Extracting the height of lettuce by using neural networks of image recognition in deep learning

Xiyue Guo[a], Yong Zhong[a], Yarong Feng[a], Ming Zhao[a], and Minjuan Wang[b]

[a,b]Key Laboratory of Modern Precision Agriculture System Integration Research, China Agricultural University, Beijing, China

ABSTRACT

The traditional method of measuring the lettuce height is a manual measurement with instruments, which is greatly affected by human error.At present, researchers have proposed to use color cameras to obtain RGB images of lettuce, and to obtain the height of lettuce from the images. However, these tasks usually require camera calibration or a reference object with a known height, which is somewhat restrictive. Considering that deep neural networks have a powerful ability to feature extraction and expression, without camera calibration and reference objects, we try to use four networks of image recognition to explore the effect of deep learning on abstracting the lettuce height from RGB images. On the test set, including 80 images and height from 0.9 cm to 7.5 cm, we achieve a good result with a mean absolute error of 1.22 mm.

Keywords: Lettuce, Plant Height, Neural Networks, Image Recognition, Deep Learning

1. INTRODUCTION

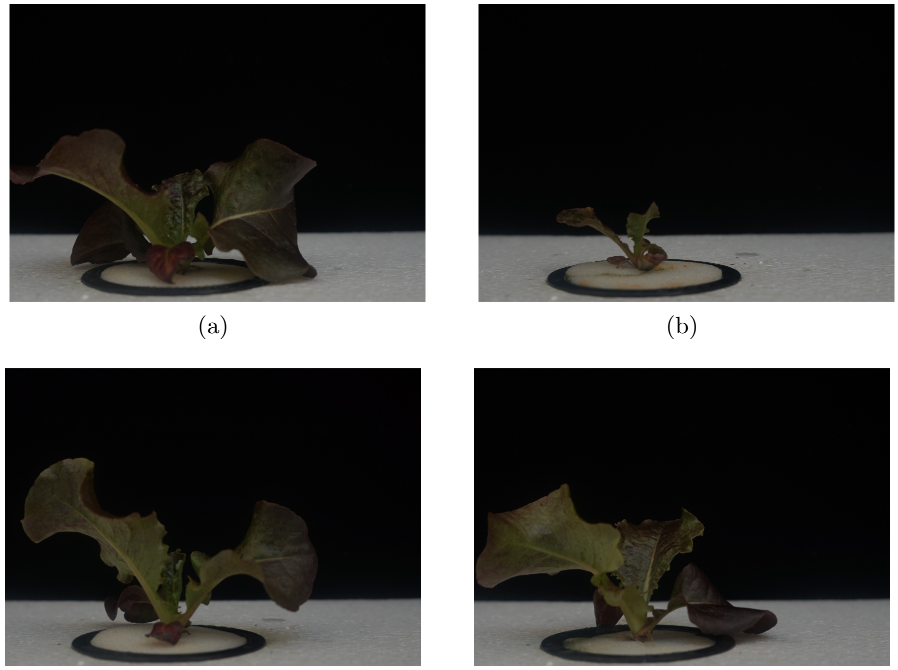
Lettuce faces many problems in planting, such as various disease and pest controls, and temperature and humidity controls. Therefore, it is necessary to monitor the growth rate of lettuce in real-time to find the abnormal growth of lettuce in time and discover the root cause.

Plant phenotype, produced by the interaction between genotype and environment, refers to some or all identifiable plant physical, physiological and biochemical characteristics, and traits that can reflect the structure and composition of a plant or reflect the growth and development of a plant[1] . As one of the phenotypic parameters of plants, plant height can not only reflect the basic growth situation of lettuce but also provide a reference for the refined planting of lettuce.

Compared with the measurement methods of plant height based on special sensors, the way of directly using RGB images captured by an ordinary digital camera to obtain plant height is not only low cost and easy to operate, but also its accuracy largely depends on image processing algorithms that can be improved to advance the outcome. However, this measurement method based on color images usually requires a reference.

Considering that Convolutional Neural Network (CNN) has a powerful ability of feature extraction and expression for images and has been widely used in kinds of fields of computer vision such as image recognition[2] and object detection[3] , this article extracts the height of lettuce from RGB images by using networks of image recognition in deep learning without camera calibration or reference objects. On lettuce datasets collected in this article, by using a powerful neural network EfficientNet-B3, we can achieve a mean absolute error of 1.35 mm,some lettuce images are shown in Fig. 1. After fusing four networks including MobileNetV1, DenseNet-121, ResNext-50 and EfficientNet-B3, we will get a better mean absolute error of 1.22 mm. These promising results exhibit that neural networks of image recognition in deep learning can be effectively applied in predicting plant height.

Further author information: (Send correspondence to Minjuan Wang)
Minjuan Wang: E-mail: minjuan@cau.edu.cn, Telephone: +86 136 0111 7404
Xiyue Guo: E-mail: gxy@163.com, Telephone: +86 173 2006 9811



(a) (b) (c) (d)

Figure 1. Images of lettuce.

2. PROPOSED METHOD

To apply the image recognition networks into regression predicting continuous labels instead of discrete classification labels, this article adjusts the number of units in the last fully connected layer as one unit. The regression network is shown in Fig. 2. The dropout layer after the global average pooling layer is added to prevent over-fitting. The first fully connected layer with 11 units followed by a Softmax activation is also added to improve networks' performance. The last fully connected layer is the prediction layer of foretelling final labels. Because the plant height is greater than zero, a ReLU activation is combined in the prediction layer.

Figure 2. The architecture of regression networks.

This paper chooses four popular image recognition networks with large differences in architecture as the backbone network, namely MobileNetV1,[4] DenseNet,[5] ResNext[6] and EfficientNet. Using model ensembling technology,[7] these four models with significant differences in architecture are combined through simple average predictions

The mean absolute error (MAE), mean relative error (MRE), maximum absolute error (HAE) and maximum relative error (HRE) are used to compare the performance of models. These four measurements can be represented as:

$$MAE = \frac{1}{N}\sum_{i}|\hat{y}_i - y_i|, \tag{1}$$

$$MRE = \frac{1}{N}\sum_{i}\frac{|\hat{y}_i - y_i|}{y_i}, \tag{2}$$

$$HAE = \max_{i}\{|\hat{y}_i - y_i|\}, \tag{3}$$

$$HSE = \max_{i}\{\frac{|\hat{y}_i - y_i|}{y_i}\}. \tag{4}$$

3. RESULT AND DISCUSSION

Because we need to get an accurate height to monitor lettuce's growth status, we pay more attention to the size of MAE. Therefore, we train each model six times and report the lowest MAE and other corresponding indicators. The results of all models on the test dataset are shown in Table 1. EfficientNet-B3 has both the lowest mean absolute error (1.35 mm) and the lowest maximum absolute error (7.28 mm), which is in line with its powerful performance on ImageNet. However, DenseNet-121 is of the smallest mean relative error (4.47%), and ResNext-50 is of the smallest maximum relative error (15.97%). The single model can achieve the best outcome on one indicator but may not reach the best results on other indicators, which demonstrates each deep neural network has its concerns in lettuce height predictions. Therefore, we need to select suitable models according to specific demands.

We also fuse these four models by a simple average of predictions. After combining them, we can significantly improve MAE by 0.13 mm and MSE by 0.56%. And the fusion model has relatively small HAE and HSE among the four original models. Although the training set has only 273 images, we can get a mean absolute error close to 1 mm on the test set with plant height from 9 mm to 75 mm. If we use more neural networks with better accuracy on ImageNet, we may achieve more minor errors and better performance. Therefore, on the relatively small dataset used in this article, the regression networks of image recognition in deep learning can be efficiently used in predicting the height of lettuce from standard RGB images. Of course, we believe it can be suitable for other plants, not just lettuce.

Table 1. The results of all models in predicting the height of lettuce.

Model	MAE (mm)	MSE (%)	HAE (mm)	HSE (%)
MobileNetV1	1.45	4.59	7.6	16.83
DenseNet-121	1.36	4.47	9.33	18.49
ResNext-50	1.42	4.55	9.71	15.97
EfficientNet-B3	1.35	4.64	7.28	16.24
Average Fusion	1.22	3.91	7.57	16.07

4. CONCLUSION

This paper shows that it is very effective to use neural networks of image recognition in deep learning to predict lettuces height from RGB images in specific application scenarios where the camera remains fixed. We believe that using those to predict other plants? height is still valid. In this paper, we only test four neural networks predicting lettuce height. And we find that if a network has higher accuracy on ImageNet, it may perform better in height prediction. To verify this, we will try more image recognition networks and search for the best model on height predictions of lettuce. It is often hard to collect much data in the actual scene. Therefore, our lettuce dataset only contains samples, which may raise doubt about whether so small data can be used by deep learning. In fact, currently, no paper defines the minimum amount of data fed to deep learning.

REFERENCES

[1] Y. Li, H. Zhang, Y. Yang. A method for obtaining plant morphological and phenotypic parameters based on image processing technology[J/OL]. Journal of Forestry Engineering: 1-9[2020-06-30].

[2] Z. Zou, Z. Shi, Y. Guo and J. Ye. Object Detection in 20 Years: A Survey, 2019; arXiv:1905.05055.

[3] A. Howard, M. Zhu, B. Chen, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications. CoRR, abs/1704.04861, 2017.

[4] G. Huang, Z. Liu, L. Maaten. Densely Connected Convolutional Networks, 2017; arXiv: 1608.06993.

[5] S. Xie, R. Girshick, P. Dollar, Z. Tu, K. He. Aggregated Residual Transformations for Deep Neural Networks, 2017; arXiv: 1611.05431.

[6] T. He, Z. Zhang, H. Zhang, Z. Zhang, J. Xie, and M. Li. Bag of tricks for image classification with convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 558-567, 2019.

[7] H. Veen, L. Dat, A. Segnini. Kaggle Ensembling Guid [EB/OL]. https://mlwave.com/kaggle-ensembling-guide/, 2015.