


Article

A Smartphone-Based Application for Scale Pest Detection Using Multiple-Object Detection Methods

Jian-Wen Chen ¹, Wan-Ju Lin ^{1,2}, Hui-Jun Cheng ³, Che-Lun Hung ^{1,4,*} , Chun-Yuan Lin ^{5,*} and Shu-Pei Chen ⁶

¹ Institute of Biomedical Informatics, National Yang Ming Chiao Tung University, Taipei 11221, Taiwan; vito9580work@gmail.com (J.-W.C.); d05522001@gmail.com (W.-J.L.)

² Department of Mechanical Engineering, National Taiwan University, Taipei 10617, Taiwan

³ Department of Computer Science and Information Engineering, Providence University, Taichung City 43301, Taiwan; hjcheng5@gm.pu.edu.tw

⁴ Department of Computer Science and Communication Engineering, Providence University, Taichung City 43301, Taiwan

⁵ Department of Computer Science and Information Engineering, Chang Gung University, No. 259, Sanmin Rd., Guishan Township, Taoyuan City 33302, Taiwan

⁶ Applied Zoology Division, Taiwan Agricultural Research Institute Council of Agriculture, Executive Yuan, Taichung City 413008, Taiwan; spchen@tari.gov.tw

* Correspondence: clhung@ym.edu.tw (C.-L.H.); cyulin@mail.cgu.edu.tw (C.-Y.L.)

Abstract: Taiwan's economy mainly relies on the export of agricultural products. If even the suspicion of a pest is found in the crop products after they are exported, not only are the agricultural products returned but the whole batch of crops is destroyed, resulting in extreme crop losses. The species of mealybugs, Coccidae, and Diaspididae, which are the primary pests of the scale insect in Taiwan, can not only lead to serious damage to the plants but also severely affect agricultural production. Hence, to recognize the scale pests is an important task in Taiwan's agricultural field. In this study, we propose an AI-based pest detection system for solving the specific issue of detection of scale pests based on pictures. Deep-learning-based object detection models, such as faster region-based convolutional networks (Faster R-CNNs), single-shot multibox detectors (SSDs), and You Only Look Once v4 (YOLO v4), are employed to detect and localize scale pests in the picture. The experimental results show that YOLO v4 achieved the highest classification accuracy among the algorithms, with 100% in mealybugs, 89% in Coccidae, and 97% in Diaspididae. Meanwhile, the computational performance of YOLO v4 has indicated that it is suitable for real-time application. Moreover, the inference results of the YOLO v4 model further help the end user. A mobile application using the trained scale pest recognition model has been developed to facilitate pest identification in farms, which is helpful in applying appropriate pesticides to reduce crop losses.

Keywords: deep learning; object detection; Faster R-CNN; SSD; YOLO v4; scale pest; pest detection system



Citation: Chen, J.-W.; Lin, W.-J.; Cheng, H.-J.; Hung, C.-L.; Lin, C.-Y.; Chen, S.-P. A Smartphone-Based Application for Scale Pest Detection Using Multiple-Object Detection Methods. *Electronics* **2021**, *10*, 372. <https://doi.org/10.3390/electronics10040372>

Academic Editor: Jordi Guitart

Received: 21 December 2020

Accepted: 29 January 2021

Published: 3 February 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Taiwan is a rich source of valuable agriculture products, and agriculture plays an important role in Taiwan's economic production. However, the climate of Taiwan is hot and humid, leading to various threats from pests in crop production. Moreover, the influence of climate change is increasingly serious and the threats from pests to crops are becoming more severe and unpredictable. Scale insects are phytophagous insects that eat green plants [1]. These insects suck the sap of plant organs, especially leaves, fruits, stems, and roots, which causes sooty mold disease. This disease affects photosynthesis and leads to tissue infection, resulting in damage to the plants and a reduction in the market value of the plant commodities in terms of quality and quantity. As farmers encounter the issue of pest attacks, they rely on their own previous experiences and knowledge to make a diagnosis. If they lack appropriate knowledge, they may fail to use appropriate medicines to control

the pests. Moreover, aged populations, agricultural labor shortage, and migration from rural areas to urban areas are critical problems across the world. The trends of population aging and exodus are getting more serious in Taiwan's rural areas; as a result, labor-based agriculture is facing a severe labor shortage. In recent years, the Taiwanese government has been encouraging younger people to return to their rural homes and engage in agricultural work. However, the younger generation lacks knowledge of the causes of pest problems and professional knowledge of agriculture, which is a big challenge in the early prevention of pest diseases.

The issue of pest disease not only affects agricultural production but also impacts economic exports. The economy of Taiwan is mainly dependent on the export of agricultural products. With the expansion of international trade, the import and export of agricultural products is becoming increasingly faster, causing the spread of pest diseases across several countries. If even the suspicion of a pest is found in the crop products after they are exported, not only are the agricultural products returned but the whole batch of crops is destroyed, resulting in extreme crop losses. To effectively control the impact of pest disease on exports, pest control strategies are becoming an important task.

Precision agriculture (PA) is a technique in which one acquires and figures out data from crops. Pest diseases can be fought by using precision agriculture. The traditional tools of operators—static stations, sensor networks, drones, and mobile robots—cannot precisely detect pests; due to this reason, these tools are not specifically employed in agriculture but are used to reduce chemical use and are costly. To execute precision farming, it is necessary for these tools to be able to process and make inferences from collected data, like an expert in the fields. Image processing integrated with machine learning is extensively applied in precision agriculture and has gained wide attention in the field of detecting plant diseases [2–6], weeds [7–11], and pests [12–16]. Ernest et al. [2] employed a linear support vector classifier and the k-nearest neighbor algorithm to diagnose diseases from plant images. They also presented different ways to extract various features from leaf images and analyze these features using the difference in the performance of the classifier. Ramesh et al. [3] used a histogram of an oriented gradient technique to extract features and applied a random forest approach to distinguish the images of healthy and diseased papaya leaves. Cheng and Maston [7] used three machine learning methods—support vector machine (SVM), neural network, and decision tree—in classifying the images of weed and rice, achieving 98.2% precision. Ahmed et al. [8] employed SVMs to recognize six species of weeds from images, achieving 97.3% accuracy by combining the extractors. Ebrahimi et al. [12] combined histogram equalization with the SVM model for detecting the pest diseases of whitefly and housefly in a strawberry greenhouse environment [13]. Lucas et al. [13] used multiple machine learning algorithms of multiple linear regression, K-neighbors regressor, random forest regressor, and artificial neural network in pest warning systems, which could improve the efficiency of the chemical control of pest diseases on coffee tree. Although machine learning has been successfully applied in agriculture fields, the limitation of image preprocessing and extracting the features in advance greatly affects recognition results. The accuracy of the recognition of the crop disease depends heavily on the characteristics of the images, such as illumination and scaling, and image preprocessing is essential before feature extraction. Moreover, feature selection has become a crucial task in the machine learning technique and significantly influences classifier accuracy. The conventional approach of image preprocessing and feature selection requires expert knowledge to manually decide and extract the features, which is tedious and time consuming.

With the rapid development of the deep learning technique, convolutional neural networks (CNNs) have been successfully applied in the field of agriculture research, which could solve the drawbacks of machine learning. The CNN model performs better in the automatic identification of pest diseases. The CNN model typically consists of two main operators, which are the convolutional layer and the pooling layer. The convolutional layer can automatically extract more complex and significant features of the image. Due to the

high computation of the convolutional network, the pooling layer reduces the number of parameters of the data. Most of the current research investigates the topic of pest image classification based on CNN models. However, detecting and localizing each pest in the images of the natural environment is more important than performing pest classification. The feature extractors of the CNN model combined with meta-architecture could deal with computer vision tasks of object detection. Methods of object detection generally include region-based convolutional neural networks (R-CNNs) [15], fast region-based convolutional networks (Fast R-CNNs) [17], faster region-based convolutional networks (Faster R-CNNs) [15,18], single-shot multibox detectors (SSDs) [15,19], and You Only Look Once (YOLO) [20,21]. Recently, numerous researchers have investigated pest identification based on object detection techniques. Fuentes et al. [15] combined the deep learning meta-learning of R-CNN, Faster R-CNN, and SSD with Visual Geometry Group network (VGG net) and residual network to recognize nine different types of tomato plant diseases and pests. Lin et al. [17] proposed an anchor-free region convolutional neural network with a Fast R-CNN as an end-to-end model to classify 24 pest categories. The result showed 56.4% mAP and 85.1% mRecall accuracy on a dataset of 24 classes of pests, which is higher than that of the Faster R-CNN and the YOLO detector. Li et al. [18] used Faster R-CNN as the object detection framework to build a real-time crop disease and pest video detection system. The result showed that the proposed system could effectively detect the untrained rice disease in the video. Zhong et al. [20] proposed a vision-based flight insect recognition system on Raspberry Pi that employed a YOLO network as a detector to roughly calculate the number of flying insects, and the SVM model was used to classify flying insects. The experimental results showed 92.50% counting accuracy and 90.18% classification accuracy. Although several researchers have been explored deep-learning-based object detection techniques for pest identification, none of the papers have discussed the recognition of scale insects. Although Taiwan is world renowned for agriculture technology, it still does not truly effectively address the pest issues. The scale insects cause major agricultural pest diseases in Taiwan, which leads to severe economic losses. Thus, the aim of this study is to use deep-learning-based object detection approaches to identify different types of scale insects.

Though the CNN model achieves high performance in computer vision applications, extrapolating the results from research datasets to actual scenarios is vital. The crucial task to implement in a realistic environment is effectively recognizing the images in real time. It is necessary to deploy the automatic detection technology on a mobile device, which provides farmers more convenience and portability in controlling pest diseases in remote conditions. Although smartphones have pervaded numerous fields, including the manufacturing industry and medicine and health, the adoption of mobile devices in the agriculture industry is slower. With progress in technology, farmers are realizing the importance of mobile agriculture (mAgri), which not only helps farmers effectively perform agricultural services using mobile phones but also turns arable farming into smart arable farming. Smart arable farming mainly uses the technologies of the Internet of Things (IOT). The main purpose of the IOT in farming is using sensing devices to collect farm data and upload these to the cloud database. Then big data analytics is used to calculate and process the stored data in the cloud, which turns the data into meaningful information for farmers. Use of the CNN model to analyze big data on the cloud has helped achieve state-of-the-art results in many tasks involved in recognizing pests and diseases. The most important aim of deploying the CNN model in an agricultural field is real-time pest disease diagnosis in a real outdoor farm environment. The existing research mostly focuses on the pest database of the laboratory for analyzing, seldom identifying the pest diseases on mobile devices in outdoor scenarios with complex environments. This paper investigates the scale pest disease diagnostics on mobile devices. We deploy state-of-the-art approaches of CNN object detection models for real-time pest recognition in agricultural fields.

2. Deep Meta-Architecture-Based Scale Pest Recognition

2.1. System Overview

The aim of this study is to design a system that farmers could use as a scale pest detector for early prevention of pest damage. The developed system is combined with the smartphone device to help farmers to improve their efficiency. However, smartphone devices of farmers have poor computing capabilities; so the cloud platform is employed in this study. By using the cloud platform, farmers can simply use the developed system without a hardware restriction. The proposed system uses the Keras platform to deploy the CNN object detection model on a mobile device, which could detect the locations of three types of scale pests in the image. Use of deep-learning-based object detection models to predict the types and locations of the scale pests appearing in the image is depicted in Figure 1. A general flowchart of this study is presented in Figures 2 and 3. Firstly, this paper uses a mobile device to collect a large number of images of scale pests. Image preprocessing of data augmentation and data annotation is performing before feeding the images into the object detection models. The prediction object detection models are trained to identify the types and locations of the scale insects appearing in the image. After training the object detection models, saving the models to the cloud platform of database is essential for further inference. This paper develops an intelligent pest identification system based on a cloud platform, which employs the Internet of Things technology to upload the trained model to a cloud storage platform through a Wi-Fi or a 4G network. The training process of the AI-based pest detection system is shown in Figure 2. To deploy and inference the trained model in a field environment, a smartphone device has been developed to combine with the cloud platform, which can be used by the user to identify the pest diseases in real time. The flowchart of inferencing the model to the smartphone device is shown in Figure 3. Next, this paper describes each component of the proposed system in detail.

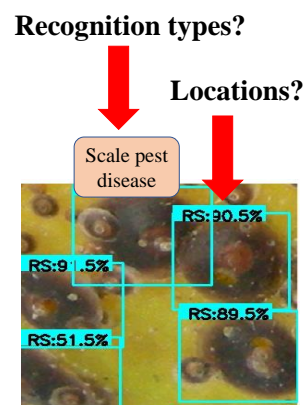


Figure 1. Predicting the types and locations of the scale pests appearing in the image.

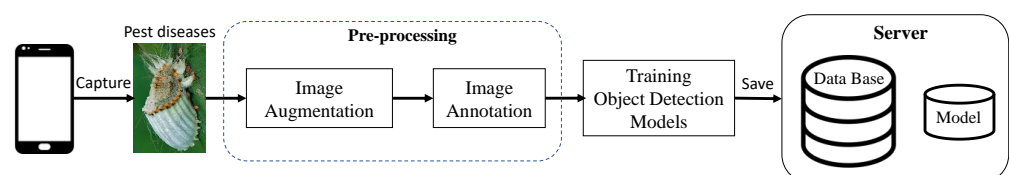


Figure 2. The training process of an AI-based pest detection system.

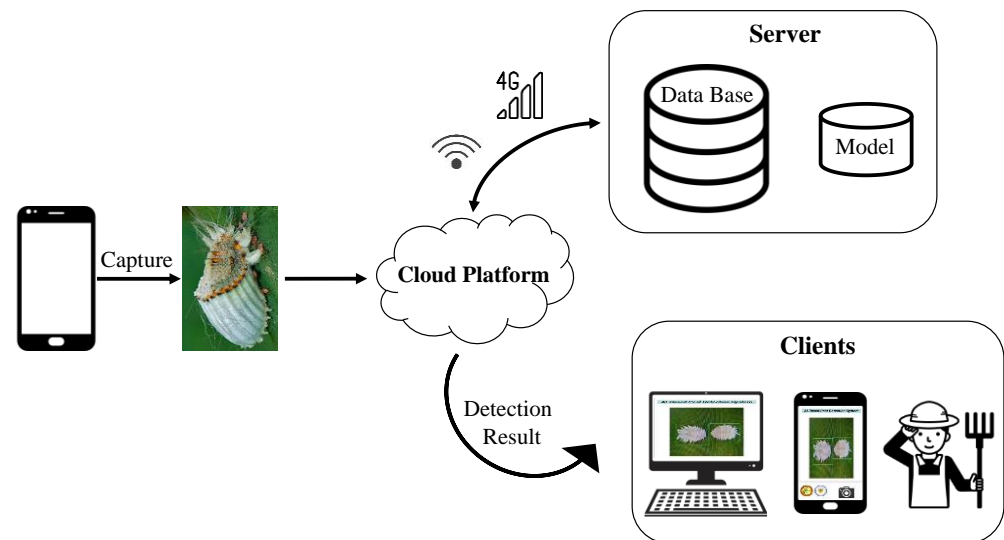


Figure 3. A flowchart of inferring the model to the smartphone device.

2.2. Data Annotation

Image annotation is an important image preprocessing step before training models. A feature is information drawn from the image, and the labeled features are assigned to the input based on the selected features. A machine can learn the labeled features during the training process. Thus, the labeling correctness of the features greatly affects the accuracy of the training model. Because many species of scale insects look quite similar, knowledge of the types of pest diseases is provided by the field of experts, which could help the computer learn features significant to different pest diseases. In this paper, the source of the scale pest dataset is Taiwan Agricultural Research Institute, Council of Agriculture, which mainly focuses on solving the issue of pest diseases in Taiwan agricultural products. The task of data annotation is mainly done by employees with more than 20 years of experience in the fields. The process of image annotation involves labeling the types and positions of the pests in the image, and the outputs of the labeled results are coordinates and bounding boxes of different types and sizes. The bounding boxes' results are also used to compare the Intersection over Union (IOU) ratio of the predicted images. Labeling is an open graphical image annotation tool that manually locates the position and category of the pest appearing in the image and further saves it as an .xml file, with the corresponding xmin, xmax, ymin, and ymax data for each bounding box. The image annotation tool is shown in Figure 4.



Figure 4. The image annotation tool of labeling.

2.3. Data Augmentation

In general, the deep learning model often executes better with more data. However, in many cases, the problem of limited data sources exists and it is difficult to collect a large amount of data. So the issue of insufficient data is often encountered in data analytics. Moreover, insufficient data will also affect the phenomenon of overfitting during the training process. Data augmentation could not only address data insufficiency but also address the overfitting problem. The present existing approaches to data augmentation include geometrical transformation. This paper uses the geometrical transformation processes of rotation and horizontal flipping. The employment of the data augmentation technique in the scale pest images is shown in Figure 5.

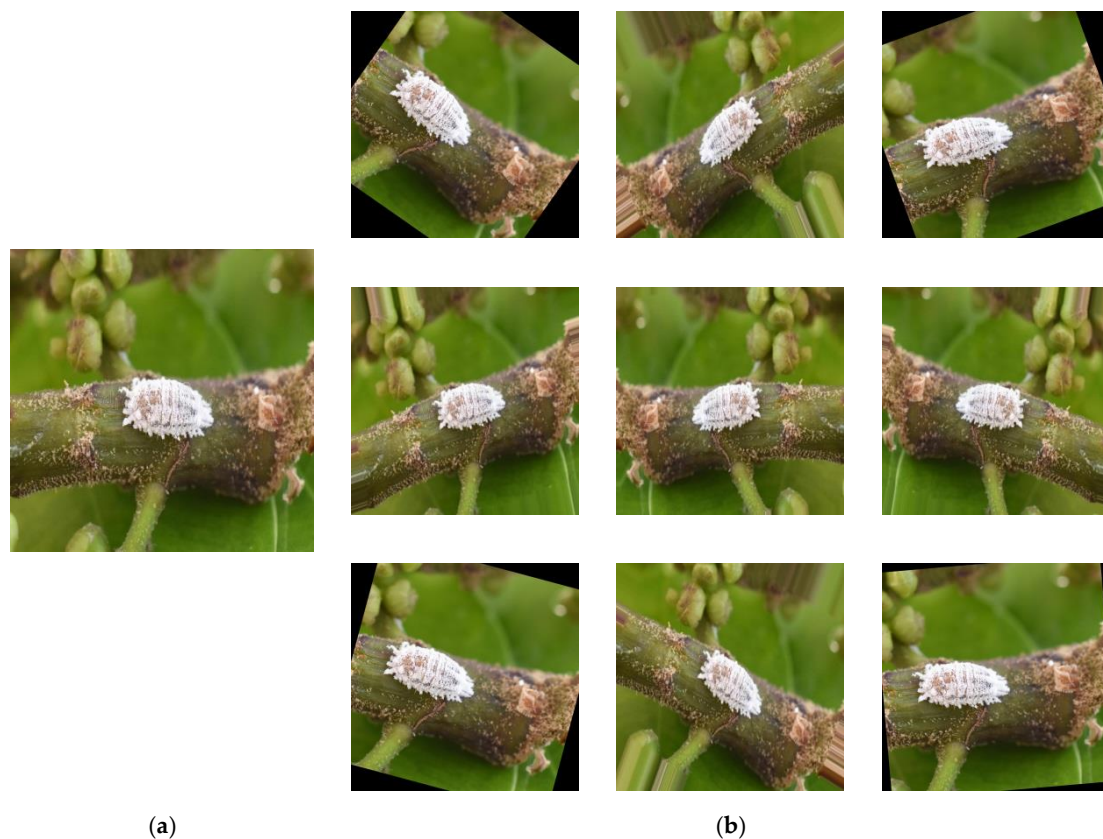


Figure 5. Applying data augmentation (i.e., shift in horizontal and vertical directions, rotation, horizontal flip) in the scale pest image. (a) Original image. (b) Data augmentation by rotation.

2.4. Object Detection Models

2.4.1. Faster Region-Based Convolutional Network (Faster R-CNN)

The Faster R-CNN [22] is the most representative method of the two-stage object detection models. A two-stage learning method involves finding the region proposal in the first stage and then performing object classification and bounding-box regression based on the region proposal in the second stage. A state-of-the-art Fast R-CNN was proposed by Ross B. Girshick in 2016. The network architecture combines the functions of feature extraction, region proposal, bounding box regression, and classification into multi-task learning. The core mechanism of the Faster R-CNN mainly includes a region proposal network and a Fast R-CNN detector. Firstly, the features are extracted by the Conv layers, and the output of the feature map is the input of the region proposal network (RPN). The region proposal network could also be considered as a feature extractor, which directly generates region proposals on the feature map to select the candidate feature region on the image. Using the region proposal network, the Faster R-CNN can efficiently replace the approach of selective search and the traditional approach of the sliding window technique.

The RPN uses the IOU between the object proposal and ground-truth to obtain better anchor boxes, which may contain an object. The RPN can simultaneously predict the object's highest scores and the bounding boxes at each location. However, the RPN can only roughly know whether the object features exist in the anchor box, and the position of the bounding box is not accurate. So the next stage is using the region proposals to train the Faster R-CNN model. The core idea of the Faster R-CNN is to perform the object classification in the anchor box of the region proposal and refine the position of the bounding box by regression.

2.4.2. Single-Shot Detector (SSD)

A single-shot detector (SSD) [15] is a classical one-step framework that uses a feed-forward convolutional network to generate a set of bounding boxes with a fixed size and predicts a confidence score of objects category in the bounding boxes. The one-step process of object detection is to directly predict the category score and location of each object. The architecture of the SSD uses the CNN model of VGG16 as a backbone structure and combines the pyramidal feature hierarchy structure to predict the multiple scale of feature maps with different resolutions. With the characteristic of ConvNet's pyramidal feature hierarchy approach, the SSD model could solve the issue of scale variation for an object detector. The SSD model is trained to have six branches, with different scales of images. Each branch has its own detector and classifier. The SSD uses the anchors concept of the Faster R-CNN to predict the offset and confidence of each default box, with different scales and aspect ratios in each feature map. Since the SSD model can detect the category and localize the coordinates of the objects in the images at the same time, the model is trained with two losses, localization loss and confidence loss. Finally, the results of the comparison between the estimated truth and the ground-truth are achieved by a non-maximum suppression (NMS) method.

2.4.3. You Only Look Once v4 (YOLO v4)

You Only Look Once (YOLO) is also a one-step framework of the object detection model. YOLO can be trained in an end-to-end way, which can deal with the position and classification task with high speed. Recently, the state-of-the-art YOLO v4 [23] network has been proposed and achieved superior performance. The YOLO v4 network is mainly composed of three parts: the backbone, the neck, and the head. The YOLO v4 network employs CSPDarknet53 as the backbone architecture to train the object detection model, spatial pyramid pooling (SPP) is combined with a path aggregation network (PANet) as the neck to extract different layers of feature maps, and the main head of the YOLO v4 network uses the YOLO head. The CSPDarknet-53 network is a type of convolutional neural network that is associated with the cross-stage partial network (CSPNet) and the DarkNet-53 model. Generally, the computation of the object detection model is higher and requires a large amount of computation time. CSPNet [24] improves the computation issue, which integrates the feature maps at the beginning and the end to reduce heavy computation. Moreover, the purpose of the neck in YOLO v4 is to fuse different scales of feature maps, which uses SPP to concatenate different scales of feature maps at the end of the network and uses PANet to insert one more layer based on the feature pyramid network.

3. Experimental Setup

The performance of the proposed system based on object detection models is evaluated and investigated in this section. First, the sources and background of the dataset in this study are described in detail. The evaluation of the experimental result mainly has two stages. In the first stage, statistical indicators of IOU, precision, recall, and F1 score are used to estimate the results of bounding box positioning. Confusion matrices further illustrate the classifier performance of the SSD, Faster R-CNN, and YOLO v4 models. In the next stage, the results of object detection models are inferenced on the mobile app. Then the outcomes of the web interface and smartphone app are provided to the end user to

perform the scale disease diagnosis in real time. The applied object detection algorithms are implemented in the graphics processing unit (GPU) with NVIDIA Geforce GTX 1050 Ti for computational acceleration. The object detection models are retrained with the machine learning backend library of Keras and Tensorflow. The smartphone used in the experiment runs Android 8.0 OS, and the hardware spec is 8 processor cores and 4 GB RAM.

3.1. Dataset Descriptions

Taiwan straddles the subtropical and tropical zones, with a hot and humid climate. Under various temperatures and humidity conditions, the spread of diseases and pests seriously affects Taiwan's valuable crops. The scale insects are major economic pests in Taiwan, with more than 7000 species. Scale insects feeding on plants cause severe damage to agricultural crops, ornamental flowers, and greenhouse crops. The species of mealybugs, Coccidae, and Diaspididae are mainly pests of the scale insect category that can lead to serious damage to the plants. This paper cooperated with Taiwan Agricultural Research Institute, Council of Agriculture, which is the authoritative organization of pest identification, to develop the proposed system. Taiwan Agricultural Research Institute, Council of Agriculture, has collected images of the three types of pests from the actual fields for decades. Figure 6 shows images of the three types of pests identified in this study. To simplify the name of each pest, this study used RF, RR, and RS to refer to mealybugs, Coccidae, and Diaspididae, respectively. Mealybugs have an oval body outline, with punctures around the body, and the body is covered in a mealy wax secretion or a white cottony coating. Coccidae secrete a wax powder covering around themselves in the egg stage, and the eggs are laid in powdery sacs, which look like a pot lid. The bodies of the Diaspididae insects are covered with a hardened and waxy scale. Different scale insects have various secretions, increasing the difficulty of prevention. This study used 600 images of scale pests in each category for model learning and validation. Among the 600 training images, there were approximately 200 images of each category for the augmentation images. The remaining 35 scale pest images were used as actual testing datasets. To attain higher performance of the model training, this paper employed the data augmentation technique of rotation, flip, and shift to increase the datasets. Moreover, with the higher resolution of the image, the detection is better. The resolution of the training and testing dataset is 416×416 .

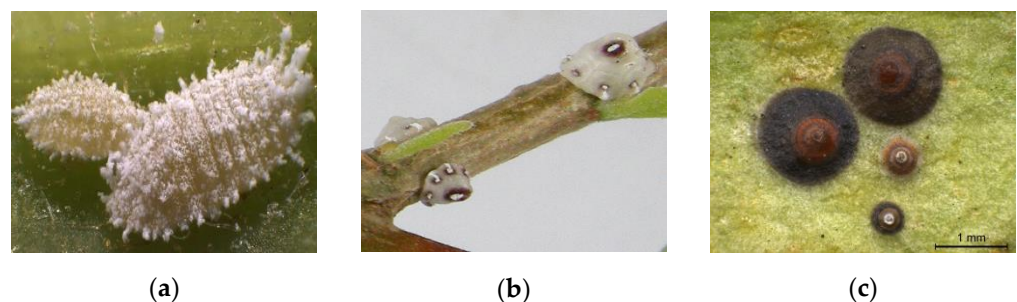


Figure 6. The three types of pest images identified in this study. (a) Mealybugs (RF). (b) Coccidae (RR). (c) Diaspididae (RS).

3.2. Performance of the Deep-Learning-Based Object Detection Model

To evaluate the performance of the applied object detection models, this paper used the unknown data of the testing dataset to assess the results of the training models. The following will elaborate the details of the experimental results. In object detection approaches, the detection result and the classifier performance are the two main indexes to evaluate the performance of the models. The standard statistical measures of Intersection over Union (IOU), recall, precision, and F1 score are usually used to assess the results of bounding box positioning. The confusion matrix is the general analytics tool to describe the performance of a classifier. For model detection, the IOU operation is to estimate the closeness between

the predicted bounding box and the actual box. It can be used to evaluate the correctness of the bounding box. The definition of the IOU is shown in Figure 7. The IOU is a ratio between the intersection and the union of the actual box and the predicted box. If the value of the IOU is higher than 0.5, the classified results of object detection are defined as true positive (TP). False positive (FP) is considered if the value of the IOU is below 0.5. False negative (FN) for object detection means that the predicting results should be positive but the models have performed incorrect detection. By the output of the IOU, the indicator of precision, recall, and F1 score can be further calculated, and it is also a popular technical indicator to evaluate the performance of the object detection models. The definitions of precision, recall, and F1 score are provided in Equation (1) to Equation (3). Precision represents the recognition ability with negative datasets. When the score of precision is higher, the ability of models to distinguish a negative dataset is stronger. Recall represents the recognition ability with positive datasets. When the score of recall is higher, the ability of models to account for positive dataset is more superior. The F1 score is an evaluated indicator to integrate the mean of the precision and the recall, which could reconcile the precision and recall of the model. A higher F1 score illustrates that the model is more robust.

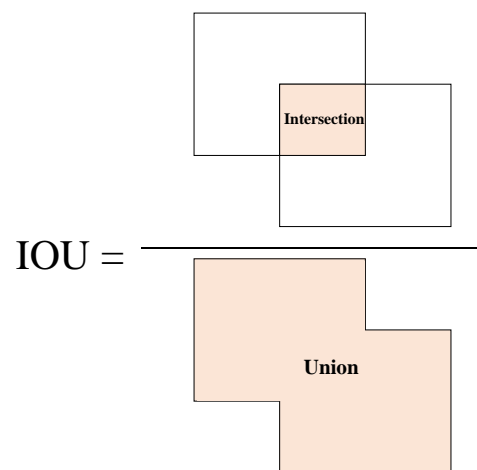


Figure 7. The definition of Intersection over Union (IOU).

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

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

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

To quantitatively evaluate the Faster R-CNN and SSD model, this paper used 70 images for each class as testing dataset to examine the experimental results. The detection results of Fast R-CNN, SSD, and YOLO.v4 models are shown in Table 1. Statistical indicators of precision, recall and F1 score have been executed to assess the robustness of three object detection models. As mentioned above, the F1 score is a comprehensive indicator to evaluate the model's robustness. Assessment of the detection performance of three object detection models based on the F1 score is presented in Table 1. The result reveals that the Faster R-CNN model detected mealybugs with an F1 score of 85%, Coccidae with that 91%, and Diaspididae with that of 83%. The SSD model achieved the same for mealybugs with an F1 score of 98%, Coccidae with that of 100%, and Diaspididae with that of 98%. YOLO v4 detected mealybugs with an F1 score of 100%, Coccidae with that of 100%, and Diaspididae with that of 100%. Moreover, the inference time of the Faster R-CNN model was 0.69 s per image, SSD took d 0.7 s per image, and YOLO v4 0.9 s per image.

Table 1. Statistical indicators for scale detection.

Testing Dataset	Mealybugs (RF)			Coccidae (RR)			Diaspididae (RS)		
	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
Faster R-CNN	75%	100%	85%	96%	88%	91%	100%	71%	83%
SSD	100%	98%	98%	100%	100%	100%	98%	100%	98%
YOLO v4	100%	100%	100%	100%	100%	100%	100%	100%	100%

Confusion matrices for the Faster R-CNN, SSD, and YOLO v4 models give a more detailed investigation of the performance of classifier models and determine which classes of pest images are easier to distinguish across different categories. Figure 8 shows the confusion matrix of images of the unlearned scale pests. The rows of the confusion matrix represent the true class, and the columns mean the predicted class of the object detection model. The context of the confusion matrix provides the quantitative accuracy for each class, which indicates whether the predicted model is misclassified or not. Based on the confusion results of Figure 8, YOLO v4 can correctly classify three types of pest diseases. Faster R-CNN and SSD models show to be confused in predicting the three types of scale pest diseases.

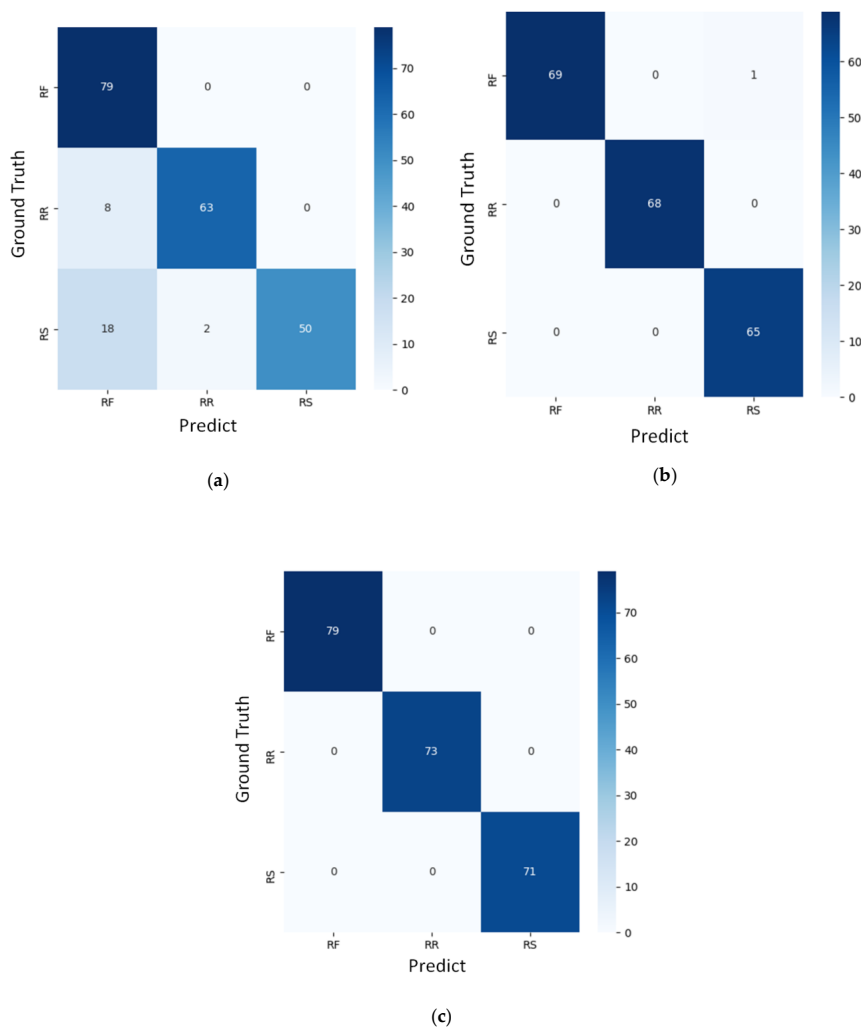
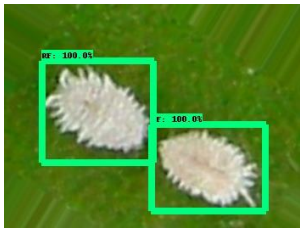

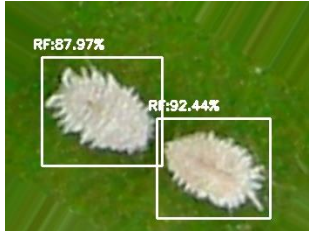
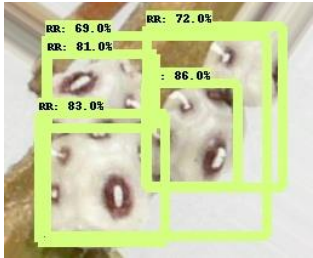

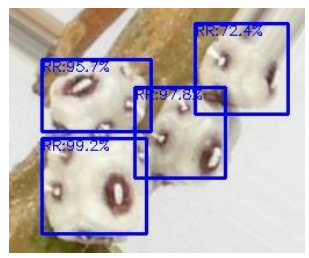
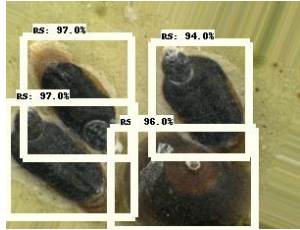
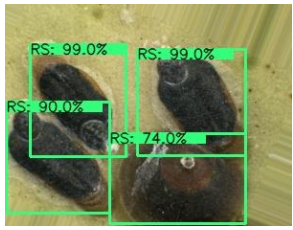
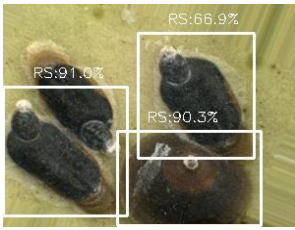


Figure 8. The confusion matrix of the faster region-based convolutional network (Faster R-CNN), single-shot multibox detector (SSD), and You Only Look Once v4 (YOLO v4) models for scale pest recognition. (a) Faster R-CNN. (b) SSD. (c) YOLO v4.

To further investigate the ability of the applied two-object detection models, this paper uses the same category of pest images to compare the results of three-object detection models. A performance comparison of three-scale recognition between Faster R-CNN, SSD, and YOLO v4 models is shown in Table 2. The results reveal that the YOLO v4 model can not only correctly localize each scale pest in the image but also achieves the highest accuracy in predicting the three types of scale pest diseases. Conversely, Faster R-CNN localizes wrong results for Coccidae pest in the image. We can conclude that the YOLO v4 model achieves a superior prediction result in recognizing and detecting the three classes of scale pests. In addition, in terms of the classification accuracy among the three categories, it is observed that Diaspididae could not be detected well. The reason is that the Diaspididae images have multiple categories and it is more difficult to train Diaspididae pests than the mealybugs. A few samples of the training dataset for mealybugs and Diaspididae are shown in Table 3.

Table 2. Performance comparison of three-scale recognition between Faster R-CNN and SSD models.

	Faster R-CNN	SSD	YOLO V4
Predicted results of mealybug pest images			
Predicted results of Coccidae pest images			
Predicted results of Diaspididae pest images			

3.3. Qualitative Results of the Pest Detection System

In this section, the qualitative results of the scale disease detection system are evaluated. First, the predicted results of the object detection model are presented. Next, the details of the web-based user interface system are given. Figure 9 shows the experimental results of the identification of the three categories of scale pests, where RF represents mealybugs, RR is Coccidae, and RS means Diaspididae. This paper randomly chooses a scale image from each category to view the predicted results of the object detection model. Each image could contain more than one predicted result.

Table 3. A few samples in the training dataset for mealybugs and Diaspididae.




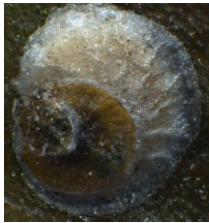


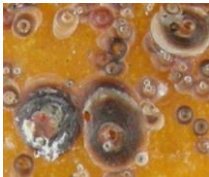
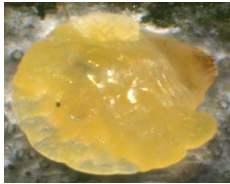




Mealybugs		Diaspididae	
			
			
			



Figure 9. Predicted result of the object detection model for detecting three pests.

After obtaining the identification outcome of the object detection models, the results of the scale disease detection system are further provided to the end user. A web interface and a smartphone are used to perform the task of pest detection application in real time. This study develops a web-based user interface system and combines the android mobile app to track the object detection outcome. The result of scale disease monitoring using the web-based user interface system is shown in Figure 10.



Figure 10. The result of scale disease monitoring using the web-based user interface system.

4. Conclusions

It is important to provide a real-time detection system for scale diseases for farmers in the agricultural fields. Therefore, this study employed object detection models to execute recognition tasks. Due to difficulty in collecting a large number of pest images in the actual field, data augmentation approaches of rotation and horizontal flipping were used to increase the number of pest images as the training data in this study. Furthermore, three object detection models, Faster R-CNN, SSD, and YOLO v4, were applied in this paper. The experimental results reveal that the YOLO v4 model achieves a superior result in recognizing the three classes of scale pests, identifying mealybugs with an F1 score of 100%, Coccidae with that of 89%, and Diaspididae with that of 97%. This study also proves that the architecture of YOLO v4 could improve the speed and increase the accuracy of real-time recognition. Moreover, this study proposed a cloud platform, which was combined with a web interface and a smartphone app to perform the task of pest detection application in real time. By using the AI-based pest detection system, farmers would not only be unhampered by the hardware but also be able to alert users of pest diseases in advance.

Author Contributions: J.-W.C. implemented algorithms and performed the experiments; W.-J.L. wrote the paper, designed the algorithms, and performed the experiments; H.-J.C. revised the paper; C.-L.H. conceived the algorithms, designed the algorithms and experiments, and revised the paper; C.-Y.L. revised the paper; S.-P.C. collected pest images and labeled them. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: This research is supported by Taiwan Agriculture Research Institute, Council of Agriculture, under the grants MOST109-2321-B-002-051, and the Chang Gung Memorial Hospital Project [grant number CORPD2J0011, CORPD2J0012].

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kondo, T.; Gullan, P.J.; Williams, D.J. Coccidology. The study of scale insects (Hemiptera: Sternorrhyncha: Coccoidea). *Cienc. Tecnol. Agropecu.* **2008**, *9*, 55–61. [[CrossRef](#)]
2. Mwebaze, E.; Owomugisha, G. Machine learning for plant disease incidence and severity measurements from leaf images. In Proceedings of the 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA), Anaheim, CA, USA, 18–20 December 2016; pp. 158–163.
3. Ramesh, S.; Hebbar, R.; Niveditha, M.; Pooja, R.; Shashank, N.; Vinod, P. Plant disease detection using machine learning. In Proceedings of the 2018 International Conference on Design Innovations for 3Cs Compute Communicate Control (ICDI3C), Bangalore, India, 25–28 April 2018; pp. 41–45.
4. Yang, X.; Guo, T. Machine learning in plant disease research. *Eur. J. Biomed. Res.* **2017**, *3*, 6–9. [[CrossRef](#)]
5. Shruthi, U.; Nagaveni, V.; Raghavendra, B. A review on machine learning classification techniques for plant disease detection. In Proceedings of the 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS), Coimbatore, India, 15–16 March 2019; pp. 281–284.
6. Padol, P.B.; Yadav, A.A. SVM classifier based grape leaf disease detection. In Proceedings of the 2016 Conference on Advances in Signal Processing (CASP), Pune, India, 9–11 June 2016; pp. 175–179.
7. Cheng, B.; Matson, E.T. A feature-based machine learning agent for automatic rice and weed discrimination. In Proceedings of the International Conference on Artificial Intelligence and Soft Computing, Zakopane, Poland, 14–18 June 2015; pp. 517–527.
8. Ahmed, F.; Al-Mamun, H.A.; Bari, A.H.; Hossain, E.; Kwan, P. Classification of crops and weeds from digital images: A support vector machine approach. *Crop Prot.* **2012**, *40*, 98–104. [[CrossRef](#)]
9. Herrera, P.J.; Dorado, J.; Ribeiro, Á. A novel approach for weed type classification based on shape descriptors and a fuzzy decision-making method. *Sensors* **2014**, *14*, 15304–15324. [[CrossRef](#)] [[PubMed](#)]
10. Saha, D.; Hanson, A.; Shin, S.Y. Development of enhanced weed detection system with adaptive thresholding and support vector machine. In Proceedings of the International Conference on Research in Adaptive and Convergent Systems, Odense, Denmark, 11–14 October 2016; pp. 85–88.
11. Bakhshipour, A.; Jafari, A. Evaluation of support vector machine and artificial neural networks in weed detection using shape features. *Comput. Electron. Agric.* **2018**, *145*, 153–160. [[CrossRef](#)]
12. Ebrahimi, M.; Khoshtaghaza, M.-H.; Minaei, S.; Jamshidi, B. Vision-based pest detection based on SVM classification method. *Comput. Electron. Agric.* **2017**, *137*, 52–58. [[CrossRef](#)]
13. de Oliveira Aparecido, L.E.; de Souza Rolim, G.; De, J.R.d.S.C.; Costa, C.T.S.; de Souza, P.S. Machine learning algorithms for forecasting the incidence of *Coffea arabica* pests and diseases. *Int. J. Biometeorol.* **2020**, *64*, 671–688. [[CrossRef](#)]
14. Kim, Y.H.; Yoo, S.J.; Gu, Y.H.; Lim, J.H.; Han, D.; Baik, S.W. Crop pests prediction method using regression and machine learning technology: Survey. *IERI Procedia* **2014**, *6*, 52–56. [[CrossRef](#)]
15. Fuentes, A.; Yoon, S.; Kim, S.C.; Park, D.S. A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition. *Sensors* **2017**, *17*, 2022. [[CrossRef](#)]
16. Silva, D.F.; De Souza, V.M.; Batista, G.E.; Keogh, E.; Ellis, D.P. Applying machine learning and audio analysis techniques to insect recognition in intelligent traps. In Proceedings of the 2013 12th International Conference on Machine Learning and Applications, Miami, FL, USA, 4–7 December 2013; pp. 99–104.
17. Jiao, L.; Dong, S.; Zhang, S.; Xie, C.; Wang, H. AF-RCNN: An anchor-free convolutional neural network for multi-categories agricultural pest detection. *Comput. Electron. Agric.* **2020**, *174*, 105522. [[CrossRef](#)]
18. Li, D.; Wang, R.; Xie, C.; Liu, L.; Zhang, J.; Li, R.; Wang, F.; Zhou, M.; Liu, W. A recognition method for rice plant diseases and pests video detection based on deep convolutional neural network. *Sensors* **2020**, *20*, 578. [[CrossRef](#)] [[PubMed](#)]
19. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. Ssd: Single shot multibox detector. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016; pp. 21–37.
20. Zhong, Y.; Gao, J.; Lei, Q.; Zhou, Y. A vision-based counting and recognition system for flying insects in intelligent agriculture. *Sensors* **2018**, *18*, 1489. [[CrossRef](#)] [[PubMed](#)]
21. Du, J. Understanding of object detection based on CNN family and YOLO. *J. Phys.* **2018**, *1004*, 012029. [[CrossRef](#)]
22. Zhao, Z.-Q.; Zheng, P.; Xu, S.-t.; Wu, X. Object detection with deep learning: A review. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *30*, 3212–3232. [[CrossRef](#)] [[PubMed](#)]
23. Bochkovskiy, A.; Wang, C.-Y.; Liao, H.-Y.M. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934.
24. Wang, C.-Y.; Mark Liao, H.-Y.; Wu, Y.-H.; Chen, P.-Y.; Hsieh, J.-W.; Yeh, I.-H. CSPNet: A new backbone that can enhance learning capability of cnn. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 16–18 June 2020; pp. 390–391.