

The Pest and Disease Identification in the Growth of Sweet Peppers Using Faster R-CNN and Mask R-CNN

Tu-Liang Lin, Hong-Yi Chang, Kai-Hong Chen

Department of Management Information Systems, National Chiayi University, Taiwan
tuliang@mail.ncyu.edu.tw, hychang@mail.ncyu.edu.tw, kaihong@mis.ncyu.edu.tw

Abstract

Early-stage control of plant pests is a crucial topic in modern agriculture. If plant pests and diseases can be identified as early as possible, farmers can prevent their occurrence in advance and avoid economic losses. Early identification of pests and diseases can minimize the cost of pesticides. However, correct identification of pests and diseases requires knowledge and corresponding expertise, and this knowledge accumulation requires time. Therefore, in this study, Faster region-convoluted neural networks (R-CNNs) and Mask R-CNNs were used to develop a knowledge-based system that can automatically identify plant pests and diseases. The Faster R-CNN exhibited a regional recognition accuracy of 89%, and the Mask R-CNN exhibited an area recognition accuracy of 81%. A pest and disease identification system was developed in this study. The developed system can be further improved by adopting the proposed reinforcement model construction flow.

Keywords: Pest and disease identification, deep learning, Convolutional neural networks, Image recognition, Object detection

1 Introduction

People consume food daily. The development of human civilization and agriculture has always been inseparable; the relationship has become more pronounced with the evolution of the times and the advancement of science and technology. Agricultural technology is constantly improving. To ensure agricultural products of a country stand out in the international market, use of technology is necessary in this age of international competition. The automation of all stages (planting, growing, and harvesting of crops) enables agriculture to flourish.

Small farmers account for approximately 85% of the total farmers of the world. Small farmers often encounter extreme weather, fall of market prices, and occurrence of pests and diseases in crops. Therefore, technology is urgently required to overcome these adversities [1]. The adoption of science and technology

in agriculture can reduce the problem of lack of work. The purpose of this study is to develop an early-stage detection system that can detect whether pests or diseases have infected colored pepper crops. Small farmers can then spray appropriate pesticides at an early stage to avoid the loss of colored pepper harvest.



Figure 1. Common pests and diseases of sweet peppers (from PlantVillage [2])

Sweet peppers grow best at temperatures between 18 and 30°C and on soil with pH between 6 and 7. Moist soil is not conducive for the growth of peppers. Therefore, they should be planted in well-drained soil or on elevated beds and in areas that receive sunlight most of the day. The weather in subtropical zones is rainy, hot, and humid throughout the year and suitable for the growth of various sweet peppers. However, such hot, humid weather is also conducive to the breeding of pests and diseases [3]. Common pests and diseases of sweet peppers include anthracnose, aphids, bacterial ulcers, bacterial wilt, and beet armyworm, as shown in Figure 1. In the prevention and control of

pests and diseases, correct diagnoses of various types of pests and diseases can directly affect prevention and treatment. However, identification and diagnosis of the many pests and diseases that occur is complex. Accurate diagnosis of pests and diseases is necessary and agricultural technicians with relevant professional knowledge are required to achieve correct diagnosis. Acquiring professional knowledge requires a certain amount of learning time and fundamental knowledge of relevant practices.

Typically, long-term diagnoses of pests and diseases are performed by experienced pest and disease research experts and scholars. Furthermore, they ensure pest and disease prevention steps are implemented successfully. Valuable experience is required for the process, which can be obtained by a large amount of plant cultivation and pest and disease knowledge. When a plant is infected by pests and diseases, a series of symptoms appear initially. These symptoms are related to each other. Several pests and diseases can have the same early symptoms, thus increasing the difficulty in the diagnosis of pests and diseases. The degree of damage to pests and diseases varies depending on the characteristics of the plants and the method of plant cultivation and management.

Because of the advancement in technology and improvement in computer computing power, many artificial intelligence or machine learning algorithms such as K-means algorithm for sugarcane disease detection [4] and machine vision for weed detection [5], are applied in cultivation. Among all machine learning algorithms, deep learning is a fast-developing machine learning technique. Google's AlphaGo, developed using deep learning technology, has defeated the Go champion. In handwriting recognition, deep learning has performed better than humans [6]. Thus, deep learning systems can recognize targets accurately.

Therefore, in this study, deep learning-based image detection technologies were used. Various pest and disease knowledge and image data were integrated with image detection technologies to develop an accurate identification model for plant disease diagnosis. The proposed method can provide instant information and is designed to assist farmers in controlling pests and diseases, and thus reduce the extent of harm and losses.

A pepper pest and disease identification expert system based on deep learning was developed to achieve early detection and control of pests and diseases. Planar imaging data and deep learning-based object detection methods, such as those based on static images, recorded videos or online streams, were integrated in the system. These methods can detect pest and disease from various input sources.

2 Literature Review

2.1 Convolutional Neural Networks (CNNs)

In recent years, deep learning algorithms have achieved considerable progress in image recognition because of automatic feature selection [7-10]. Manual design features, such as features from scale-invariant feature transform [11] and speeded up robust features [12], are used in traditional image recognition algorithms. After the features have been captured, the captured features are input into the machine learning algorithms to perform image recognition. The accuracy of the traditional image recognition algorithm heavily depends on the quality of the extracted features. However, feature extraction is complex and cumbersome. For multiple problems or applications, it is usually necessary to redesign the features. This is called feature engineering. Feature engineering requires expert knowledge and many algorithms for identifying the relevant targets (in this case, pests and diseases). Traditional machine learning techniques and image features are required to be manually redesigned and reselected, respectively, for various pests or crop images. In recent years, because of the rise of deep learning, CNN can be used to automatically perform feature extraction without artificial feature design.

CNN feature extracted algorithms have performed better traditionally designed feature extraction algorithms in many ImageNet image recognition competitions. Therefore, applying CNNs to image detection of crop pests and diseases results in better accuracy than applying traditional algorithms.

CNN is a major component of deep learning neural networks. It has been proven that CNN-algorithms can perform better than humans in some image recognition applications [9]. It is desired that machine learning models can correctly predict labels regardless of image translation, reduction, rotation, or deformation. Convolution is performed in a CNN to filter image blocks and calculate the degree of similarity, and pooling is used to obtain the maximum values from the blocks. Pooling is an image compression method, and after pooling, the total number of pixels can be reduced. CNNs can be divided into two layers, namely the convolution layer and the pooling layer. Neurons are used in each layer to handle weights and outputs. Figure 2 depicts a schematic of a CNN [8].

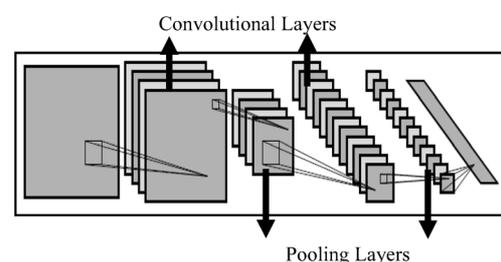


Figure 2. Schematic illustration of a CNN [8]

CNN methods can be considered as machine learning methods developed from neural networks. The original concept and basic structure of a neural network mimics the learning mechanism of neurons in neurobiology, which learns patterns through a series of simple, fast calculations [13]. Similar to real neurons, the role of neurons in the artificial neural network is to sum the input values according to the weights and subsequently transform the sum to output a value according to the trigger function, such as the sigmoid or hyperbolic tangent function. The calculation of neural networks can be easily processed simultaneously, and it can learn according to the given samples. Therefore, data analysis is not limited by the premise of selecting the samples [14]. Thus, many artificial neural network models have been successfully developed. For example, a combination of deep learning network and Support Vector Machine has been applied to the recognition of hyperspectral images [15]. CNN also can be used to learn image representations and a vision learning framework to generate compact binary hash codes for quick vision search using convolution neural networks has been proposed [16].

2.2 Object Detection Using CNNs

Although CNNs can perform image recognition and output the labels, a CNN alone cannot indicate the location of the identified objects. More than one type of pest and disease can occur in the crop images. Therefore, it is necessary to use object detection technology to mark the position and size of multiple objects in the image and perform multicategory identification for the cropped images. The most intuitive method to perform this task is to use sliding windows that use fixed-size frames to sweep through the entire image one-by-one; the framed images are presented to CNN to determine the categories. Because the size of the object is unpredictable, it is necessary to use different sizes of frames for detection. However, the sliding window is a brute force method that requires scanning whole images several times. Because of the sliding window concept, the whole process consumes large amounts of computing resources and is slow. Therefore, R-CNN was proposed [17]. Instead of sweeping through the whole image, approximately 2000 possible areas are prescreened and subsequently these possible areas were predicted individually.

2.3 Faster R-CNN

Faster R-CNN is an improved version of R-CNN. Instead of prescreening region proposals, region proposals are selected in Faster R-CNN directly from the feature map calculated from the CNN. Faster R-CNN uses a region proposal network (RPN) that is a CNN. The input of the RPN is the feature map from the first CNN and the output of the RPN is a bounding box and the probabilities that the bounding box

contains an object. From the RPN, the most likely bounding boxes can be obtained. Although these bounding boxes are not highly accurate, they can be processed with the region of interest (RoI) pooling. After RoI pooling, each region can be swiftly classified to determine the most accurate bounding box coordinates.

Although Faster R-CNN achieves good accuracy, the inherited two stage design makes the detection speed slow. A different design approach SSD (Single Shot multibox Detector), a one stage method which does not re-sample the bonding box proposals, was introduced and SSD has significant improvement in speed [18].

2.4 Mask R-CNN

Mask R-CNN is the pixel-wise object detection method proposed by He et al. in 2017 [19]. This method won the COCO 2016 Challenge. Mask R-CNN extends the network framework of Faster R-CNN, adding a mask branch to detect the category of each box in the images. In a fully convolution network (FCN), full convolution is added to perform segmentation. Therefore, in Mask R-CNNs, the object detection task is transformed into classification, regression, and segmentation. Therefore, Mask R-CNN can be understood as a combination of Faster R-CNN and FCN. Mask R-CNN can be used to identify the pixel-wise location of the target, which belongs to the instance segmentation model category. The instance segmentation refers to classifying each object in the image, segmenting each object, and marking the category to which each box belongs, for example, identifying a single vehicle, creature.

2.5 Pest and Disease Identification Using Deep Learning Methods

Previously, many traditional machine learning methods, such as Support Vector Machines (SVM) [20], k Means [21] and Decision Trees [22], have been applied for pest and disease identification. However, due to the rapid development of computer vision and object recognition using deep learning approach, automatic pest identification using deep learning methods has become an active research topic in recent years. A convolutional neural network (CNN) based approach is proposed to classify plant diseases from images, especially the leaf images, but the research did not work on pest identification [23]. In this work, deep learning-based object detection approaches, such as Faster R-CNN and Mask R-CNN, are adopted to construct the pest and disease identification system.

3 Problem Formulation

The crop pest and disease identification problem can be divided into the training and testing phases. In the training phase, given n sets of image data, area of

infection, and infection types as the training data, the training data can be formulated as follows:

$$T = \{Img_1, \{(bb_{11}, l_{11}), (bb_{12}, l_{12}), \dots, (bb_{1m_1}, l_{1m_1})\}\}, \dots, \{Img_2, \{(bb_{21}, l_{21}), (bb_{22}, l_{22}), \dots, (bb_{2m_2}, l_{2m_2})\}\}, \dots, \{Img_i, \{(bb_{i1}, l_{i1}), (bb_{i2}, l_{i2}), \dots, (bb_{im_i}, l_{im_i})\}\}, \dots, \{Img_n, \{(bb_{n1}, l_{n1}), (bb_{n2}, l_{n2}), \dots, (bb_{nm_n}, l_{nm_n})\}\} \quad (1)$$

where Img_i is the i th image, bb_{im_i} is the m_i th infection areas, which contains the coordinates of the infection areas and l_{im_i} is the infection type of the m_i th infection area in the i th image. The final training model is the input of an image and output the predicted infection areas and infection types. The mathematical formula can be represented as follows:

$$\{(bb_{j1}, l_{j1}), (bb_{j2}, l_{j2}), \dots, (bb_{jm_j}, l_{jm_j})\} = F(Img_j) \quad (2)$$

Given an image Img_j as input, the trained model $F(Img_j)$ should output several infection areas that indicate the infection area and the types of the infection. The verification system determines whether the predicted infection area and the pest and disease type matches the given infection area and infection types.

When detecting the pest and disease in the identification system, the infection area should be marked. A mathematical method to compare the labeled ground-truth infection area and the predicted infection area is the intersection over union (IoU) [24], as shown in Figure 3. The IoU can be used as a foundation to measure the accuracy of the detected infection area of a corresponding infection crop in a particular image. As long as an infection area is derived from the output, IoU can be calculated using the following formula (4).

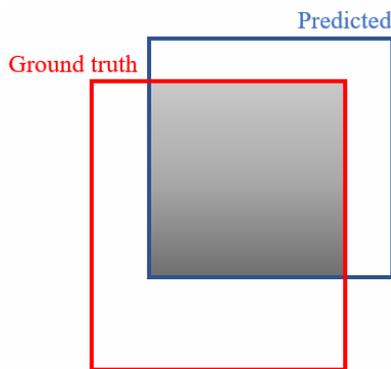


Figure 3. Intersection of Union

$$IOU(A, B) = \left| \frac{A \cap B}{A \cup B} \right| \quad (4)$$

where A represents the ground-truth infection area, and B represents the predicted infection area. Therefore, to

measure IoU of any size and shape, the following are required.

- (1) ground truth of the infection area
- (2) predicted results of the infection area

To simplify the computation of IoU, rectangular boxes or bounding boxes instead of irregular shapes are used to perform the IoU calculation for comparison.

4 Problem Solution

To identify plant diseases and pests, image detection technology is adopted. Although the CNN can perform image detection, given the input image, CNN can also determine the class of pests and diseases. However, only category information can be identified in the CNN identification process.

However, only achieving plant pest and disease category identification is not sufficient. It is also necessary to mark the position and size of multiple infected area in the image. Thus, the object detection method was adopted for multicategory and area identification.

Many multiple object detection methods, such as the R-CNN [17], Fast R-CNN [25], Faster R-CNN [26], and Mask R-CNN [19], have been proposed. However, irrespective of the R-CNN or Fast R-CNN, region proposal preselection through selective search [27] is still necessary, and selective searching is a slow process. Therefore, in this study, we used the Faster R-CNN and Mask R-CNN to construct the core of the pest and disease identification system. Faster R-CNN skips region proposal selection and selects region proposals directly from feature maps of CNNs. Faster R-CNN adopts an RPN that is a specific CNN used for a specialized purpose. The feature map from the first CNN is input into the RPN, and the rectangular infection areas and the probabilities of the infections can be obtained at the final stage. Therefore, the most likely infection areas and infection types can be identified. Mask R-CNNs extend the network framework of Faster R-CNNs, adding a mask branch to detect the infection areas in the images. The Mask R-CNN is the pixel-wise object detection method, which can be applied to pest and disease identification. The pixel-wise infection areas and infection types can be obtained from Mask R-CNNs.

A pest identification and pesticide spraying system is illustrated in Figure 4.



Figure 4. Pest identification and pesticide spraying system

Stage 1: Use the wireless camera to capture the video stream from the experimental field.

Stage 2: Use the real-time streaming protocol to obtain real-time images and perform real-time pest and disease identifications using the deep learning network models.

Stage 3: Use the deep learning network model to determine the infection type and area. Select the sprayed pesticide and output the coordinates to the robotic arm and move the arm to the infection area. Spray pesticides through the nozzle.

In this study, we used mean average precision (mAP) to compare various object detection methods. The operation of mAP involves average-precision (AP) calculation for all categories and then obtaining the average. Thus, the AP is the average precision of one category and mAP is the mean average precision of all categories. The AP can be obtained from the following formula (3).

$$AP = \frac{1}{11} \sum_{r \in \{0,0.1,\dots,1\}} P_{interp}(r) \tag{3}$$

where $P_{interp}(r) = \max_{\tilde{r} \geq r} p(\tilde{r})$

Typically, in mAP calculation, the precision recall curve is plotted first. In formula (3), r is recall and P is precision. The calculation determines the maximum precision when recall is higher than a certain threshold. In this study, we set the threshold to 0.1 for a superior result. Normally, the threshold is set to 0.5. Because the threshold is set to 0.1 and the increment is 0.1, eleven recall values occur from 0 to 1. Therefore, the summation is divided by 11 in formula (3).

Figure 5 illustrates the design of the pest and disease identification system. The system contains two infection detection algorithms, Faster R-CNN and Mask R-CNN, and three network models, Inception v2, ResNet-50 and ResNet-101. Six combinations can be used to detect the pest and disease infections. Given original images and labeled infection areas, the models are trained to identify the pests and diseases. When images, videos and online streams are provided, the trained model can output the infection areas and types.

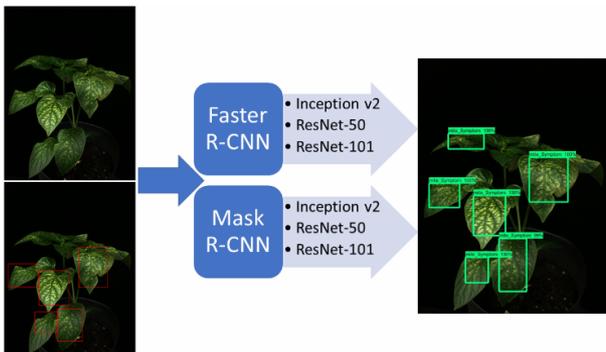


Figure 5. Pest and disease identification system

5 Experimental Results

5.1 Sweet Pepper Disease and Pest Dataset

Our data set includes a total of 1239 images of sweet pepper strains cultivated from greenhouses in Chiayi City, Taiwan. The categories and the number of annotated samples used in this research are listed in Table 1.

Table 1. List of categories in our sweet peppers diseases and pests dataset and annotated samples

Class	Number of Images	Number of Annotated Samples ^a
Whitefly Symptom	118	349
Mite Symptom	382	1038
Thrips	739	1243
Total	123	2630

The symptoms of the pests and diseases of sweet pepper are depicted in Figure 6. Three pests, namely spider mite, whitefly, and thrips were vaccinated to nine yellow pepper and red pepper plants. Three types of symptoms are present in this experimental data set: mite_Symptom, whitefly_Symptom, and thrips. Data were classified separately and randomly divided into sets containing 70%, 15%, and 15% of all data for training, verification, and test data.

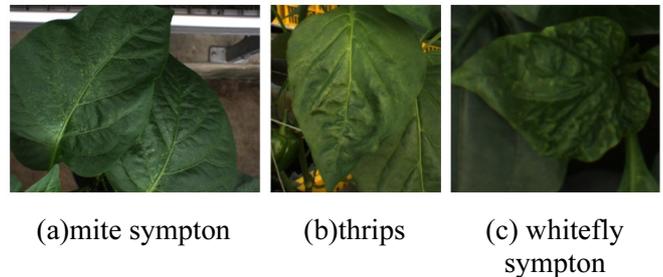


Figure 6. Symptoms of diseases and pests that affect sweet pepper plants

The proposed models (Faster R-CNN and Mask R-CNN) were trained and tested end-to-end using NVidia GeForce GTX 2080ti GPUs on an Intel Core I5 8500 3.0 GHz Processor personal computer. Figure 7 depicts the Faster R-CNN loss function (Losses_TotalLoss) of the three network models, Inception v2, ResNet_50 and ResNet_101 during the training process. Figure 7 indicates that the longer the model training is, the lower the total loss is. Thus, the training is effective.

When training neural networks, repeated learning and memorizing is necessary to ensure the final trained model can detect the infections in the crops when images, videos, and online streams are inputted to the models. Figure 8 displays the final detection results and the marked area in the plant leave image. In Figure 8, the upper left image is the original image, lower left

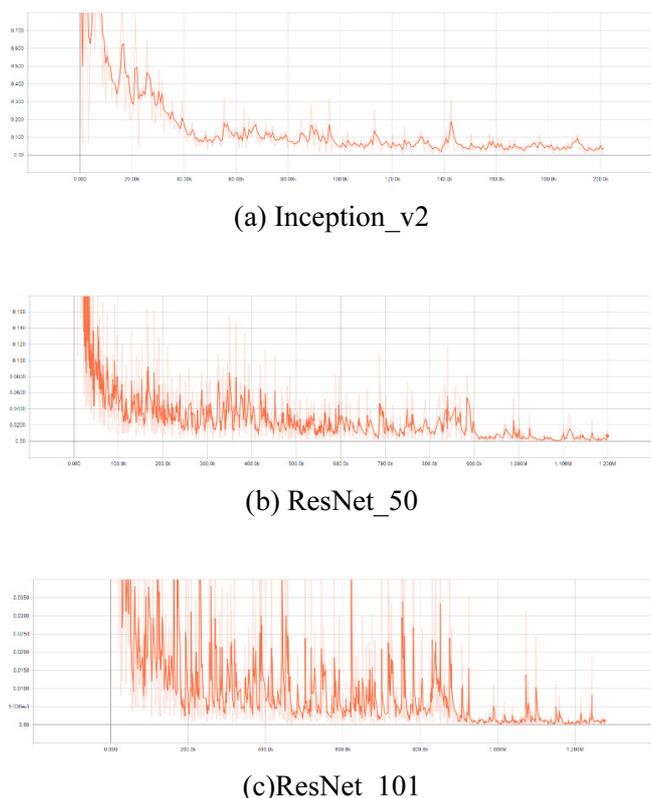


Figure 7. Loss function of three network models

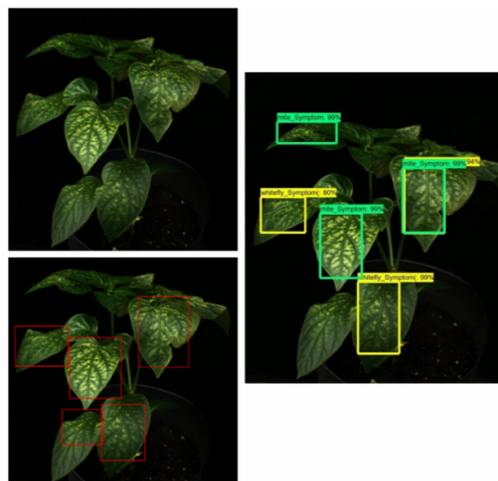


Figure 8. Detection results of multiple symptoms using Faster R-CNN with ResNet_101 network model (upper left: original image, lower left: human labeled image, right: machine prediction)

image contains the infection areas labeled by experts and the right image is the detection result. The mite infections are labeled in green boxes and whitefly infections are labeled in yellow boxes.

Table 2 lists the results of mAP calculations for Faster R-CNN and Mask R-CNN. The network models used in this project are Inception v2, ResNet50, ResNet101.

Table 2. The mAP results of Faster R-CNN and Mask R-CNN

Class/Feature Extractor	Faster R-CNN			Mask R-CNN		
	Inception v2 ^a	ResNet-50 ^a	ResNet-101 ^a	Inception v2 ^a	ResNet-50 ^a	ResNet-101 ^b
mite_Symptom	0.66	0.66	0.58	0.12	0.15	0.34
Whitefly_Symptom	0.65	0.54	0.34	0.16	0.09	0.58
Thrips	0.38	0.35	0.33	0.14	0.15	0.38
Total mean AP	0.56	0.52	0.42	0.14	0.13	0.43

Note. ^a TensorFlow model; ^b Keras model.

With the exception for the results of Mask R-CNN with the ResNet-101 network model, which were obtained using the Keras framework, the rest of the results were obtained using the Tensorflow framework. The Faster R-CNN with the Inception v2 network model exhibited the best performance. Overall, the test can be improved.

As listed in Table 2, the mAP values of the thrips are relatively small when compared with other categories. This is because the ground-truth region in the image is relatively large, and the predicted range is relatively small. Although the mAP values of the thrips were relatively low, the categories were still predicted accurately. The average mAP was approximately 56% for all categories when the Faster R-CNN was used with the Inception v2 network model.

The application scenario of this study was to use a robotic arm to shoot and spray pesticides. When identifying the infection areas of pests and diseases, the pesticide was sprayed in a large area around the

infection, as shown in Figure 9. Because pesticide was sprayed in the region around the infection, we could expand the identification area to the spray region and calculate accuracy. We named this type of accuracy as the identification accuracy of the spray region. Table 3 displays the result of the identification accuracy of the spray region. In terms of identification accuracy of the spray region, the Faster R-CNN using ResNet-101 exhibited the highest accuracy rate of 89%, followed by the Faster R-CNN using Inception v2 and Mask R-CNN using ResNet-101.



Figure 9. Infection area (Green box) and the spray area (Red box)

Table 3. Identification accuracy of the spray region

Class/Feature Extractor	Faster R-CNN			Mask R-CNN		
	Inception v2 ^a	ResNet-50 ^a	ResNet-101 ^a	Inception v2 ^a	ResNet-50 ^a	ResNet-101 ^b
mite_Symptom	95%	95%	95%	62%	62%	83%
Whitefly_Symptom	75%	64%	93%	61%	50%	82%
Thrips	80%	70%	80%	39%	42%	79%
Total mean AP	83%	76%	89%	51%	54%	81%

Note. ^a TensorFlow model; ^b Keras model.

Figure 10 displays the detection result using the Faster R-CNN with the ResNet-50 network model under sunlight. A thrips infection area was correctly identified. Therefore, the system can be applied to different luminosity backgrounds.

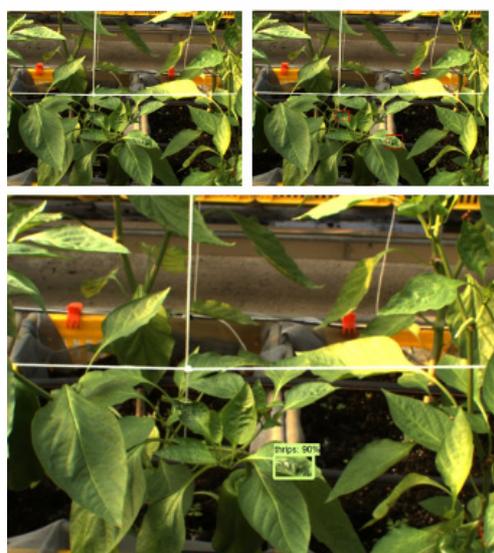


Figure 10. Thrips symptom detection results under sunlight using the Faster R-CNN with ResNet-50 network model (upper left: original image, upper right: human labeled image, below: machine prediction)

Figure 11 depicts the detection result using the Faster R-CNN with the ResNet-101 network model under sunlight. Two infection areas with two infection types, namely mite and whitefly, were marked in the original image. Multiple infection areas were correctly identified. Surprisingly, another suspicious infection was identified by the identification system without human intervention. This finding can further improve the precision of the pest and disease identification system. Therefore, based on this result a reinforcement model construction flow was designed as depicted in Figure 12. The data set was roughly divided into training data and testing data. Both training and testing data were labeled images with infection areas and infection types. The training data were then used to train the model. The trained model performed the tests on the testing data. The testing results were provided to experts for further verification. The controversial cases in the testing results were re-examined and verified by the experts and the modified cases were added to the



Figure 11. Detection results of mite symptom and whitefly symptom using the Faster R-CNN with the ResNet-101 network model (upper left: original image, upper right: human labeled image, below: machine prediction)

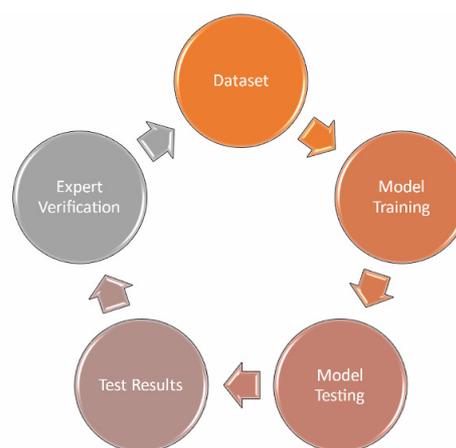


Figure 12. Inspired reinforcement model construction flow for the pest and disease identification system

data set. In general, numerous labeling errors can be corrected and the precision of a typical model can be increased using several reinforcement model construction cycles. In this study, the disagreements between machine and human answers, such as the results depicted in Figure 12, were minimized using the proposed reinforcement model construction flow.

Figure 13 illustrates the detection results using the Mask R-CNN with the Inception v2 network model. Two areas with thrips infection were identified. This result is controversial. The identification system also identified a suspected infection area. Accuracy and precision were improved using the proposed reinforcement model. Mask R-CNNs are a pixel-wise object detection algorithm. Therefore, the identified infection area is irregular in shape, as depicted in Figure 13. For comparison, a rectangular bounding box was used to perform IoU calculation in the Mask R-CNN. The results showed that Faster R-CNN has higher precision than the Mask R-CNN irrespective of mAP or identification accuracy of the spray region when adopting the rectangular bounding box. However, the Mask R-CNN has the advantage of mapping the irregular shape of the infection area, which is useful for precision pesticide spraying if the pesticide nozzle can be controlled precisely.

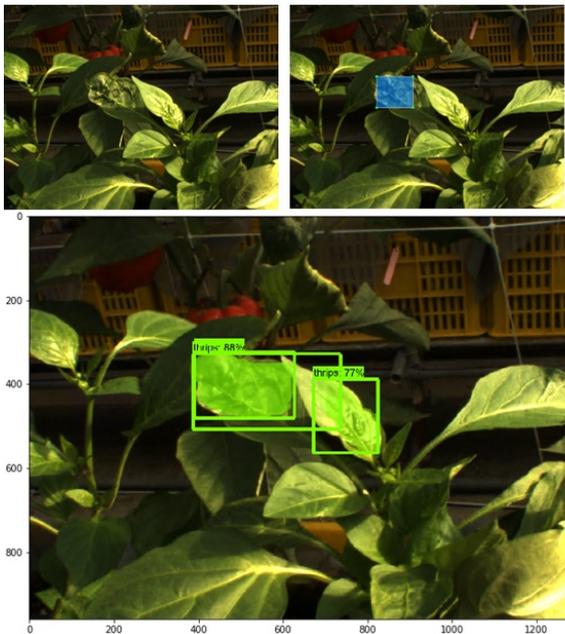


Figure 13. Thrip symptom detection results using the Mask R-CNN with the Inception v2 network model (upper left: original image, upper right: human labeled image, below: machine prediction)

Figure 14 depicts the detection result using the Mask R-CNN with the Inception v2 network model. Multiple mite infection areas were identified. Because the Mask R-CNN is a pixel-wise object detection algorithm, the identified infection areas were irregularly shaped, as shown in the Figure 14. A rectangular bounding box was used to perform IoU calculation in the Mask R-CNN and the results indicated that Faster R-CNN has higher precision than the Mask R-CNN. This phenomenon can be explained by thorough examination of Figure 14. Figure 14 indicates that many ground-truth bounding boxes were combined into one large bounding box in the final prediction. Therefore, in this study, we determined that the

combination of many irregular infection areas may have impaired accuracy in the Mask R-CNN.

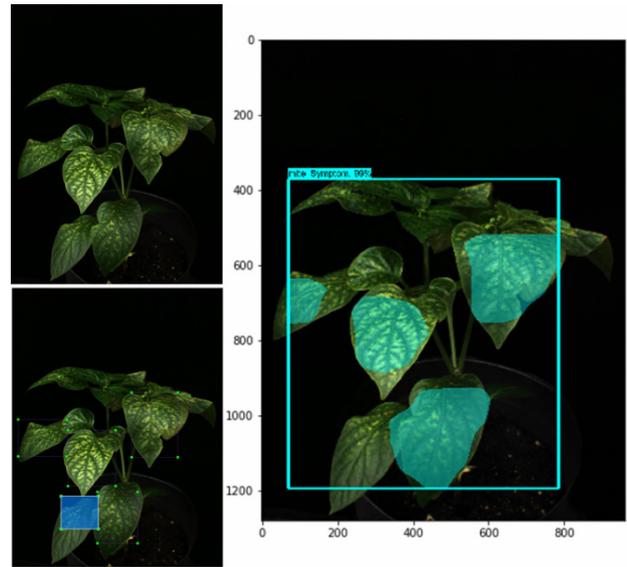


Figure 14. Detection results of mite symptoms using the Mask R-CNN with the Inception v2 network model (upper left: original image, lower left: human labeled image, right: machine prediction)

6 Conclusion

To prevent the spread of the pests and diseases, we developed an automatic identification system for early-stage detection of pests and diseases of sweet peppers without professional knowledge and expertise. This can assist farmers and reduce the production costs of pesticides and achieve early prevention of pests and diseases, thus increasing harvests. The Faster R-CNN and Mask R-CNN were used to develop a knowledge base system that can automatically identify plant pests and diseases. The Faster R-CNN has a regional recognition accuracy of 89%, and the Mask R-CNN has an area recognition accuracy of 81%. We also discovered that the pest and disease identification system can be further improved by using the proposed reinforcement model construction flow. The controversial cases obtained from the testing results can be further re-examined and verified by the experts in the reinforcement cycle and the final modified cases can be further added to the data set. After several reinforcement model construction cycles, many image labeling errors can be corrected and model accuracy can be further improved. Differences in the answers given by machines and humans can be resolved in the proposed reinforcement model construction flow.

Currently, the training data set has 1239 images. Although the detection results are satisfactory, if the number of images can be increased, the detection rate can be increased to achieve superior results. The variety of pest and disease images should be increased. A diversified data set would be useful in the future.

Acknowledgements

This study was supported by the Ministry of Science and Technology (MOST) of Taiwan. MOST provided the research funding and devices. The related project number of this work is MOST- 107-2321-B-415 -004 and 107-2221-E-415 -007.

References

- [1] N. Oksana, Small Farms: Current Status and Key Trends, *Information Brief Prepared for Workshop on the Future of Small Farms Research*, Wye College, 2005.
- [2] D. Hughes, M. Salathé, *An Open Access Repository of Images on Plant Health to Enable the Development of Mobile Disease Diagnostics*, <https://arxiv.org/abs/1511.08060>, 2015.
- [3] K. A. Garrett, S. P. Dendy, E. E. Frank, M. N. Rouse, S. E. Travers, Climate Change Effects on Plant Disease: Genomes to Ecosystems, *Annual Review of Phytopathology*, Vol. 44, pp. 489-509, 2006.
- [4] A. Khan, M. S. Yadav, S. Ahmad, Image Processing Based Disease Detection for Sugarcane Leaves, *International Journal of Advance Research, Ideas And Innovations In Technology*, Vol. 3, No. 4, pp. 497-502, March, 2017.
- [5] K. Sabanci, C. Aydin, Smart Robotic Weed Control System for Sugar Beet, *Journal of Agricultural Science and Technology*, Vol. 19, No. 1, pp. 73-83, February, 2017.
- [6] L. Chen, S. Wang, W. Fan, J. Sun, S. Naoi, Beyond Human Recognition: A CNN-based Framework for Handwritten Character recognition, *3rd IAPR Asian Conference on Pattern Recognition*, Kuala Lumpur, Malaysia, 2015, pp. 695-699.
- [7] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet Classification with Deep Convolutional Neural Networks, *Advances in Neural Information Processing Systems*, Lake Tahoe, Nevada, USA, 2012, pp. 1097-1105.
- [8] M. D. Zeiler, R. Fergus, Visualizing and Understanding Convolutional Networks, *European Conference on Computer Vision*, Zurich, Switzerland, 2014, pp. 818-833.
- [9] K. Simonyan, A. Zisserman, *Very Deep Convolutional Networks for Large-scale Image Recognition*, <https://arxiv.org/abs/1409.1556>.
- [10] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going Deeper with Convolutions, *2015 IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, USA, 2015, pp. 1-9.
- [11] D. G. Lowe, Distinctive Image Features from Scale-invariant Keypoints, *International Journal of Computer Vision*, Vol. 60, No. 2, pp. 91-110, January, 2004.
- [12] H. Bay, T. Tuytelaars, L. Van Gool, Surf: Speeded up Robust Features, *European Conference on Computer Vision*, Graz, Austria, 2006, pp. 404-417.
- [13] H. A. Rowley, S. Baluja, T. Kanade, Neural Network-based Face Detection, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 1, pp. 203-208, January, 1996.
- [14] M. W. Craven, J. W. Shavlik, Using Neural Networks for Data Mining, *Future Generation Computer Systems*, Vol. 13, No. 2, pp. 211-229, November, 1997.
- [15] Y. Li, J. Li, J.-S. J. J. o. I. T. Pan, Hyperspectral Image Recognition Using SVM Combined Deep Learning, *Journal of Internet Technology*, Vol. 20, No. 3, pp. 851-859, May, 2019.
- [16] J.-y. Li, J.-h. J. J. o. I. T. Li, Prompt Image Search with Deep Convolutional Neural Network via Efficient Hashing Code and Addictive Latent Semantic Layer, *Journal of Internet Technology*, Vol. 19, No. 3, pp. 949-957, May, 2018.
- [17] R. Girshick, J. Donahue, T. Darrell, J. Malik, Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation, *The IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, 2014, pp. 580-587.
- [18] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, A. C. Berg, *SSD: Single Shot Multibox Detector*, <https://arxiv.org/abs/1512.02325>.
- [19] K. He, G. Gkioxari, P. Dollár, R. Girshick, Mask R-CNN, *The IEEE International Conference on Computer Vision*, Venice, Italy, 2017, pp. 2961-2969.
- [20] T. Rumpf, A.-K. Mahlein, U. Steiner, E.-C. Oerke, H.-W. Dehnb, L. Plümer, Early Detection and Classification of Plant Diseases with Support Vector Machines based on Hyperspectral Reflectance, *Computers and Electronics in Agriculture*, Vol. 74, No. 1, pp. 91-99, October, 2010.
- [21] H. Al-Hiary, S. Bani-Ahmad, M. Reyalat, M. Braik, Z. ALRahamneh, Fast and Accurate Detection and Classification of Plant Diseases, *International Journal of Computer Applications*, Vol. 17, No. 1, pp. 31-38, March, 2011.
- [22] A. Nithya, V. Sundaram, Classification Rules for Indian Rice Diseases, *International Journal of Computer Science*, Vol. 8, No. 1, pp. 444-448, January, 2011.
- [23] M. Brahimi, K. Boukhalfa, A. Moussaoui, Deep Learning for Tomato Diseases: Classification and Symptoms Visualization, *Applied Artificial Intelligence*, Vol. 31, No. 4, pp. 299-315, May, 2017.
- [24] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, A. Zisserman, The Pascal Visual Object Classes (voc) Challenge, *International Journal of Computer Vision*, Vol. 88, No. 2, pp. 303-338, September, 2010.
- [25] R. Girshick, Fast R-CNN, *The IEEE International Conference on Computer Vision*, Santiago, Chile, 2015, pp. 1440-1448.
- [26] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: Towards Real-time Object Detection with Region Proposal Networks, *Advances in Neural Information Processing Systems*, Montréal, Canada, 2015, pp. 91-99.
- [27] J. R. Uijlings, K. E. Van De Sande, T. Gevers, A. W. Smeulders, Selective Search for Object Recognition, *International Journal of Computer Vision*, Vol. 104, No. 2, pp. 154-171, April, 2013.

Biographies



Tu-Liang Lin is an assistant professor in the Department of Management Information Systems at National Chiayi University, Taiwan. He obtained the Ph.D. degree from the Iowa State University in 2011, under the supervision of Prof. Guang Song. His main research interests include robotics, wireless network, network security, bioinformatics and machine learning. He applied the robotic motion planning to discover the ligand migration path of dynamic proteins in his doctoral work. Currently, he is working on applying the deep learning techniques to solve the medical and biology problems.



Hong-Yi Chang received the Ph.D. degree from the National Tsing-Hua University, Hsin-Chu, Taiwan, in 2010, under the supervision of Prof. Nen-Fu Huang. His research topics are mainly related to Wireless Sensor Networks, Cloud Computing and Resource Management, Machine Learning, Deep Learning, Big Data. From 2011 to 2017, he was an Assistant Professor and Associate Professor at the Department of Management Information Systems, National Chiayi University, Taiwan. Dr. Chang is currently a Professor in the Department of Management Information Systems, National Chiayi University, Taiwan, since Aug 2017. He is a member of the IEEE.



Kai-Hong Chen is a graduated student in the Department of Management Information Systems at National Chiayi University, Taiwan. He is under the supervision of Dr. Tu-Liang Lin. His main research interests include wireless network and machine learning. Currently, he is working on applying the deep learning techniques to solve the traffic flow prediction problems.