Monitoring Growth Rate of Lettuce Using Deep Convolutional Neural Networks

Conference Paper · January 2019

DOI: 10.13031/aim.201900341

Chung-Liang Chang
National Pingtung University of Science and Technology

Yan-Fu Kuo
National Taiwan University

2950 Niles Road, St. Joseph, MI 49085-9659, USA
269.429.0300 fax 269.429.3852 hq@asabe.org www.asabe.org

An ASABE Meeting Presentation
DOI: https://doi.org/10.13031/aim.201900341
Paper Number: 1900341

Monitoring Growth Rate of Lettuce Using Deep Convolutional Neural Networks

[1]Jie-Yan Lu, [2]Chung-Liang Chang, [1]Yan-Fu Kuo
[1]Department of Bio-Industrial Mechatronics Engineering,
[1]National Taiwan University, Taiwan
[2]Department of Biomechatronics Engineering,
[2]National Pingtung University of Science & Technology, Taiwan

Written for presentation at the
2019 ASABE Annual International Meeting
Sponsored by ASABE
Boston, Massachusetts
July 7–10, 2019

ABSTRACT. ***Lettuce is one majorly consumed vegetable worldwide. According to the United States Department of Agriculture, lettuce accounted for 7% of total vegetable consumption in 2013. Monitoring the growth of lettuces is crucial for ensuring the quality and quantity of lettuce. This study proposed to monitor the growth of lettuces in greenhouses using time-lapse images and deep learning. An imaging system was constructed to acquire time-lapse images of lettuces in greenhouses. A mask region-based convolutional neural network (Mask R-CNN) model was next developed to localize the lettuces in the images and segment the leaf areas simultaneously. The growth rates were then determined as the leaf areas of the lettuces versus time. Experimental results showed that the Mask R-CNN model achieved an accuracy reached 97.63% on estimating leaf area.***

Keywords. Deep learning, growth monitoring, mask region-based convolutional neural network.

Introduction

Vegetable is essential in human diet. Although accounting for a low percentage to human calories intake, vegetable contains crucial nutrients, such as dietary fiber and diverse vitamins. However, the production of vegetable in the world is generally not enough due to the unique growing conditions. World Health Organization reported that inadequate vegetable consumption attributed to approximately 3.4 million deaths in 2013 (McColl, 2016). Greenhouse cultivation is an effective way to alleviate the problem. Greenhouse is equipped with various environmental sensors that can monitor and control the environmental factors (e.g., temperature, humidity, and CO_2) to optimize the conditions for plant growth. Still, the monitoring on growth of plants still relies on manual observation. Therefore, this study proposed to automatically monitor the growth of lettuce in greenhouse using image-based approaches.

Recently, a deep learning approach - convolutional neural network (CNN) has become a powerful tool for tackling image problems. Many CNN-based approaches, such as deconvolution network (Noh et al., 2015), fully convolutional network (Long et al., 2015), and SegNet (Badrinarayanan et al., 2017), were developed to solve *semantic segmentation* problem.

These networks are able to separate the objects from the background but lack of the ability of classification. *Instance segmentation* was proposed to perform semantic segmentation and object classification simultaneously, allowing users to segment and detect objects with clear area and contour. Mask region-based convolutional neural network (He et al., 2017) is a state-of-the-art instance segmentation method. Gan et al. (2019) applied Mask R-CNN on detection of surface area to predict the volume of ellipsoidal ham. By using Mask R-CNN, Mulay et al. (2019) presented an accuracy of 0.88 on the area detection of the retinal ridge, an important landmark in specific medical diagnosis.

This study proposed to monitor the growth rates of lettuces using Mask R-CNN. The specific objectives were to (1) identify the leaf area of lettuce (Fig. 1a) using Mask R-CNN, (2) verify the accuracy of area estimation, and (3) monitor the growth rates of the lettuce.

(a) (b)

Figure 1. Experimental images of lettuce: (a) original image, (b) labeled image.

Materials and Methods

Image Acquisition and Preprocessing

The images of lettuce plants were collected using a customized imaging system (Fig. 2; Fu et al., 2018). The system was installed in a greenhouse (Intelligent System & Signal Processing Lab., Department of Biomechatronics Engineering, National Pingtung University of Science and Technology, Taiwan) that cultivated lettuce using a PVC hydroponic system (Fig. 2). The imaging system automatically captured images of vegetables from the PVC hydroponic system. The imaging system was composed of a Raspberry Pi (Raspberry Pi 3, Raspberry Pi Foundation, UK) and a webcam (BRIO 4K HD, Logitech Inc., Taiwan). The Raspberry Pi was used to control the webcam to acquire images automatically. In this study, images of lettuces were recorded every half an hour from 6 am to 6 pm for a complete growth period of the plants (almost twenty days). Totally, there were 1218 images collected from 5 lettuces. The images were first resized from 3280 × 2464 pixels to 800 × 600 pixels to expedite model training. Then, the images were divided into training set, test set, and a set for monitoring the growth rate with the numbers of 850, 200, and 168, respectively. The set for monitoring growth rate contained 168 images collected from the same lettuce. The annotations of lettuces in the images of training and test sets were achieved using labelme (Russell et al., 2008).

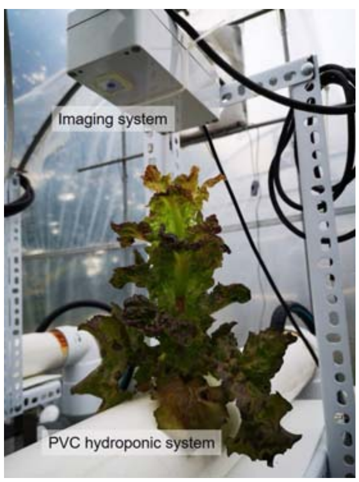

Figure 2. Imaging system and PVC hydroponic system.

Lettuce Segmentation

Mask R-CNN was used to automatically segmented the leaf area of the lettuce in the image. Mask R-CNN is composed of a base CNN, a feature extractor network, a region proposal network (RPN; Ren et al. 2015), region of interest (RoI) alignment, bounding-box regression (BBR), object classification (OC), and mask prediction (MP). ResNet-50 (He et al., 2015) was used as the base CNN. The architecture of ResNet-50 was constructed by five convolution blocks, C1 to C5. C1 was a single convolution layer. Blocks C2 to C5 were composed of 3, 4, 6, and 3 convolutional layers, respectively. Feature pyramid network (FPN; Lin et al., 2016) was used as the feature extraction network in the base CNN. FPN merged the outputs of C2, C3, C4, and C5 to create feature maps of various scales. RPN was used to propose RoIs in 5 sizes (16 × 16, 32 × 32, 64 × 64, 128 × 128, and 256 × 256 pixels) and 3 aspect ratios (1:2, 1:1, and 2:1) from the feature maps. An RoI was positive if the intersection over union (IoU) between and RoI and a ground truth was over 0.7. An RoI was negative when the IoU between the RoI and all the ground truths were less than 0.3. The RoI alignment was used to segment the RoI accurately in the feature maps using bilinear interpolation (Kirkland, 2010). The outputs of the RoI alignment were the spatial feature maps of 7 × 7 pixels. The feature maps were then fed into three branches: BBR, OC, and MP. The BBR outputted the coordinates of the bounding boxes and the OC outputted the confidence scores of the objects being a class (e.g., vegetable). The BBR and OC were composed of fully connected layers. The MP outputted the binary masks of the objects. The MP was a fully convolutional network that composed of four consecutive 2 × 2 convolution layers, one 2 × 2 deconvolution layer, and one convolution layer in sequence.

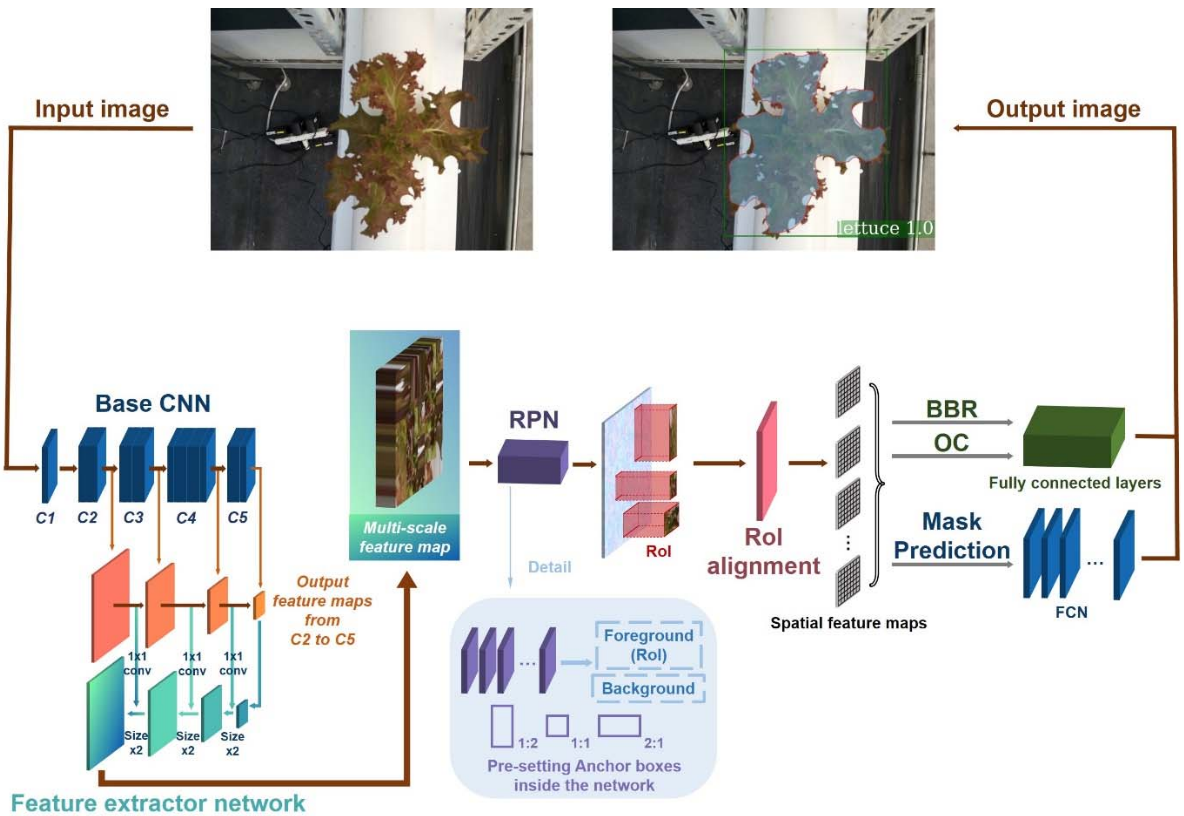

Figure 3. Structure of Mask R-CNN.

The training, test, and the instance detection processes were executed on an open-source python environment (Van, 1995) and an open-source package Detectron (Girshick et al., 2018). In the training process, the model was first initialized with the model with weights pre-trained using the COCO dataset and was trained for 18000 iterations. The initially learning rate was set to 0.0001 and the decay (Bhatnagar et al., 2013) by a factor of 10 was set at the 12000 and 16000 iterations, respectively. Two graphics processing units (GPU; GeForce GTX 1080Ti, Nvidia; Santa Clara, USA) were used. Each GPU processed 8 images in every epoch. In the testing process, the *validation monitor* performed the validation in common training process to save the training time.

Leaf Area Verification

The verification of leaf area was executed by computing the detection error. Each image in test images proposed an error by comparing the ground truth area and detected area. The number of pixels in a ground truth area was manually decided in the labeled image (Fig. 1b). The number of pixels of detected area was automatically decided by the output image in the ID process. Therefore, 200 errors from all test images were recorded and an average error was viewed as the index of realibility of the model. The errors in pixel were then converted to cm^2 using the dimension of the PVC tube (10 cm in diameter) in the images.

Growth Rate Estimation

The growth rate of lettuce was monitored based on the growth of leaf area. Image set for monitoring the growth rate was used as the input of the ID process. The outputs of ID process were further recorded to estimate the growth rate, computed by the exponential growth equation (Contributors, 2019),

$$x_t = x_o(1 + r)^t \quad (1)$$

where

x_t = the value in time t
x_0 = the value in time 0
r = the growth rate
t = time in discrete intervals (e.g., 0, 1, 2, 3, …)

Result

In section 'Detection of Leaf Area', the area detection and precision of the developed Mask R-CNN are demonstrated. In section 'Accuracy of Estimated Leaf Area', the accuracy of the developed Mask R-CNN is demonstrated. In section 3.3, the developed Mask R-CNN is applied on time-lapse images to measure the grow rate of the lettuce in images.

Detection of Leaf Area

The detection of lettuce by Mask R-CNN is shown in Fig. 4(c). The red contour presents the detected leaf area. The green box is the bounding box with the name of detected object and its score during the ID process. The AP in this detection model reached 0.8234.

(a) (b) (c)

Figure 4. Leaf area detection: (a) labeled area, (b) original image, and (c) detection area.

Accuracy of Estimated Leaf Area

The performance of the trained Mask R-CNN model was evaluated using the 200 test images (Fig. 5). The model reached a mean accuracy of 97.63% in leaf area estimation. The error of the estimation was relatively large when the leaf area was small. Still, the maximum error was only 0.25 cm^2.

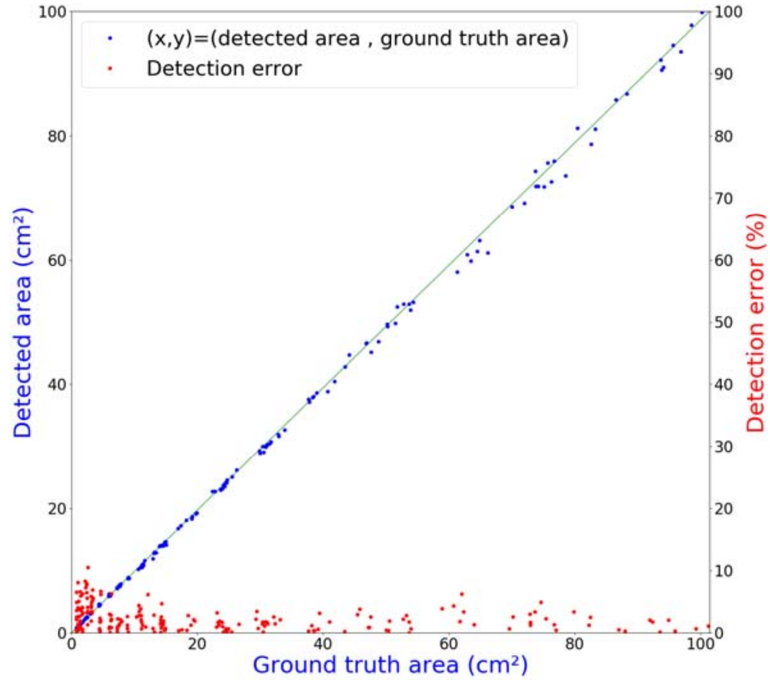

Figure 5. Ground truth area-detected area graph of test images. The blue points represent the images with two different area data. The red points represent the errors (%) in detected area and ground truth area.

Growth Rate of Lettuce

The detection of leaf area of the same lettuce was evaluated using time-lapse images (Fig. 6). The mean area in a day was used for calculating the growth rate. The growth rate of the lettuce was 23.25% per day (Fig. 7).

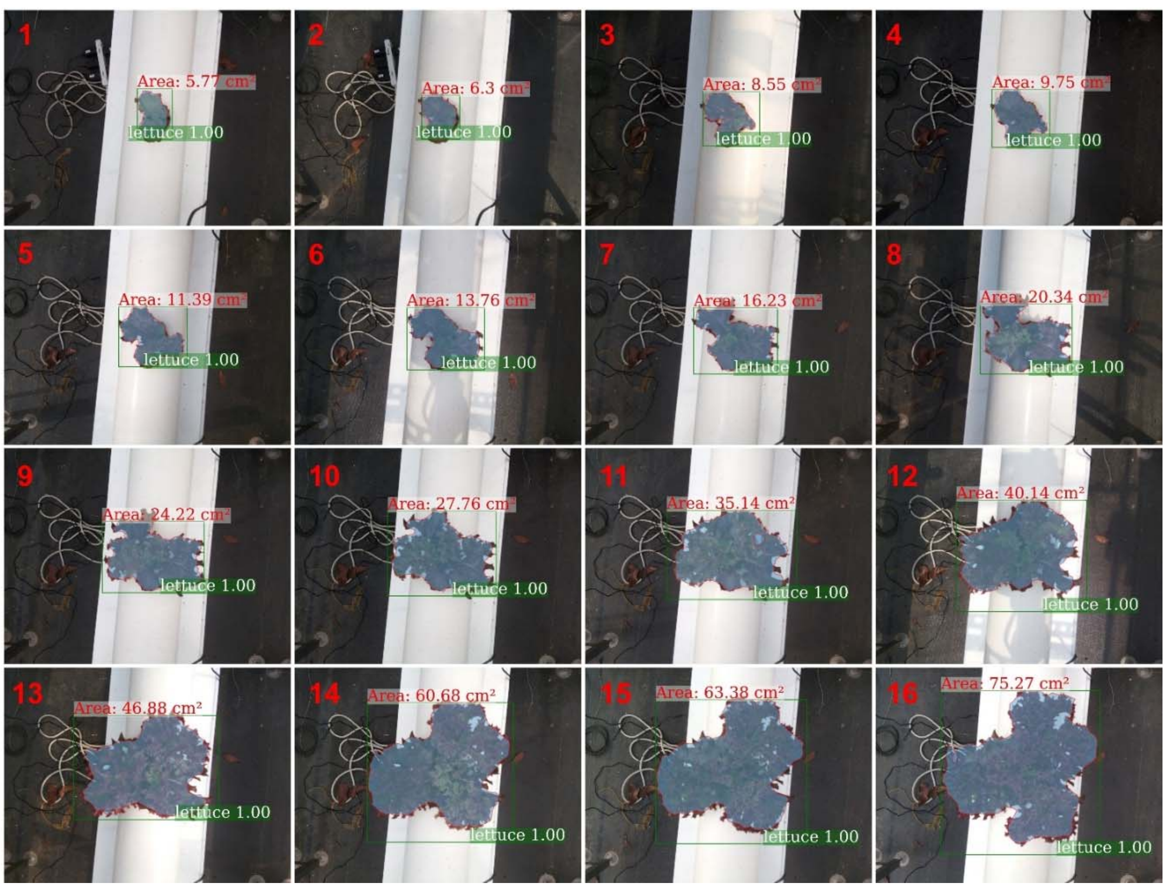

Figure 6. Leaf area (cm^2) detection. The numbers on the upper left corner correspond to the order of image token.

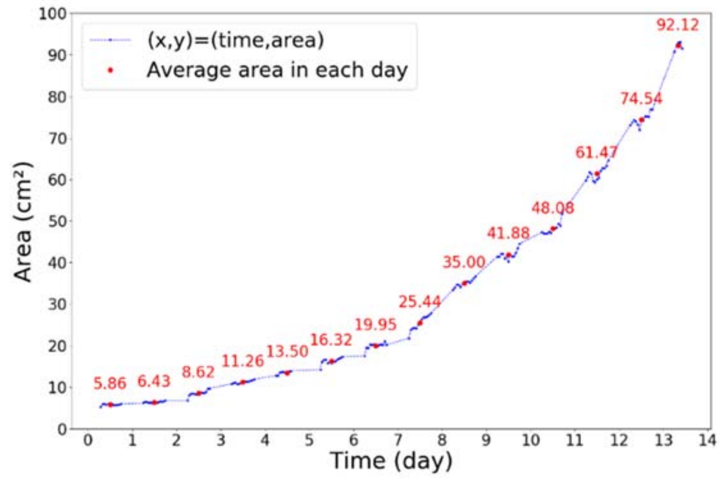

Figure 7. Area-time graph of time-lapse images.

Conclusion

This study proposed to automatically monitor the growth of lettuce using time-lapse images and Mask R-CNN. An imaging system was customized to acquire top-view time-lapse images of lettuces in a greenhouse. A Mask R-CNN model was then trained to measure the top-view areas of the lettuces in the images. The approach reached a mean accuracy of 97.63%. The growth rates of the lettuces were also presented.

Acknowledgement

This research was supported by the Council of Agriculture, Executive Yuan, Taiwan, under the grant 108AS-13.2.11-ST-a9.

References

Badrinarayanan, V., et al. (2017). SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *Ieee Transactions on Pattern Analysis and Machine Intelligence, 39*(12), 2481-2495. doi:10.1109/Tpami.2016.2644615

Bhatnagar, S., et al. (2013). Stochastic Approximation Algorithms. In S. Bhatnagar, H. L. Prasad, & L. A. Prashanth (Eds.), *Stochastic Recursive Algorithms for Optimization: Simultaneous Perturbation Methods* (pp. 17-28). London: Springer London.

Contributors, W. (2019). Exponential growth. Retrieved from https://en.wikipedia.org/w/index.php?title=Exponential_growth&oldid=891549829

Fu, W. L., Huang, C. C., & Chang, C.L. (2018). Design and implementation of moving platform for image acquisition and processing in a hydroponic greenhouse [In Chinese]. *Proc. 2018 Conf. Agricultural Machinery and Bio-Mechatronics Engineering*, pp. 815-822. Taiwan: CIAM.

Gan, Y. S., et al. (2019). Automatic Surface Area and Volume Prediction on Ellipsoidal Ham using Deep Learning. *eprint arXiv:1901.04947*, arXiv:1901.04947.

Girshick, R., et al. (2018). Detectron.

He, K., et al. (2017). Mask R-CNN. *eprint arXiv:1703.06870*, arXiv:1703.06870.

He, K., et al. (2015). Deep Residual Learning for Image Recognition. *eprint arXiv:1512.03385*, arXiv:1512.03385.

Kirkland, E. J. (2010). Bilinear Interpolation. In E. J. Kirkland (Ed.), *Advanced Computing in Electron Microscopy* (pp. 261-263). Boston, MA: Springer US.

Lin, T.-Y., et al. (2016). Feature Pyramid Networks for Object Detection. *eprint arXiv:1612.03144*, arXiv:1612.03144.

Lin, T.-Y., et al. (2014, 2014//). *Microsoft COCO: Common Objects in Context.* Paper presented at the Computer Vision – ECCV 2014, Cham.

Long, J., et al. (2015, 7-12 June 2015). *Fully convolutional networks for semantic segmentation.* Paper presented at the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

McColl, K. (2016). Increasing fruit and vegetable consumption to reduce the risk of noncommunicable diseases. Retrieved from https://www.who.int/elena/titles/commentary/fruit_vegetables_ncds/en/

Mulay, S., et al. (2019). *Early detection of retinopathy of prematurity stage using deep learning approach*.

Noh, H., et al. (2015). Learning Deconvolution Network for Semantic Segmentation. *arXiv e-prints*. https://ui.adsabs.harvard.edu/abs/2015arXiv150504366N

Ren, S., et al. (2015). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *eprint arXiv:1506.01497*, arXiv:1506.01497.

Russell, B. C., et al. (2008). LabelMe: A Database and Web-Based Tool for Image Annotation. *International Journal of Computer Vision*, *77*(1), 157-173. doi:10.1007/s11263-007-0090-8

van G, R. (1995). *Python Tutorial*.