Computers and Electronics in Agriculture 207 (2023) 107734

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Green fruit segmentation and orientation estimation for robotic green fruit thinning of apples

Magni Hussain [a], Long He [a,b,*], James Schupp [b,c], David Lyons [d], Paul Heinemann [a]

[a] Department of Agricultural and Biological Engineering, The Pennsylvania State University, University Park, PA, USA
[b] Fruit Research and Extension Center, The Pennsylvania State University, Biglerville, PA, USA
[c] Department of Plant Science, The Pennsylvania State University, University Park, PA, USA
[d] Department of Autonomy, Perception, and Cognition, The Pennsylvania State University, University Park, PA, USA

ARTICLE INFO

Keywords:
Apple
Robotics
Green Fruit Thinning
Computer Vision
Instance Segmentation
Principal Component Analysis

ABSTRACT

Apple is a highly valued specialty crop in the U.S. Green fruit thinning is an important operation of apple production, which is the removal of excess fruitlets in the early summer. The task ensures that remaining fruits at harvest time grow to have good size and quality while reducing the risk of biennial bearing. Current methods of thinning include hand, chemical, and mechanical. However, hand thinning generally requires a large labor force to implement, chemical thinning is non-selective and dependent on timing and weather during application, and mechanical thinning is also non-selective and destructive. A robotic green fruit thinning system could possibly be implemented that does not exhibit the drawbacks of current methods. A vision system is an essential component for a robotic green fruit thinning system that is responsible for green fruit detection and segmentation, decision-making on which fruit to remove, and environment reconstruction for path planning. This study took the first step towards developing a vision system for robotic green fruit thinning. First, green fruit and stem instance segmentation was applied using Mask R-CNN. Then, green fruit and stem orientation estimation was applied using Principal Component Analysis (PCA). Average precision scores for green fruit and stem segmentation on all mask sizes were 83.4% and 38.9%, respectively, whereas these increased to 91.3% and 67.7% if only considering the fruits and stems with mask sizes greater than 32^2 pixels. Green fruit orientation estimation with correction made 89.3% and 75.5% of estimates accurate within 30° of actual orientations for ground-truth and segmentation-generated masks, respectively. Performances respectively were 97.4% and 84.0% when only unoccluded masks are considered. Orientation correction resulted in considerable improvements in all cases of green fruit orientation estimation, with the greatest improvement seen on unoccluded ground truth masks where estimates accurate within 30° of ground truth orientations increased by 23.9%. Stem orientation estimation achieved very high accuracies with corresponding scores of 99.8% and 99.7%. The outcomes provided guideline information for developing a robust machine vision system for robotic green fruit thinning.

1. Introduction

Apples are a highly-valued and produced crop in the U.S., with 9.56 billion pounds valued at $3.03 billion produced in 2021 (USDA, 2021). Green fruit thinning is an essential task for apple production, which is the process of removing excess fruitlets in May or June. Green fruit is thinned to increase size and quality of remaining fruit, as well as reduce the likelihood of biennial bearing, i.e., the occurrence of a heavy crop load in one season, and a light crop load the next (Vanheems, 2015). Manual green fruit thinning selectively removes unwanted fruits from apple trees. However, manual thinning is a labor-intensive task, and the shrinking labor force in apple orchards makes manual thinning difficult for fruit growers to implement. Manual thinning is more often used as follow-up to chemical thinning or mechanical thinning, as opposed to being the primary thinning method.

Chemical thinning has been studied on numerous tree fruit crops, such as apples, peaches, and citrus (Farias et al., 2019; Gonzalez et al., 2020; Stander et al., 2018). Chemical thinning can be applied much more quickly than manual thinning by using air blast sprayers. However, it is climate and cultivar dependent, as well as time sensitive (Schupp et al., 2017; Tyagi et al., 2017). Meanwhile, mechanical thinners have been studied on apple and peach crops at the bloom and green

* Corresponding author.
E-mail address: luh378@psu.edu (L. He).

https://doi.org/10.1016/j.compag.2023.107734
Received 21 December 2022; Received in revised form 20 February 2023; Accepted 21 February 2023
Available online 1 March 2023

fruit stages (Auxt Baugher et al., 2010; Kon et al., 2013; Miller et al., 2011; Reighard & Henderson, 2012). Similarly to chemical thinning, mechanical blossom thinning can reduce fruit load quickly, but it is non-selective, can increase the risk and transfer of disease, and can cause significant damage to spur leaf tissue (Kon & Schupp, 2018). A green fruit thinning solution that is more precise and selective would greatly benefit fruit growers. In particular, based on recent studies implementing robotic systems for tree fruit production (J. R. Davidson & Mo, 2015; Onishi et al., 2019; Ye et al., 2021; Zahid et al., 2021; Zhao et al., 2016), a robotic green fruit thinning system could potentially be developed that selectively removes fruit while exhibiting less of the drawbacks that exist in current methods.

One essential component of a robotic green fruit thinning system is a vision system, which is used to detect green fruit, determine the ones to remove, and generate a point cloud of the tree environment to be used for path planning purposes. Extensive work has been done recently in computer vision for object detection in specialty crops. Sa et al. (2016) implemented Faster Region-based CNN (Faster R-CNN) for apple detection using color and near-infrared images and obtained an F1 score of 0.838 for sweet pepper detection. Bargoti & Underwood (2017) implemented Faster R-CNN for the detection of mangoes, almonds, and apples in orchards, and achieved F1 scores of over 0.9 for apples and mangoes. Ganesh et al. (2019) implemented the instance segmentation framework, Mask R-CNN, to obtain pixel-wise masks of oranges using RGB + HSV images to obtain a precision value of 0.9753. With most research focusing on the detection of mature fruit, relatively little work has investigated the detection of green fruit in apple orchards. Wang & He (2021) implemented the YOLO V5s deep learning algorithm for the detection of apple fruitlets with recall, precision, F1 score, and false detection rate of 87.6 %, 95.8 %, 91.5 % and 4.2 %, respectively. However, currently no work has reported on instance segmentation for green fruit of apples. Furthermore, there is no known work in either detection or instance segmentation for green fruit stems, which is important information for end-effector positioning for green fruit thinning.

Instance segmentation is the process of detecting pixel-wise masks of objects in images. Several algorithms for instance segmentation have been proposed recently, particularly including the well-known Mask R-CNN (He et al., 2017) and YOLACT (Bolya et al., 2019). Two dimensional (2D) instance segmentation is potentially important for completing robotic tasks, as it can allow for 3D point cloud instance segmentation to be conducted from RGB-D images obtained with stereo cameras instead of more expensive lidar sensors, as demonstrated by Wang et al. (2021). Also, masks generated from instance segmentation can be readily utilized for orientation estimation algorithms, and can help determine more precise 3D features of objects when mapped into a 3D point cloud using depth images obtained with stereo-vision camera. Ultimately, instance segmentation could outperform standard detection algorithms by providing more important information for robotic green fruit thinning.

One of the main challenges in green fruit detection is that green fruit color is often similar to that of the background canopy. This can make green fruit segmentation difficult using methods that rely simply on color thresholding. Other features, such as green fruit and canopy texture and shape, are necessary to properly differentiate green fruit from canopy. Classical machine learning methods traditionally used hand-crafted features, i.e., features manually designed by humans, such as Histogram of Oriented Gradients and Scale Invariant Feature Transform (Dalal & Triggs, 2005; Lowe, 1999). However, deep neural networks, which have been recently introduced in computer vision, are able to leverage the most important features for classification and segmentation tasks without requiring features to be manually indicated. Several studies have shown success in applying segmentation to persimmons and green apples using deep neural network-based methods (Jia, Liu, et al., 2022; Jia, Wei, et al., 2022; Liu et al., 2022). It is expected that segmentation of apple in the green fruit (fruitlet) stage should work similarly well.

End-effectors for use in robotic systems have been developed for a large variety of specialty crops including apples, oranges, cucumbers, and tomatoes (J. Davidson et al., 2020). Hussain et al. (2022) describe development of a green fruit thinning end effector prototype. To properly position an end-effector for green fruit removal, target fruit and stem locations and orientations are required. The type of end-effector used will determine which information will be the most useful. Knowing the stem orientation is important for orienting a snipper-like end-effector perpendicular to the stem for an optimal cut. For a pulling end-effector, the most efficient direction for pulling green fruit is directly in-line with the fruit, as sideways motions are more inefficient; in this case, it is most important to know the orientation of the fruit. Several methods exist for determining the orientation of objects within a scene, some of which are based on deep learning (Choi et al., 2016; Hara et al., 2017). One relatively simple method for determining the orientation of an object is based on principal component analysis (PCA) (Pearson, 1901). Principal component analysis is traditionally used for applications such as dimensionality reduction. However, the generation of eigenvectors through principal component analysis can be leveraged for determining an object's orientation. Furthermore, the ellipsoid nature of green fruit can lend itself to providing accurate estimates of green fruit orientations using PCA without requiring large and complex neural networks to be trained and implemented for this purpose.

This study aimed to implement and evaluate fruit and stem segmentation and orientation estimation algorithms for their feasibility of use in a machine vision system for robotic green fruit thinning in apple orchards. The main objectives were 1) to apply transfer learning on the Mask R-CNN algorithm to perform instance segmentation for green fruits and stems; and 2) to implement an orientation estimation algorithm using Principal Component Analysis to estimate the orientation of fruits and stems based on their generated masks. A correction algorithm is applied after orientation estimation to obtain the correct sense of the resulting vector. The performance of segmentation and orientation is then evaluated using a dataset consisting of Golden Delicious, GoldRush, and Fuji cultivars during the green fruit stage containing ground truth segmentation and orientation annotations.

2. Methodology

2.1. Image dataset

A set of images were acquired on May 18 (Fuji and GoldRush) and May 23, 2021 (Golden Delicious). All images were obtained at the Penn State Fruit Research and Extension Center (Biglerville, PA). The green fruit diameters during this time broadly ranged from 10 to 30 mm. The images were obtained using an iPhone 12 and a Samsung Note 10 + at resolution 3024 × 4032. Images were taken at distances between 1 and 5 ft, each varying in the number of fruit clusters. Images were resized for training and testing to 1024 × 1024. A total of 521 images including Fuji, Golden Delicious, and GoldRush cultivars were obtained for instance segmentation algorithm training and evaluation. The images contained 5,683 green fruit and 4,302 stem masks.

The dataset was split into training, validation, and test datasets using a 70/15/15 ratio. The training, validation, and testing datasets were randomly and uniformly sampled from the dataset. The training dataset consisted of 365 images, which contained a total of 3918 green fruit masks and 2960 stem masks. All fruits and stems within the training dataset images were annotated using the VGG Image Annotator (VIA) (Dutta & Zisserman, 2019). The validation and testing datasets each contained 78 images. The validation dataset contained 918 fruit masks and 693 stem masks, and the test dataset contained 847 fruit masks and 649 stem masks. Further information on the number of images, the number of fruit/stem masks, and average number of fruit/stem masks per image for each dataset/cultivar combination, and the average fruit/stem mask size for each dataset/cultivar combination can be found

respectively in Tables 1-4. Table 5 Shows fruit/stem mask size category distributions for each cultivar.

2.2. Green fruit and stem instance segmentation

2.2.1. Mask R-CNN model development

The instance segmentation algorithm, Mask R-CNN (He et al., 2017), was used to obtain pixel-wise masks of green fruits and stems in the RGB images. The flowchart of Mask R-CNN applied to fruit and stem segmentation is illustrated in Fig. 1. The first stage of Mask R-CNN consists of a convolutional neural network, which is used to generate a set of feature maps from an input image. The feature maps are then processed by a region proposal network. This network proposes bounding box regions which may contain green fruit or stem masks. The RoIAlign operation is applied to these regions of interest to generate a small feature map for each. Each feature map is processed by a fully-connected neural network to generate predicted bounding boxes and corresponding classes, while another set of convolutional neural networks is used to generate pixel-wise masks.

The Mask R-CNN model was trained using transfer learning, with the ResNet-101 backbone and the original model of MS COCO (Lin et al., 2014). All parameters of the backbone were trained. A workstation graphics card (Quadro P5000, Nvidia Corporation, USA) was used for training. Horizontal flipping and 90°, 180°, and 270° rotations were applied to the training dataset during training as data augmentation to effectively multiply the training dataset size by a factor of eight. Training continued until convergence of the algorithm was observed at 64 epochs. The training parameters, learning rate for first 20 epochs, learning rate after 20 epochs, learning momentum, weight decay, training steps per epoch, and validation steps per epoch, were 0.001, 0.0001, 0.9, 0.0001, 365, and 78, respectively. A batch size of one image was used for training.

2.2.2. Segmentation evaluation

Average precision (AP) (IoU = 0.5) was used to evaluate the performance of instance segmentation for each mask type. The metrics are first calculated for the following precision and recall values at 101 evenly-spaced threshold values between 0 and 1, as in Equations (1) and (2).

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

$$R_i = \frac{TP_i}{TP_i + FN_i} \tag{2}$$

Where P_i and R_i are the precision and recall values at threshold index i, respectively, and TP_i, FP_i, and FN_i are the number of true positives, false positives, and false negatives at threshold index i, respectively. After these values are calculated, average precision is then calculated through Equation (3).

$$AP = \sum_{n=1}^{101} (Recall_n - Recall_{n-1}) Precision_n \tag{3}$$

The performance of the Mask R-CNN tends to increase as the size of the fruit/stem mask increases (He et al., 2017). The fruit and stem mask sizes may vary considerably, which in turn will result in varying instance segmentation performances. Thus, the AP scores are also obtained for the following mask sizes: Mask $> 32^2$, $27^2 <$ Mask $< 32^2$, $20^2 <$ Mask $< 27^2$, and Mask $< 20^2$, all in pixel units. Segmentation performance is also evaluated for each cultivar across the different size categories. Since it is apparent that individual size categories in the test dataset for fruit and stem masks for a given category shown in Table 5 can have insufficiently few masks for evaluation, mask size categories are combined for each mask type accordingly for more reliable evaluation: $<32^2$ and $> 32^2$ for green fruit masks, and $< 20^2$ and $> 20^2$ for stem masks.

Table 1
The number of images per dataset and cultivar.

Images/cultivar(s)	All	training	validation	Testing
All	521	365	78	78
GoldRush	228	164	33	31
Fuji	180	119	28	33
Golden Delicious	113	82	17	14

Table 2
The number of fruit/stem masks, respectively, per dataset and cultivar.

Fruit/stems per cultivar(s)	All	training	validation	Testing
All	5683/4302	3918/2960	918/693	847/649
GoldRush	2659/1993	1879/1429	438/318	342/246
Fuji	1878/1450	1224/920	294/236	360/294
Golden Delicious	1146/859	815/611	186/139	145/109

Table 3
The average number of fruit/stem masks per image for each dataset/cultivar subset.

Avg. fruit/stems per image	All	training	validation	Testing
All	10.9/8.2	10.7/8.1	11.8/8.9	10.9/8.2
GoldRush	11.7/8.7	11.5/8/7	13.3/9.6	11.0/7.9
Fuji	10.4/8.1	10.3/7.7	10.5/8.4	10.9/8.9
Golden Delicious	10.1/7.6	9.9/7.5	10.9/8.2	10.4/7.8

Table 4
The average fruit/stem mask size (pixels) for each dataset/cultivar subset.

Fruit/stem size avg	all	training	validation	Testing
All	2023.5/ 329.0	2033.0/ 323.5	2008.3/ 330.0	1996.1/ 3535
GoldRush	2033.8/ 303.7	2038.0/ 306.0	1972.7/ 298.0	2089.1/ 297.2
Fuji	1564.9/ 302.3	1482.7/ 272.3	1833.1/ 334.1	1625.0/ 371.0
Golden Delicious	2751.1/ 433.0	2847.8/ 441.4	2368.9/ 396.1	2697.7/ 433.3

Table 5
The number of fruit/stem masks per mask-size subset for each cultivar.

Fruit/stem masks per cultivar (test dataset)	Overall	$> 32^2$	27^2–32^2	20^2–27^2	$< 20^2$
All		526/ 26	104/26	120/ 108	97/ 489
GoldRush		248/7	32/5	42/35	20/ 199
Fuji		171/ 14	56/6	63/42	70/ 232
Golden Delicious		107/5	16/15	15/31	7/58

2.3. Orientation estimation

2.3.1. Principal component analysis (PCA)

An orientation estimation algorithm from the OpenCV Library (Bradski, 2000) based on principal component analysis (PCA) (Pearson, 1901) was used to estimate the orientation of fruit and stem masks. The green fruit and stem orientation estimation process is illustrated in Fig. 2. The principal idea of PCA applied to orientation estimation is to first consider a fruit or stem mask as a set of points, then find a weight vector that maps the set of mask points into a new set of points with maximal variance. This vector is also the eigenvector for the set of points that corresponds to the largest eigenvalue. The obtained weight vector is then assigned as the orientation vector for a given fruit or stem mask.

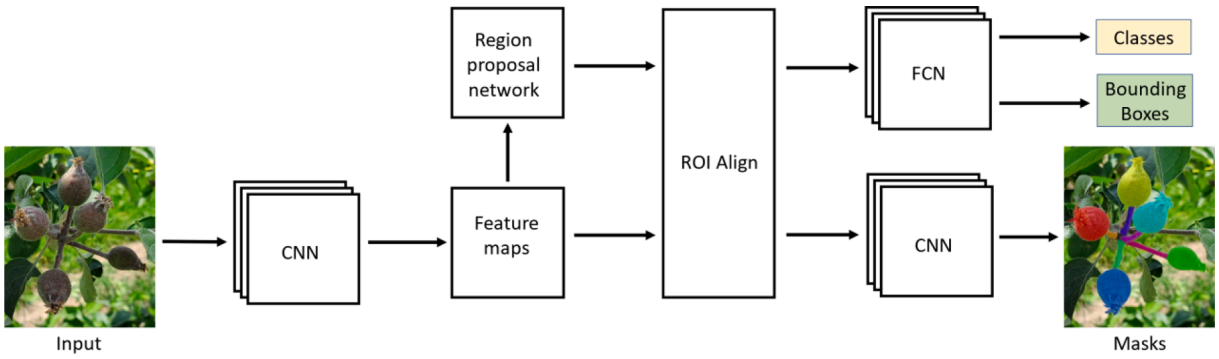

Fig. 1. Flowchart of Mask R-CNN model for green fruit and stem segmentation. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

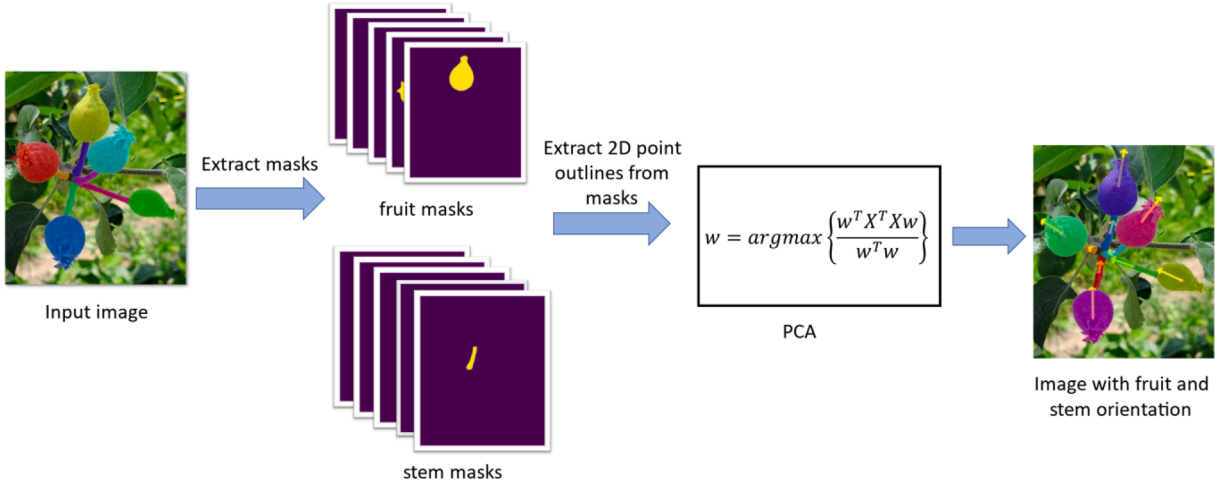

Fig. 2. The procedure of green fruit and stem orientation estimation with Principal Component Analysis (PCA) method. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The orientation vector w is obtained by solving for the following equation:

$$w = argmax\left\{\frac{w^T X^T X w}{w^T w}\right\} \tag{4}$$

Where X is the matrix containing the set of points in a given mask. Finally, orientation correction is applied to the resulting image.

2.3.2. Orientation correction

The orientation vector, once generated, should be aligned with approximately the correct orientation of the mask. However, PCA has no inherent way of determining the "sense" of the orientation vector, i.e., it cannot determine if the direction in which the orientation vector faces is in one direction along the orientation line or the opposite. While the OpenCV implementation used for PCA returns an orientation with both sense and direction, it is not clear how the sense is obtained, which can make the generated sense unreliable. For stems, this does not matter so much for the purposes of green fruit thinning, as any thinning end-effector that is designed to interact with the stem, e.g., a snipping end-effector, should likely be able to align with the stem properly regardless of whether the orientation vector faces one way or the opposite. Thus, orientation correction for stems is not applied. However, for end-effectors that rely on the fruit orientation, e.g., a pulling end-effector, knowing the correct direction of the orientation vector is essential for ensuring that an end-effector approaches fruit from the calyx end, rather than the stem end, which will likely result in stem collision or ineffective removal action, depending on the end-effector design. Thus, orientation correction needs to be applied for green fruit. The correct orientation of a green fruit was defined as one that points from the green fruit centroid towards its calyx. A method was developed to correct green fruit orientations with correct directions but incorrect senses, i.e., the ones whose orientation vectors point towards the stem end of the green fruit.

The proposed orientation correction method relies on the "jagged" nature of the calyx of green fruit when viewed as a mask in an image. Due to this, a fruit mask typically has more apparent corners in its calyx than in the stem end. First, the outline of each green fruit mask within an image is converted into a set of points in 2D. Then, the orientation of the point set is aligned with the x-axis. Afterwards, orientation correction applies the Harris corner detection (Harris & Stephens, 1988) for the fruit mask to detect all corners. The Harris corner detection first starts with the sliding-window function as shown in Equation (5).

$$E(u,v) = \sum_{x,y} w(x,y)[I(x+u, y+v) - I(x,y)]^2 \tag{5}$$

Where u and v are window displacements in the $\times$ and y directions, respectively; $w(x,y)$ is the window function, which is a 3 $\times$ 3 unity window for this application; and $I(x,y)$ is the input image value at coordinates (x, y). The function can be approximated as the following using the Taylor Series:

$$E(u,v) \approx [uv]M\begin{bmatrix} u \\ v \end{bmatrix} \tag{6}$$

$$M = \sum_{(x,y)\in W} \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \tag{7}$$

Where I_x and I_y are partial derivatives of the input image. The eigenvalues of M, λ_1 and λ_2, determine whether a particular patch in an image is flat, an edge, or a corner. Specifically, if both eigenvalues are large, then a patch is a corner. If one eigenvalue is considerably greater than the other, then the patch is an edge. If both eigenvalues are small, then the patch is flat. Instead of calculating the eigenvalues of M directly, the following metric R can be used for each location:

$$R = \det(M) - k(trace(M))^2 \tag{8}$$

$$\det(M) = \lambda_1 \lambda_2 \tag{9}$$

$$trace(M) = \lambda_1 + \lambda_2 \tag{10}$$

Where k is a user-chosen parameter selected to obtain the best corner-detection performance in a given application. The chosen k value for orientation correction was 0.04. R exhibits the following behavior based on the above eigenvalue response: if $|R|$ is small, which corresponds to small eigenvalues of M, then the patch is flat; if $R < 0$, which occurs when one eigenvalue is considerably greater than the other, then the patch is an edge; if $R \gg 0$, which occurs when both eigenvalues are large, then the patch is a corner. The chosen corner threshold was 10^{-11}. Finally, after the Harris corner detection is complete, the detected points are mapped onto the x-axis, and the sense of the orientation vector is determined to be towards the side of the origin with more corner points. The orientation correction process is illustrated in Fig. 3.

2.3.3. Orientation evaluation

The performance of the orientation estimation algorithm needs to be evaluated to determine its sufficiency for green fruit and stem orientation estimation. To determine this, the estimated orientations of the fruit/stem masks are compared to the corresponding ground truth orientations by obtaining the absolute angle error, which is defined as follows:

$$|error_i| = |GT_i - Est_i| \tag{11}$$

Where i is the mask orientation index, GT_i is the ground truth orientation for mask i (in °), and Est_i is the orientation estimation for mask i (in °). A ground truth orientation label for each mask is created in VIA by adding a single line to the mask that corresponds to its correct orientation vector.

For green fruit masks, angular errors were classified to various groups from 0°-180° in 15° increments. For stem masks, however, angular errors were only classified to various groups from 0°-90° in 15° increments. In other words, evaluation for orientation estimation considers the sense of green fruit masks while it does not for stem masks. For end-effectors that need to approach the calyxes of target green fruit, having the correct senses of the fruit is required. Meanwhile, whether the sense of a stem faces one way or another is not believed to matter for robotic green fruit thinning, as the stem orientation will unlikely affect the positioning of a pulling end-effector. The stem sense will not affect a stem-cutting end-effector whose blade approaches a stem perpendicularly. A cumulative histogram plot was applied to the data to illustrate the distribution of the angular errors for both mask types.

The algorithm performance for orientation estimation and orientation correction was evaluated on both ground truth masks (manually labeled) and Mask R-CNN-generated masks for green fruits and stems. Only generated masks with an IoU value>0.5 with their corresponding ground truth mask are used, as generated masks with lower IoU values are likely too dissimilar to their ground truth masks to provide reliable estimates.

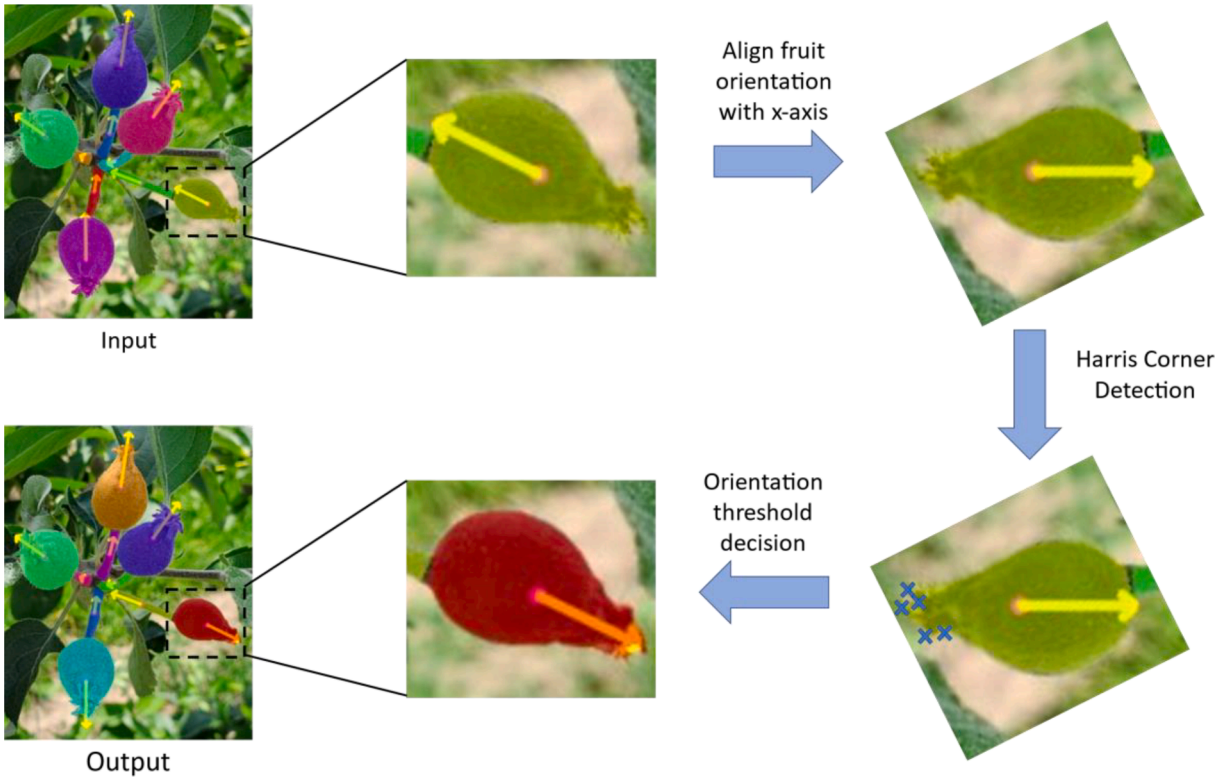

Fig. 3. Green fruit orientation correction using the Harris corner detection. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3. Results and discussions

3.1. Green fruit and stem segmentation

Table 6 shows the average precision of the green fruit and stem segmentation. Overall, the average precision for green fruit segmentation is better than that for stem. Green fruit segmentation results on all test dataset masks are reasonably high at average precision of 83.4 %. Stem segmentation results on all masks, however, are notably lower at 38.9 %. When all masks are divided into bins according to their sizes, varying performances for each mask type is apparent. In general, performance for segmentation on each mask type is greatest for the largest mask sizes ($>32^2$), with green fruit and stem segmentation performances being 91.3 % and 67.7 %, respectively. The increase in performance when compared to using all masks is particularly notable for stems. Performance tends to trend downward as size decreases from largest to smallest. The decrease in performance becomes dramatic once mask sizes are less than 20^2, particularly for green fruit masks, which decreases to 24.2 % from 85.4 % from the next largest mask bin (20^2-27^2).

Due to the performance degradation of segmentation with decreasing mask size, especially for stems, it is better to take closer images for segmentation of fruit for the purposes of robotic green fruit thinning such that the minimum size of each mask is at least 202 pixels. Otherwise, segmentation performance will not be sufficient for the purpose. Since a robotic green fruit thinning system will likely use a stereo camera to obtain 3D information of an environment, the closest distance from which images can be taken to obtain reliable RGB and depth images will depend on the baseline (the distance between the two lenses) as well as other factors. Minimum distances range of current popular stereo cameras range from tens to hundreds of centimeters.

When inspecting the mask sizes training dataset size in Table 4, it is clear that the average fruit mask size is considerably larger than then average stem size. Training the Mask R-CNN on larger stem masks or making parameter adjustments could help increase the performance of stem segmentation. Furthermore, a green fruit image containing multiple green fruit clusters could first be split into multiple smaller images, each containing one of the clusters. Then, these images could be enlarged to increase the apparent size of each fruit and stem, thereby increasing segmentation performance. However, the interpolation operation used to enlarge the cluster images could also introduce artifacts into the images, which could negatively affect segmentation performance.

Table 7 shows the average precisions of green fruit and stem segmentation for each cultivar and stem size. For each mask type, it is still apparent that for each cultivar, the larger mask size category shows superior performance in comparison to the smaller mask size category. Again, for each cultivar, green fruit segmentation performance is superior in all cases when compared to stem segmentation performance. For both fruit and stem masks, there is a notable pattern in which overall performance ranks from worst to best in the following order: Fuji, GoldRush, and Golden Delicious. When inspecting Table 5, it is apparent that the average size of fruit masks for each cultivar in the test dataset also follows this order, i.e., Golden Delicious green fruit masks are overall the largest, while Fuji ones are the smallest. Thus, the variation in green fruit segmentation performance between cultivars may in considerable part be due to their variation in mask sizes. However, this correlation between performance and mask size is not apparent for stem masks. While stem mask segmentation performance for Golden Delicious is the best of all cultivars and it has the largest stem mask sizes, Fuji stem masks have the worst segmentation performance despite having larger stem masks. Thus, other factors such as color and texture may be affecting stem segmentation performance.

Table 6
Results for green fruit and stem segmentation in terms of average precision (AP).

Objects	Average precision (AP) (%)					
	Mask size (pixel)	Overall	$> 32^2$	27^2 –32^2	20^2 –27^2	$< 20^2$
Green Fruit		83.4	91.3	80.0	85.4	24.2
Stem		38.9	67.7	64.5	57.0	30.5

Table 7
Segmentation performances for each cultivar and size category.

Objects	Average precision (AP) (%)					
	Mask size (pixel)	Overall	$> 32^2$	$< 32^2$	$> 20^2$	$< 20^2$
Green Fruit	All	83.4	91.3	66.0	–	–
	GoldRush	85.2	91.1	63.9	–	–
	Fuji	79.5	89.6	68.8	–	–
	Golden Delicious	88.2	94.3	56.1	–	–
Stem	All	38.7	–	–	60.6	30.3
	GoldRush	42.3	–	–	62.6	37.2
	Fuji	34.1	–	–	58.8	26.4
	Golden Delicious	44.2	–	–	61.2	25.4

Overall, all of these scores can be considered acceptable for segmentation, considering the reported results for the Mask R-CNN on the MS COCO dataset. However, there are some explanations that can be identified for lower segmentation performance for stems. First, it is apparent from inspection of green fruit images that green fruit are more distinct from the background than stems, which may have similar appearances to other background objects such as branches or green shoots. This may cause more false positives for stem segmentation. Second, the average size of stem masks is considerably smaller than the average size of green fruit masks, which is apparent from the green fruit and stem mask size distributions shown in Fig. 4. This means that less details may have been available for training of the Mask R-CNN for stem detection. Also, less training data for large stem masks will likely result in decreased performance for these masks. Third, the MS COCO dataset parameters, which are utilized for Mask R-CNN transfer learning, was originally trained to detect apples but not stems. This may indicate the MS COCO dataset parameters are far away from the optimal solution for stem segmentation, which could result in in either local convergence to a suboptimal solution, or at least extremely slow convergence of the Mask R-CNN to the optimal solution.

While segmentation results are acceptable for both green fruit and stem, occlusion is believed to cause performance degradation, particularly in cases where masks are occluded in such a way that the mask is split into two disconnected components. An example of this can been seen in Fig. 5.

Another concern regarding Mask R-CNN is on the segmentation performance for the calyxes of green fruit. While the algorithm does well in obtaining the general shape of green fruit, it shows less sufficiency in preserving the sharper details present in the calyx. This can be particularly problematic for orientation estimation, especially for occluded fruit, which is discussed later. An example of this is shown in Fig. 6. In the manually labelled fruits (Fig. 6 left), the calyx portion of the mask was presented with more detail when compared to that of the segmented fruit mask (Fig. 6 right).

3.2. Fruit orientation estimation and correction

To evaluate the performance of the fruit orientation estimation with the segmented fruit masks by the developed Mask R-CNN algorithm, the fruit orientation estimation was conducted with both segmented fruit masks and the ground truth masks (manually labelled). The fruit orientation estimation results were calculated using the PCA method under the two conditions before and after correction was applied. Histograms were used to illustrate the instances of different angular error levels and the accumulated percentage of these angular errors as shown in Fig. 7 and Fig. 8. Overall, orientation estimation using the PCA method presented good performance with the majority of orientations

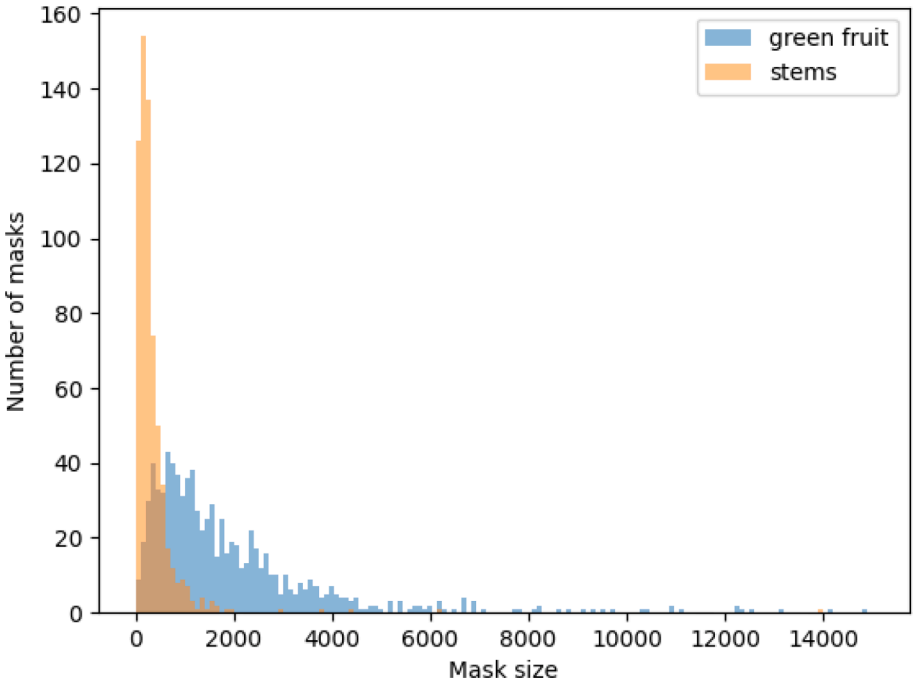

Fig. 4. The distribution of mask sizes (pixels) for each mask type including green fruits and stems in the test dataset. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 5. Occlusions in ground truth masks (left) vs segmentation masks (right).

being estimated within 15° of ground truth orientations.

Before correction, 65.5 % and 61.9 % of fruit orientation estimates were accurate within 15° using ground truth masks and Mask R-CNN segmented masks respectively. Correspondingly, 73.1 % and 71.6 % of the estimates were accurate within 30°. Very few estimates fell into the error range of 30-165° for both situations. Meanwhile, 19.8 % and 18.6 % of estimates had errors in largest error category of 165-180°, correspondingly. In other words, a certain portion of the fruits were estimated to almost the opposite orientation. Therefore, orientation estimation using PCA without correction is indeed incorrectly estimating the senses of some green fruit.

Orientation correction is shown to increase the accuracy of orientation estimation of green fruit for both mask types. The correction algorithm improved the orientation estimation accuracy, with 80.5 % and 65.6 % of estimates falling within 15° of ground truth orientations for ground truth masks and generated masks, respectively. Meanwhile, 89.3 % and 75.5 % of estimates fell within 30° of ground truth orientations for ground truth masks and generated masks, respectively. The percentage of estimates within the greatest error bins decreased to 4.8 % and 15.0 %, respectively. It is evident that the increase in orientation estimation accuracy for green fruit masks corresponds to a considerable decrease in the number of estimates in the largest error category, which includes the estimates that face the opposite way of their corresponding ground truth values. This improvement is considerably larger for ground truth masks than it is for masks generated by Mask R-CNN. The greater improvement for ground truth masks relative to generated masks is not surprising. The proposed orientation correction method relies on the jagged features of the calyx; these features are purposefully preserved in the annotation of green fruit images, while they are not preserved as well by Mask R-CNN, which decreases the reliability of orientation correction and increases the chance of falsely corrected and uncorrected orientations. Fig. 9 shows an example of orientation correction difference between the ground truth mask and corresponding segmented mask.

Overall orientation estimation is shown to perform very well on stem masks, with 97.6 % and 98.3 % of stem orientations being estimated within 15° of ground truth orientations for ground truth and generated masks, respectively. Meanwhile, 99.8 % and 99.7 % of stem orientations were estimated within 30° of ground truth orientations for ground truth and generated masks, respectively. No stems orientations were

Fig. 6. Green fruit calyx features in ground truth masks (left) vs segmentation masks (right). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

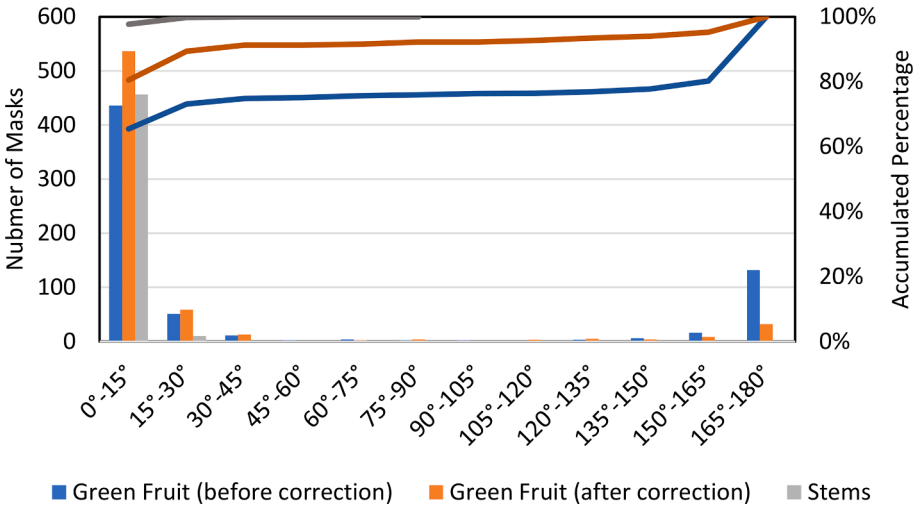

Fig. 7. Histogram of orientation estimation errors with all the ground truth (GT) fruit masks. Bars are the number of masks, and the lines are the accumulated percentages.

estimated with errors>45°. Unlike green fruit masks, orientation estimates for stem masks are shown to be very accurate while being robust to both ground truth and segmentation masks without the requirement of a correction algorithm.

One natural weakness to the proposed method of orientation estimation is that, if a green fruit is facing directly or nearly directly toward the camera, then its orientation estimate will not be practically meaningful, as it will considerably differ from the ground truth orientation in 3D space. In a future study, depth images and point clouds obtained using a stereo vision camera will be used to develop and evaluate an algorithm for 3D orientation estimation of green fruit.

During the process of the orientation estimation, some occluded masks have been observed which could potentially affect the accuracy. Therefore, a set of calculations was conducted to estimate the fruit orientations only for these unoccluded fruits. The results of these calculations are illustrated in Fig. 10 and Fig. 11.

When occluded masks are excluded, there are notable performance increases in orientation estimation. Before correction, 69.3 % and 68.0 % of fruit orientation estimates were accurate within 15° for unoccluded ground truth masks and Mask R-CNN segmented masks, respectively. Correspondingly, 73.5 % and 72.5 % of the estimates were accurate within 30°. Very few estimates fell into the error range of 30-165° for both situations. There were 25.2 % and 23.7 % of estimates with errors that fell in the largest error category of 165-180°. Interestingly, before correction, an increase in orientation estimates with correct directions but incorrect senses occur with the exclusion of occluded masks.

Orientation correction causes an even greater performance increase on unconcluded masks relative to the use of all masks. The correction algorithm improved the orientation estimation accuracy, with 92.7 % and 78.9 % of estimates falling within 15° of ground truth orientations

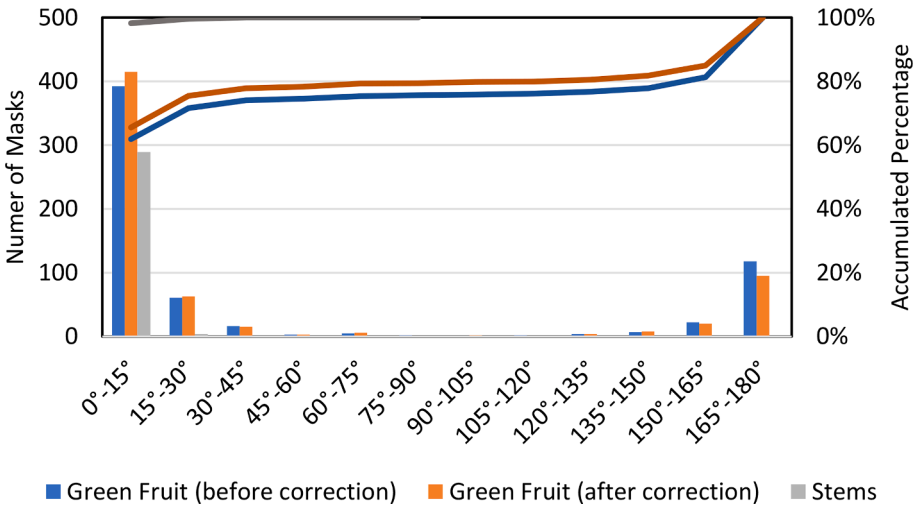

Fig. 8. Histogram of orientation estimation errors with all the Mask R-CNN algorithm segmented masks. Bars are the number of masks, and the lines are the accumulated percentages.

Fig. 9. Orientation correction performance example on ground truth mask (left) vs corresponding segmentation mask.

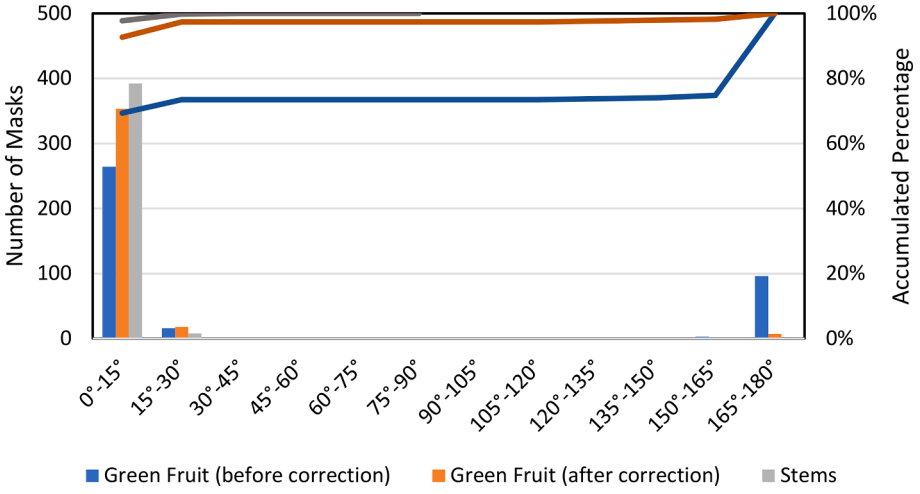

Fig. 10. Histogram of orientation estimation errors with GT masks and using only unoccluded masks. Bars are the number of masks, and the lines are the accumulated percentages.

for ground truth masks and generated masks, respectively. Meanwhile, 97.4 % and 84.0 % of estimates fell within 30° of ground truth orientations for ground truth masks and generated masks, respectively. The estimates within the greatest error bins decreased to 1.8 % and 12.8 %, respectively.

Occlusions decrease the performance and reliability of the

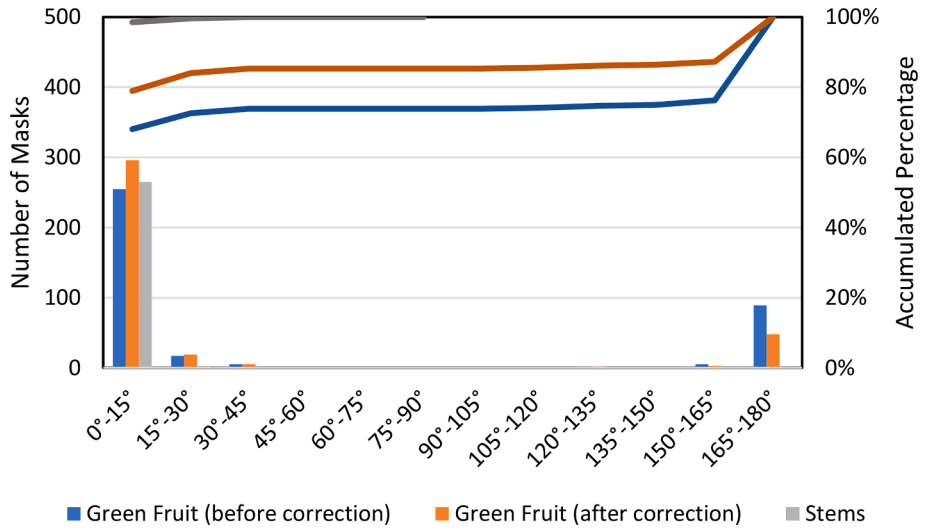

Fig. 11. Histogram of orientation estimation errors with Mask R-CNN masks and using only unoccluded masks. Bars are the number of masks, and the lines are the accumulated percentages.

orientation correction by introducing sharp corners towards the stem-end of green fruit, which increases the number of false corrections. The type of occlusion is also likely to be a factor in performance. In cases where occlusions do not block out the calyx of the fruit and do not considerably change the apparent mask shape, orientation estimates are still reasonably accurate.

Orientation estimation overall is shown to perform very well on stem masks when occlusions are omitted, with 97.8 % of stem orientations being estimated within 15° of ground truth orientations for both ground truth and generated masks. Meanwhile, 99.8 % and 99.6 % of stem orientations were estimated within 30° of ground truth orientations for ground truth and generated masks, respectively. No stems orientations were estimated with errors>45°. Performance of orientation estimation for unoccluded stems is similarly very good when compared to orientation estimation performance for all stems. Even when stems are occluded, their length generally stays high relative to their width, which likely explains the high performance of stem orientation estimation even in the presence of occlusions.

3.3. Discussion

Orientation correction for green fruit orientation estimates resulted in considerably large estimation improvements in all cases, with the greatest improvement seen on unoccluded ground truth masks. Overall, while orientation estimation using the PCA method and orientation correction based on Harris Corner Detection is seen to work reasonably well, its performance for green fruit is not very robust in cases of occlusions and masks generated without pixel-accurate calyxes. There are at least two possible ways to improve this: 1) utilize other orientation estimation methods, such as deep neural network-based ones, which may show better performances in such cases, as they may be more robust to the lack of apparent corners in generated Mask R-CNN masks; 2) generate masks with higher pixel accuracy around calyxes, either through parameter or structure changes in Mask R-CNN, or potentially a different instance segmentation algorithm altogether that is able to do so. In future work, such improvements will be investigated.

The research revealed some additional limitations. This study does not consider potential variations in performances between cultivars for orientation estimation. Future studies may look into any potential differences in performances between cultivars using the method discussed in this study as well as additional orientation estimation methods. While segmentation performance has a notable a correlation with fruit/stem mask size, there are other factors that may factor into segmentation performance, such as green fruit color, texture, and how they vary from the background. While evaluation of this may be somewhat more challenging in comparison to performance according to size, there may be merit in future studies investigating this. Furthermore, this study does not consider the ground truth sizes of green fruits and stems, i.e., their diameters and lengths. However, this is not believed to matter as much, since segmentation tends to depend more on the apparent in-image mask size of an object rather than its real-life size. Nonetheless, fruit and stem dimensions should still be considered in an overall robotic green fruit thinning system, as they may be an important factor for making thinning decisions.

This study does not quantitatively investigate how occlusions may affect segmentation performance. For segmentation, occlusions are not believed to have a significant effect on performance unless a fruit or stem is occluded such that the occlusion splits the object into two disconnected masks, which may be difficult for current segmentation algorithms to properly segment. Future work may consider how performance differs on occluded masks versus unoccluded masks. Furthermore, smart phone images were used for both segmentation and orientation estimation. However, smart phone images may have characteristics that differ from those usually used in robotic systems, e.g., stereo cameras. Nonetheless, the model trained for green fruit and stem instance segmentation on smart phone images could be used for transfer learning to obtain a model that works well on stereo camera images.

As stated earlier, instance segmentation performed well on green fruit and stems, particularly when medium or large mask sizes are used. However, instance segmentation does not intrinsically pair together corresponding fruits and stems, i.e., it does not indicate whether a stem belongs to one green fruit or another. Regardless of its type, when an end-effector for green fruit thinning is engaging with a target fruit, knowing where the corresponding stem of the fruit is located is likely important for properly positioning an end-effector for fruit removal. Thus, an additional algorithm for fruit and stem pairing is ultimately needed for a robotic green fruit thinning vision system.

In an integrated robotic green fruit thinning system, the vision system will be used to obtain important information of a tree being thinned that will be used to decide which green fruit to thin and determine collision-free paths for the robotic manipulator to take in thinning target fruit. The vision system will mainly consist of a camera, which will likely be a stereo camera that can obtain 3D point clouds of the tree, and a graphics processing unit that can quickly implement computer vision algorithms for green fruit thinning. The first algorithm the vision system needs to run before further steps can be taken is an instance

segmentation algorithm for detecting and generating pixel-wise masks of green fruits, which was investigated here. The initial step of segmenting individual green fruit in images is important for being able to implement further vision system algorithms that determine other important information required for robotic green fruit thinning, which includes orientation estimation for end-effector alignment. Precise segmentation of green fruit masks can provide guidance for green fruit and stem pairing, green fruit clustering, and decision making for fruit removal. Decision making for fruit removal will depend on the heuristic used; one example is to thin fruit to leave one fruit for every-six inches of branch (Renquist, 2018). These heuristics in general may utilize a variety of information including green fruit clusters, tree structure, fruit distribution and size, and fruit distancing future vision system development will be done to obtain these features. Green fruit segmentation can also be integrated with point cloud information to perform 3D instance segmentation of green fruit in point clouds, which enables a robotic green fruit thinning system to navigate to target fruit in the 3D environment using path planning and sequencing algorithms.

4. Conclusions

Algorithms for instance segmentation and orientation estimation on green fruit and their stems were implemented and evaluated. Mask R-CNN was used for instance segmentation, and PCA was used for orientation estimation. An orientation correction algorithm based on Harris corner detection was developed and implemented.

In summary, the Mask R-CNN model was able to obtain good results for green fruit segmentation, particularly on larger masks, the average precision for green fruit and stem detection were 91.3 % and 67.7 % for the mask size greater than 32^2. This will allow for sufficient pixel-wise mask detection of green fruit and stems, which can be used to determine fruit locations, as well as extract shape and orientation-based information about the fruit that would not be as readily obtainable from object bounding boxes alone. Furthermore, orientation estimation with correction obtained accurate results for green fruit with the accuracies of 89.3 % and 75.5 % within 30° of their actual orientations with ground-truth and segmentation-generated masks respectively. Fruit and stem orientation estimation is important for properly position a robotic-thinning end-effector for green fruit removal.

While green fruit segmentation and orientation estimation are important first steps for developing a robotic green fruit thinning vision system, there are further components that need to be developed. First, while instance segmentation can be used to generate the pixel-wise masks of green fruit and stems, this does not pair corresponding fruit and stem masks. Thus, an additional algorithm is required for this process. Second, decision making will be required to determine which green fruit to remove from clusters. An algorithm for this may utilize information such as green fruit distribution and size, tree structure, location relative to center of a cluster, and the presence of disease or damage. Third, since a green fruit end-effector is ultimately required to engage and remove target green fruit in 3D space, 3D vision algorithms will need to be implemented that can detect target fruit green fruit in 3D space and determine optimal target end-effector positions based on this information. Ultimately, a robotic green fruit thinning system is intended to work on all apple cultivars, with perhaps minor alterations needed between cultivars.

CRediT authorship contribution statement

Magni Hussain: Conceptualization, Investigation, Methodology, Validation, Writing – original draft. **Long He:** Conceptualization, Supervision, Writing – review & editing, Funding acquisition. **James Schupp:** Supervision, Writing – review & editing. **David Lyons:** Supervision, Writing – review & editing. **Paul Heinemann:** 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 research was partially supported in part by the United States Department of Agriculture (USDA)'s National Institute of Food and Agriculture (NIFA) Federal Appropriations under Project PEN04653 and Accession No. 1016510. We would like to give our special thanks for the support from the USDA NIFA AFRI Foundational and Applied Science Program Grant 2020-67021-31959 and the USDA NIFA Specialty Crop Research Initiative Grant 2020-51181-32197.

References

Auxt Baugher, T., Schupp, J., Ellis, K., Remcheck, J., Winzeler, E., Duncan, R., Johnson, S., Lewis, K., Reighard, G., Henderson, G., Norton, M., Dhaddey, A., Heinemann, P., 2010. String blossom thinner designed for variable tree forms increases crop load management efficiency in trials in four united states peach-growing regions. HortTechnology 20 (2), 409–414. https://doi.org/10.21273/horttech.20.2.409.

Bargoti, S., Underwood, J., 2017. Deep fruit detection in orchards. Proceedings - IEEE Int. Conference on Robotics and Automation 3626–3633. https://doi.org/10.1109/ICRA.2017.7989417.

Bolya, D., Zhou, C., Xiao, F., & Lee, Y. J. (2019). YOLACT: Real-time instance segmentation. *Proceedings of the IEEE International Conference on Computer Vision*, *2019-Octob*(Iccv), 9156–9165. Doi: 10.1109/ICCV.2019.00925.

Bradski, G., 2000. The OpenCV Library. Dr. Dobb's J. Software Tools.

Choi, J., Lee, B.-J., & Zhang, B.-T. (2016). Human Body Orientation Estimation using Convolutional Neural Network. *CoRR*, *abs/1609.0*. http://arxiv.org/abs/1609.01984.

Dalal, N., & Triggs, B. (2005). Histograms of Oriented Gradients for Human Detection. *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference On*, *1*, 886–893. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1467360.

Davidson, J. R., & Mo, C. 2015 Mechanical Design and Initial Performance Testing of an Apple-Picking End-Effector. *ASME International Mechanical Engineering Congress and Exposition, Proceedings (IMECE)*, *4A-2015*(November). Doi: 10.1115/IMECE2015-50482.

Davidson, J., Bhusal, S., Mo, C., Karkee, M., Zhang, Q., 2020. Robotic manipulation for specialty crop harvesting: a review of manipulator and end-effector technologies. Global J. Agricultural and Allied Sci. 2 (1), 25–41. https://doi.org/10.35251/gjaas.2020.004.

Dutta, A., & Zisserman, A. 2019 The VIA Annotation Software for Images, Audio and Video. *Proceedings of the 27th ACM International Conference on Multimedia*. Doi: 10.1145/3343031.3350535.

Farias, R.de.M., Martins, C.R., Barreto, C.F., Giovanaz, M.A., Malgarim, M.B., Mello-Farias, P., 2019. Time of metamitron application and concentration in the chemical thinning of 'Maciel' peach. Revista Brasileira de Fruticultura 41 (4). https://doi.org/10.1590/0100-29452019017.

Ganesh, P., Volle, K., Burks, T.F., Mehta, S.S., 2019. Deep Orange: mask R-CNN based Orange Detection and Segmentation. IFAC-PapersOnLine 52 (30), 70–75. https://doi.org/10.1016/j.ifacol.2019.12.499.

Gonzalez, L., Torres, E., Àvila, G., Bonany, J., Alegre, S., Carbó, J., Martín, B., Recasens, I., Asin, L., 2020. Evaluation of chemical fruit thinning efficiency using Brevis® (Metamitron) on apple trees ('Gala') under Spanish conditions. Scientia Horticulturae 261, 109003. https://doi.org/10.1016/j.scienta.2019.109003.

Hara, K., Vemulapalli, R., & Chellappa, R. 2017 Designing Deep Convolutional Neural Networks for Continuous Object Orientation Estimation. *CoRR*, *abs/1702.0*. http://arxiv.org/abs/1702.01499.

Harris, C., & Stephens, M. 1988. A Combined Corner and Edge Detector. *Proceedings of the 4th Alvey Vision Conference*, 147–151.

He, K., Gkioxari, G., Dollar, P., & Girshick, R. 2017. *Mask R-CNN*. 2980–2988. Doi: 10.1109/ICCV.2017.322.

Hussain, M., He, L., Schupp, J., Heinemann, P., 2022. Green fruit removal dynamics for development of robotic green fruit thinning end-effector. J. ASABE 65 (4), 779–788. https://doi.org/10.13031/ja.14974.

Jia, W., Liu, J., Lu, Y., Liu, Q., Zhang, T., Dong, X., 2022a. Polar-Net: Green fruit instance segmentation in complex orchard environment. Front. Plant Sci. 13, 1–13. https://doi.org/10.3389/fpls.2022.1054007.

Jia, W., Wei, J., Zhang, Q., Pan, N., Niu, Y., Yin, X., Ding, Y., Ge, X., 2022b. Accurate segmentation of green fruit based on optimized mask RCNN application in complex orchard. Front. Plant Sci. 13 https://doi.org/10.3389/fpls.2022.955256.

Kon, T.M., Schupp, J.R., 2018. Apple crop load management with special focus on early thinning strategies: A US perspective. Hortic. Rev. 46 (46), 255–298. https://doi.org/10.1002/9781119521082.ch6.

Kon, T.M., Schupp, J.R., Edwin Winzeler, H., Marini, R.P., 2013. Influence of mechanical string thinning treatments on vegetative and reproductive tissues, fruit set, yield, and fruit quality of "Gala" apple. HortSci. 48 (1), 40–46. https://doi.org/10.21273/hortsci.48.1.40.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L., 2014. Microsoft COCO: common Objects in Context. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (Eds.), Computer Vision – ECCV 2014. Springer International Publishing, pp. 740–755.

Liu, J., Zhao, Y., Jia, W., Ji, Z., 2022. DLNet: Accurate segmentation of green fruit in obscured environments. J. King Saud University - Computer and Information Sci. 34 (9), 7259–7270. https://doi.org/10.1016/j.jksuci.2021.09.023.

Lowe, D. G. 1999 Object recognition from local scale-invariant features. *Proceedings of the Seventh IEEE International Conference on Computer Vision*, *2*, 1150–1157 vol.2. Doi: 10.1109/ICCV.1999.790410.

Miller, S.S., Schupp, J.R., Baugher, T.A., Wolford, S.D., 2011. Performance of mechanical thinners for bloom or green fruit thinning in peaches. HortSci. 46 (1), 43–51. https://doi.org/10.21273/hortsci.46.1.43.

Onishi, Y., Yoshida, T., Kurita, H., Fukao, T., Arihara, H., Iwai, A., 2019. An automated fruit harvesting robot by using deep learning. ROBOMECH J. 2–9 https://doi.org/10.1186/s40648-019-0141-2.

Pearson, K., 1901. LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of*. Science 2 (11), 559–572. https://doi.org/10.1080/14786440109462720.

Reighard, G.L., Henderson, W.G., 2012. Mechanical Blossom Thinning in South Carolina Peach Orchards. Acta Horticulturae 965, 117–122. https://doi.org/10.17660/actahortic.2012.965.14.

Renquist, S. 2018 *Fruit Thinning*. https://extension.oregonstate.edu/gardening/berries-fruit/fruit-thinning#:~:text=A third reason to thin,respond positively to fruit thinning.

Sa, I., Ge, Z., Dayoub, F., Upcroft, B., Perez, T., McCool, C., 2016. Deepfruits: a fruit detection system using deep neural networks. Sensors (Switzerland) 16 (8). https://doi.org/10.3390/s16081222.

Schupp, J.R., Winzeler, H.E., Kon, T.M., Marini, R.P., Baugher, T.A., Kime, L.F., Schupp, M.A., 2017. A method for quantifying whole-tree pruning severity in mature tall spindle apple plantings. HortSci. 52 (9), 1233–1240.

Stander, O.P.J., Botes, J., Krogscheepers, C., 2018. The potential use of metamitron as a chemical fruit-thinning agent in mandarin. HortTechnology 28 (1), 28–34. https://doi.org/10.21273/HORTTECH03913-17.

Tyagi, S., Sahay, S., Imran, M., Rashmi, K., Mahesh, S.S., 2017. Pre-harvest factors influencing the postharvest quality of fruits: a review. Curr. J. Appl. Sci. Technol. 23 (4), 1–12.

Usda, 2021. National agricultural statistics database. USDA National Agricultural Statistics Service, Washington, DC https://quickstats.nass.usda.gov.

Vanheems, B. 2015 *How to Thin Fruit for a Better Harvest*. GrowVeg. https://www.growveg.com/guides/how-to-thin-fruit-for-a-better-harvest/.

Wang, D., He, D., 2021. Channel pruned YOLO V5s-based deep learning approach for rapid and accurate apple fruitlet detection before fruit thinning. Biosystems Engineering 210, 271–281. https://doi.org/10.1016/j.biosystemseng.2021.08.015.

Wang, Z., Xu, Y., Yu, J., Xu, G., Fu, J., Gu, T., 2021. Instance segmentation of point cloud captured by RGB-D sensor based on deep learning. Int. J. Comput. Integr. Manuf. 34 (9), 950–963. https://doi.org/10.1080/0951192X.2021.1946853.

Ye, L., Duan, J., Yang, Z., Zou, X., Chen, M., Zhang, S., 2021. Collision-free motion planning for the litchi-picking robot. Comput. Electron. Agric. 185 (483), 106151 https://doi.org/10.1016/j.compag.2021.106151.

Zahid, A., Mahmud, M.S., He, L., Heinemann, P., Choi, D., Schupp, J., 2021. Technological advancements towards developing a robotic pruner for apple trees: a review. Comput. Electron. Agric. 189, 106383 https://doi.org/10.1016/j.compag.2021.106383.

Zhao, Y., Gong, L., Liu, C., Huang, Y., 2016. Dual-arm robot design and testing for harvesting tomato in greenhouse. IFAC-PapersOnLine 49 (16), 161–165. https://doi.org/10.1016/j.ifacol.2016.10.030.