Computers and Electronics in Agriculture 201 (2022) 107333

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

A novel labeling strategy to improve apple seedling segmentation using BlendMask for online grading

Rui Suo [a], Longsheng Fu [a,b,c,d,*], Leilei He [a], Guo Li [a], Yaqoob Majeed [e], Xiaojuan Liu [a], Guanao Zhao [a], Ruizhe Yang [a], Rui Li [a,d]

[a] College of Mechanical and Electronic Engineering, Northwest A&F University, Yangling, Shaanxi 712100, China
[b] Key Laboratory of Agricultural Internet of Things, Ministry of Agriculture and Rural Affairs, Yangling, Shaanxi 712100, China
[c] Shaanxi Key Laboratory of Agricultural Information Perception and Intelligent Service, Yangling, Shaanxi 712100, China
[d] Northwest A&F University Shenzhen Research Institute, Shenzhen, Guangdong 518000, China
[e] Faculty of Agricultural Engineering and Technology, University of Agriculture, Faisalabad 38000, Pakistan

ARTICLE INFO

Keywords:
Apple seedling
BlendMask
Segmental labeling
Segmentation

ABSTRACT

A large number of apple seedlings are planted in orchards each year, where accurate and fast seedling grading to ensure their quality before planting has become a crucial problem. However, seedling grading by manual measurement of morphological indicators is laborious and inaccurate, and it's thus highly desirable to be replaced by machine vision. Seedling segmentation is one of the key steps of measuring morphological indicators and grading by machine vision. Therefore, a segmentation method of apple seedlings based on BlendMask with ResNet-101 to do transfer learning was proposed. A total of 450 original images were captured with Azure Kinect DK sensor. Root, rootstock, graft union, and scion of apple seedlings were labeled using a novel labeling strategy, which probably affect segmentation of thin and long objects. Scion was labeled with three different strategies, namely whole labeling (WL), segmental labeling (SL), and segmental-end-merge labeling (SEML). Results showed that the most suitable strategy was the SL for scion, which obtained a mean average precision of 91.2 % and the highest mean intersection over union of 79.3 % in the three labeling strategies. The average precisions of root, rootstock, graft union, and scion with the SL were 98.9 %, 89.3 %, 90.6 %, and 85.6 %, respectively. Intersection over unions of root, rootstock, graft union, and scion by the SL were 87.2 %, 75.8 %, 69.3 %, and 84.9 %, respectively. And it cost about 285 ms on average to process an image with resolution 3840 × 2160 pixels. The above results illustrated that the SL strategy is conducive to improve segmentation precision of thin and long objects. Moreover, apple seedlings can be effectively segmented, which is beneficial for the machine vision to measure morphological indicators and grade.

1. Introduction

Quantities of apple seedlings need to replenish dead seedlings and rebuild new orchards every year. Apple cultivation area has reached 1.91 million hectares in China, accounting for more than 41 % of world's total area (UN Food and Agriculture Organization, 2020). And about 153 million apple seedlings need to be replanted on approximately 5 % of the cultivation area (Sun et al., 2020). However, different planting regions have different requirements for seedling grades (Clark et al., 2015). Therefore, apple seedlings are required to be graded to ensure their quality before planting (Hou et al., 2018; Wang et al., 2011). Currently, accurate and fast evaluation of seedling grades has becoming desired in fruit industry.

It has been proved that machine vision has a potential of measurement morphological indicators and grading of apple seedlings. The apple seedlings are generally including root, rootstock, graft union and scion (Eaton and Robinson, 1977; Khanizadeh et al., 2005), as shown in Fig. 1. And they are mainly graded by morphological indicators (Gong et al., 2019), such as height (distance from bottom of rootstock to top of scion), diameter (at approximately 10 cm above graft union), and rootstock length of apple seedlings. Measurement and grading of apple seedlings by labors is time-consuming, laborious and non-objective. In recent years, applications of machine vision have been studied deeply in various fields, for instance, in the measurement of morphological indicators (Kim et al., 2021; Li et al., 2020; Sun et al., 2022; Wu et al.,

* Corresponding author at: College of Mechanical and Electronic Engineering, Northwest A&F University, Yangling, Shaanxi 712100, China.
E-mail addresses: fulsh@nwafu.edu.cn, longsheng.fu@wsu.edu (L. Fu).

https://doi.org/10.1016/j.compag.2022.107333
Received 1 June 2022; Received in revised form 12 August 2022; Accepted 22 August 2022
Available online 27 August 2022
0168-1699/© 2022 Elsevier B.V. All rights reserved.

Nomenclature

AP	Average Precision
AP^d	Average Precision of Detection
AP^s	Average Precision of Segmentation
Fast R-CNN	Fast Region Convolutional Neural Networks
FN	False Negatives
FP	False Positives
IoU	Intersection over Union
Mask R-CNN	Mask Region Convolutional Neural Networks
mAP	mean Average Precision
mAP^d	mean Average Precision of Detection
mAP^s	mean Average Precision of Segmentation
mIoU	mean Intersection over Union
P	Precision
R	Recall
ResNet	Residual Neural Network
RGB	a colour space, R for red color component, G for green color component, and B for blue color component
SEML	Segmental-end-merge Labeling
SL	Segmental Labeling
SOLO	Segmenting Objects by Locations
TP	True Positives
WL	Whole Labeling

2022). Compared with manual measurements, morphological indicators measured by machine vision is faster and less prone to error (McGuinness et al., 2021; Su et al., 2018). Consequently, it is promising to achieve rapid and accurate measurement and grading of apple seedlings by machine vision.

Accurate and robust plant segmentation based on deep learning is a primary technology for the measurement of morphological indicators by machine vision. Compared with traditional methods in plant segmentation, deep learning methods have great advantages in many aspects, including speed and precision (Gao et al., 2020; Kang et al., 2020; Parvathi and Tamil, 2021; Wang and He, 2021). Hao et al. (2021) applied Mask R-CNN (Mask Region Convolutional Neural Networks) with ResNet-50 (Residual Neural Network 50) to segment China firs, which achieved a F1 score of 84.7 % and an IoU (Intersection over Union) of 91.3 %. Perez-Borrero et al. (2021) employed Mask R-CNN with ResNet-50 to perform strawberry segmentation, which reached a mIoU (mean Intersection over Union) of 93.4 % and a segmentation speed of 33 ms per image with resolution of 1008 × 756 pixels. Tian et al. (2020) used Mask RCNN with ResNet-101 to segment apple flowers, which obtained a mIoU of 91.5 %. Chen et al. (2021) adopted DeepLabV3+ with ResNet-50 to segment occluded apple trees, which reported an IoU of 83.7 %, and required 36 ms to process an image with resolution of 640 × 480 pixels. Li et al. (2021) used U-Net to segment green apples, which obtained a precision rate 95.8 % and cost 350 ms on a 1920 × 1200 pixels' image. The above researches of plant segmentation adopting deep learning achieved high accuracy with fast speed. However, as far as author's best knowledge, there is no report about segmentation of apple seedlings up to now.

In terms of deep learning methods, segmentation of thin and long objects (such as scion in this study) may be affected by labeling strategy. General labeling strategy of target segmentation is to label an object with a tightly surrounded polygon. However, this labeling strategy is not suitable for thin and long objects. For instance, Li et al. (2021) only obtained a correct rate of 19.9 % using Mask R-CNN to segment transmission lines that labeled with thin and long polygons. Lin et al. (2021) applied Mask R-CNN with ResNet-101 to segment guava branches that labeled with thin and long polygons, which only achieved a precision of 49.9 %. What's more, some researchers tried different labeling strategies on thin and long objects. Yang et al. (2020) employed Mask R-CNN to train and segment citrus branches that labeled by segmental labeling and overall labeling, and obtained a precision of 96.2 % with grid labeling, which was 28.6 % higher than that with overall labeling. Wang et al. (2020) labeled branches and trunks of citrus with segmental labeling for training Mask R-CNN to segment citrus trees, and achieved an AP (average precision) of 96.8 %. Hence, a suitable labeling strategy is necessary to obtain good segmentation for thin and long scion.

BlendMask has made remarkable progress in image instance segmentation. Instance segmentation can provide multiple labels for separate instances of objects belonging to the same object class (Hafiz and Bhat, 2020), and is suitable for different labeling strategies. Currently, instance segmentation networks are divided into single-stage and two-stage (Roy and Bhaduri, 2022), where two-stage networks generally have higher accuracy but slower speed, while single-stage networks have the opposite situation (Kang and Chen, 2020). Recently, some single-stage networks have surpassed two-stage networks in terms of speed and precision, such as BlendMask (Gao et al., 2022). Quan et al. (2021) applied BlendMask with ResNet-101 for weed segmentation, which achieved a F1 score of 93.3 % and a segmentation speed of 115 ms per image with resolution of 1,024 × 1,024 pixels. Xi et al. (2021) used BlendMask with ResNet-101 to segment crowns of ginkgo trees in unmanned aerial vehicle multispectral images, which obtained an AP of 95.3 %. On the MS COCO dataset, BlendMask achieved higher AP and faster segmentation speed than other instance segmentation networks, such as Mask R-CNN, SOLO (Segmenting Objects by Locations) and TensorMask (Chen et al., 2020). In consequence, BlendMask was put into use to perform segmentation of apple seedlings.

In this study, root, rootstock, graft union, and scion of apple seedlings were labeled, trained, and segmented for the measurement of morphological indicators and grading. Among them, three different strategies were applied to label thin and long scion. Instance segmentation network, BlendMask with ResNet-101, was employed to do transfer learning for segmenting apple seedlings. The most suitable labeling strategy was selected to realize segmentation of apple seedlings by comparing AP and IoU based on BlendMask.

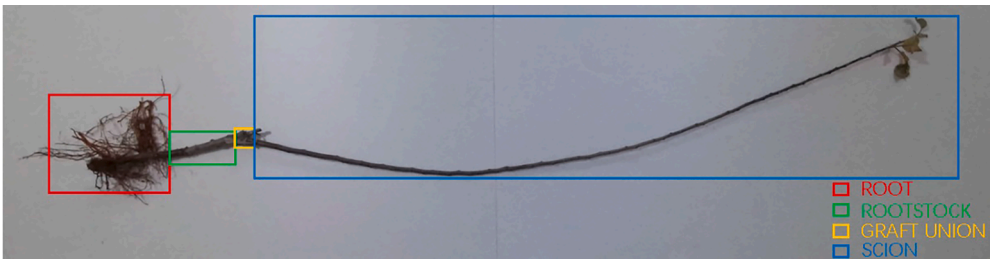


Fig. 1. Example of composition of the apple seedling. Root, rootstock, graft union, and scion are in red, green, yellow, and blue rectangles, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2. Materials and methods

2.1. Image acquisition

Image data of 'Yanfu 6′ apple seedlings was collected indoors in December 2021 and August 2022 from Northwest A&F University, Shaanxi, China. Apple seedlings of different heights, diameters, and shapes were pulled out from field. And after simple cleaning, they were placed on white background boards in different postures to take images, as shown in Fig. 2. RGB (Red, Green, and Blue) images were acquired using Azure Kinect DK sensor (MIC2574, Microsoft Inc., Redmond, WA), which was mounted on a support frame. The sensor lens was parallel to the background boards, and approximately 1.4 m away from the white background boards. In total, 450 original images with 3840 × 2160 pixels were collected, and then saved as JPG format. The dataset was uploaded on https://github.com/fu3lab/apple-seedling-images.

2.2. Labeling strategy and dataset augmentation

LabelMe 4.6.0 (an image labeler toolbox developed by Computer Science and Artificial Intelligence Laboratory of Massachusetts Institute of Technology) was used to label manually seedling images. Root, rootstock, graft union, and scion of apple seedlings were labeled according to the morphological indicators that need to be measured. Among them, root was labeled as 'RO' with a red rectangle, containing lateral and main roots, as shown in Fig. 3a. Rootstock was labeled as 'RS' with a green polygon, as shown in Fig. 3b. Occasionally, it does not exist on some apple seedlings. Graft union was slightly thicker than other parts, and was labeled as 'GU' with a yellow polygon, as shown in Fig. 3c. Root, rootstock, and graft union were labeled using above labeling methods in different labeling strategies.

Three different labeling strategies were used to label scion, as shown by yellow rectangles in Fig. 4. The first strategy was whole labeling (WL), which indicated that scion was labeled as 'SC' with a thin and long polygon, as shown in Fig. 4a. The second strategy was segmental labeling (SL), which referred to that scion was labeled as 'SC-*i*' (*i* represented the sequential number of polygon) by several polygons with 300 pixels long, as shown in Fig. 4b. On some seedlings, there were small forks on scion, as shown by red circles in Fig. 4c. In this dataset, the longest distance between the fork and front end of scion was 241 pixels. If a fork was labeled in two polygons, it will affect network to learn features of scion. Hence, length of the polygon was set to 300 pixels to ensure forks within the first polygon. The length of the last polygon was usually less than 300 pixels, as shown by green rectangle in Fig. 4b. However, if scion end in the last polygon was too small, it may be difficult for models to segment. Accordingly, the third strategy, segmental-end-merge labeling (SEML), was designed, aiming to improve segmentation precision of scion end. It was similar to the SL in that it also used multiple polygons with 300 pixels long to label scion. The difference between the SL and SEML was that the last polygon shorter than 300 pixels long was merged with previous polygon, as shown by green rectangle in Fig. 4d.

Once apple seedlings were labeled, corresponding annotated files, containing labeling name, point coordinates of the polygons, image information, etc., were generated and saved as JSON format file. Image dataset was randomly divided into training set (360 images) and testing set (90 images) with a ratio of 4:1. The training set and testing set were exclusive mutually, which ensured reliability of later evaluation standards.

Vertical mirroring, brightness transformation, and contrast transformation were adopted for dataset augmentation to improve robustness of segmentation models and to prevent it from over-fitting. In this study, Imgaug (https://imgaug.readthedocs.io/en/latest/index.html), a library for image augmentation in machine learning experiments was used. Vertical mirroring used 'flip' to transform the upper and lower sides of the image center on the horizontal centerline of the image. 'Multiply Brightness' was adopted to adjust brightness, and its thresholds were set as 1.1 and 0.7, respectively. 'Linear Contrast' was employed to contrast transformation, and its thresholds were set as 1.1 and 0.8, respectively. After data augmentation, the number of images in training set was increased from 360 to 2520.

2.3. Networks and training hyperparameters

BlendMask with ResNet-101 (Chen et al., 2020) was employed to train segmentation models of apple seedlings. The training of BlendMask network was conducted in a computer with AMD Ryzen 7 5800X 8-Core Processor (3.80 GHz) CPU, NVidia GeForce GTX 3080 Ti 12 GB GPU (10,240 CUDA cores), and 64 GB of memory, running on Ubuntu 18.04 system. Software included OpenCV 4.5.4, CUDA 11.0, cuDNN 8.2.1, AdelaiDet 0.2.0, detectron2 0.6, and Python 3.8. Batch size was set to 2, initial learning rate of 0.001 was used for all layers in the network, and a total of 400,000 iterations were set to observe training situation. The same parameters were used in three training of different labeling strategies. Transfer learning was employed for training network of seedling segmentation, which can adapt to new tasks and contribute to more accurate and faster training results (Li et al., 2022; Suo et al., 2021).

2.4. Performance evaluation

Four-evaluation metrics, namely average precision (AP_i) of each part, mean average precision (mAP), intersection over union (IoU_i) of each part, and mean intersection over union (mIoU) were used to evaluate the results of segmentation and detection achieved by trained models. The AP_i included AP_i^s and AP_i^d, where AP_i^s was calculated by segmentation masks, while AP_i^d by detection rectangles. The AP_i was calculated by precision (P_i) and recall (R_i) of segmentation or detection, which were defined in Eqs. (1), (2), (3), and (4), respectively. The *i* value represented the i^{th} part of apple seedlings: root ($i = 1$), rootstock ($i = 2$), graft union ($i = 3$), and scion ($i = 4$).

$$P_i^s = \frac{TP_i^s}{TP_i^s + FP_i^s} \tag{1}$$

$$P_i^d = \frac{TP_i^d}{TP_i^d + FP_i^d} \tag{2}$$



Fig. 2. Images of different apple seedlings in different postures.

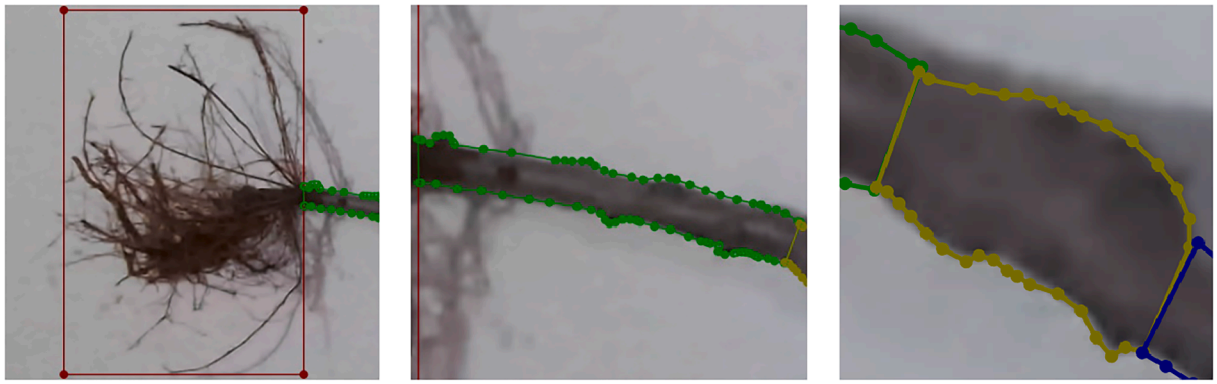


Fig. 3. Examples of (a) root labeled with a red rectangle; (b) rootstock labeled with a green polygon; and (c) graft union labeled with a yellow polygon. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

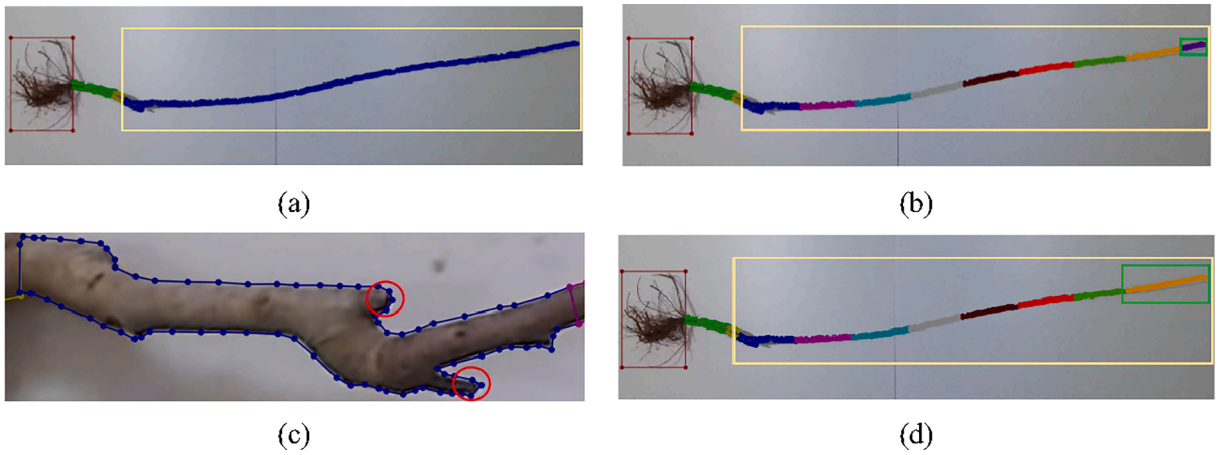


Fig. 4. Examples of different labeling strategies for scion. Scions in yellow rectangles were labeled with the (a) WL, (b) SL, and (d) SEML. Green rectangles in (b) and (d) represented the difference of the SL and SEML. Red circles in (c) represented the small forks within scion. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

$$R_i^s = \frac{TP_i^s}{TP_i^s + FN_i^s} \tag{3}$$

$$R_i^d = \frac{TP_i^d}{TP_i^d + FN_i^d} \tag{4}$$

Where TP (True Positives) indicates the number of segmented or detected correctly i^{th} part of apple seedlings, FP (False Positives) refers to the number of segmented or detected falsely i^{th} part of apple seedlings, and FN (False Negatives) represents the number of i^{th} part missed.

The AP_i is the area under the P_i and R_i curve, and can be expressed as Eqs. (5) and (6). AP is a measure for sensitivity of the network to instance segmentation, and is also an index that reflects performance of network. The mAP is the average value of AP_i for apple seedling, and defined in Eqs. (7) and (8). The IoU_i was the overlap ratio between ground truth images and the predicted segmentation maps, which was defined in Eq. (9). And the mIoU was the average IoU over all parts and defined in Eq. (10).

$$AP_i^s = \int_0^1 P_i^s(R_i^s) dR_i^s \tag{5}$$

$$AP_i^d = \int_0^1 P_i^d(R_i^d) dR_i^d \tag{6}$$

$$mAP^s = \frac{1}{4}\sum_{i=1}^{4} AP_i^s \tag{7}$$

$$mAP^d = \frac{1}{4}\sum_{i=1}^{4} AP_i^d \tag{8}$$

$$IoU_i = \frac{|A_i \cap B_i|}{|A_i \cup B_i|} \tag{9}$$

$$m\text{IoU} = \frac{1}{4}\sum_{i=1}^{4} IoU_i \tag{10}$$

Where A and B denote the ground truth images and the predicted segmentation maps of apple seedlings, respectively.

3. Results and discussion

3.1. Training evaluation

Training loss curves of the three labeling strategies by BlendMask were convergent, as shown in Fig. 5, where different line colors represented trained models of different labeling strategies, respectively. Abscissa and ordinate represented the number of iterations of training and loss values. At initial stage of the loss curve, the loss values of three

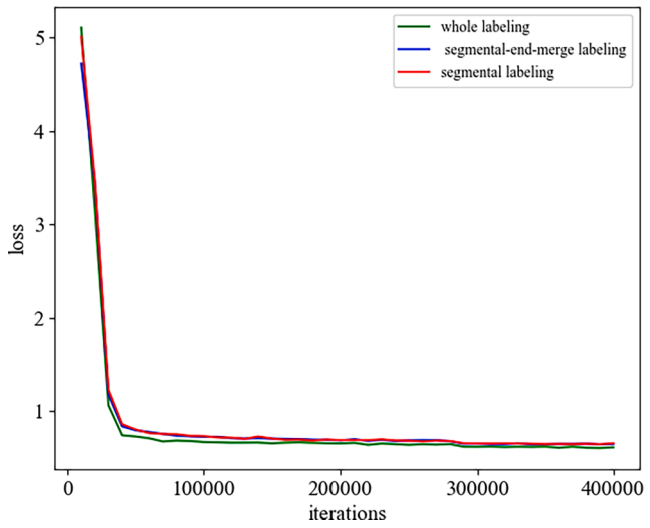


Fig. 5. Training loss curves by BlendMask with the three labeling strategies.

models rapidly decreased within approximately 40,000 iterations. After about 310,000 iterations, the WL, SL, and SEML began to reach stable loss values of 0.605, 0.646, and 0.647, respectively. The convergent curves with low loss values demonstrated that three models efficiently learn target features, which was promising to achieve desired segmentation and detection.

3.2. Segmentation results of WL and SL

The segmentation results of rootstock and graft union with the WL were superior to that with the SL. For rootstock, the AP^s of the SL was 89.3 %, and 3.5 % lower than that of the WL, as shown in Table 1. The same image in testing set was segmented by BlendMask, where rootstock with the SL was falsely segmented as scion (Fig. 6a), while correctly segmented with the WL in Fig. 6b. Same situation occurred in Fig. 6c, where a part of the scion was falsely segmented as rootstock with the SL. The reason for this result is that rootstock and scion have similar features (such as length and color) after labeling with the SL, which easily leads to false segmentation. However, after labeling with the WL, rootstock and scion have different length features, and were not easy to segmented falsely, as shown in Fig. 6b. In terms of graft union, the AP^s and IoU of the SL were 90.6 % and 69.3 %, which were 3.3 % and 1.1 %, respectively, lower than that of the WL. One correct and one false segmentation of graft union by the SL were in Fig. 6c, while only a correct segmentation was in Fig. 6d. Since graft union is generally between scion and rootstock, fork may be segmented falsely as graft union if scion is falsely segmented as rootstock. Although only scion is labeled with different strategies, it also has an impact on other parts, such as rootstock and graft union.

The segmentation results of scion with the SL had greatly improved by comparison with the WL. The AP^s and IoU of scion with the SL were 85.6 % and 84.9 %, which were 34.6 % and 41.2 %, respectively, higher than that with the WL, as shown in Table 1. The results were similar to Yang et al. (2020), which showed that precision on citrus branches with the SL based on Mask R-CNN with ResNet-101 was 26.63 % higher than that with overall labeling. The low AP^s and IoU of the WL come from a high number of miss segmentation. In total, 18 scions of apple seedlings were missed, accounting for 20.0 % of testing set. Scion was segmented correctly by several different color polygons with the SL (Fig. 7a), while it was missed with the WL, as shown in Fig. 7b. The reason is due to that the scion features become simple after labeling with the SL, which is beneficial for models to achieve more precise segmentation. Moreover, the number of samples in training set for scion increased from 2520 to 21,735 after using the SL, which is helpful for network to learn scion features, and improve generalization ability of models.

Overall, the SL demonstrated positive impact on segmentation of apple seedling. For root, the AP^ss of the SL and WL were same basically, and reached more than 98 %. The AP^ss of rootstock and graft union with the SL were near 90 % or above, which is acceptable segmentation. However, the AP^s, IoU, mAP^s, and mIoU of scion with the SL were 34.4 %, 41.2 %, 8.3 %, and 9.9 %, respectively, higher than that with the WL. The above results indicated that the SL minimized miss segmentation of scion, which will contribute to measurement of seedling height. Therefore, the SL is suitable for seedling segmentation in comparison to the WL.

3.3. Segmentation results of SL and SEML

For root, rootstock, and graft union, there was little difference in segmentation results of the SL and SEML due to the similar labeling strategies. The IoU and AP^ss of root, rootstock, and graft union by the SL and SEML were very close, as shown in Table 1. Additionally, the AP^ss of root, rootstock, and graft union were about 90 % or above, which meant that the three parts were segmented precisely by the two labeling strategies, as shown in Fig. 8.

In terms of scion, segmentation results with the SL outperformed that with the SEML. The SEML was expected to solve the problem that scion end with the SL was too small, and probably easily to be missed, as shown by black rectangle in Fig. 8a. However, the results were not work well, where scion end was also missed with the SEML, as shown by black rectangle in Fig. 8b. The IoU of scion with the SL was 84.9 %, which was 2.7 % higher than that with the SEML. The longer scion end was missed (Fig. 8b), while the comparatively shorter was segmented correctly (the second white rectangle in Fig. 8d). Longer scion end has more complex features, which is difficult to segment for network models (Li et al., 2021). In addition, the segmentation results of middle scion with the SL (black rectangle Fig. 8c) were worse than that with the SEML (the first white rectangle Fig. 8d). And this is the reason that the AP^ss of scion by the SL and SEML were same when the SL had better segmentation results on scion end than the SEML.

In general, the SL was more suitable for seedling segmentation compared to the SEML. For root, rootstock, and graft union, segmentation results of the two labeling strategies were very close. Although the AP^ss of scion with the SL and SEML were same, the SL had a higher IoU on scion, and showed the highest mIoU during the three strategies. Moreover, the SL had better segmentation results on scion end, which had positive influence on calculating seedling height. Although the number of segmentation targets was different in the three labeling strategies, average segmentation time of three models was all about 285 ms per image, which is consistent with the statement that inference time does not increase with the increasing number of segmentation targets (Chen et al., 2020).

Table 1
Segmentation results with different labeling strategies of apple seedlings.

Strategies	AP^s				mAP^s	IoU				mIoU
	Root	Rootstock	Graft Union	Scion		Root	Rootstock	Graft Union	Scion	
WL	98.8 %	92.8 %	93.9 %	51.0 %	84.1 %	87.5 %	75.8 %	70.4 %	43.7 %	69.4 %
SL	**98.9 %**	**89.3 %**	**90.6 %**	**85.6 %**	**91.1 %**	**87.2 %**	**75.8 %**	**69.3 %**	**84.9 %**	**79.3 %**
SEML	98.9 %	89.7 %	90.4 %	85.6 %	91.1 %	86.9 %	75.9 %	69.6 %	82.2 %	78.7 %

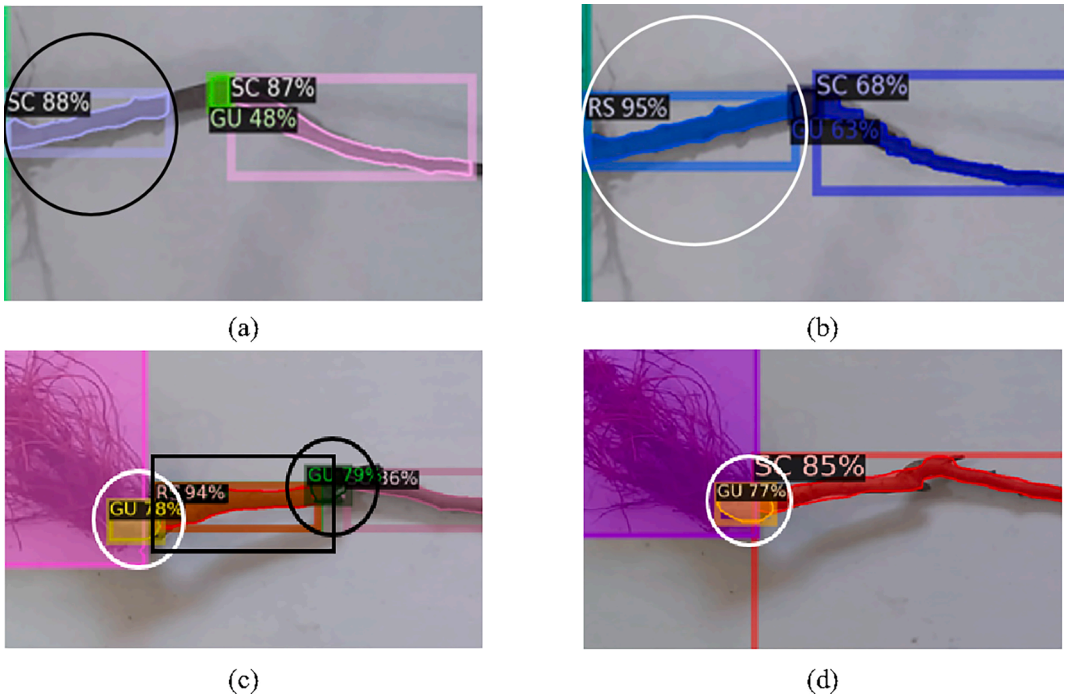


Fig. 6. Examples of rootstock and graft union segmentation by BlendMask with the WL and SL. Rootstock with (a) the SL and (b) WL; Graft union with (c) the SL and (d) WL. Segmented falsely parts were marked by black circles manually drawn, and examples of segmented correctly parts by white circles. The number in each bounding box referred to confidence.

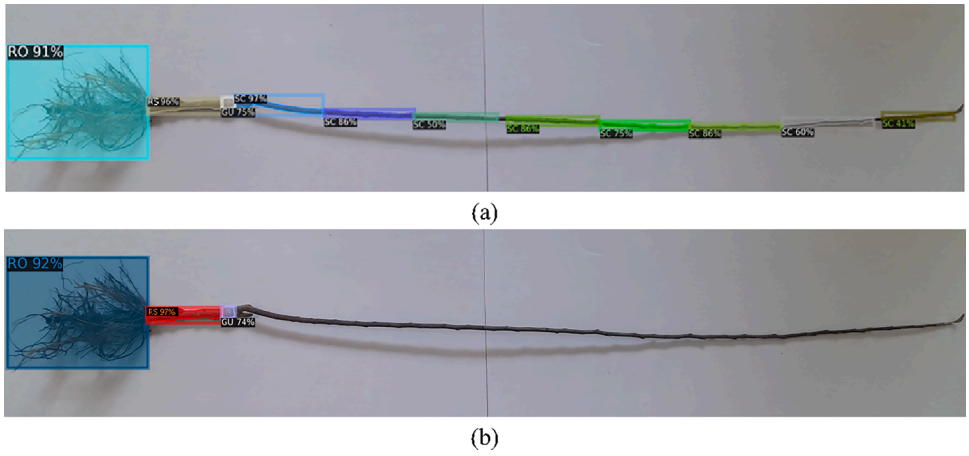


Fig. 7. Examples of seedling segmentation by BlendMask with the WL and SL. (a) scion correct segmentation with the SL; and (b) scion miss with the WL.

3.4. Detection results of the three labeling strategies

The three labeling strategies achieved high detection accuracy, as shown in Table 2. The mAP^ds of the WL, SL, and SEML were 95.4 %, 96.1 %, and 96.7 %, respectively. The AP^ds of each part with the three labeling strategies were more than 92 %. Accurate detection can help models to predict top-level attentions, which will generate top-level coarse instance information (Chen et al., 2020), and improve segmentation precision of apple seedling.

Detection results were also reflected the advantage of the SL on scion. The AP^d of scion with the WL was 92.9 %, which was 2.0 % and 3.2 % lower than that with the SL and SEML, respectively. Furthermore, the SL and SEML had better results of correct segmentation on scion than WL, as shown in Fig. 9, which illustrated that the SL also beneficial to for detection of thin and long objects. The SL was also employed in other studies, and obtained high mAP^ds, as shown in Table 3. However, networks and objects were all different between this study and other studies. Hence, Mask R-CNN with ResNet-101 was used for training the SL dataset to facilitate the comparison results.

A small proportion of target pixels in images led to poor detection results. Wang et al. (2020) and Yang et al. (2019) employed Mask R-CNN to segment citrus branches, and obtained mAP^ds of 98.2 % and 96.8 %, respectively. However, the mAP^d of Mask R-CNN on apple seedling dataset only reached 80.5 %. The reason is that the small proportion of target pixels makes it difficult for network to learn target features more precisely. Song et al. (2021) showed that wires were segmented difficultly than branches due to the smaller proportion of wire pixels. And similar conclusions could be drawn in Zhang et al. (2021). Pixel proportion of apple seedling in our study (Fig. 10c) was obviously smaller than that of citrus branches in others' studies (Fig. 10a and b). As a consequence, increasing proportion of target pixels is a good way to improve detection precision.

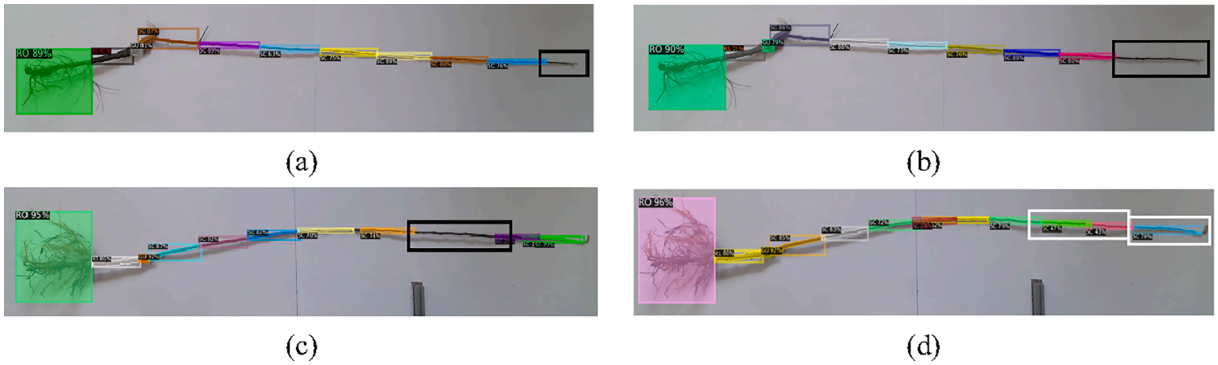


Fig. 8. Examples of seedling segmentation with the SL and SEML. Scion end was missed with (a) the SL and (b) the SEML. A part of scion middle was (c) missed with the SL; and (d) segmented correctly with the SEML. Hand-drawn black rectangles marked miss segmentation of scion, and white rectangles marked example of correct segmentation of scion.

Table 2
Detection results with different labeling strategies of apple seedlings.

	AP^d				mAP^d
Strategies	Root	Rootstock	Graft Union	Scion	
WL	98.9 %	97.4 %	92.5 %	92.9 %	95.4 %
SL	98.9 %	97.5 %	93.1 %	94.9 %	96.1 %
SEML	98.9 %	97.9 %	93.8 %	96.1 %	96.7 %

Compared with Mask R-CNN, BlendMask achieved a higher mAP^d and a faster processing speed. The mAP^d of BlendMask was 96.1 %, and 15.6 % higher than that of Mask R-CNN. BlendMask mixes features both from top-down and bottom-up (Zhang et al., 2021), which can help to find location of the bounding box more accurate than Mask R-CNN (Xi et al., 2021). Moreover, it cost 0.285 s for BlendMask to process an image, while 7.5 s for Mask R-CNN, which is not friendly for real-time applications, such as the measurement of morphological indicators and online grading of apple seedlings.

4. Conclusions

The primary goal of this study was to segment apple seedling accurately. Root, rootstock, graft union, and scion of apple seedlings were labeled, trained, and segmented, where scion was labeled with three different strategies (WL, SL, and SEML). Experimental results showed that the most suitable strategy was the SL, which achieved a mAP^s of 91.1 % and a highest mIoU of 79.3 % using BlendMask, and took about 285 ms on average to process an image with resolution 3,840 × 2,160 pixels. The AP^ss of root, rootstock, graft union, and scion with the SL were 98.9 %, 89.3 %, 90.6 %, and 85.6 %, respectively. The IoUs of root, rootstock, graft union, and scion with the SL were 87.2 %, 75.8 %, 69.3 %, and 84.9 %, respectively. The results indicated that measurement and grading of apple seedling could potentially achieve by machine vision. Additionally, the AP^s and IoU of scion with the SL were 34.4 % and 41.2 %, respectively, higher than that with the WL, which provides a research basis for improving the segmentation precision of thin and long objects based on deep learning. In the future, the segmentation results can be used to measure morphological indicators of apple seedlings, such as height, diameter, and rootstock length, which will be used to design and develop an algorithm for online seedling grading.

Table 3
Detection results from other studies with the SL.

	Method	Detection target	mAP^d	Processing speed (s / image)
Wang et al. (2020)	Mask R-CNN	Citrus branches	96. 8 %	8.2
Yang et al. (2019)	Mask R-CNN	Citrus branches	98.2 %	–
Proposed method	Mask R-CNN	Apple seedling	80.5 %	7.5
	BlendMask	Apple seedling	96.1 %	0.285

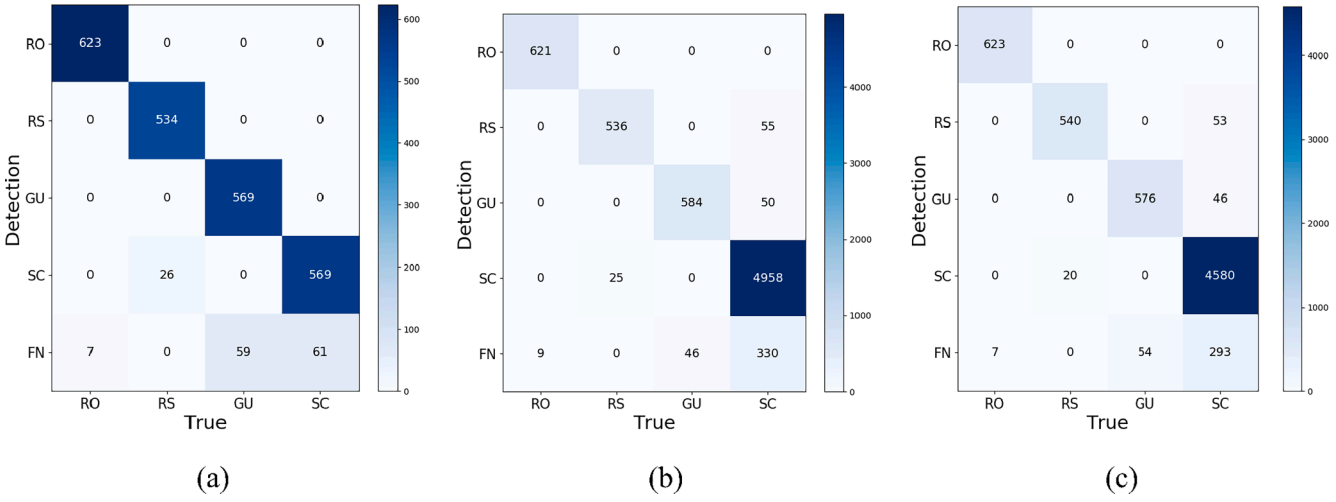


Fig. 9. Confusion matrixes of (a) WL, (b) SL, and (c) SEML.

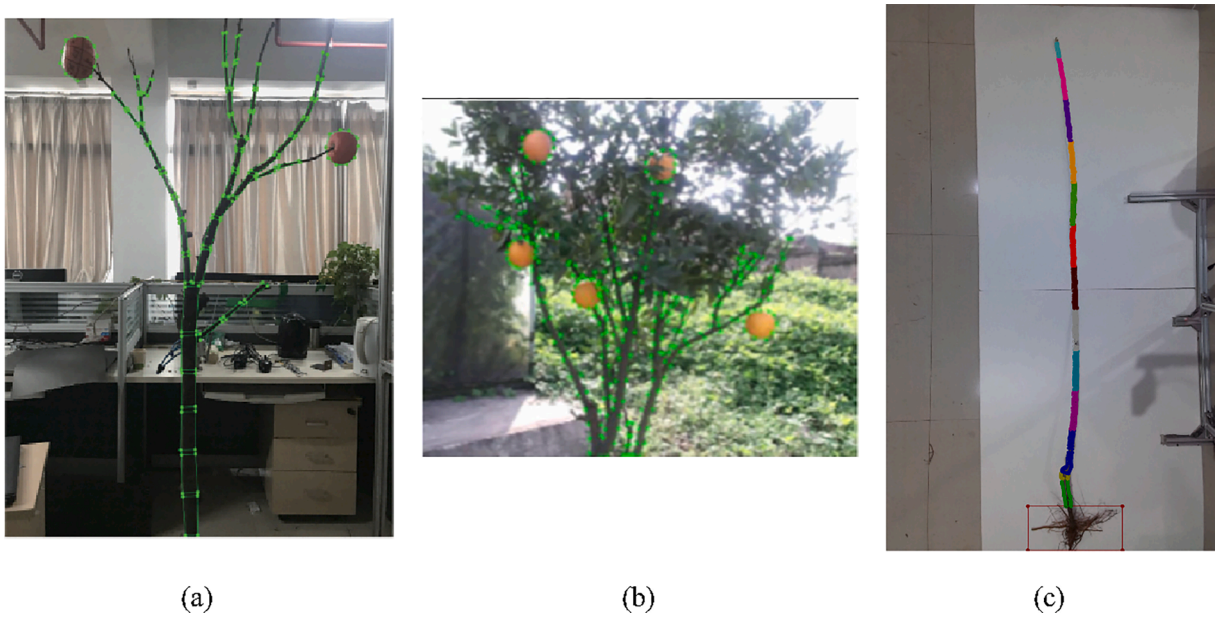


Fig. 10. Proportion of target pixels of (c) proposed method is smaller than that of (a) Wang et al. (2020) and (b) Yang et al. (2019).

CRediT authorship contribution statement

Rui Suo: Data curation, Investigation, Methodology, Writing – original draft. **Longsheng Fu:** Conceptualization, Methodology, Supervision, Writing – review & editing. **Leilei He:** Conceptualization, Methodology, Writing – review & editing. **Guo Li:** Conceptualization, Methodology, Writing – review & editing. **Yaqoob Majeed:** Investigation, Conceptualization, Writing – review & editing. **Xiaojuan Liu:** Investigation, Methodology, Writing – review & editing. **Guanao Zhao:** Methodology, Supervision, Writing – review & editing. **Ruizhe Yang:** Conceptualization, Methodology, Writing – review & editing. **Rui Li:** Methodology, Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (32171897); Youth Science and Technology Nova Program in Shaanxi Province of China (2021KJXX-94); Science and Technology Promotion Program of Northwest A&F University (TGZX2021-29); National Foreign Expert Project, Ministry of Science and Technology, China (DL2022172003L, QN2022172006L).

References

Chen, H., Sun, K., Tian, Z., Shen, C., Huang, Y., & Yan, Y. (2020). Blendmask: top-down meets bottom-up for instance segmentation. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8570–8578. doi: 10.1109/CVPR42600.2020.00860.

Chen, Z., Ting, D., Newbury, R., Chen, C., 2021. Semantic segmentation for partially occluded apple trees based on deep learning. Comput. Electron. Agric. 181, 105952 https://doi.org/10.1016/j.compag.2020.105952.

Clark, S.L., Schlarbaum, S.E., Schweitzer, C.J., 2015. Effects of visual grading on northern red oak (Quercus rubra L.) seedlings planted in two shelterwood stands on the Cumberland Plateau of Tennessee, USA. Forests 6 (10), 3779–3798. https://doi.org/10.3390/f6103779.

Eaton, G.W., Robinson, M.A., 1977. Interstock effects upon apple leaf and fruit mineral content. Can. J. Plant Sci. 57 (1), 227–234. https://doi.org/10.4141/cjps77-031.

Gao, F., Fu, L., Zhang, X., Majeed, Y., Li, R., Karkee, M., Zhang, Q., 2020. Multi-class fruit-on-plant detection for apple in SNAP system using Faster R-CNN. Comput. Electron. Agric. 182, 106052 https://doi.org/10.1016/j.compag.2020.105634.

Gao, Y., Qi, Z., Zhao, D., 2022. Edge-enhanced instance segmentation by grid regions of interest. The Visual Computer 1–12.

Gong, B., Wang, N., Zhang, T., Wu, X., Lü, G., Chu, X., Gao, H., 2019. Selection of tomato seedling index based on comprehensive morphology and leaf chlorophyll content. Trans. Chin. Soc. Agric. Eng. 35 (8), 237–244. https://doi.org/10.11975/j.issn.1002-6819.2019.08.028.

Hafiz, A.M., Bhat, G.M., 2020. A survey on instance segmentation: state of the art. Int. J. Multimedia Information Retrieval 9, 171–189. https://doi.org/10.1007/s13735-020-00195-x.

Hao, Z., Lin, L., Post, C.J., Mikhailova, E.A., Li, M., Chen, Y., Yu, K., Liu, J., 2021. Automated tree-crown and height detection in a young forest plantation using mask region-based convolutional neural network (Mask R-CNN). ISPRS J. Photogramm. Remote Sens. 178, 112–123. https://doi.org/10.1016/j.isprsjprs.2021.06.003.

Hou, J., Guo, H., Du, T., Shao, S., Zhang, Y., 2018. Effect of seedling grade standard on improving the quality of licorice (Glycyrrhiza uralensis F.): changes in the seven bioactive components and root biomass during two-year growth. Food Sci. Biotechnol. 27 (4), 939–945. https://doi.org/10.1007/s10068-018-0333-1.

Kang, H., Chen, C., 2020. Fast implementation of real-time fruit detection in apple orchards using deep learning. Comput. Electron. Agric. 168, 105108 https://doi.org/10.1016/j.compag.2019.105108.

Kang, H., Zhou, H., Wang, X., Chen, C., 2020. Real-time fruit recognition and grasping estimation for robotic apple harvesting. Sensors 20 (19), 14248220. https://doi.org/10.3390/s20195670.

Khanizadeh, S., Groleau, Y., Levasseur, A., Granger, R., Rousselle, G.L., Davidson, C., 2005. Development and evaluation of St Jean-Morden apple rootstock series. HortScience 40 (3), 521–522. https://doi.org/10.21273/hortsci.40.3.521.

Kim, W.S., Lee, D.H., Kim, Y.J., Kim, T., Lee, W.S., Choi, C.H., 2021. Stereo-vision-based crop height estimation for agricultural robots. Comput. Electron. Agric. 181, 105937 https://doi.org/10.1016/j.compag.2020.105937.

Li, B., Cockerton, H.M., Johnson, A.W., Karlström, A., Stavridou, E., Deakin, G., Harrison, R.J., 2020. Defining strawberry shape uniformity using 3D imaging and genetic mapping. Hortic. Res. 7, 115. https://doi.org/10.1038/s41438-020-0337-x.

Li, B., Chen, C., Dong, S., Qiao, J., 2021a. Transmission line detection in aerial images: An instance segmentation approach based on multitask neural networks. Signal Process. Image Commun. 96, 116278 https://doi.org/10.1016/j.image.2021.116278.

Li, Q., Jia, W., Sun, M., Hou, S., Zheng, Y., 2021b. A novel green apple segmentation algorithm based on ensemble U-Net under complex orchard environment. Comput. Electron. Agric. 180, 105900 https://doi.org/10.1016/j.compag.2020.105900.

Li, X., Pan, J., Xie, F., Zeng, J., Li, Q., Huang, X., Liu, D., Wang, X., 2021c. Fast and accurate green pepper detection in complex backgrounds via an improved Yolov4-tiny model. Comput. Electron. Agric. 191, 106503 https://doi.org/10.1016/j.compag.2021.106503.

Li, G., Suo, R., Zhao, G., Gao, C., Fu, L., Shi, F., Dhupia, J., Li, R., Cui, Y., 2022. Real-time detection of kiwifruit flower and bud simultaneously in orchard using YOLOv4 for

robotic pollination. Comput. Electron. Agric. 193, 106641 https://doi.org/10.1016/j.compag.2021.106641.
Lin, G., Tang, Y., Zou, X., Wang, C., 2021. Three-dimensional reconstruction of guava fruits and branches using instance segmentation and geometry analysis. Comput. Electron. Agric. 184, 106107 https://doi.org/10.1016/j.compag.2021.106107.
McGuinness, B., Duke, M., Au, C.K., Lim, S.H., 2021. Measuring radiata pine seedling morphological features using a machine vision system. Comput. Electron. Agric. 189, 106355 https://doi.org/10.1016/j.compag.2021.106355.
Parvathi, S., Tamil, S.S., 2021. Detection of maturity stages of coconuts in complex background using Faster R-CNN model. Biosyst. Eng. 202, 119–132. https://doi.org/10.1016/j.biosystemseng.2020.12.002.
Perez-Borrero, I., Marin-Santos, D., Vasallo-Vazquez, M.J., Gegundez-Arias, M.E., 2021. A new deep-learning strawberry instance segmentation methodology based on a fully convolutional neural network. Neural Comput. Appl. 33 (22), 15059–15071. https://doi.org/10.1007/s00521-021-06131-2.
Quan, L., Wu, B., Mao, S., Yang, C., Li, H., 2021. An instance segmentation-based method to obtain the leaf age and plant centre of weeds in complex field environments. Sensors 21 (10), 3389. https://doi.org/10.3390/s21103389.
Roy, A.M., Bhaduri, J., 2022. Real-time growth stage detection model for high degree of occultation using DenseNet-fused YOLOv4. Comput. Electron. Agric. 193, 106694 https://doi.org/10.1016/j.compag.2022.106694.
Song, Z., Zhou, Z., Wang, W., Gao, F., Fu, L., Li, R., Cui, Y., 2021. Canopy segmentation and wire reconstruction for kiwifruit robotic harvesting. Comput. Electron. Agric. 181, 105933 https://doi.org/10.1016/j.compag.2020.105933.
Su, Q., Kondo, N., Li, M., Sun, H., Firmanda, A., Habaragamuwa, H., 2018. Potato quality grading based on machine vision and 3D shape analysis. Comput. Electron. Agric. 152, 261–268. https://doi.org/10.1016/j.compag.2018.07.012.
Sun, X., Fang, W., Gao, C., Fu, L., Majeed, Y., Liu, X., Gao, F., Yang, R., Li, R., 2022. Remote estimation of grafted apple tree trunk diameter in modern orchard with RGB and point cloud based on SOLOv2. Comput. Electron. Agric. 199, 107209 https://doi.org/10.1016/j.compag.2022.107209.
Sun, X., Zou, Z., Yu, X., Shan, Y., 2020. Present situation and suggestions of apple nursery stock production in Yantai City. Deciduous Fruits 52 (4), 31–33.
Suo, R., Gao, F., Zhou, Z., Fu, L., Song, Z., Dhupia, J., Li, R., Cui, Y., 2021. Improved multi-classes kiwifruit detection in orchard to avoid collisions during robotic picking. Comput. Electron. Agric. 182, 106052 https://doi.org/10.1016/j.compag.2021.106052.
Tian, Y., Yang, G., Wang, Z., Li, E., Liang, Z., 2020. Instance segmentation of apple flowers using the improved Mask R-CNN model. Biosyst. Eng. 193, 264–278. https://doi.org/10.1016/j.biosystemseng.2020.03.008.
UN Food and Agriculture Organization. Production of apple by countries. Retrieved 2022-03-25 (2020).
Wang, Y., He, Y., Wang, K., Xiong, L., Wang, Z., Zhang, Y., 2020. Method of getting complete surface information of citrus by inspection robot. Trans. Chin. Soc. Agric. Mach. 51 (4), 232–240. https://doi.org/10.6041/j.issn.1000-1298.2020.04.027.
Wang, D., He, D., 2021. Channel pruned YOLO V5s-based deep learning approach for rapid and accurate apple fruitlet detection before fruit thinning. Biosyst. Eng. 210, 271–281. https://doi.org/10.1016/j.biosystemseng.2021.08.015.
Wang, F., Kang, S., Du, T., Li, F., Qiu, R., 2011. Determination of comprehensive quality index for tomato and its response to different irrigation treatments. Agric. Water Manag. 98 (8), 1228–1238. https://doi.org/10.1016/j.agwat.2011.03.004.
Wu, Z., Li, G., Yang, R., Fu, L., Li, R., Wang, S., 2022. Coefficient of restitution of kiwifruit without external interference. J. Food Eng. 327, 111060 https://doi.org/10.1016/j.jfoodeng.2022.111060.
Xi, X., Xia, K., Yang, Y., Du, X., Feng, H., 2021. Evaluation of dimensionality reduction methods for individual tree crown delineation using instance segmentation network and UAV multispectral imagery in urban forest. Comput. Electron. Agric. 191, 106506 https://doi.org/10.1016/j.compag.2021.106506.
Yang, C., Wang, Z., Xiong, L., Liu, Y., Kang, X., Zhao, W., 2019. Identification and reconstruction of Citrus branches under complex background based on Mask R-CNN. Trans. Chin. Soc. Agric. Mach. 50 (8), 22–30. https://doi.org/10.6041/j.issn.1000-1298.2019.08.003.
Yang, C., Xiong, L., Wang, Z., Wang, Y., Shi, G., Kuremot, T., Zhao, W., Yang, Y., 2020. Integrated detection of citrus fruits and branches using a convolutional neural network. Comput. Electron. Agric. 174, 105469 https://doi.org/10.1016/j.compag.2020.105469.
Zhang, X., Karkee, M., Zhang, Q., Whiting, M.D., 2021b. Computer vision-based tree trunk and branch identification and shaking points detection in Dense-Foliage canopy for automated harvesting of apples. J. Field Rob. 38 (3), 476–493. https://doi.org/10.1002/rob.21998.
Zhang, H., Sun, H., Ao, W., Dimirovski, G., 2021a. A survey on instance segmentation: Recent advances and challenges. Int. J. Innovative Comput. Inform. Control 17 (3), 1041–1053. https://doi.org/10.24507/ijicic.17.03.1041.