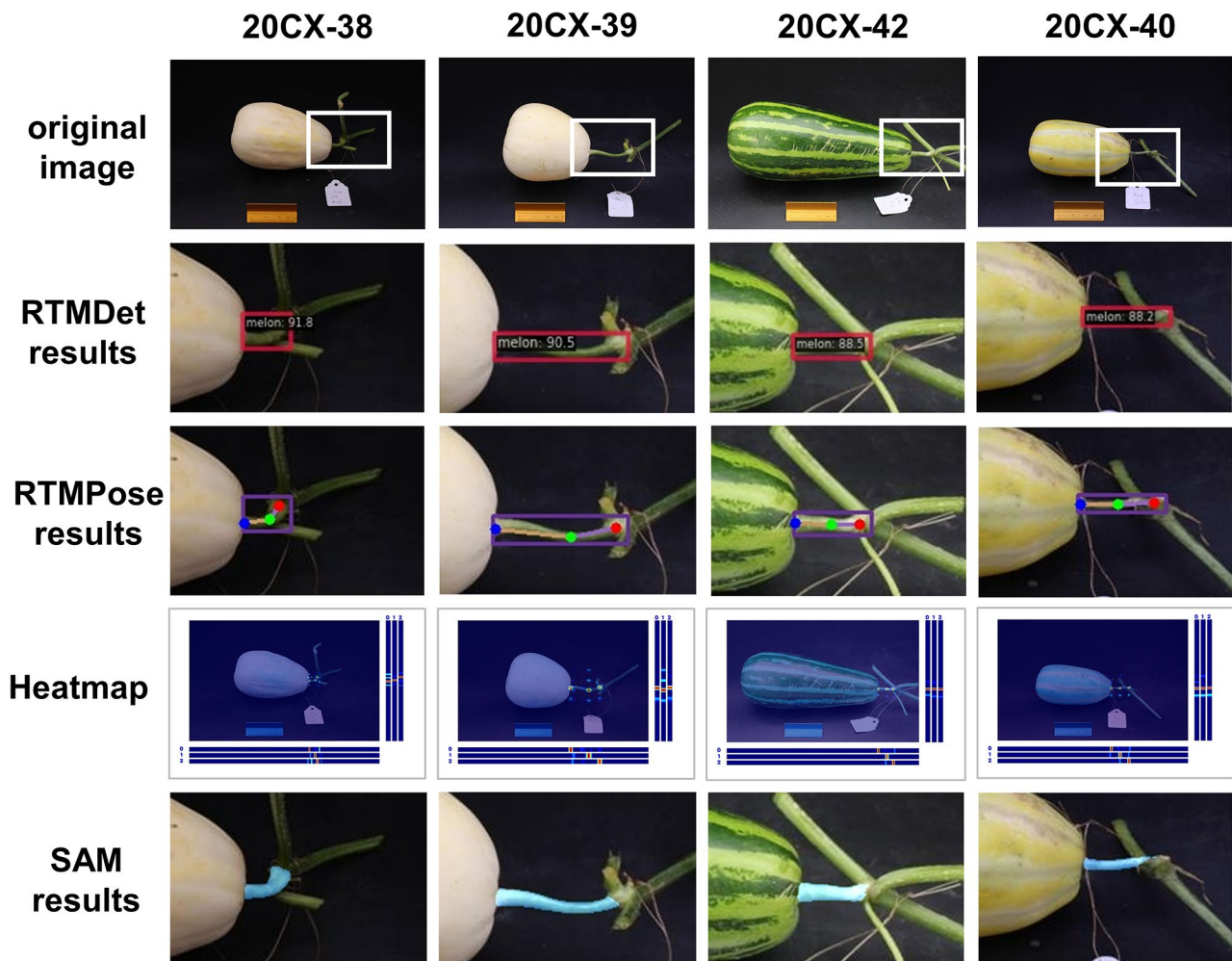


Fig. 5 Results of object detection, keypoint detection, and segmentation by MobileSAM for melon pedicel. From top to bottom: the original image, the predicted results from object detection, the predicted results from keypoint detection, the heatmaps for both axis directions in keypoint detection, and the results from SAM segmentation

it accurately located the bending points using keypoints, which allowed the skeleton formed by connecting keypoints to better represent the length of the pedicel. The high-precision detection results also provided strong support for subsequent segmentation.

The MobileSAM model, which does not require additional training, can directly segment the pedicel based on the coordinates of the aforementioned detection boxes and keypoints, demonstrating high accuracy on the melon dataset. Compared to directly using semantic segmentation models for segmentation, this approach clearly achieves higher precision, better meeting our task requirements. Additionally, due to the lightweight processing of MobileSAM compared to SAM, its inference speed is entirely sufficient for real-time extraction of phenotypic traits.

Phenotypic traits extraction

After completing the construction of melon segmentation algorithm, we extracted 11 phenotype parameters from all melon images, including the horizontal diameter, vertical diameter, fruit shape index of melon fruit, length and width of melon pedicel, as well as RGB and HSV values of fruit skin.

In order to validate the accuracy of the model, we manually measured the fruit length and width of 35 varieties of melons from the year 2020 and calculated their fruit shape index. Subsequently, we extracted the predicted values of three corresponding phenotypic traits obtained by the model from the images of melons in the year 2020 and separately fitted them with the true values (Fig. 6a). The data fitting and image plotting were performed in Origin (Version 2021, OriginLab Corporation, USA), using linear fitting, and a 95% confidence band was indicated. As depicted in the figure, the Pearson correlation coefficients between the true values and predicted values

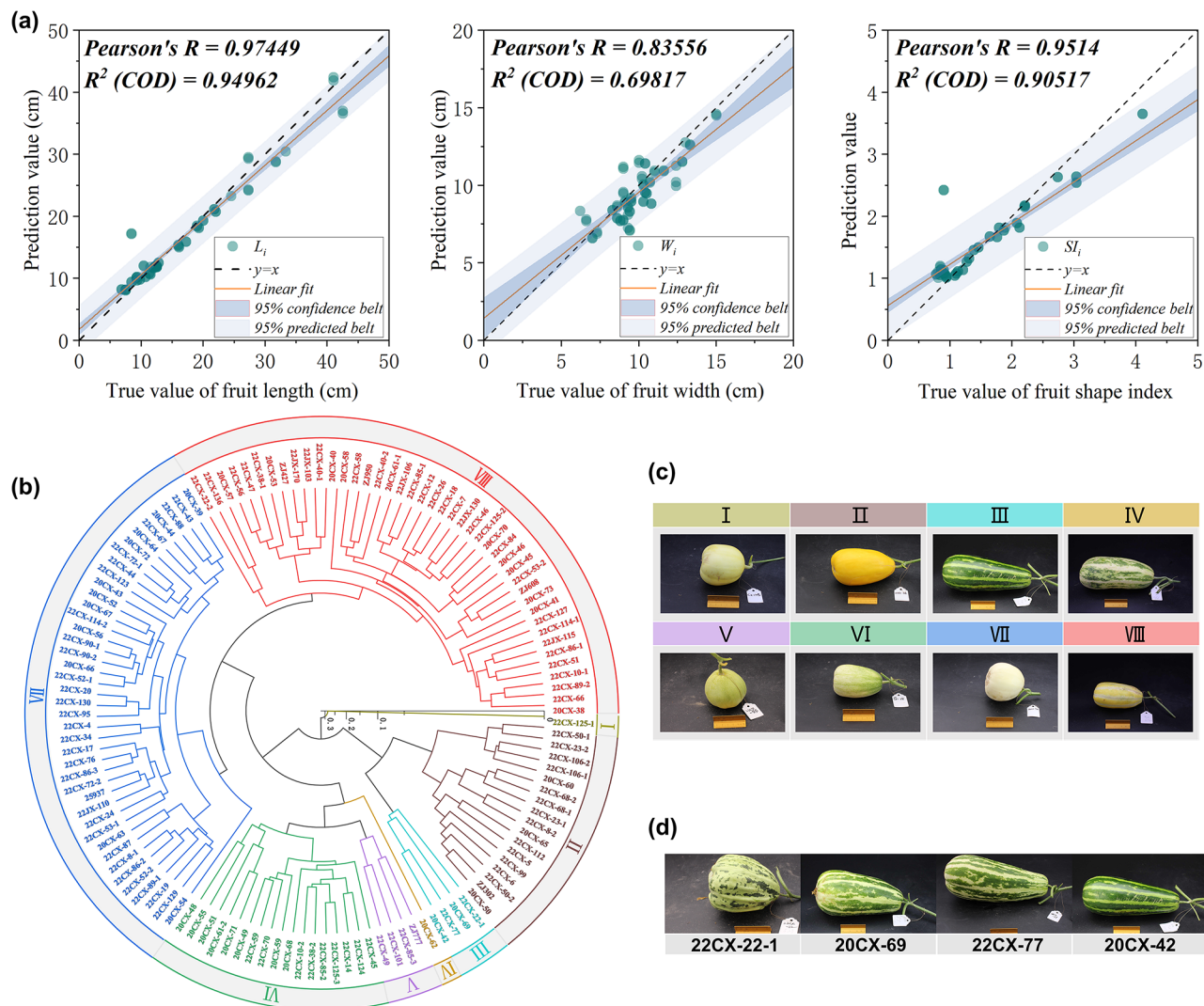


Fig. 6 Results of cluster analysis and linear regression. **(a)** Linear regression results for the predicted values of transverse diameter, longitudinal diameter, and fruit shape index compared to manually measured ground truth values for selected melon fruits. **(b)** Cluster results for all melon varieties across all phenotypic traits, with a clustering number of 8 classes. **(c)** Representative images of melons for each of the 8 clusters obtained through cluster analysis. **(d)** Images of all melon varieties contained within the third cluster

for both length and fruit shape index parameters are greater than 0.9. Compared to length, the width of melons is relatively similar among most varieties, with the high concentration of data amplifying minor errors, thus affecting the fitting performance to some extent. Nevertheless, the Pearson correlation coefficient remains above 0.8, indicating a good fit between the predicted and true values, suggesting a high degree of correlation between them. This further underscores the feasibility and accuracy of utilizing the algorithm proposed in our study for automated melon phenotypic traits acquisition.

Meanwhile, we conducted cluster analysis on all varieties of melons using the obtained 11 phenotypic traits in Origin. Specifically, for each melon variety with multiple images captured, the average values among different images were taken as representations of its phenotypic

traits. Eventually, all varieties were clustered into 8 clusters (Fig. 6b), each with its representative melon (Fig. 6c). It can be observed that melon varieties from different clusters exhibit obvious visual differences in characteristics such as fruit skin color, size, and shape, while those within the same cluster demonstrate a high degree of similarity. For instance, in the third cluster depicted in the figure, it comprises 4 varieties of melons (Fig. 6d), all displaying elongated melons with mottled dark green skin, a feature not found in other clusters. Furthermore, upon comparing the results of cluster analysis with those of genetic sequencing, we found them to be largely consistent. This suggests, on one hand, the existence of a correspondence between melon phenotypic traits and genotypes, indicating the feasibility of studying melon's genetic characteristics through phenotype analysis. On

the other hand, it underscores that the selected phenotypic traits in this study adequately capture the differences among different melon varieties and reflect the quality of melons, which is important for melon breeding.

Software development

To apply the findings of our study practically, we have integrated all the deep learning models mentioned before to develop a user-friendly phenotypic traits extraction software. The software consists of two pages: a login page and a phenotypic traits extraction page. Upon entering user credentials on the login page, the software automatically redirects to the phenotypic traits extraction page. Here, users can import local images that meet the specified requirements. The software supports various image formats for input. Once imported, the image is displayed in the display box on the left side of the page. Users can then perform various operations by clicking on the corresponding buttons, including fruit segmentation, pedicel keypoint detection, and pedicel segmentation. The segmented image results along with the values of corresponding phenotypic traits will be displayed in real-time below the page. Finally, all obtained phenotypic traits can be exported to an Excel spreadsheet for further analysis.

Conclusion

This study presents a deep learning framework that efficiently and accurately retrieves melon phenotypic traits based on a dataset of 117 melon varieties' images. On the one hand, we comprehensively compared the segmentation performance of six classic semantic segmentation models—ANNet, DANet, FCN, DeepLabV3, DeepLabV3+, and PSPNet—on melon fruits. Results demonstrate that DANet exhibits the best robustness and accuracy, outperforming other models with an F-score of 98.33% for fruit segmentation on the test set. On the other hand, we achieved high-precision pedicel keypoint localization and segmentation by integrating the object detection algorithm RTMDet-tiny, keypoint detection algorithm RTMPose-s, and MobileSAM model. Building upon this, we designed a series of extraction algorithms based on fruit and pedicel mask information, as well as pedicel keypoint coordinates, to obtain 11 phenotypic traits for melons. Linear fitting of model-predicted values and manually measured true values for some parameters revealed Pearson correlation coefficients greater than 0.8, confirming the feasibility and accuracy of the algorithms. Additionally, clustering analysis was conducted on all 117 varieties based on the obtained 11 phenotypic traits, revealing a close relationship between the selected phenotypic traits and melon genotypes, thus characterizing the genetic properties of different melon varieties and aiding in the selection of high-quality resources. Finally, we integrated various algorithms and developed a simple

and user-friendly software. By applying the research findings to the breeding process of melons, this software contributes to deeper investigations into the relationship between melon phenotypes and genotypes, benefiting experts in related fields.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13007-024-01293-1>.

Supplementary Material 1

Acknowledgements

Not applicable.

Author contributions

S.X. designed experiments with the advice of Y.W., H.H. and wrote the draft. J.S. collected and processed the data. X.F. guided experiments and edit the manuscript. Y.H. and Y.L. review the manuscript.

Funding

This work was supported by Key R&D Program of Zhejiang (Grant No. 2022C02032) and the Fundamental Research Funds for the Central Universities (226-2024-00038).

Data availability

No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 29 April 2024 / Accepted: 21 October 2024

Published online: 30 October 2024

References

1. Zhao G, Lian Q, Zhang Z, Fu Q, He Y, Ma S, et al. A comprehensive genome variation map of melon identifies multiple domestication events and loci influencing agronomic traits. *Nat Genet.* 2019;51(11):1607–15.
2. Wang X, Zeng H, Lin L, Huang Y, Lin H, Que Y. Deep learning-empowered crop breeding: intelligent, efficient and promising. *Front Plant Sci.* 2023;14:1260089.
3. Weyler J, Magistri F, Seitz P, Behley J, Stachniss C. In-Field Phenotyping Based on Crop Leaf and Plant Instance Segmentation. In 2022 [cited 2022 Oct 26]. pp. 2725–34. https://openaccess.thecvf.com/content/WACV2022/html/Weyler_In-Field_Phenotyping_Based_on_Crop_Leaf_and_Plant_Instance_Segmentation_WACV_2022_paper.html
4. Pieruschka R, Schurr U. Plant Phenotyping: Past, Present, and Future. *Plant Phenomics* [Internet]. 2019 Mar 26 [cited 2023 Apr 22];2019. <https://doi.org/10.34133/2019/7507131>
5. Tong H, Nikoloski Z. Machine learning approaches for crop improvement: leveraging phenotypic and genotypic big data. *J Plant Physiol.* 2021;257:153354.
6. Song P, Wang J, Guo X, Yang W, Zhao C. High-throughput phenotyping: breaking through the bottleneck in future crop breeding. *Crop J.* 2021;9(3):633–45.

7. Gongal A, Amatya S, Karkee M, Zhang Q, Lewis K. Sensors and systems for fruit detection and localization: a review. *Comput Electron Agric*. 2015;116:8–19.
8. Turgut K, Dutagaci H, Rousseau D, RoseSegNet. An attention-based deep learning architecture for organ segmentation of plants. *Biosyst Eng*. 2022;221:138–53.
9. Pound MP, Atkinson JA, Townsend AJ, Wilson MH, Griffiths M, Jackson AS, et al. Deep machine learning provides state-of-the-art performance in image-based plant phenotyping. *GigaScience*. 2017;6(10):gix083.
10. Liu X, Li N, Huang Y, Lin X, Ren Z. A comprehensive review on acquisition of phenotypic information of Prunoideae fruits: Image technology. *Front Plant Sci* [Internet]. 2023 [cited 2023 Dec 22];13. <https://www.frontiersin.org/article/https://doi.org/10.3389/fpls.2022.1084847>
11. Liu X, Xu Z, Editorial. Machine vision and machine learning for plant phenotyping and precision agriculture. *Front Plant Sci*. 2023;14:1331918.
12. Tu S, Xue Y, Zheng C, Qi Y, Wan H, Mao L. Detection of passion fruits and maturity classification using Red-Green-Blue depth images. *Biosyst Eng*. 2018;175:156–67.
13. Wu W, Liu T, Zhou P, Yang T, Li C, Zhong X, et al. Image analysis-based recognition and quantification of grain number per panicle in rice. *Plant Methods*. 2019;15(1):122.
14. Ni X, Li C, Jiang H, Takeda F. Deep learning image segmentation and extraction of blueberry fruit traits associated with harvestability and yield. *Hortic Res*. 2020;7(1):110.
15. Li S, Yan Z, Guo Y, Su X, Cao Y, Jiang B, et al. SPM-IS: an auto-algorithm to acquire a mature soybean phenotype based on instance segmentation. *Crop J*. 2022;10(5):1412–23.
16. Ho TT, Hoang T, Tran KD, Huang Y, Le NQK. Non-destructive classification of melon sweetness levels using segmented rind properties based on semantic segmentation models. *J Food Meas Charact*. 2023;17(6):5913–28.
17. Qian C, Sun S, Dong C, Chen C, Liu W, Du T. A study on phenotypic micro-variation of stored melon based on weight loss rate. *Postharvest Biol Technol*. 2023;204:112464.
18. Cho BH, Lee KB, Hong Y, Kim KC. Determination of Internal Quality indices in oriental melon using snapshot-type Hyperspectral Image and Machine Learning Model. *Agronomy*. 2022;12(9):2236.
19. Kalantar A, Edan Y, Gur A, Klapp I. A deep learning system for single and overall weight estimation of melons using unmanned aerial vehicle images. *Comput Electron Agric*. 2020;178:105748.
20. Sun Q, Chai X, Zeng Z, Zhou G, Sun T. Multi-level feature fusion for fruit bearing branch keypoint detection. *Comput Electron Agric*. 2021;191:106479.
21. Zheng C, Chen P, Pang J, Yang X, Chen C, Tu S, et al. A mango picking vision algorithm on instance segmentation and key point detection from RGB images in an open orchard. *Biosyst Eng*. 2021;206:32–54.
22. Wu Z, Xu D, Xia F, Suyin ZA, Keypoint-Based NY. 2022 [cited 2022 Oct 26]. <https://papers.ssrn.com/abstract=4199859>
23. Zhu Z, Xu M, Bai S, Huang T, Bai X. Asymmetric Non-local Neural Networks for Semantic Segmentation [Internet]. arXiv; 2019 [cited 2023 Dec 5]. <http://arxiv.org/abs/1908.07678>
24. Fu J, Liu J, Tian H, Li Y, Bao Y, Fang Z et al. Dual Attention Network for Scene Segmentation [Internet]. arXiv; 2019 [cited 2023 Apr 23]. <http://arxiv.org/abs/1809.02983>
25. Chen LC, Papandreou G, Schroff F, Adam H. Rethinking Atrous Convolution for Semantic Image Segmentation [Internet]. arXiv; 2017 [cited 2023 Dec 5]. <http://arxiv.org/abs/1706.05587>
26. Chen LC, Zhu Y, Papandreou G, Schroff F, Adam H. arXiv.org. 2018 [cited 2023 Dec 5]. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. <https://arxiv.org/abs/1802.02611v3>
27. Long J, Shelhamer E, Darrell T. Fully Convolutional Networks for Semantic Segmentation. In 2015 [cited 2023 Dec 5]. pp. 3431–40. https://www.cv-foundation.org/openaccess/content_cvpr_2015/html/Long_Fully_Convolutional_Networks_2015_CVPR_paper.html
28. Zhao H, Shi J, Qi X, Wang X, Jia J. Pyramid Scene Parsing Network [Internet]. arXiv; 2017 [cited 2023 Dec 5]. <http://arxiv.org/abs/1612.01105>
29. Lyu C, Zhang W, Huang H, Zhou Y, Wang Y, Liu Y et al. RTMDet: An Empirical Study of Designing Real-Time Object Detectors [Internet]. arXiv; 2022 [cited 2023 Apr 24]. <http://arxiv.org/abs/2212.07784>
30. Zhang J, Zhang J, Zhou K, Zhang Y, Chen H, Yan X. An improved YOLOv5-Based underwater object-detection Framework. *Sensors*. 2023;23(7):3693.
31. Yang X, Bist RB, Subedi S, Chai L. A computer vision-based Automatic System for Egg Grading and defect detection. *Animals*. 2023;13(14):2354.
32. Li X, Sun K, Fan H, He Z. Real-time cattle pose estimation based on Improved RTMPose. *Agriculture*. 2023;13(10):1938.
33. Jiang T, Lu P, Zhang L, Ma N, Han R, Lyu C et al. RTMPose: Real-Time Multi-Person Pose Estimation based on MMPose [Internet]. arXiv; 2023 [cited 2023 Apr 24]. <http://arxiv.org/abs/2303.07399>
34. Li Y, Yang S, Liu P, Zhang S, Wang Y, Wang Z et al. SimCC: a Simple Coordinate Classification Perspective for Human Pose Estimation [Internet]. arXiv; 2022 [cited 2024 Mar 19]. <http://arxiv.org/abs/2107.03332>
35. Zhang C, Han D, Qiao Y, Kim JU, Bae SH, Lee S et al. Faster Segment Anything: Towards Lightweight SAM for Mobile Applications [Internet]. arXiv; 2023 [cited 2023 Dec 5]. <http://arxiv.org/abs/2306.14289>
36. Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L et al. Segment Anything [Internet]. arXiv; 2023 [cited 2023 Apr 24]. <http://arxiv.org/abs/2304.02643>

Publisher's note

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