



HAL
open science

CNN-Based Growth Prediction of Field Crops for Optimizing Food Supply Chain

Shunsuke Iitsuka, Nobutada Fujii, Daisuke Kokuryo, Toshiya Kaihara,
Shinichi Nakano

► **To cite this version:**

Shunsuke Iitsuka, Nobutada Fujii, Daisuke Kokuryo, Toshiya Kaihara, Shinichi Nakano. CNN-Based Growth Prediction of Field Crops for Optimizing Food Supply Chain. IFIP International Conference on Advances in Production Management Systems (APMS), Sep 2019, Austin, TX, United States. pp.148-154, 10.1007/978-3-030-30000-5_20 . hal-02419248

HAL Id: hal-02419248

<https://inria.hal.science/hal-02419248v1>

Submitted on 19 Dec 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

CNN-based Growth Prediction of Field Crops for Optimizing Food Supply Chain

Shunsuke Iitsuka¹, Nobutada Fujii¹, Daisuke Kokuryo¹,

Toshiya Kaihara¹, Shinichi Nakano²

¹ Graduate School of System Informatics, Kobe University
1-1 Rokkodai-cho, Kobe, Hyogo 657-8501, Japan
iitsuka@kaede.cs.kobe-u.ac.jp, nfujii@phoenix.kobe-u.ac.jp,
kokuryo@port.kobe-u.ac.jp, kaihra@kobe-u.ac.jp

² Hyogo Prefectural Technology Center for Agriculture, Forestry and Fisheries
560-1 Yagiyoginaka, Minamiawaji, Hyogo 656-0442, Japan
shinichi_nakano@pref.hyogo.lg.jp

Abstract. Along with the aging of the Japanese agricultural population in recent years, the nation's food self-sufficiency rate has declined. The vitality of agriculture has also declined. This study specifically examines the food supply chain to integrate farm production, manufacturing, and sales. For the food supply chain, it is necessary to ascertain the harvest time of field crops for establishing proper sales strategies. A method to predict harvest time from field crop images using convolutional neural networks (CNN) is proposed. Then the effectiveness of the proposed method is verified using computer experiments. Results show that the discrimination rate is high in the early stage of the growth, even if misclassification occurs. Labels close to the correct answer are predicted. However, the discrimination rate is not so high and the deviation from the correct answer becomes greater in later stages of growth.

Keywords: agriculture, convolutional neural networks, food supply chain.

1 Introduction

Along with the aging of Japanese agricultural population in recent years, the food self-sufficiency rate has also declined. The vitality of agriculture has also declined [1]. Revitalizing agriculture requires optimization of the food supply chain, which combines agriculture as a primary industry, manufacturing industry as a secondary industry, and retail industry as a tertiary industry [2]. The food supply chain has four main actors: farmer, farmer manager, vendor, and consumer. To fulfill their role of establishing proper sales strategies for sellers, managers must infer the harvest time of field crops. At present, managers visit farmers to infer harvest times. Such visits entail time costs and other costs. Therefore, a method to predict field crops' growth from these images is proposed to support managers.

In recent years, deep learning has prospered. Convolutional neural networks (CNN) are said to be effective for image recognition [3]. For this study, a CNN is used to address difficulties of growth inference for head lettuce.

An earlier report [4] described studies of crop classification, disease detection, crop counting, etc. Nevertheless, few reports have described studies that have predicted growth. Furthermore, no report of the relevant literature has described a study of head lettuce. This study specifically examines head lettuce growth prediction.

2 Proposed method

An overview of the proposed method is portrayed in Fig. 1. The lettuce image is input with all pixels for each channel of RGB. Then the leaf age is output, which represents the number of leaves.

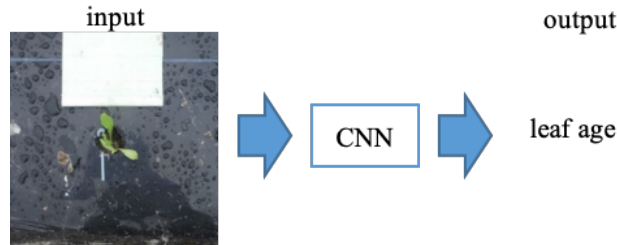


Fig. 1. Overview of the proposed method.

The proposed method has two stages: the image processing stage before forming input images and the image discrimination stage in CNN. In the image processing stage, all image ratios are adjusted, the target ranges of all images are cut out. The image size is changed. Furthermore, the image is classified into G group ($G = 21$ in this research) according to the leaf age. Subsequently, one image is selected randomly in each leaf age group and is used as a testing image. The remaining images are used as training images. To improve the discrimination accuracy, training images in each group are increased to N ($N = 100$ for this study). Subsequently, the image is input to CNN for discrimination.

The breed of target lettuce is “elegant,” cultivated in Minamiawaji city during 2015. The photograph frequency is once weekly.

2.1 Image processing

Adjusting image ratio and cutting out. The lettuce’s image is adjusted and cut to $30 \text{ cm} \times 26 \text{ cm}$ size to fit the gap separating two lettuces, as presented in Fig. 2.

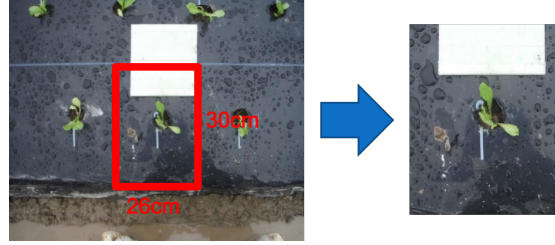


Fig. 2. Trimming to 30 cm × 26 cm.

Image size change. The image size is changed to input CNN. The 30 cm × 26 image size is changed to 75 pixels × 65 pixels.

Image classification. A label is added to the image. In the previous research, the number of weeks after planting was given as a label [5]. At this time, labels should be grouped according to the leaf age. The number of groups is 21 from four leaf ages to 30 leaf ages. Table 1 shows the number of images for each leaf age.

Table 1. Number of images for each leaf age.

Leaf age	4	5	6	7	8	9	10	11	12	13	
No. images	48	46	35	13	28	19	24	28	19	32	
14	15	16	17	18	21	22	23	24	27	30	sum
22	10	6	5	4	2	2	2	3	2	2	352

Data augmentation. In CNN, a large dataset must be used to obtain high recognition accuracy [3]. To get a large dataset, data augmentation that changes a few images to increase the images is said to be effective [6]. The number of training images is increased until the number of images in each leaf age group reaches 100 by data augmentation. To increase the number, we reverse the image, and change the brightness and contrast of the image until the number of images reaches 100.

2.2 Image discrimination

The CNN used in the proposed method is depicted in Fig. 3. As depicted in Fig. 3, there are two convolutional layers, two pooling layers, and three fully connected layers.

The convolutional layer performs an operation to convolve the filter on the input. The input sized $H \times W$ in each channel is convolved with $L \times L$ sized filters. The output for one channel is calculated such as Equation (1) if the input is set as $y_{ijk}((i, j, k) \in [1, H] \times [1, W] \times [1, N])$ and the filter as $h_{ijk}((i, j, k) \in [1, L] \times [1, L] \times [1, N])$.

In this study, the input size is set to $75 \times 65 \times 3$. The first convolutional filter size is $5 \times 5 \times 40$. The second convolutional filter size is $5 \times 5 \times 50$.

$$x_{ij} = \sum_{k=1}^N \left[\sum_{p=1}^L \sum_{q=1}^L y_{i+p,j+q,k} h_{pqk} \right] + b_k \quad (1)$$

Pooling is defined as discarding extra information for recognition from extracted features and converting it into a new representation that retains information necessary for recognition. Unit(i, j) of the pooling layer collects the output y_{pq} of the internal unit(p, q) $\in P_{ij}$ for a small area P_{ij} of a part of the input layer. In this study, maximum pooling is used as expressed by Equation (2). In this study, the size of P_{ij} is set to 2×2 in each pooling layer.

$$\tilde{y}_{ijk} = \max_{(p,q) \in P_{ij}} y_{pqk} \quad (2)$$

In addition, the ReLU function expressed by Equation (3) is used after the output of the convolutional layers and the fully connected layers. In Equation (3), x_j represents the input.

$$f(x_j) = \max(x_j, 0) \quad (3)$$

Furthermore, in the classification problem such as the one examined for this study, the final output is assigned the same number of units as the classification group. The output of these units is not the activation function but the softmax function expressed by Equations (4) and (5). At the time of discrimination, the index $j = \underset{j}{\operatorname{argmax}} p_j$, in which p_j takes the maximum value, is output as the expected group.

$$p_j = \frac{e^{x_j}}{\sum_{k=1}^n e^{x_k}} \quad (4)$$

$$\sum_{j=1}^n p_j = 1 \quad (5)$$

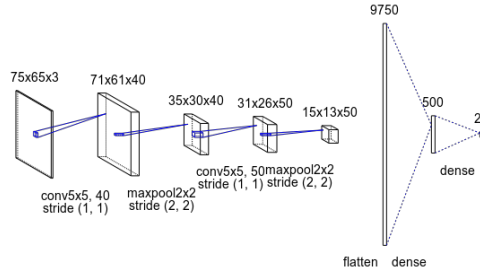


Fig. 3. CNN used in the proposed method.

Adam [7] is used as a learning algorithm. By this learning method, batch processing is used to reduce the processing time for each image. The algorithm is not used to input images one by one, but to input and process plural images. At this time, the number of input images is called the batch size. In addition, one learning term is found from the input of the image for the batch size until the update of the weight is completed. Furthermore, the number of times during which all training images are input and the weight update is completed is called the epoch.

3 Computational experiment

To confirm the effectiveness of the proposed method, computer experiments were conducted. The leaf age of the 21 groups above is discriminated using the proposed method. One image randomly extracted from each group is used as the testing image. The remaining images are used as the training image. Learning is performed using the training image. The training images and the test images are discriminated.

3.1 Experiment conditions

The experiments were performed under the following conditions.

- Number of groups (G): 21
- Number of training images: 2100
- Number of testing images: 21
- Batch size: 32
- Number of epochs: 20
- Number of trials: 5
- Learning rates: 0.001

3.2 Result

First, results for the discrimination rate are shown. Table 2 presents the mean and standard deviation of the discrimination rates for the five trials, and shows the results for the 20th epoch.

Table 2. Discrimination result (%).

	avg.	S.D.
training	99.8	0.003
testing	30.5	0.06

From Table 2, the discrimination rate at the end of the 20th epoch was 99.8% for the training image and 30.5% for the testing image. It is apparent that the testing image does not improve the discrimination rate even as the learning of the training image progresses.

Table 3 shows the discrimination rate for each leaf age in the testing images.

Table 3. Discrimination rate for each leaf age (%).

leaf age	4	5	6	7	8	9	10	11	12	13	
avg.	100	60	80	40	60	20	40	0	20	80	
S.D.	0	0.49	0.4	0.49	0.49	0.4	0.49	0	0.4	0.4	
	14	15	16	17	18	21	22	23	24	27	30
60	20	20	0	0	0	0	0	0	40	0	0
0.49	0.4	0.4	0	0	0	0	0	0	0.49	0	0

Table 3 shows that the discrimination rate is high when the leaf age is less and not so high when the leaf age is greater. When the leaf age is less, i.e., less than leaf age 14, the difference from the correct answer is within 2, even if it is misclassified. By contrast, when the leaf age is greater, i.e., leaf age 15 or more, the label which is far from the correct answer is predicted, when it is misclassified. It is considered that the original number of images affects the results obtained for each leaf age. As Table 1 shows, it is apparent that the images of leaf age 15 or greater are fewer than the images of leaf age 14 or less. Learning is insufficiently achieved because the original images are few.

The anomaly of the discrimination rate for leaf age 11 and 24 is explained next. The reason for the low discrimination rate even though the number of leaf age 11 images is 28 might be an error in labeling. As described above, the image of leaf age 11 is also predicted within 2 before and after the correct label. The discrimination rate of leaf age 24 is higher than that of the front and rear leaves because the number of images is greater than that of the front or rear leaves: CNN can learn the training images on a relative basis.

4 Conclusion

A method to predict leaf age has been proposed. Predicting leaf age engenders knowledge of the harvest time to optimize the food supply chain. The results obtained using the computer experiments showed a high discrimination rate for less leaf age. However, for greater leaf age, the discrimination rate is not so high, partly because the number of original images is small.

Future studies will include consideration of the increase in the number of images and application to the food supply chain.

References

1. Ministry of Agriculture, Forestry and Fisheries Homepage, http://www.maff.go.jp/j/wpaper/w_maff/h21_h/trend/part1/chap4/c4_10.html, last accessed 2019/3/27.
2. Policy Research Institute, Ministry of Agriculture, Forestry and Fisheries: Logic and development direction of sixth industrialization – Construction of value chain and promotion of innovation –. Research paper (2015).
3. A. Krizhevsky, I. Sutskever and G.E. Hinton: Imagenet classification with deep convolutional neural networks, *Advances in Neural Information Processing Systems 25* (2012), pp. 1097-1105.
4. A. Kamilaris, F. X. Prenafeta-boldu: Deep learning in agriculture: A survey, *Computers and Electronics in Agriculture 147* (2018), pp.70-90.
5. S. Iitsuka, N. Fujii, D. Kokuryo, T. Kaihara, S. Nakano, and R. Santo: Growth prediction of farm products using convolutional neural networks, *The Japan Society of Mechanical Engineers* (2018).
6. J. Wang and L. Perez: Effectiveness of Data Augmentation in Image Classification using Deep Learning, *arXiv:1712.04621* (2017).
7. D. Kingma and J. Ba.: Adam, A method for stochastic optimization, *arXiv:1412.6980* (2014).