COTTON STAND COUNT USING UAS IMAGERY AND DEEP LEARNING

Zhe Lin
Wenxuan Guo
Department of Plant and Soil Science, Texas Tech University
Lubbock, TX
Ahmed Harb Rabia
Damanhour University
Damanhour El-Behera, Egypt

Abstract

Stand count is critical for growers to make decisions for replanting and other site-specific management to avoid yield loss. This study applied and compared two object detection models, MobileNets and CenterNet, in cotton stand count using unmanned aerial system (UAS) images. The results showed that the overall mean precision and recall for the CenterNet model were higher than those for the MobileNets model. The CenterNet trained model had an R2 value of 0.60 and the MobileNets trained model had an R2 value of 0.48. The accuracy for the CenterNet model was 61% and for the MobileNets model was 46%. The results indicate that the CenterNet model has a better overall performance on cotton plant detection and counting.

Introduction

In cotton (Gossypium hirsutum L.) production, stand count is critical for growers to make decisions for replanting and other site-specific management to avoid yield loss (Boman and Lemon, 1990; Godfrey et al., 2015). The cotton yield will significantly decrease if the plant density is below five plants per linear meter of a row in Texas High Plains (Hopper and Supak, 1993). Traditional methods to evaluate plant stand count is manually counting the number of plants per unit area. However, conventional methods are time-consuming and labor-intensive with sampling bias. It is necessary to develop methods to determine stand count accurately and efficiently. The UAS technology can provide high-resolution images to perform high-throughput plant phenotyping through vision detection. Researchers have been using UAS equipped with RGB cameras, multispectral and thermal sensors to estimate plant or fruit counts. Deep learning object detection algorithms offered opportunities for high-throughput plant phenotyping in recent years. The deep networks have been tested to learn complex models that involve crop phenotypic attributes. For example, Lin and Guo (2020) proposed the integration of image segmentation and the U-Net CNN model using UAV images for sorghum panicle detection and localization. However, the effectiveness of different models has not been tested using the same training datasets. Various improved and customized models have been developed and tested successfully on object detection. This study applied and compared two object detection models, MobileNets (Howard et al., 2017) and CenterNet (Duan et al., 2019), in cotton stand count using UAS images

Materials and methods

Experimental Site

This study was conducted in a research field (33° 35' 50.53" N, 101° 54' 27.30" W) in New Deal, Texas, in 2020. In total, there were 208 plots, each 8 m long and eight rows wide in a north-south direction. A 1.5-m alley was arranged between plots. A subsurface drip irrigation system was used for irrigation in this field during the growing season.

UAS Image Acquisition

A DJI Phantom 4 Pro (DJI, Shenzhen, China) with a 4K RGB camera was applied in image acquisition. The flight plan was created using the DJI GSPro software (DJI, Shenzhen, China). The flight plan included 80% front overlap and 80% side overlap. The angle of the camera was set at 90 degrees to the land surface during flight. The UAS was flown at an altitude of 20 m at 2.4 m s-1. The spatial resolution was 4.3 mm for 20 m altitude. Two image datasets were acquired on June 8, and June 14, 2020. All image acquisitions were completed under sunny conditions with light to moderate wind around local solar noon. Raw images were stitched using the Pix4DMapper software (Pix4D S.A., Switzerland).

Training and Testing Images

The training images containing 400 images were prepared by randomly cropping the raw UAS images and the stitched image (Figure 1). We used LabelImg tool (Tzutalin, 2015) for image annotation. Individual cotton plant with two or four leaves was labeled with a rectangular bounding box. Each training image has a corresponding xml file in PASCAL VOC format containing the filename, path, object and bounding box's top left hand and bottom right-hand corners information, height and width of the image. Both training images and their corresponding xml files would be used in model training.

The test dataset containing 100 images was prepared for accuracy assessment. The images in the test dataset were different from the images in the training dataset in dimensions and size. Each test image covered one row of the cotton plot that had a 1-m length in the field. Cotton plants in each test image were manually counted, and the number of cotton plants in these test images varied from 8 to 21.

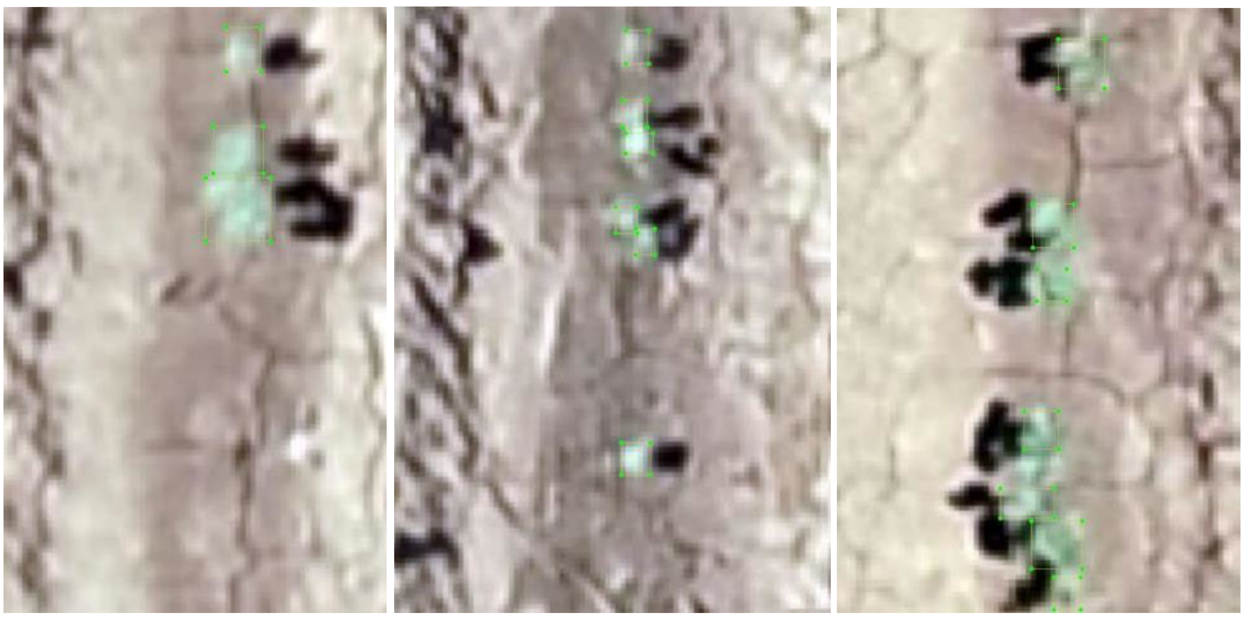

Figure 1. Examples of training images with a bounding box for cotton plant detection and counting

MobileNets

We used the pre-trained SSD-MobileNet-V2 -FPNlite 320x320 model to train the dataset with Tensorflow Object Detection API (Abadi et al., 2016). The single shot detector (SSD) architecture aims to predict bounding box locations and classify these boxes in one pass. The SSD consists of the MobileNets as the base architecture. The MobileNets model reduced the network and model size comparing with other object detection models. The MobileNets model uses only a single convolution network that applies to all the channels of the input image and slides the weighted sum to the next pixel (Sandler et al., 2018). The model used a random normal initializer and momentum optimizer with a learning rate base of 0.1. The COCO mAP is 28.2.

CenterNet

Another pre-trained model we selected was Centernet Resnet50 from Tensorflow Object Detection API. CenterNet is an object detection architecture based on a deep convolution neural network trained to detect each object as a triplet, rather than a pair, of keypoints (Duan et al., 2019). It only focuses on the center information, which minimizes the computation cost for this approach. The backbone used in this model was ResNet50. Center pooling, which helped to better detection of center keypoints in both horizontal and vertical directions, aims to capture more recognizable visual patterns. Cascade corner pooling focus on determining the corners of the bounding box by finding the maximum values on the boundary directions. Both cascade corner pooling and center pooling could be computed by combining corner pooling at different directions based on various situations (Law and Deng, 2018). The model used a random normal initializer and the adam optimizer with a learning rate base of 0.001. The COCO mAP was 31.2.

Evaluations

Precision, recall, and F-score were used in this study to determine the effectiveness and accuracy of cotton plant detection. Precision and recall are the most commonly used indicators to evaluate the performance of object detection models. F1-score aims to balance the two indicators (Zhao et al., 2018). Precision, recall, and F1-score are defined in terms of true positive (TP), false positive (FP), and false negative (FN) as follows:

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

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

$$F1 = 2 \times \frac{\mathrm{Precision} \times \mathrm{Recall}}{Precision + Recall} \quad (3)$$

The mean absolute error (MAE), mean absolute percentage error (MAPE), accuracy (ACC), coefficient of the determination (R2), and the root mean squared error (RMSE) were used as evaluation metrics to assess the performance of the cotton stand counting.

$$MAE = \frac{1}{n}\sum_{1}^{n}|m_i - c_i| \quad (4)$$

$$MAPE = \frac{1}{n}\sum_{1}^{n}\left|\frac{m_i - c_i}{m_i}\right| \quad (5)$$

$$ACC = (1 - \frac{1}{n}\sum_{1}^{n}\frac{|m_i - c_i|}{m_i}) \times 100\% \quad (6)$$

$$R^2 = 1 - \frac{\sum_1^n (m_i - c_i)^2}{\sum_1^n (m_i - \overline{m_i})^2} \quad (7)$$

$$RMSE = \sqrt{\frac{\sum_1^n (m_i - c_i)^2}{n}} \quad (8)$$

where mi, $\overline{\mathrm{m}}$i, and ci represent the manually counted cotton plants for the ith image, the mean manual counts, and the predicted count for the ith image, respectively. n is the number of test images.

Hardware and Libraries

The algorithm was implemented using the Python programming language (Version 3.7, Python Software Foundation). The integrated development environment (IDE) used for our study was Google Colaboratory. Training, evaluation, and testing were performed using the TensorFlow2 (Abadi et al., 2016) high-level neural networks application programming interface (API). The execution of the algorithm was performed under GPUs with 12GB RAM. Both models resized input images to 512 x 512. 30000 total steps were run for each model.

Results and Discussion

Training Period and Cotton Plant Detection

The mean per-step training time was 1.03 seconds for the MobileNets model and 0.84 seconds for the CenterNet model. With 30000 total training steps, the CentetNet model saved about 2 hours of training time. The trained model size was about 30 MB for the MobileNets and about 220 MB for the Centernet. Table 1 showed the mean precision and recall under various Intersection over Union (IoU) thresholds and target area sizes for the two models. Both mean precision and recall for the CenterNet model were higher than those for the MobileNets model, especially for small and medium target area sizes. The mAP for the CenterNet model was 31.2 and 28.2 for the MobileNets model. The results indicated the CenterNet model had a better performance for cotton plant detection.

Table 1. Mean precision and recall values in various IoU thresholds and target area sizes for the MobileNets and CenterNet models

	IoU Threshold	Area size	MobileNets	CenterNet
Mean Precision	0.50:0.95	All	0.284	0.382
	0.50	All	0.667	0.904
	0.75	All	0.220	0.143
	0.50:0.95	Small	0.297	0.378
	0.50:0.95	Medium	0.278	0.442
Mean Recall	0.50:0.95	All	0.392	0.475
	0.50:0.95	Small	0.400	0.470
	0.50:0.95	Medium	0.357	0.511

Cotton Stand Count

Table 2 shows the evaluation metrics for the performance of cotton stand count with MobileNets and CenterNet models. With 400 training images, the CenterNet trained model had an R2 value of 0.60, and the MobileNets trained model had an R2 value of 0.48. The CenterNet model showed lower values of mean absolute error, mean absolute percentage error, and root mean square error as compared with the MobileNets model. On the other hand, the accuracy for the CenterNet model was higher than the MobileNets model. The overall counting performance of the CenterNet model was better than the MobileNets model.

Table 2. Coefficient of determination (R2), Mean absolute error (MAE), mean absolute percentage error (MAPE), accuracy (ACC) and root mean squared error (RMSE) for cotton stand count with MobileNets and CenterNet models using unmanned aerial system images.

	R2	MAE	MAPE	ACC	RMSE
MobileNets	0.48	7.48	0.54	46%	7.81
CenterNet	0.60	5.39	0.39	61%	6.08

The models had challenges on test images with high brightness and low contrast environment. Both models could not detect cotton plants (Figure 2A). The orange box pointed that CenterNet model detected and separated smaller cotton plants while the MobileNets model only detected one cotton plant. Figure 2D showed that both models detected cotton plants in conditions with high contrast. Furthermore, these two models could not detect overlapping cotton leaves or cotton plants in high-density situations (Figure 2B). The red box labeled on the image showed that both models did not separate two cotton plants. The test results showed that both models worked better with test images from June 14, 2020, because the cotton plant was relatively larger. Based on the target object size, CenterNet model had a better performance with smaller cotton plants (Figure 2C).

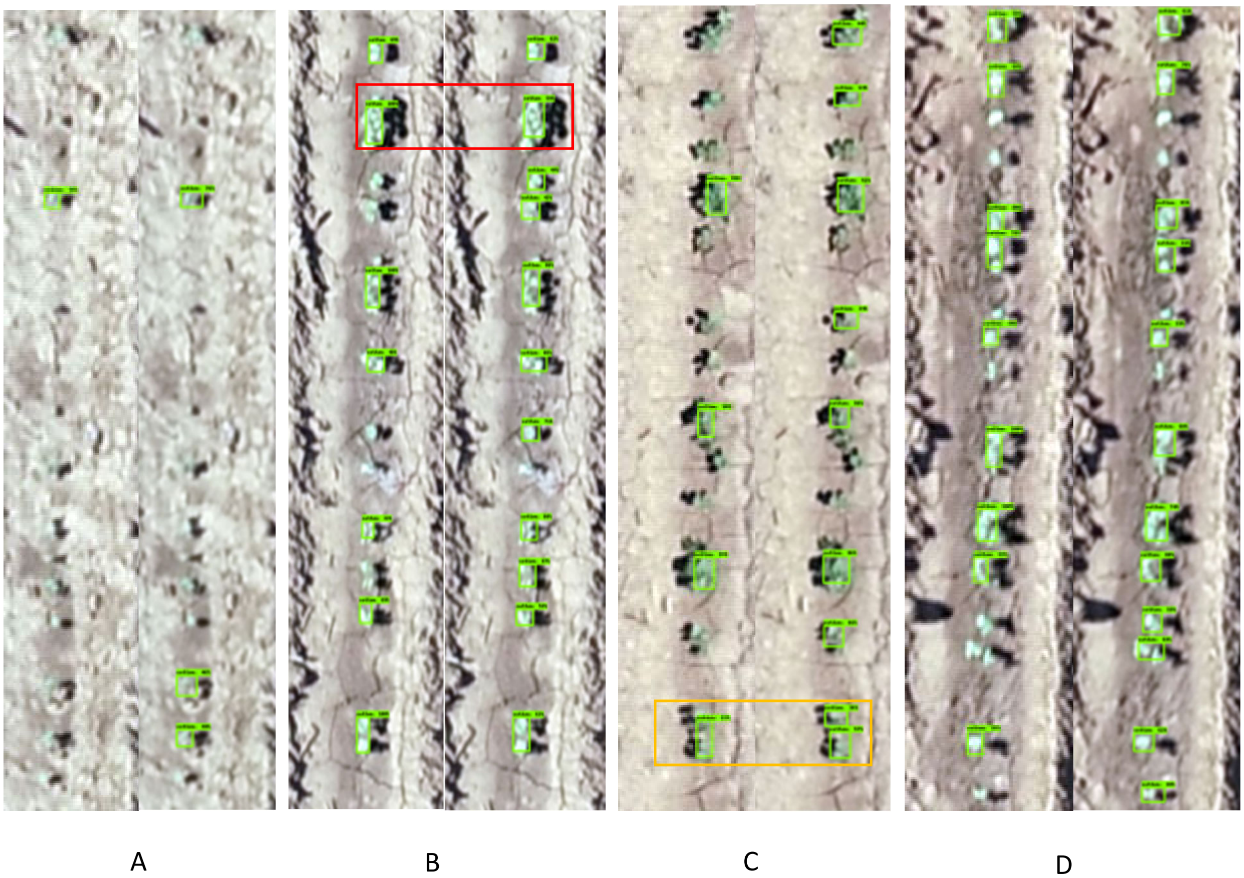

Figure 2. Examples of cotton plant detection results (Left images are from the MobileNets model, right images are from the CenterNet models, A and B represent images from June 8, 2020, C and D represent images from June 14, 2020)

Summary

This study compared two object detection models, MobileNets and CenterNet, in cotton stand count. The results showed that the overall mean precision and recall for the CenterNet model were higher than those for the MobileNets model. The CenterNet trained model had an R2 value of 0.60 and the MobileNets trained model had an R2 value of 0.48. The accuracy for the CenterNet model was 61% and for the MobileNets model was 46%. The results indicated that the CenterNet model had a better overall performance on cotton plant detection and counting. For both models, there were still challenges in detecting small cotton plants under high brightness and low contrast conditions. More training images are required to improve the accuracy and robustness of deep learning object detection models. Besides, cotton stand count accuracy as influenced by environmental factors, including image resolution, soil background, and illumination levels, requires further evaluation.

References

Abadi, M., A. Agarwal, P. Barham, E. Brevdo, Z. Chen, et al. 2016. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. www.tensorflow.org. (accessed 20 November 2019).

Boman and Lemon. 1990. Making Replant Decisions in Cotton.

Duan, K., S. Bai, L. Xie, H. Qi, Q. Huang, et al. 2019. CenterNet: Keypoint triplets for object detection. Proceedings of the IEEE International Conference on Computer Vision. p. 6568–6577

Godfrey, L.D., P.B. Goodell, E.T. Natwick, D.R. Haviland, and V.M. Barlow. 2015. UC IPM Pest Management Guidelines:cotton. University of California Division of Agriculture and Natural Resources. http://ipm.ucanr.edu/PMG/r3300311.html (accessed 16 November 2020).

Hopper, N.W., and J. Supak. 1993. Fungicide Treatment Effects on Cotton (Gossypium Hirsutum) Emergence, Establishment and Yield. Texas Journal of Agriculture and Natural Resources 6: 69–80. http://txjanr.agintexas.org/index.php/txjanr/article/view/374 (accessed 16 November 2020).

Howard, A.G., M. Zhu, B. Chen, D. Kalenichenko, W. Wang, et al. 2017. MobileNets: Efficient convolutional neural networks for mobile vision applications. arXiv.

Law, H., and J. Deng. 2018. CornerNet: Detecting Objects as Paired Keypoints. International Journal of Computer Vision 128(3): 642–656. http://arxiv.org/abs/1808.01244 (accessed 5 January 2021).

Lin, Z., and W. Guo. 2020. Sorghum Panicle Detection and Counting Using Unmanned Aerial System Images and Deep Learning. Frontiers in Plant Science 11. doi: 10.3389/fpls.2020.534853.

Sandler, M., A. Howard, M. Zhu, A. Zhmoginov, and L.C. Chen. 2018. MobileNetV2: Inverted Residuals and Linear Bottlenecks. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. p. 4510–4520

Tzutalin, D. 2015. LabelImg. Git code. https://github.com/tzutalin/labelImg.

Zhao, B., J. Zhang, C. Yang, G. Zhou, Y. Ding, et al. 2018. Rapeseed seedling stand counting and seeding performance evaluation at two early growth stages based on unmanned aerial vehicle imagery. Frontiers in Plant Science 9: 1362. doi: 10.3389/fpls.2018.01362.