# Comparing U-Net convolutional networks with fully convolutional networks in the performances of pomegranate tree canopy segmentation

**5 authors**, including:

Tiebiao Zhao
University of California, Merced
**34** PUBLICATIONS   **160** CITATIONS

SEE PROFILE

Haoyu Niu
University of California, Merced
**8** PUBLICATIONS   **5** CITATIONS

SEE PROFILE

YangQuan Chen
University of California, Merced
**908** PUBLICATIONS   **23,755** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project   Applied Fractional Calculus (AFC) View project

Project   fractional calculus and its applications View project

# Comparing U-Net convolutional network with mask R-CNN in the performances of pomegranate tree canopy segmentation

Tiebiao  Zhao, Yonghuan  Yang, Haoyu  Niu, Dong Wang, YangQuan  Chen

**SPIE.**

# Comparing U-Net Convolutional Network with Mask R-CNN in the Performances of Pomegranate Tree Canopy Segmentation

Tiebiao Zhao[a], Yonghuan Yang[b], Haoyu Niu[a], Dong Wang[c], and YangQuan Chen[a]

[a]University of California, Merced, 5200 Lake Rd, Merced, California, USA
[b]Santa Clara University, 500 El Camino Real, Santa Clara, California, USA
[c]USDA-ARS Water Management Research Unit, San Joaquin Valley Agricultural Sciences Center, Parlier, California, USA

## ABSTRACT

In the last decade, technologies of unmanned aerial vehicles (UAVs) and small imaging sensors have achieved a significant improvement in terms of equipment cost, operation cost and image quality. These low-cost platforms provide flexible access to high resolution visible and multispectral images. As a result, many studies have been conducted regarding the applications in precision agriculture, such as water stress detection, nutrient status detection, yield prediction, etc. Different from traditional satellite low-resolution images, high-resolution UAV-based images allow much more freedom in image post-processing. For example, the very first procedure in post-processing is pixel classification, or image segmentation for extracting region of interest(ROI). With the very high resolution, it becomes possible to classify pixels from a UAV-based image, yet it is still a challenge to conduct pixel classification using traditional remote sensing features such as vegetation indices (VIs), especially considering various changes during the growing season such as light intensity, crop size, crop color etc. Thanks to the development of deep learning technologies, it provides a general framework to solve this problem. In this study, we proposed to use deep learning methods to conduct image segmentation. We created our data set of pomegranate trees by flying an off-shelf commercial camera at 30 meters above the ground around noon, during the whole growing season from the beginning of April to the middle of October 2017. We then trained and tested two convolutional network based methods U-Net and Mask R-CNN using this data set. Finally, we compared their performances with our dataset aerial images of pomegranate trees. [Tiebiao- add a sentence to summarize the findings and their implications to precision agriculture]

**Keywords:** Canopy Segmentation, U-net CNN, Mask R-CNN, Instance-aware Segmentation, Pomegranate Trees

## 1. INTRODUCTION

With the development of sensors and cameras, images with higher resolution can be obtained easily and flexibly at a very low cost for precision agriculture applications. Published studies includes fruit detection such as mangoes, almonds, apples,[1] melons,[2] sweet peper,[3] weed detection[4] and tree and nursery management.[5] In our study of almond tree water stress quantification,[6–11] high resolution images make it possible to detect an individual tree from aerial images and even measure size of tree canopies in the tree level.

As a typical crop feature, canopy size can help estimate water use,[12] pesticide usage[13] and yield.[14] The most popular method to estimate canopy size is based 3D point-cloud data obtained from Lidar.[13, 15, 16] However, because of high cost and complex post-processing, its usage is still limited. Recently with the advent of unmanned

---

Further author information: (Send correspondence to YangQuan Chen)
Tiebiao Zhao: E-mail: tzhao3@ucmerced.edu
Yonghuan Yang: E-mail: yyonghuan@gmail.com
Haoyu Niu: E-mail: hniu2@ucmerced.edu
Dong Wang: E-mail: dong.wang@ars.usda.gov
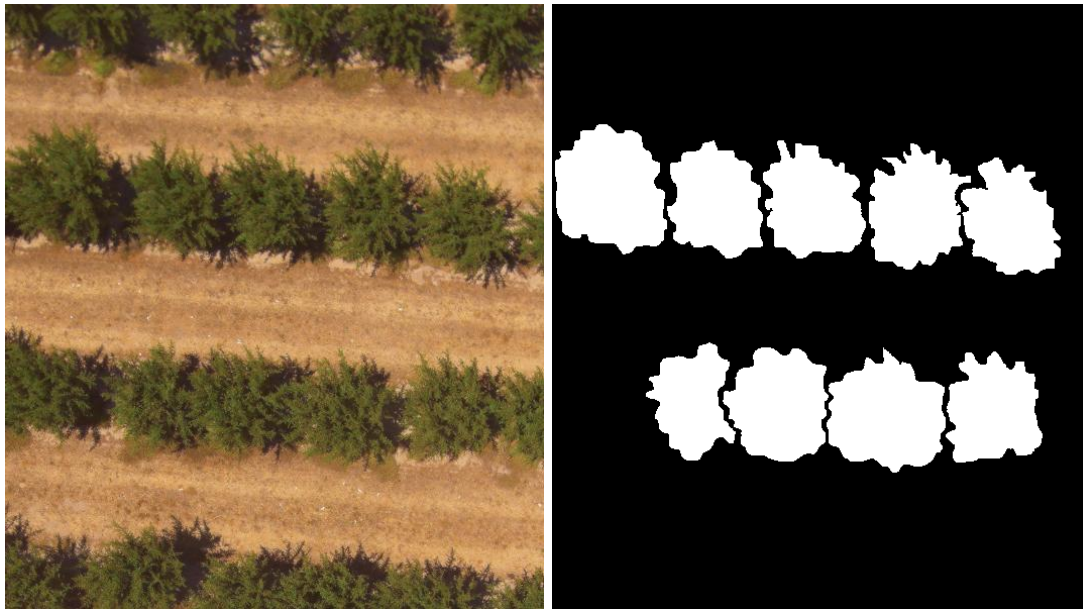YangQuan Chen: E-mail: ychen53@ucmerced.edu

Figure 1. Dataset samples (size:534*600), where the left is an aerial image and the right is its ground truth label

aerial vehicles and minimized high resolution cameras, it is also possible to estimate canopy size using 2D images.[5, 14, 17–21] In these studies, canopy classification and detection were based on man-made features, such threshold determination,[14] shape and compactness parameters[5, 18, 21] and watershed methods.[19] Although these feature are very specific designed, they are not robust to changes of objects, background or camera setting.

Deep learning based segmentation has been proposed in tree segmentation[22] using instance-sensitive segmentation.[23] With the development of new methods in deep learning technology, we evaluated two most recent different deep learning based methods for segmentation of tree canopies, U-net[24] and Mask R-CNN.[25] These two segmentation models are used to segment tree canopy pixels from pixels of soil, grass and their neighbor trees. Some potential application scenarios of these tasks are detecting and counting trees, quantifying canopy size in the block level, and further in the single-tree level.

## 2. MATERIAL AND METHODS

### 2.1 Aerial Image Collection and Labeling

The test field is in the USDA-ARS, San Joaquin Valley Agricultural Sciences Center(36.594N, 119.512W), Parlier, California. The size of this experiment field is 1.3 ha. The pomegranates trees were planted in 2010 with the space of 5 m between rows and 2.75 m within rows.[26]

Aerial images of these trees were collected during the growing season of 2017 using both ELPH110HS (Canon, Japan) and Survey2 (MAPIR, USA). Two different cameras were flown to test the robustness of the algorithm to imaging sensors. The original aerial images were cropped with the size of 534*600, to make sure the number of trees in each patch is around 10. These small patches were then manually labeled to obtain ground-truth mask. We chose the patches that are covering trees with representative colors and shapes so we can minimize the labeling effort and maximize the dataset variability. Only those trees with full part of canopies were masked in the purpose of simplifying the learning task. As see in the right mask of Figure 1, the trees of four sides are not labeled.

### 2.2 U-Net Based Segmentation

U-Net[24] is first introduced in biomedical image segmentation in 2015, which is built upon fully convolutional network.[27] The main idea of fully convolutional network is to supplement a usual contracting network by

successive layers, where pooling operators are replaced by upsampling operators. By increasing of the resolution of the output, a successive layer can learn to assemble a more precise output. U-Net architecture extends the feature channels to the upsampling layers, which allow the network to propagate context information to higher resolution layers. The network is more or less a u-shaped architecture, as a result of the symmetric between the expansive path and the contracting path.

## 2.3 Mask R-CNN Based Segmentation

Mask R-CNN[25] is built upon Faster R-CNN.[28] Faster R-CNN has two outputs for each candidate object, a class label and a bounding-box offset. The author of Mask R-CNN added a third branch to Faster R-CNN architecture, which output the object mask. The additional mask output requires extraction of much finer spatial layout of an object, which is distinct from the class and box outputs. There are basically two architectures in Mask R-CNN framework. One is the convolutional backbone architecture used for feature extraction over an entire image. The other one is the head network for bounding box recognition and mask prediction which is applied separately to each ROI.

# 3. RESULTS AND DISCUSSIONS

## 3.1 Evaluation Metrics

All the metrics we used in the experiments are listed as follows:

- True Positive(TP): A prediction result is a correct detection, which means the IOU between prediction region and the ground truth region is greater than or equal to the threshold(depending on the metric, we use 0.5, 0.6, 0.7, 0.8 in the experiment).

- False Positive(FP): A prediction result is a wrong detection, which means the IOU between prediction region and the ground truth region is less than the threshold.

- False Negative(FN): A ground truth region is not detected.

- True Negative(TN): A prediction result is a correct mis-detection.

- Precision: The ability of a model to identify only the correct objects. It is given by:

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

- Recall: The ability of a model to detect all the correct objects. It is also referred to true positive rate or sensitivity and given by:

$$Precision = \frac{TP}{TP + FN} \tag{2}$$

- Average Precision(AP): It is averaged Precision over all the categories, which is only one object in our case.

- Average Recall(AR): The maximum Recall given a fixed number of detections per image over all the categories, which is one category in this case.

- Mean Average Precision(mAP): It is averaged AP over IOUs for each image, when IOU equals to 0.5, 0.6, 0.7, 0.8.

- Mean Average Recall(mAR): It is averaged AR over IOUs for each image, when IOU equals to 0.5, 0.6, 0.7, 0.8.

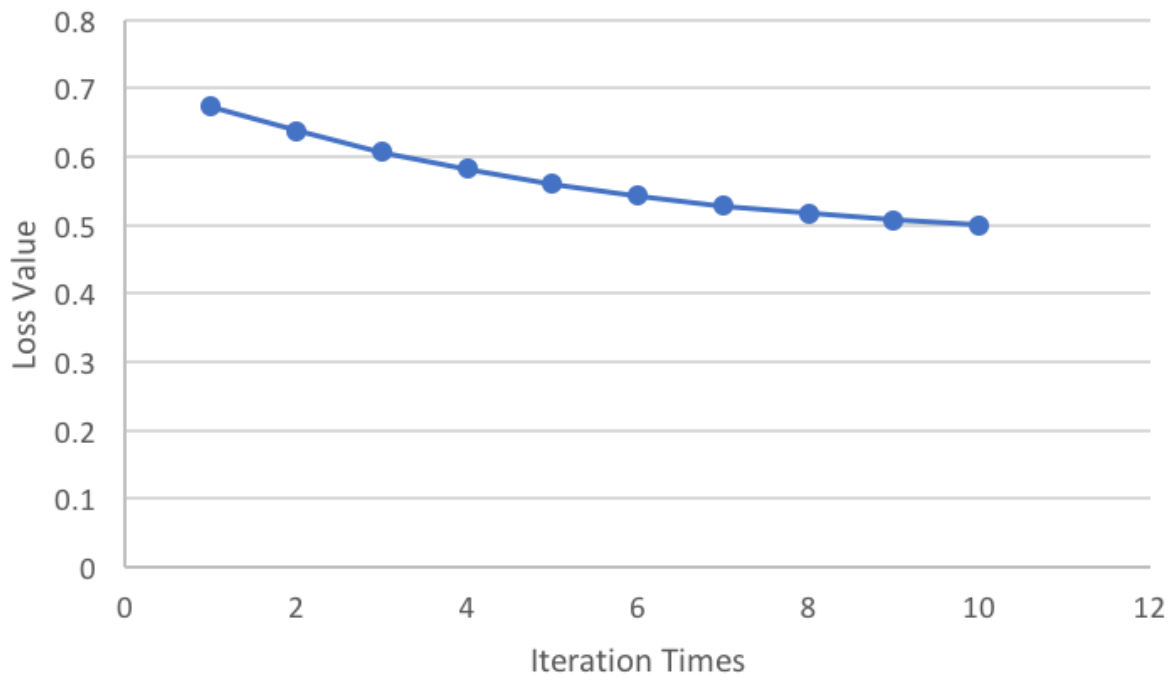In Table 1, Table 2 and Table 3, all the values are averaged over all images in the corresponding dataset.
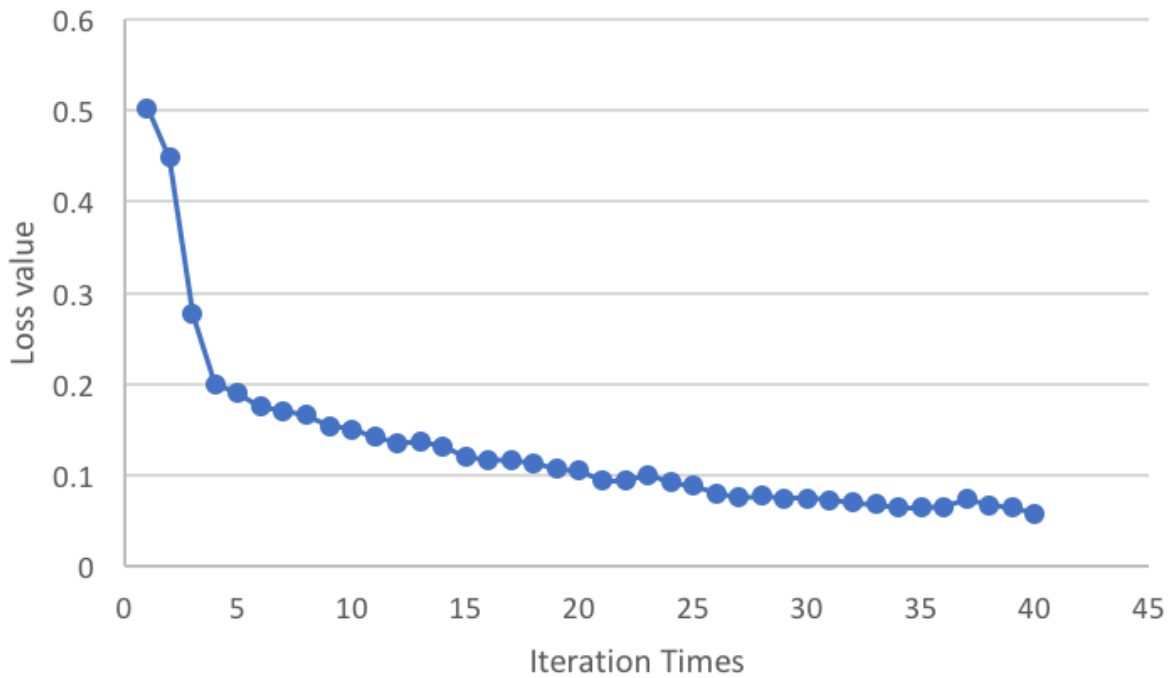
Figure 2. Training loss of U-Net


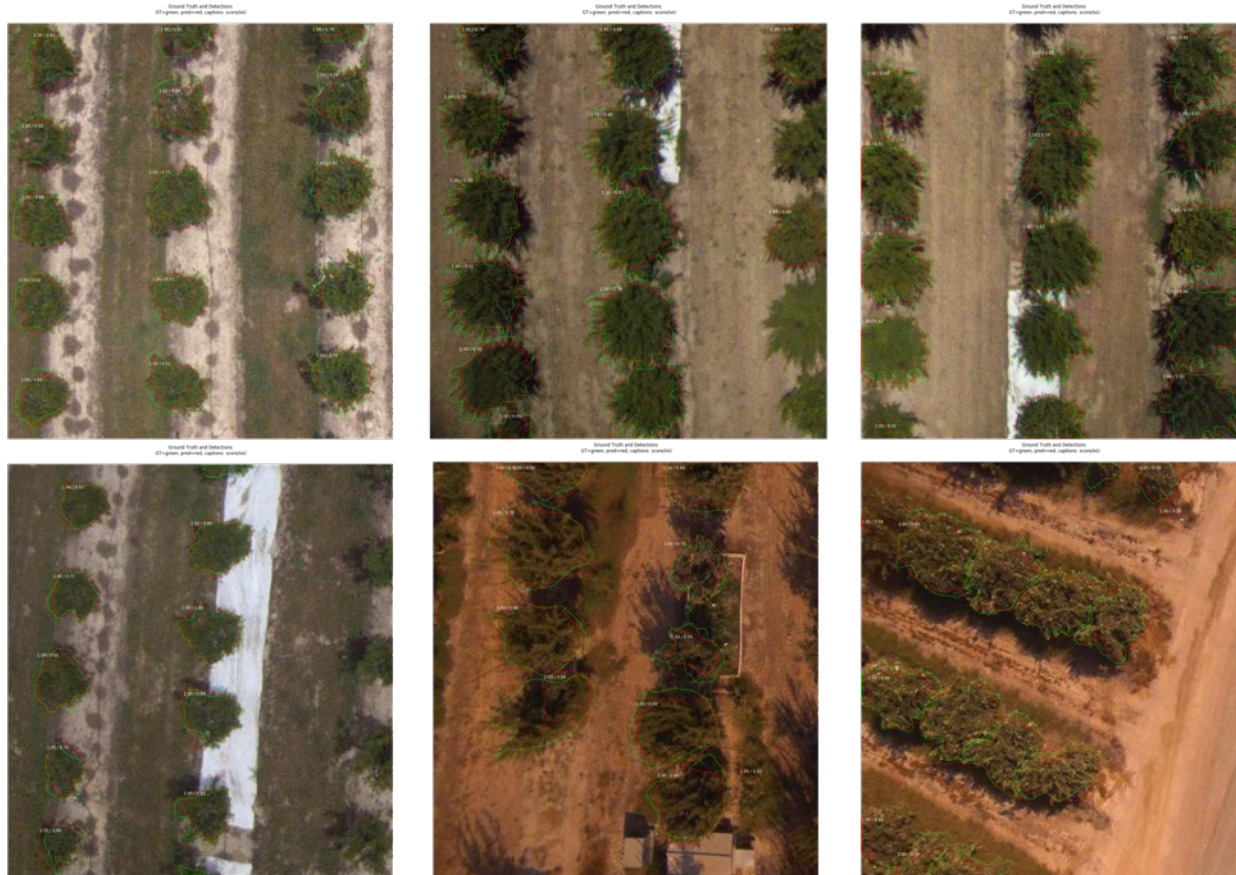
Figure 3. Training loss of Mask-RCNN

Figure 4. U-Net Segmentation results. Groundtruth is marked as green and prediction is in red.

## 3.2 Result Analysis of U-Net

The initial size of images is 534*600. In the U-Net experiment, we crop the images to 512*512 as the input of the network. The loss function used during training is binary cross-entropy and we use accuracy and hamming distance as validation metrics. The training loss curve is shown in Figure 2.

The AP and AR result of U-Net is listed in Table 1. We got 68.7 on $AP_{0.5}$, which means the average precision when ROI is set as 0.5.

As shown in Figure 4, U-Net performs good when the trees are sparse in the images. However, there are still some problems of the prediction results. First, the model can detect few unlabeled instances. But the number is low. Second, the ROI between ground truth and prediction is not big enough. And the contour of the crown is not accurate enough. Last but not the least, some trees are not totally segmented. There are more than one tree connecting with each other sometimes.

Table 1. U-Net result(AP and AR)

| Data Sets | **mAP** | $AP_{0.5}$ | $AP_{0.6}$ | $AP_{0.7}$ | $AP_{0.8}$ | **mAR** | $AR_{0.5}$ | $AR_{0.6}$ | $AR_{0.7}$ | $AR_{0.8}$ |
|-----------|---------|-----------|-----------|-----------|-----------|---------|-----------|-----------|-----------|-----------|
| Train | **42.9** | 74.9 | 70.1 | 57.6 | 25.5 | **66.0** | 80.5 | 75.9 | 65.6 | 41.8 |
| Test | **36.2** | 68.7 | 63.1 | 50.0 | 17.9 | **61.2** | 76.5 | 70.4 | 61.8 | 36.0 |

Figure 5. Mask-RCNN segmentation result(IOU=0.5). Groundtruth is marked as green and prediction is in red.

## 3.3 Result Analysis of Mask R-CNN

In Mask R-CNN experiment, each label in the initial ground truth image is isolated in the same size single image. We use pre-trained weights[29] for MS COCO to initiate the Mask R-CNN model. The training loss curve is shown in Figure 3.

In Table 2, the mAP and mAR result of Mask R-CNN in this experiment is 57.5 and 98.5. And it shows that Mask R-CNN model performs well at AP and AR when IOU varies from 0.5 to 0.7. From testing samples shown in Figure 5, we can notice: (i) the IOU of ground truth and prediction label is big; (ii) the model can predict lots of unlabeled trees; (iii) there are rarely multiple connected trees.

Table 2. Mask R-CNN Result (AP and AR)

| Data Sets | mAP | $AP_{0.5}$ | $AP_{0.6}$ | $AP_{0.7}$ | $AP_{0.8}$ | mAR | $AR_{0.5}$ | $AR_{0.6}$ | $AR_{0.7}$ | $AR_{0.8}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| Train | **58.6** | 96.5 | 95.0 | 84.2 | 41.2 | **98.7** | 98.7 | 98.7 | 98.7 | 98.7 |
| Test | **57.5** | 96.2 | 94.4 | 82.1 | 38.4 | **98.5** | 98.5 | 98.5 | 98.5 | 98.5 |

## 3.4 Comparing Those Two Results

The comparing of the testing result between U-Net and Mask R-CNN is shown in Table 3. In our pomegranate tree canopy segmentation experiments, it is clear that Mask R-CNN performed better in mAP and mAR, as

well as AP and AR when ROI is from 0.5 to 0.8. As discussed in Section 3.2 and 3.3, Mask R-CNN performs better in three aspects: (i) the IOU of ground truth and prediction label is bigger; (ii) the model can predict more unlabeled trees; (iii) there are rarely multiple connected trees. The downside of Mask R-CNN is its longer training time, which can be remediated by using pre-training weights.

Table 3. U-net VS. Mask R-CNN(AP and AR)

| Methods | **mAP** | $AP_{0.5}$ | $AP_{0.6}$ | $AP_{0.7}$ | $AP_{0.8}$ | **mAR** | $AR_{0.5}$ | $AR_{0.6}$ | $AR_{0.7}$ | $AR_{0.8}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| U-net | **36.2** | 68.7 | 63.1 | 50.0 | 17.9 | **61.2** | 76.5 | 70.4 | 61.8 | 36.0 |
| Mask R-CNN | **57.5** | 96.2 | 94.4 | 82.1 | 38.4 | **98.5** | 98.5 | 98.5 | 98.5 | 98.5 |

# 4. CONCLUSIONS

In order to segment pomegranate tree canopy for next step researches, we did experiments on two different CNN networks, which are U-Net and Mask R-CNN. The Mask R-CNN architecture achieves much better performance on our application. After using pre-trained weights to initiate the model, it has a very reasonable training time as well.

# ACKNOWLEDGMENTS

# REFERENCES

[1] United States Department of Agriculture Economic Research Service, "The Fruit and Tree Nuts Yearbook." https://www.ers.usda.gov/data-products/fruit-and-tree-nut-data/yearbook-tables/. (Accessed: March 10th,2017).

[2] Zhao, T., Wang, Z., Yang, Q., and Chen, Y., "Melon yield prediction using small unmanned aerial vehicles," in [*Autonomous Air and Ground Sensing Systems for Agricultural Optimization and Phenotyping II*], **10218**, 1021808, International Society for Optics and Photonics (2017).

[3] Sa, I., Ge, Z., Dayoub, F., Upcroft, B., Perez, T., and McCool, C., "Deepfruits: A fruit detection system using deep neural networks," *Sensors* **16**(8), 1222 (2016).

[4] Pérez-Ortiz, M., Peña, J. M., Gutiérrez, P. A., Torres-Sánchez, J., Hervás-Martínez, C., and López-Granados, F., "Selecting patterns and features for between-and within-crop-row weed mapping using uav-imagery," *Expert Systems with Applications* **47**, 85–94 (2016).

[5] She, T., Ehsani, R., Robbins, J., Leiva, J. N., and Owen, J., "Applications of small uav systems for tree and nursery inventory management," in [*12th International Conference on Precision Agriculture. Sacramento, CA*], (2014).

[6] Zhao, T., Stark, B., Chen, Y., Ray, A. L., and Doll, D., "A detailed field study of direct correlations between ground truth crop water stress and normalized difference vegetation index (ndvi) from small unmanned aerial system (suas)," in [*Unmanned Aircraft Systems (ICUAS), 2015 International Conference on*], 520–525, IEEE (2015).

[7] Zhao, T., Stark, B., Chen, Y., Ray, A., and Doll, D., "More reliable crop water stress quantification using small unmanned aerial systems (suas)," *IFAC-PapersOnLine* **49**(16), 409–414 (2016).

[8] Zhao, T., Stark, B., Chen, Y., Ray, A. L., and Doll, D., "Challenges in water stress quantification using small unmanned aerial system (suas): Lessons from a growing season of almond," *Journal of Intelligent & Robotic Systems* **88**(2-4), 721–735 (2017).

[9] Zhao, T., Doll, D., Wang, D., and Chen, Y., "A new framework for uav-based remote sensing data processing and its application in almond water stress quantification," in [*Unmanned Aircraft Systems (ICUAS), 2017 International Conference on*], 1794–1799, IEEE (2017).

[10] Zhao, T., Chen, Y., Ray, A., and Doll, D., "Quantifying almond water stress using unmanned aerial vehicles (uavs): correlation of stem water potential and higher order moments of non-normalized canopy distribution," in [*ASME 2017 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*], V009T07A058–V009T07A058, American Society of Mechanical Engineers (2017).

[11] Zhao, T., Doll, D., and Chen, Y., "Better almond water stress monitoring using fractional-order moments of non-normalized difference vegetation index," in [*2017 ASABE Annual International Meeting*], 1, American Society of Agricultural and Biological Engineers (2017).

[12] Marsal, J., Johnson, S., Casadesus, J., Lopez, G., Girona, J., and Stöckle, C., "Fraction of canopy intercepted radiation relates differently with crop coefficient depending on the season and the fruit tree species," *Agricultural and forest meteorology* **184**, 1–11 (2014).

[13] Llorens, J., Gil, E., Llop, J., et al., "Ultrasonic and lidar sensors for electronic canopy characterization in vineyards: Advances to improve pesticide application methods," *Sensors* **11**(2), 2177–2194 (2011).

[14] Sola-Guirado, R. R., Castillo-Ruiz, F. J., Jiménez-Jiménez, F., Blanco-Roldan, G. L., Castro-Garcia, S., and Gil-Ribes, J. A., "Olive actual on year yield forecast tool based on the tree canopy geometry using uas imagery," *Sensors* **17**(8), 1743 (2017).

[15] Strîmbu, V. F. and Strîmbu, B. M., "A graph-based segmentation algorithm for tree crown extraction using airborne lidar data," *ISPRS Journal of Photogrammetry and Remote Sensing* **104**, 30–43 (2015).

[16] Alonzo, M., Bookhagen, B., and Roberts, D. A., "Urban tree species mapping using hyperspectral and lidar data fusion," *Remote Sensing of Environment* **148**, 70–83 (2014).

[17] Leiva, J. N., Robbins, J., Saraswat, D., She, Y., and Ehsani, R., "Evaluating remotely sensed plant count accuracy with differing unmanned aircraft system altitudes, physical canopy separations, and ground covers," *Journal of Applied Remote Sensing* **11**(3), 036003 (2017).

[18] Khalid, N., Hamid, J. A., and Latif, Z. A., "Accuracy assessment of tree crown detection using local maxima and multi-resolution segmentation," in [*IOP Conference Series: Earth and Environmental Science*], **18**(1), 012023, IOP Publishing (2014).

[19] Yang, J., He, Y., and Caspersen, J., "A multi-band watershed segmentation method for individual tree crown delineation from high resolution multispectral aerial image," in [*Geoscience and Remote Sensing Symposium (IGARSS), 2014 IEEE International*], 1588–1591, IEEE (2014).

[20] Kang, J., Wang, L., Chen, F., and Niu, Z., "Identifying tree crown areas in undulating eucalyptus plantations using jseg multi-scale segmentation and unmanned aerial vehicle near-infrared imagery," *International Journal of Remote Sensing* **38**(8-10), 2296–2312 (2017).

[21] Díaz-Varela, R. A., de la Rosa, R., León, L., and Zarco-Tejada, P. J., "High-resolution airborne uav imagery to assess olive tree crown parameters using 3d photo reconstruction: application in breeding trials," *Remote Sensing* **7**(4), 4213–4232 (2015).

[22] Zhao, T., Niu, H., de la Rosa, E., Doll, D., Wang, D., and Chen, Y., "Tree canopy differentiation using instance-aware semantic segmentation," in [*2018 ASABE Annual International Meeting*], 1, American Society of Agricultural and Biological Engineers (2018).

[23] Dai, J., He, K., Li, Y., Ren, S., and Sun, J., "Instance-sensitive fully convolutional networks," in [*European Conference on Computer Vision*], 534–549, Springer (2016).

[24] Ronneberger O., Fischer P., B. T., "U-net: Convolutional networks for biomedical image segmentation," in [*Medical Image Computing and Computer-Assisted Intervention*], **9351**, 234–241 (2015).

[25] Kaiming He, Georgia Gkioxari, P. D. R. G., "Mask r-cnn," in [*arXiv preprint arXiv:1703.06870*], (2017).

[26] Zhang, H., Wang, D., Ayars, J. E., and Phene, C. J., "Biophysical response of young pomegranate trees to surface and sub-surface drip irrigation and deficit irrigation," *Irrigation Science* **35**(5), 425–435 (2017).

[27] Long, J., Shelhamer, E., and Darrell, T., "Fully convolutional networks for semantic segmentation," in [*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*], 3431–3440 (2015).

[28] Ren, S., He, K., Girshick, R., and Sun, J., "Faster R-CNN: Towards real-time object detection with region proposal networks," in [*Advances in Neural Information Processing Systems (NIPS)*], (2015).

[29] Matterport, I., "Mask r-cnn for object detection and instance segmentation on keras and tensorflow." https://github.com/matterport/Mask_RCNN.