HIGH-THROUGHPUT PHENOTYPING OF WALNUT YIELD

Brian Bailey, Mason Earles, Pat J. Brown, Maryia Halubok, Kaiming Fu

ABSTRACT

The overarching goal of this project has been to develop a new high-throughput phenotyping approach for walnut yield that uses ground-based imagery to estimate tree-level yield. During Year 1, we focused on collection of walnut imagery in the field with the aim of developing, testing, and assessing different methods of walnut detection in the images. Imagery collection was scaled up in Year 2 and resulted in collection of 5,016 images, out of which 3,871 were RGB images from a consumer-grade Nikon camera and 1,145 images from an RGB+NIR multispectral camera. Visible walnuts in each image were manually labeled with rectangular bounding boxes. Additionally, measurements of total walnut count were collected on the imaged trees, and tree harvested walnut weight was measured in the trees of one block. Data was analyzed directly to determine the correlation between nut weight and count, and visible to total nut count. Correlation between tree-level nut weight and count within a block was high, with an observed Pearson's *r* value of 0.977. Correlation between total nut count and visible nuts in a single image was relatively high regardless of camera viewing direction, but was highest when the tree was viewed from the row-perpendicular direction ($r \approx 0.85$). For the trees measured in this work, roughly 30-40% of the total nuts were visible in a single image. Aggregating multiple viewing angles increased the percentage of total visible nuts (potentially with some double-counting), and only marginally increased the correlation between total and visible nuts. These results suggested that a single image perpendicular to the row direction provided the best 'bang-for-the-buck' in terms of capturing relative differences in nut count between trees, and using two opposing image views per tree provided marginal improvement. Finally, we used machine learning to determine whether visible nut detection could be automated using a model in order to eliminate the need for tedious hand-labeling of images. We evaluated the model's ability to detect walnuts in high-resolution RGB images collected from the relatively inexpensive consumer-grade camera, and using images from the 4-channel RGB+NIR camera (but with lower overall resolution). The detection precision for the consumer RGB camera was 0.86, while the precision for the RGB channels of the multispectral camera was 0.81. Adding the NIR channel to the multispectral RGB images decreased performance relative to the RGB channels only, which resulted in a precision of 0.78. Overall, using hand-annotated images from a consumer-grade camera, or using a machine learning model to detect nuts in images, was able to reasonably predict tree-to-tree variability in walnut count and yield for relatively young trees such as that characteristic of a breeding trial. Older trees with a closed canopy present additional challenges, which requires additional work to fully evaluate. As part of this project, we also developed a simulation-based approach for generating artificial walnut tree imagery that can be auto-annotated, which may help to improve results in mature orchards with high occlusion.

OBJECTIVES

Specific objectives for the entire project are as follows:

1. Collect RGB and multispectral imagery on unripe walnuts in the field at the pre-harvest stage of development

2. Develop and assess various potential methods for image-based detection and quantification of unripe nuts
3. Investigate, assess, and address the issues of occlusion, clustering and other factors on image-based yield estimation
4. Validate the developed methods against field data

SIGNIFICANT FINDINGS

Objective 1

1) In Year 1, we collected around 500 RGB and multispectral images, which were hand labeled and used to train a neural network model for nut identification.
2) In Year 2, we collected more than 5,000 images including RGB and multispectral images that were then labeled for further use in analyzing the relationships between visible walnut count and total walnut count and factors that can potentially influence such relationships, such as location and angle of the camera used for imagery collection.

Objective 2

1) We evaluated the model's ability to detect walnuts in high-resolution RGB images collected from a relatively inexpensive consumer-grade camera, and using images from a 4-channel RGB+NIR camera (but with lower overall resolution). The detection precision for the consumer RGB camera was 0.86, while the precision for the RGB channels of the multispectral camera was 0.81. Thus, for RGB images only, the image quality had a significant effect on model performance.
2) Adding the NIR channel to the multispectral RGB images decreased performance relative to the RGB channels only, which resulted in a precision of 0.78.

Objectives 3 and 4

1) Correlation between tree-level nut yield and count within a block was high, with an observed Pearson's *r* value of 0.977.
2) Correlation between total nut count and visible nuts in a single image was relatively high regardless of camera viewing direction, but was highest when the tree was viewed from the row-perpendicular direction ($r \approx 0.85$). For the trees measured in this work, roughly 30-40% of the total nuts were visible in a single image. Aggregating multiple viewing angles increased the percentage of total visible nuts (potentially with some double-counting), and only marginally increased the correlation between total and visible nuts.
3) These results suggested that a single image perpendicular to the row direction provided the best 'bang-for-buck' in terms of capturing relative differences in nut count between trees, and using two opposing views per tree provided marginal improvement.

PROCEDURES

AUTOMATED NUT DETECTION ALGORITHMS

We used the YOLOv5 neural network model based on PyTorch with the aim of autonomously identifying and localizing nuts within tree images (https://pytorch.org/hub/ultralytics_yolov5/). We created a custom dataset by manually labeling walnuts in the selected images (see below) with the use of the Computer Vision Annotation Tool (CVAT) tool, then trained the object detection model on this custom dataset, and tested it by applying the trained model to other images. The labeled images were sub-divided into three sets: training, validation, and test. The YOLO model is trained using the ‘training’ image set, which is iteratively evaluated against the ‘validation’ set to determine the ‘best’ model. This best model is then evaluated against the ‘test’ image set, from which the following metrics were calculated: precision, recall, and mean average precision with an intersection-over-union threshold of 0.5 (mAP_0.5).

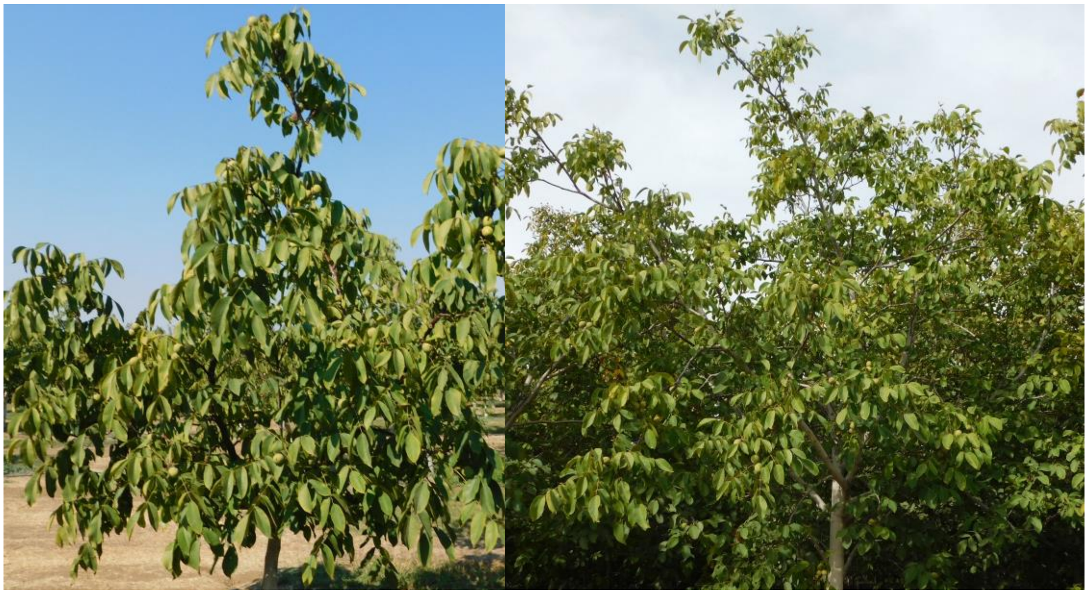
FIELD DATA COLLECTION FOR VALIDATION PURPOSES

Imagery for methodological development, testing and validation were collected in two research orchards within the UC Davis pomological research fields located on Hutchison Drive (Fig. 1). Imagery was collected from early September to early October 2021. One orchard, which we will call “J15A” was from a 5-year-old breeding trial with high genotypic diversity. The second, which we will call the “Mature Orchard”, consisted of a mixture of 9-year-old Chandler and Howard scions on either NCB or Paradox rootstock, which provided good tree-to-tree variability. This orchard had a closed canopy with notably larger trees than J15A.

Images of the walnut trees were collected with the use of two camera types: 1) a NIKON B500 16MP camera, which is a consumer-grade camera that retails for around $300. It has relatively high resolution of 4608x3456 = 16MP; 2) a Spectral Devices MSC-RGBN-1-A multispectral camera with co-located channels of red-green-blue+NIR. This multispectral camera has been modified with a custom filter that includes visible (RGB) and 955 nm long pass filter in order to allow simultaneous collection of images in the RGB spectrum and near 955±10 nm with aligned pixels. This camera had lower resolution of 1024x1024 = 1MP.

The idea behind using the NIR camera was that, because of the very large difference in light transmission between the leaves (about 30% in the visible) and nuts (opaque), nuts tend to show up as dark objects in the NIR images even when occluded by leaves (Fig. 5). Results from Year 1 with a different NIR camera showed similar model performance between RGB-only and NIR-only images. We hypothesized that combining RGB and NIR images would allow for better overall detection of visible+occluded nuts.

Figure 1. Sample RGB image from the experimental J15A orchard (left) and “Mature” orchard (right).

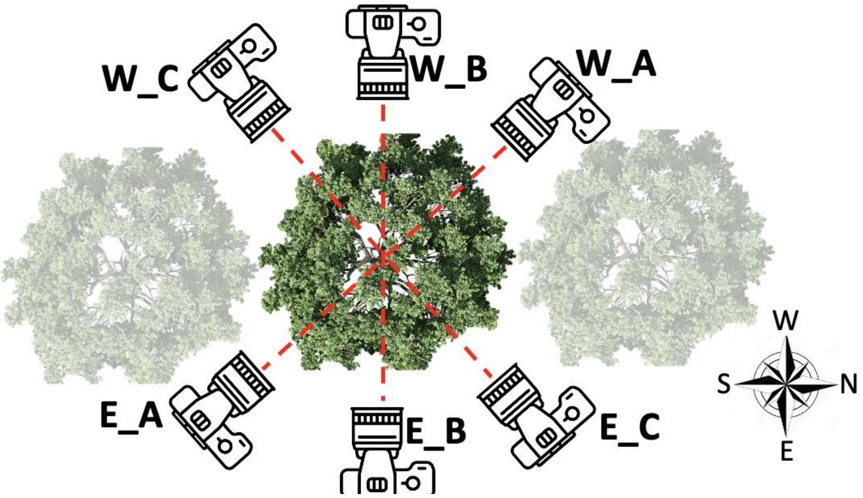


Images were collected from ground and mid-canopy height, at a distance of 1-3 m, and at 6 azimuthal viewing directions for each tree (Fig. 2). More specifically, the imagery was collected with the cameras being placed on the eastern (case E) and western sides (case W) of the tree row at 45 degrees to the left (case A), looking directly at a tree of interest (case B), and 45 degrees to the right of the tree (case C). Images were collected such that the entire crown was in the image. Neighboring trees in view of the camera were manually cropped out to the extent possible in a postprocessing step.

Overall, we collected about 1,145 RGB+NIR multispectral images and 4,200 RGB images (including images from both 2020 and 2021) from a range of horizontal and vertical positions throughout the orchards. To test the effect of the camera platform and addition of an NIR channel, the image sets were aggregated into four groups: 1) Nikon RGB, 2) Multispectral RGB channels only, 3) Multispectral NIR channel only, 4) Multispectral RGB+NIR channels.

We imaged and manually counted walnuts on ~80 trees in the J15A orchard without harvesting. We then proceeded to image and shake-harvest ~30 walnut trees in the "Mature" orchard. After shaking the trees and collecting the walnuts, we measured field weight and placed them in a drying bin for two days. Then, dry weight was measured, and in addition the collected walnuts were counted manually.

Figure 2. Schematic of image collection locations/directions relative to each tree. Six images were collected of each tree: two on each side of the tree facing perpendicular to the row direction (W_B and E_B), and four at 90 degrees from each other and offset 45 degrees from the row direction (W_A, W_B, E_A, and E_C).



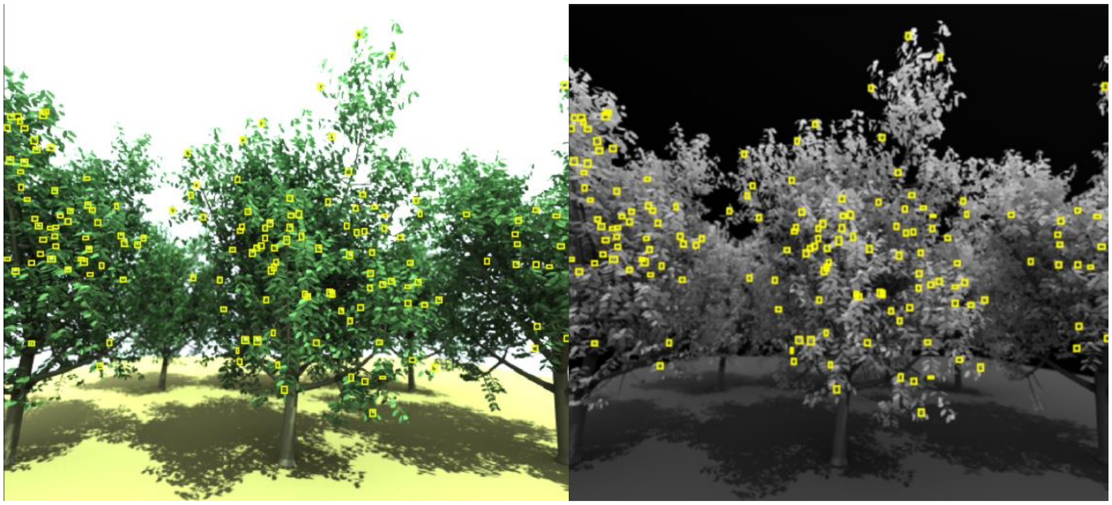
SYNTHETIC IMAGE DATA GENERATION

While we have collected over 5,000 RGB+NIR images and manually counted nuts for over 100 trees, this dataset is still relatively limited in terms of the full breadth of conditions that may be encountered in real walnut orchards. We developed an approach for generation of simulated images that can be automatically labeled and used to efficiently enhance model training (Fig. 3), in addition to allowing for testing of various assumptions related to image labeling. The Helios 3D plant modeling framework (baileylab.ucdavis.edu/software/helios) has been enhanced to allow for automatic annotation of images based on arbitrary classification schemes.

The model considers the full simulated 3D orchard geometry, as well as the radiative properties of every surface in the simulated domain. Using models of radiative transfer physics, along with information about the intrinsics and extrinsics of the camera, Helios is able to realistically simulate camera images. Because everything about the model geometry is known, it can automatically label the locations of nuts in the images (or any other element in the images). This allows for rapid creation of a large number of images that can help supplement model training. Additionally, it can be used to assess errors/biases in manual image labeling since the 'true' nut count in each image is exactly known.

We have not yet been able to include the synthetic imagery in the model training as part of this project, but this is a direction for future work. We would also like to analyze the accuracy of manually labeled images as part of future work.

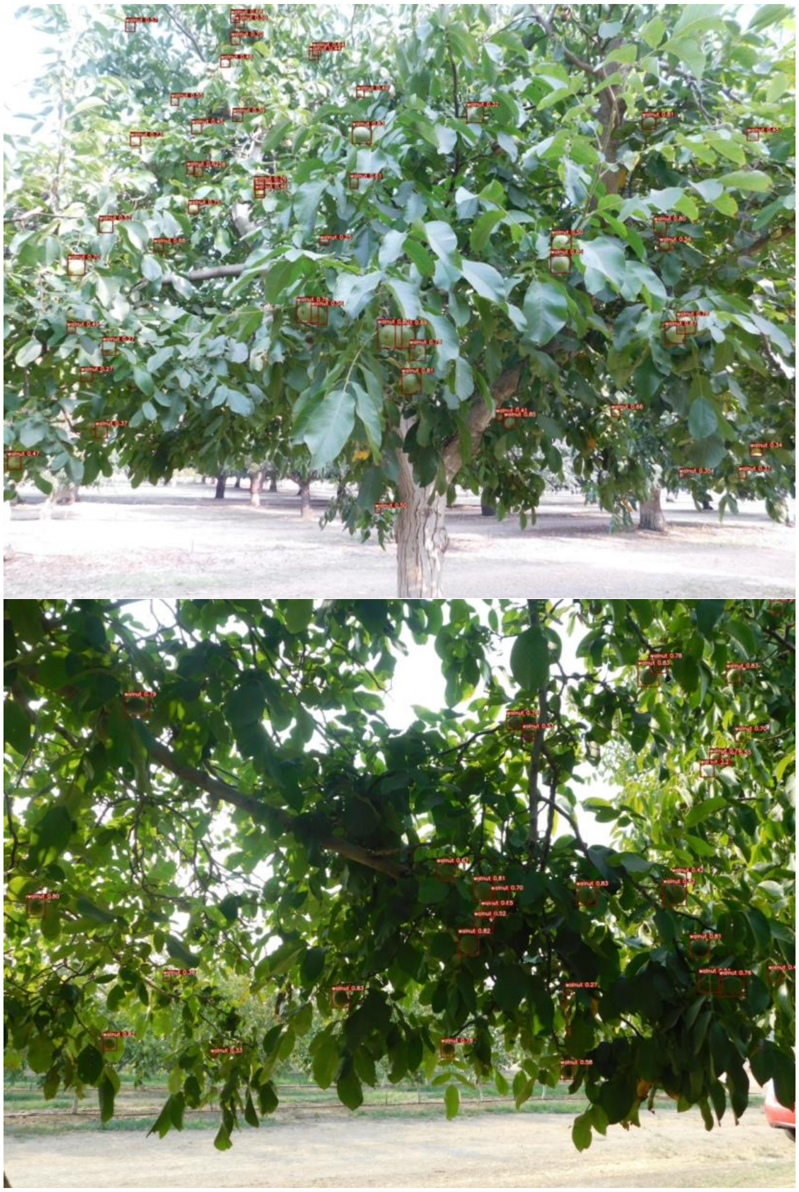
Figure 3. Synthetic (simulated) images of a walnut orchard with nuts automatically labeled. Left: RGB image; right: NIR (970 nm) image.



RESULTS AND DISCUSSION

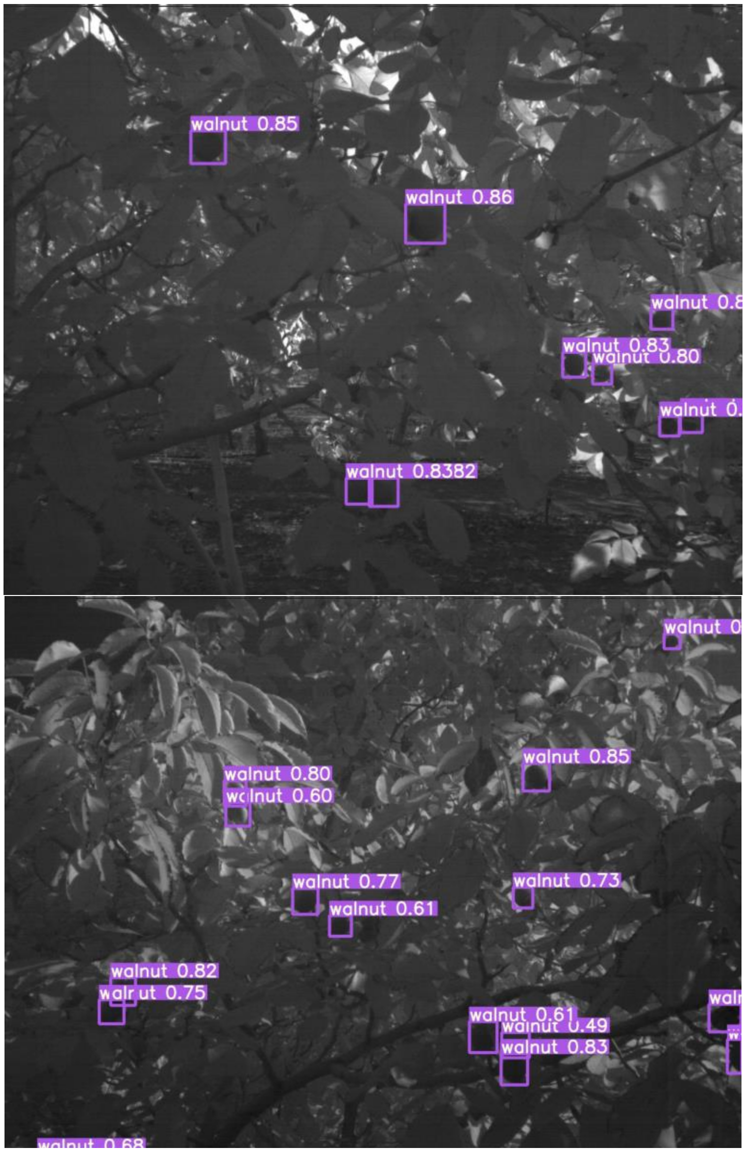
NUT DETECTION ALGORITHMS (Objective 2)
An illustration of results from autonomous nut detection is shown in Figs. 4 and 5, which is based on the initial exploratory dataset previously collected in Year 1. We have tested the YOLOv5 model performance on 4608*3456px RGB images and 1024*1024px NIR images.

Figure 4. Example Nikon RGB images with walnuts identified by the YOLOv5 method. Note the differences in sun-sensor position on the two images.



As a means of testing the performance of this nut detection approach, we have calculated several performance metrics based on the “test” datasets (Table 1). Overall, the model performed best with images from the Nikon RGB camera, which had the highest resolution and image quality. Precision of the model for this image set was relatively high and indicated that when the model detects a nut, there is about an 86% change it is actually a nut. This suggests the model is good at not mistaking leaves for nuts, for example. The recall value was significantly lower at 0.630, which suggested the model could correctly identify about 63% of the nuts in images.

Figure 5. Example NIR images with walnuts identified by the YOLOv5 method.



Performance metrics were generally lower for the multispectral camera with RGB channels only. This is perhaps expected given that the clarity of nuts in the images was lower for the multispectral camera than the Nikon camera. The recall value was slightly higher for the multispectral camera than the Nikon camera, which could be due to the fact that the higher resolution Nikon images allowed human labelers to identify small nuts that were difficult for the model to detect.

Surprisingly, the addition of NIR images – whether considering NIR images only or combining RGB and NIR channels – decreased model performance relative to the multispectral RGB images only. This is in contrast to our results from Year 1 in which we used a different NIR-only multispectral camera, and found similar model performance between RGB-only images and NIR-only images. It could be that the quality of the NIR images with the new multispectral camera was lower, making it more difficult to detect the nuts. The decrease in quality of this camera is a trade-off related to the advantage that the RGB and NIR channels are perfectly aligned.

Table 1. Metrics assessing model performance for the four combinations of image collection platforms. Precision essentially quantifies probability that an identified object is actually a nut. Recall essentially quantifies the probability that nuts in an image were correctly identified. mAP_0.5 incorporates the level of overlap between the predicted and ground truth nut bounding boxes.

Metric	**Nikon Camera - RGB**	**Multispectral Camera – RGB only**	**Multispectral Camera – NIR only**	**Multispectral Camera – RGB+NIR**
Precision	0.856	0.809	0.709	0.783
Recall	0.630	0.649	0.519	0.577
mAP_0.5	0.729	0.739	0.631	0.682

INVESTIGATING THE RELATIONSHIP BETWEEN YIELD, TOTAL WALNUT COUNT AND VISIBLE WALNUT COUNT (Objectives 3 and 4)

Relationship between total and visible walnut count: We first sought to answer the question of how representative the number of walnuts visible in an image reflected the total number of nuts on a tree, and the improvement (if any) that could be gained by incorporating more images of a single tree from different viewing angles.

Figure 6 shows the relationship between the visible and total walnut count when only a single image is used. The relationship is plotted for the 6 different viewing positions in the J15A orchard in Fig. 2 (Nikon RGB images). The correlation between total and visible nut count is relatively high for each of the 6 viewing locations, with a Pearson's *r* correlation coefficient ranging from 0.778 to 0.875. The highest correlations were observed when images were collected from the row-perpendicular viewing direction, which produced *r* values of 0.849 and 0.875. As was expected, the ratio of visible to total nut count is significantly less than 1 with a ratio of 0.3-0.4. This indicates that around 30-40% of the total nuts on the tree are visible in a single image. It is expected that this ratio will go down along with correlation coefficients as the orchard matures, but we have yet to complete processing the data from the "Mature" orchard block.

Adding additional images for a given tree in order to estimate visible nut count increased the ratio of visible to total nut count, but only marginally increased correlation coefficients (Figs. 7 and 8). As is to be expected, the ratio of visible to total nut count roughly doubled when opposing viewing directions were used to estimate visible nut count. However, correlation coefficients were only slightly higher with values ranging from *r* = 0.886 to 0.892 (Fig. 7). It could be argued that using two images per tree did reduce variability between viewing positions. Using 4 or 6 images per tree did continue to increase the ratio of visible to total nut count, as well as produce a small increase in the correlation coefficient. When 4 opposing images were used, *r* = 0.916 (Fig. 8a), and *r* increased to 0.920 when all 6 images were used. However, the ratio of visible to total nuts increased to greater than 100% when both 4 and 6 images were used, suggesting there was significant double-counting of nuts.

Results indicated that, if only relative differences in nut count between trees are desired, then the procedure for image collection does not matter too much provided that it is consistent between all trees. Adding more images eventually results in some double counting of nuts, but it also tends to reduce random variability and increase the overall correlation between visible and total nut counts. If the absolute value of total nut count is desired, then it is recommended to use two opposing viewing positions. This does relatively well at predicting the total nut count for low nut count, but does under-estimate the nut count at high nut counts.

Relationship between total nut count and yield: The next question we sought to answer was how well nut count (which is estimated from images) predicts total yield. Figure 8 shows this relationship for the "Mature" orchard. It was determined that the yield (lbs.) was about 0.023 times the nut count, with an *r* value of 0.977. It was surprising that the correlation between yield and count was so high, given that there was very high variability in tree size and yield resulting from the different scion-rootstock combinations.

Figure 6. Scatterplot of the visible walnut count (as seen by a **single Nikon RGB image** taken from each of the 6 viewing positions) and total walnut count (manual) in J15A: a) case E_A; b) E_B; c) E_C; d) W_A; e) W_B; and f) W_C, respectively. Note that visible nut count was determined from manual image labels, and not using the machine learning model.

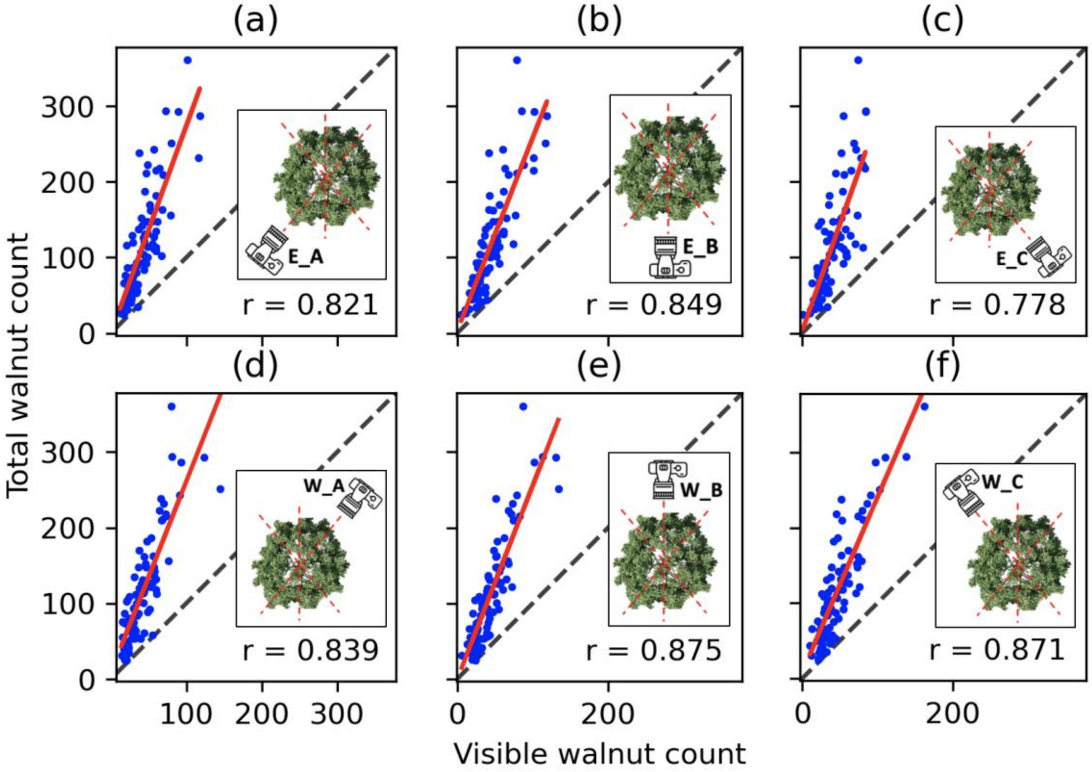


Figure 7. Scatterplot of the visible walnut count (as seen by **pairs of opposing <u>Nikon RGB images</u>** taken from each of the 6 viewing positions) and total walnut count (manual) in J15A: a) VWC summed for cases E_A and W_A; b) VWC summed for cases E_B and W_B; c) VWC summed for cases E_C and W_C. Note that visible nut count was determined from manual image labels, and not using the machine learning model.

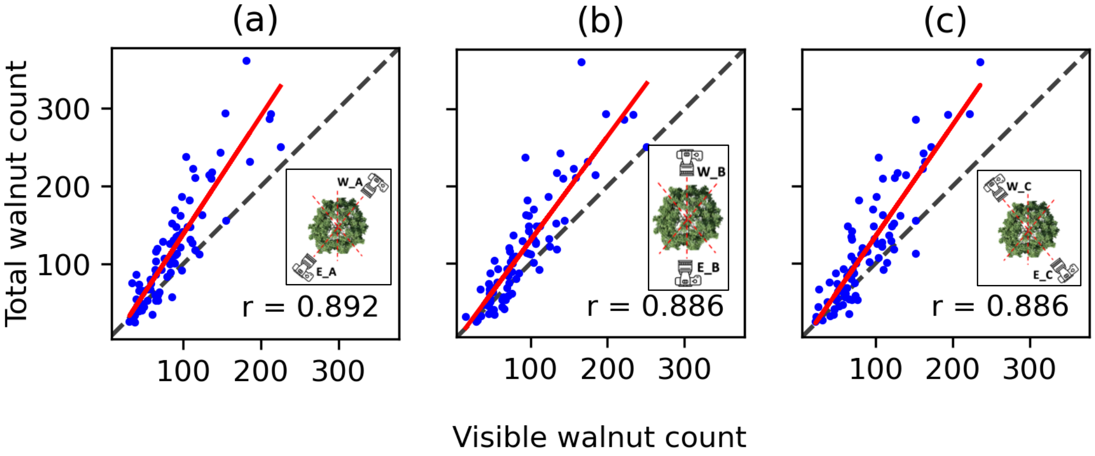


Figure 8. Scatterplot of the visible walnut count (as seen on the Nikon RGB imagery taken in the field conditions) and total walnut count (manual) in J15A: a) VWC summed for cases E_A, W_A, E_C, and W_C b) VWC summed for all 6 cases. Note that visible nut count was determined from manual image labels, and not using the machine learning model.

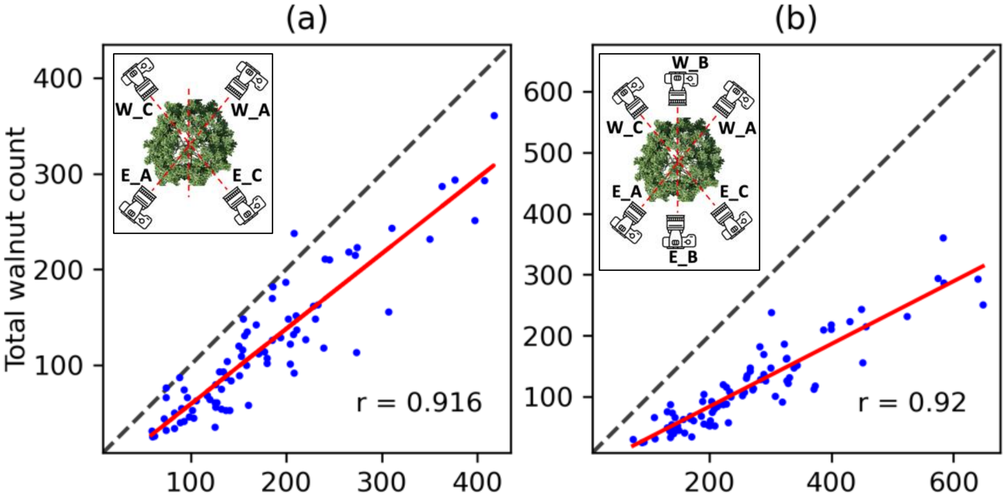
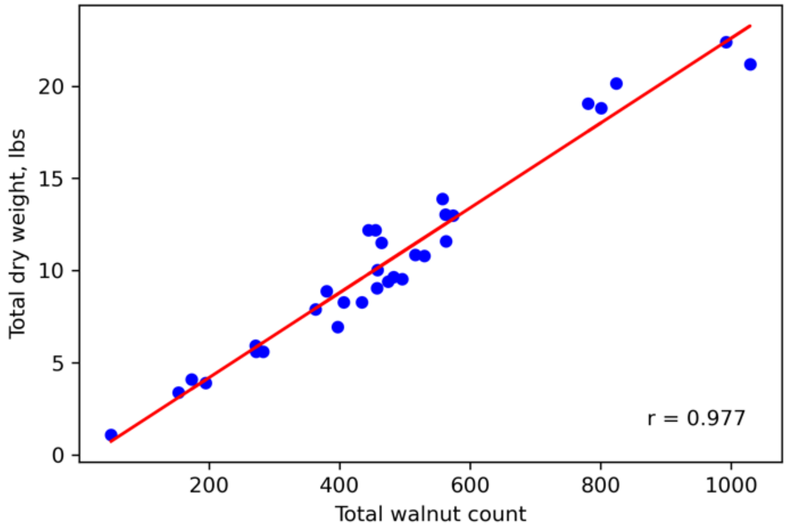


Figure 9. Scatterplot of total walnut count and dry weight for walnuts collected in the "Mature" orchard.



SOME PRACTICAL CONSIDERATIONS

It was usually not possible to position the camera such that only the tree crown of interest is in the image. Usually, there will be some branches from neighboring trees, and trees in other rows in the camera view, which may have nuts on them. In order to get accurate tree-level visible nut counts, it was important to crop the image to the tree of interest as best as possible and only label nuts on the tree of interest. However, since the machine learning model sees the entire image, it may still detect nuts on neighboring trees, which could have affected the performance of the model.

It is best to collect images at a time of day when the weight = 0.023(count) of the photographer, such that the visible tree crown is as illuminated as possible. This maximizes contrast between leaves and nuts in the images.

Manually labeling an image takes a few minutes on average. It is best to zoom into sub-portions of the image to avoid missing small nuts. When labeling images, there will inevitably be many instances in which the identity of some objects is unclear. One can take a conservative approach and only label nuts in which the labeler is very confident it is an actual nut, or a liberal approach in which anything that resembles a nut is labeled, or some approach in between. It is important to clearly determine the approach to be taken and be consistent throughout, especially if there are multiple labelers of a given image set.