ISSN 1870-4069

Pests Detection in Agricultural Crops using Computer Vision

Lauro Reyes-Cocoletzi[1], Miguel Angel Ortega-Palacios[2], Luis A. Cuecuecha-Sánchez[1]

[1] Universidad Autónoma de Tlaxcala, Facultad de Ciencias Básicas, Ingeniería y Tecnología, Mexico

[2] Benemérita Universidad Autónoma de Puebla, Complejo Regional Centro, San José Chiapa, Mexico

lauro.reyesco@uat.mx, miguel.ortega@correo.buap.mx, luisangel.cuecuecha.s@uatx.mx

Abstract. In the future, computer vision systems that connect machines with fungicides, insecticides and herbicides to be used on a regular basis will be needed. In the long term the systems will have autonomous monitoring of crop health and taking timely action against factors that damage crops. The problem to be solved consists of the recognition and classification of pest affectations in leaves of agricultural crops by means of machine learning algorithms ResNet18 based on the use of computer vision. Advances were made in a computer vision system for the detection of a specific pest that affects corn crops, Spodoptera frugiperda. In addition to the detection of the worm present in the captured images, plant damage is detected to infer the presence of the pest damaging the crop. This project also aims to make contributions in models of recognition systems and computer vision applicable in the prevention and reduction of the impact of pests present in agriculture of the countries.

Keywords: Computer vision, pest recognition, machine learning, plant damage.

1 Introduction

Agriculture plays an important role in food security, alleviating poverty and driving economic development. The world's population is expected to reach 9.7 billion by 2050 and 11.2 billion by the end of this century [9], so food production must increase despite various factors affecting crop yields, such as pests, weeds, pathogens, nutrients, water, sunlight, soil moisture, soil fertility nutrients, water, sunlight, soil degradation, environmental impact and scarcity of arable land.

Manual crop inspection is time consuming, prone to human error and some parts of the field may be difficult to access, reducing inspection efficiency.

Technology adaptation is crucial for food production, machine vision systems (MVS) can automate crop inspection with the help of on-site or off-site imaging techniques to improve overall crop yield [1]. Compared to human vision, MVS can predict crop problems more accurately by analyzing information acquired from images.

In recent years, the industrialized countries have had an accelerated growth in the use of applied agricultural technologies, which has led to the automation of plantations through the implementation of sensors for monitoring and determining the conditions in which crops are developing. The use of technology for automation, including novel precision agriculture techniques, has increased due to the need to determine and provide better conditions for good crop quality [1]. A pending problem to be solved is the increase of pest populations in agricultural crops (e.g., wheat, barley, corn and oats) because they can cause a significant reduction in grain yields due to their excessive spread.

A useful tool is the automation of industrial processes by means of computer vision, because it represents reliability, efficiency and speed of information processing of the environment, so the agricultural industry will use more this type of technology to monitor relevant aspects of crops.

Computer vision as mentioned is a powerful tool for crop health monitoring and with the right sensors it is possible to locate areas of affected crops in a given location in the field. With respect to specific tasks that machine vision can perform in agriculture, is the recognition and morphology of plants, so that new methods can be implemented for the detection of pests and diseases [11]. New methods for pest detection can be implemented, and advances in machine learning and high-performance computing can create efficient solutions for identifying crop diseases in order to create management, monitoring and control alternatives to reduce decision-making time once the pest or disease has been detected.

The proposal of this work focuses on computer vision techniques for the identification of pests in agricultural crops. Artificial neural networks (ANN) have a great potential in the identification of natural resources, precision agriculture, product quality assessment, sorting, grading, etc., ANN can recognize the color, shape, size and texture of an object and can find the point of interest from them (regions of interest).

The designed algorithm in this work of investigation is focused on the detection of the pest Spodoptera frugiperda (worm) and the detection of the damage caused by this animal in corn crops.

This paper is organized as follows, in section 2, relevant related works are mentioned, in section 3, details of the proposed methodology and relevant information are presented, in section 4, the experiments and the results obtained are described. Finally, Section 5 presents the conclusions and future work.

2 Related Work

The following are some related works to be taken into consideration with respect to the problem to be solved.

Yao et al. [14] propose a rice pest identification system using two 12MP digital cameras, the cameras are placed on a glass plate with 4 black light sources to attract the pests. The main objective was to detect four different rice pests of lepidopteran species. Their work achieves an accuracy of 90.5% without cross-validation and 97.5% by cross-validation. The main problem is the overlapping of insects, in these cases, manual separation is performed.

Vakilian et al. [3] developed a system to identify beet armyworm (Spodoptera Exigua), a pest of vegetable, field, and flower crops. Images were captured with a digital camera together with an illumination module, images utilized for training ANN classifier and remaining for evaluation. Convolutional neural network (CNN) classifier was able to classify armyworms with an accuracy of 90%.

Qing et al. [15] proposed a technique to measure the population density of white-backed grasshopper (WBPH) in rice paddies. A digital camera attached to an extendable pole was used to detect the pest on rice stalks. Detection was done in three-layer mechanism, the first layer is an AdaBoost classifier, the second is a support vector machine (SVM) classifier based on the histogram of oriented gradient (HOG), and the third layer used threshold based on one color and three shape features. They achieved a detection accuracy of 90.7% with 4.9% false detection rate.

Rajan et al. [16] proposed an automatic pest identification system to detect whiteflies, aphids, and cabbage moths. Digital camera was used to capture images of the crop which may have pests on their leaves. SVM classifier was used to train with threshold values and the slack variables of the images in the database collected. The threshold value was used to distinguish the object from the background and classification of the pests was done using slack variables. They achieved a detection accuracy of 95%.

The above-mentioned works are some of those already carried out; a summary of other proposals is shown in Table 1.

In this context, progress is presented in the research project to develop a system for the detection and subsequent monitoring of pests in agricultural fields by applying machine learning models through computer vision.

Specifically, the aim is to identify those pests that damage or alter the surface of the crop leaves, because the determination of the existence of a pest can be detected directly or indirectly by computer vision. Direct detection of the pest involves observing the insect or worm or feeding on crop leaves; on the other hand, indirect observation involves detecting damage or discoloration on crop leaves without the presence of the animal that causes it.

It is important remember that the artificial vision system designed in this work focuses on detecting a specific worm (Spodoptera frugiperda), in addition to detecting damage to the leaves of the crop (specifically corn) to infer the presence of pests damaging the plants.

Table 1. Computer vision based methods for pest detection.

Reference	Type of crop	Pest (name)	Method	Accuracy
[18]	Corn	Corn disease	ResNet	97.5%
[11]	Multiple crops	Beet armyworm	ANN	90.0%
[12]	Paddy	WBPH	AdaBoost & SVM Classifiers	85.2%
[13]	Paddy	Brown plant hopper (BPH)	One-way ANOVA	< 70.5%
[15]	Multiple crops	Codling moth	ConvNets (CNN)	93.4%
[14]	Multiple crops	Whiteflies	SVM	95.0%
[17]	Strawberry	Thrips	SVM	> 97.5%

3 Methodology

The methodological basis of this research project is the observation of the environment through the implementation of computer vision with previously captured images. The methodological basis of computational processing applies object detection and color segmentation algorithms for the analysis of the information of interest.

A Machine Learning approach is used for the analysis of the information of interest, which allows the classification of the observed in the crop. As mentioned the processed images include crop leaves in different health conditions (different colors), the number of images of the different classes will be balanced in an acceptable range of samples to validate the training of the ANN.

Ideally, pests should be detected as early as possible, but when their small size, e.g. at the egg stage, macro lenses have to be used to obtain images, this is not practical in field applications.

Preprocessing includes considerations of crop leaf damage distribution as they may occupy only a small portion of pixels in the captured images and may not be suitable enough for ANN model training. Regions of interest (ROI) are highlighted from the original images (data labeling).

Figure 1 shows in general the blocks corresponding to the algorithm to be developed, which is explained in more detail below.

3.1 Information Collection (dataset) and ROI

The system dataset consists of one general scenario, which consists of the identification of leaf, stem and fruit damage in the corn crop caused by the corn worm pest. IP102 dataset [5] has 737 images of interest, also we utilize the internet as the primary source to collect images, which is widely used to build datasets such as the ImageNet [12] and the Microsoft COCO [10]. The first collection step relies on common image search engines, including kaggle,the

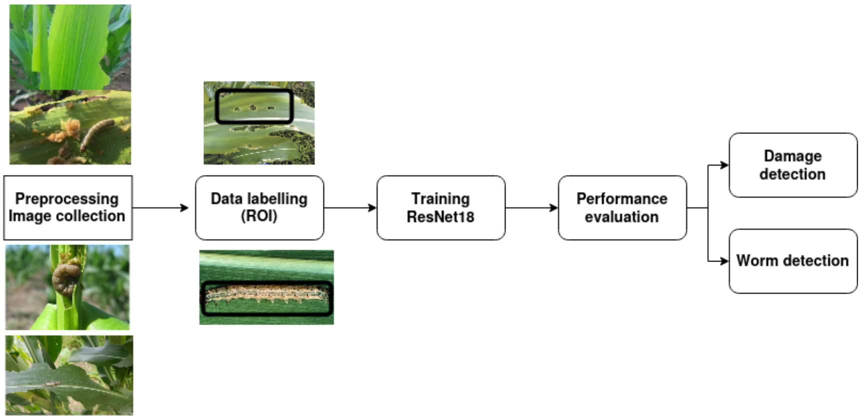

Fig. 1. Stages of the algorithm design.

Fig. 2. Example of crop damage (a) leaf affected by worm, (b) worm in the stem.

total number of images collected that make up the data set is about 7,000. Figure 2 shows some of these samples.

Regions of interest are delimited in the images to train with samples of worm-damaged leaves as well as the presence of the worm in the stems or leaves, the identification (labeling) was done with labelImage [7]. The cropped images have sizes from 328 × 328 pixels to 600 × 600 pixels and the regions of interest can be clearly observed in the images.

According to the proposed methodology, the first step involves preprocessing the information to facilitate the characterization of the leaves of the crops, a filtering process is performed within the image obtained, with the purpose of eliminating as much noise as possible, present in the images after their acquisition. This stage focuses on suppressing excess lighting, unwanted shadows and elements that are not part of the leaf, highlighting in turn, the information necessary for further analysis.

3.2 Learning Stage

Characterization of the relevant aspects of the image, since this stage of the system is very important for an adequate final classification.

The basis for the detection of damage in agricultural crops is based on the implementation of a ANN. This network will focus on the detection of

Layer	Output size	18 layer
Conv_1	112x112	7x7, 64 stride 2
Conv_2	56x56	3x3 max pool, stride 2 [3x3, 64 3x3, 64] x2
Conv_3	28x28	[3x3, 128 3x3, 128] x2
Conv_4	14x14	[3x3, 256 3x3, 256] x2
Conv_5	7x7	[3x3, 512 3x3, 512] x2

Fig. 3. Architecture ResNet18 [8].

affectations in crop leaves and presence of worm, as an initial development, the use of a Residual Network architecture based on ResNet18 [8] was proposed to deal with the vanishing gradient problem.

The core idea of ResNet18 is to introduce hop connections or residual connections, which allow network layers to learn differences rather than learning entire functions. These residual connections allow gradients to propagate more easily through the network, which helps to avoid the problem of gradient fading [4].

The ResNet architecture is based on residual blocks, each block contains a series of convolutional layers and can be stacked to form a deeper network [2]. The residual block has a shortcut path structure that adds the output of one layer to the output of a subsequent layer, known as the skip connection operation (Figure 3).

As usual after convolutional layers, convolutional filters are used to extract local features from the images, however, after passing through several convolutional and clustering layers, the resulting representation may still have local features and not be completely related to the final classification. Dense layers are added to perform a global classification and combine the extracted features into a more complete and global representation of the image.

The model assigns a probability to each class that represents its confidence that the example belongs to that class.

On the other hand, we have the actual labels that indicate the true class to which each example in the data set belongs; cross-entropy is used to quantify the difference between the prediction probabilities generated by the model and the actual probabilities or labels in the data set.

ResNet18 uses a training loop to adjust the network parameters using the training set, at each iteration, it performs the following steps:

– Passes a batch of images through the network to obtain the predictions.

- Calculates the loss using the predictions and the actual labels.
- Performs backpropagation of the error to compute the gradients.
- Updates the network parameters using the optimizer and the calculated gradients.

Cross entropy measures how different these two probability distributions are, the probabilities predicted by the model and the actual probabilities. In particular, it is a measure of the loss of information or uncertainty in the model prediction compared to the actual labels.

In the case of binary classification (worm, damaged leaf), the Equation 1 is applied:

$$H(p,q) = -[p \cdot log(q) + (1-p) \cdot (log(1-q))] \ , \tag{1}$$

where p is the actual probability of the class and q is the probability predicted by the model for the class.

The optimization during training used is the ADAM algorithm (Adaptive Moment Estimation) to adapt the size of the learning steps (learning rate) for each parameter as a function of the first and second moment estimates of the gradients [6]. The initialization of parameters and hyperparameters is of great importance to efficiently perform the iterative loop.

It is important to note that the choice of these hyperparameters is not trivial and requires adjustments by experimentation. In general, a hyperparameter search was performed using random search techniques to find combinations that work well for our specific problem resulting in $\beta_1 = 0.85$, $\beta_2 = 0.94$.

At each iteration, the gradient of the objective function with respect to the parameters is calculated, gradient indicates the direction in which the parameters should be adjusted to reduce the loss. Parameter updates are calculated using the corrected first-order and second-order moments and the adaptive learning rate, the moments are corrected to compensate for initial biases.

3.3 Performance Evaluation

Generally, the transfer learning method discards the last layer of a pre-trained model and adds a fully connected layer where the neurons correspond to the number of predicted classes. During the training stage the last layer is trained from scratch, while the others are initialized from the pre-trained model and updated.

To provide a direct observation of the classification results confusion matrices will be calculated in addition statistics will be used to evaluate the performance of the models, training accuracy as the percentage of correctly classified samples in the training data set and similarly validation on the data set given the epochs run when the model begins to converge.

Detect and classify the observed pest to assign a label to it compare with a number of relevant samples the system carried out.

Since the model returns probabilities for each class instead of direct labels, it is necessary to convert the probabilities to predicted labels. This is usually

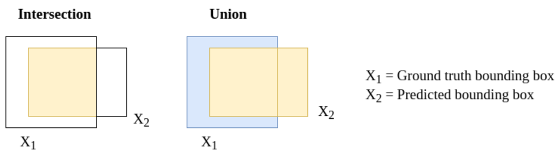

Fig. 4. Bounding box detection relation CM.

done by taking the index of the highest probability as the predicted label for each example.

The mean average precision (mAP) was used as the validation metric [18] for damage leaf and pest detection, mAP score was calculated as follows: average across the number of classes of the true positive divided by the true positives plus false positive as in the Equation 2:

$$mAP = \frac{1}{\#classes} \sum_{1}^{\#classes} \frac{\#TP}{\#TP + \#FP} \ . \tag{2}$$

In addition to mAP score, we also computed a confusion matrix (CM), for each detection, the algorithm mines all the ground-truth boxes and classes, along with the detected boxes, classes, and scores (probability of success in the bounding box). Only detections with a score ≥ 0.5 were considered and anything under this threshold were excluded. The list of matches was trimmed to remove duplicates (ground-truth boxes that match with more than one detection box or viceversa), if there are duplicates, the best match was continually selected.

The CM was updated to refect the resultant matches between ground-truth and detections, a detected box was refected as correct where the intersection over union (IoU) of that box and the corresponding ground-truth box was $\leq$ 0.5. Explanation for calculating IoU [17] is shown in Figure 4 and Equation 3, the CM was normalized:

$$IoU(X_1, X_2) = \frac{X_1 \cap X_2}{X_1 \cup X_2} \ . \tag{3}$$

4 Results

Implement a functional and reliable system for the detection and classification of pests affecting agricultural crop fields by means of novel machine learning techniques, which include the characterization of ResNet18 architecture. The convolutional neural network in addition to detecting the ROI (affected areas on the leaf) will also perform a membership prediction with respect to a number of possible classes to identify the type of affectation and/or pest.

This prediction is indicated through an enveloped frame of the object detected in the scene as well as the percentage of class membership using a

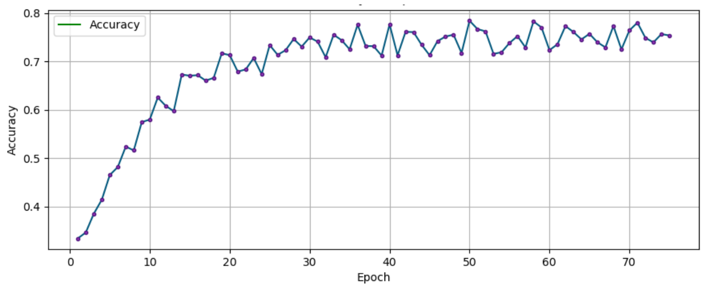

Fig. 5. Accuracy vs epochs.

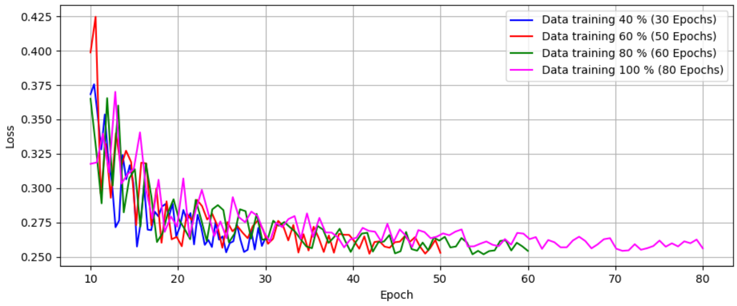

Fig. 6. Training loss curves of diferents epochs.

multi-label classification. During training, binary cross-entropy loss is used for the class predictions by classifying the class predictions by means of independent logistic classifiers.

After more than 60 epochs of training, Figure 5 and Figure 6 depict the accuracy and loss curves in training. With different amounts of training data-samples according to the number of epochs to be performed, in Figure 6 shows the loss value of ResNet18 reaching 0.26. After training, the maximum accuracy of our proposed method on the validation set can reach 0.75, and the minimum loss is 0.25.

Figure 7 shows examples of results obtained, showing the percentage and label of belonging to the two classes to be detected: worms and damaged leaves.

The performance shown by the proposed algorithm obtained for the detection of damaged leaves in the test data set a mAP rate of around 70%, for the case of worm detection in the case of detection of leaves damaged by pest the mAP rate obtained was 75%. Based on the results obtained on the test dataset, the confusion matrix associated with the results obtained is presented in Table 2.

Set the CM M_{ij}, in which each column (Table 2) of the matrix M_j (where i = 1, 2) represents the class prediction of the sample by the classifier, and each row of the matrix M_i (j= 1, 2) represents the ground truth to which the sample

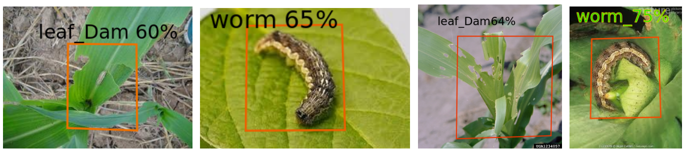

Fig. 7. Results obtained in pest detection.

Table 2. Results of confusion matrix.

	Prediction worn	Prediction damage leaf
Worn	656	86 Damage leaf
75	495	

belongs. Three general metrics for evaluating the performance of class models can be obtained from the CM.

The accuracy is percentage of correctly labeled samples in classified samples. It can reflect the classification performance of the model on data (Equation 4):

$$accuracy = \frac{\sum_{i=1}^{2} M_{ii}}{\sum_{i=1}^{2} \sum_{j=1}^{2} M_{ij}} . \tag{4}$$

Equation 5 defines precision, which measures the probability of correctly predicted samples in all predicted i-type samples. It denotes the classification effect of the algorithm:

$$precision = \frac{M_{ii}}{\sum_{j=1}^{2} M_{ij}} . \tag{5}$$

The recall is used to measure the probability that the prediction is correct in the instances labeled as i. It can express the effect of a certain type of recall, this calculation process is described in Equation 6:

$$recall = \frac{M_{ii}}{\sum_{i=1}^{2} M_{ij}} . \tag{6}$$

The f1-score (f1) is calculated by taking the weighted average of precision and recall (Equation 7). In other words, f1 conveys a balance between precision and recall. Although it is not as intuitive as accuracy, f1 is generally more valuable than accuracy, mainly when the class distribution is uneven:

$$f1 = 2 * \frac{precision * recall}{precision + recall} . \tag{7}$$

With the information from the CM, the results obtained (expressed in percentages) for the metrics of interest are listed below. For precision a value of 89% was obtained, for recall a value of 88% was obtained and finally for f1 a value of 88% was obtained.

It is worth mentioning that the confusion matrix presents the results where the mAP corresponding to the detection in the image of the detected class is greater than 65%, however there are images that were provided to the algorithm where it was not able to locate worm or damaged leaf when in fact there was.

According to the related work, an initial comparison can be made of the performance presented in this work vs. the works found in the state of the art. In general terms the performance of the proposed system has an accuracy of 75% and the work with the best performance has 97.5% [18], however the algorithm of our proposal in addition to detecting the animal (pest) additionally takes into consideration the damage that this pest causes in the plant.

These results are susceptible to improvement, the main problems to overcome being background noise in field environment, substantial overlapping of multiple leaves and scattered symptoms on different leaves. To handle these issues, we are currently collecting and labeling images of early stage damage leaf for improving the accuracy of the model, and the ability to generalize, because the dataset is not big enough.

5 Conclusions

This research work presents initial results with respect to crop pest detection, the evaluation metrics provide promising data feasible to improve by specific changes to the proposed algorithm.

Take the feature map (ResNet18) of the layers and increment by two, use a feature map from a previous network layer and merge with the particulars of the up-sampling using concatenation to predict a similar tensor, but now twice the size. This method will allow us to obtain more meaningful semantic information of the sampled distinguishing features and more detailed information of the previous feature map.

Increment more convolutional layers under the addition + concatenation model to predict boxes for the final scale. In this way, the predictions for the third scale benefit from all previous computation, as well as from the fine-grained characteristics of the first stages of the network.

With respect to the ANN, it is expected to predict boxes at 3 different scales in order to extract distinctive qualities of those scales using a concept similar to that of feature pyramid networks. From the base feature extraction, convolutional layers are added to extract the relevant qualities, in particular the last layer predicts a tridiagonal layer predicts a three-dimensional tensor that encodes the bounding box and class predictions.

ADAM optimization model used also has areas of opportunity such as momentum decay factors (β_1 and β_2) to aid model convergence without negatively affecting the rate (slower approximation to the learning rate).

The results obtained by ADAM are relevant, however, other optimizers such as SGD (stochastic gradient descent) could be used with momentum to verify the increase in algorithm performance.

Detection performance can be improved by using an architecture with more layers, in this case a network superior to ResNet18, such as ResNET34, ResNet50 or higher, ResNet18 was used because the capabilities of the computer equipment used are limited, with the right hardware a more robust architecture can be implemented.

At this point and with the obtained data, the results need to increase in terms of the performance obtained, the points mentioned with respect to the changes in the ANN will help improve the evaluation metrics of the algorithm.

As additional work, field tests or at least processing videos taken from real environments are also contemplated to verify and compare the results obtained with the images from the repositories.

The results obtained in this work for the moment are preliminary, they are advances corresponding to an initial stage of development of the proposed algorithm that can be improved.

References

1. Abbaspour-Gilandeh, Y., Aghabara, A., Davari, M., Maja, J. M.: Feasibility of using computer vision and artificial intelligence techniques in detection of some apple pests and diseases. Applied Sciences, vol. 12, no. 2, pp. 906 (2022)
2. Albanese, A., Nardello, M., Brunelli, D.: Automated pest detection with dnn on the edge for precision agriculture. IEEE Journal on Emerging and Selected Topics in Circuits and Systems, vol. 11, no. 3, pp. 458–467 (2021)
3. Asefpour Vakilian, K., Massah, J.: Performance evaluation of a machine vision system for insect pests identification of field crops using artificial neural networks. Archives of phytopathology and plant protection, vol. 46, no. 11, pp. 1262–1269 (2013)
4. Barraza, J. A., Espinoza, E. J., Espinos, A. G., Serracin, J.: Precision agriculture with drones to control diseases in the rice plant, (2020)
5. Bojer, C. S., Meldgaard, J. P.: Kaggle forecasting competitions: An overlooked learning opportunity. International Journal of Forecasting, vol. 37, no. 2, pp. 587–603 (2021)
6. Ermoliev, Y. M., Wets, R.-B.: Numerical techniques for stochastic optimization. Springer-Verlag (1988)
7. Gao, B.-B., Zhou, H.-Y.: Learning to discover multi-class attentional regions for multi-label image recognition. IEEE Transactions on Image Processing, vol. 30, pp. 5920–5932 (2021)
8. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
9. Júnior, T. D. C., Rieder, R., Di Domênico, J. R., Lau, D.: Insectcv: A system for insect detection in the lab from trap images. Ecological Informatics, vol. 67, pp. 101516 (2022)
10. Liu, H., Chahl, J. S.: A multispectral machine vision system for invertebrate detection on green leaves. Computers and Electronics in Agriculture, vol. 150, pp. 279–288 (2018)
11. Liu, H., Chahl, J. S.: Proximal detecting invertebrate pests on crops using a deep residual convolutional neural network trained by virtual images. Artificial Intelligence in Agriculture, vol. 5, pp. 13–23 (2021)

12. Muppala, C., Guruviah, V.: Machine vision detection of pests, diseases and weeds: A review. J. Phytol, vol. 12, pp. 9–19 (2020)
13. Partel, V., Nunes, L., Stansly, P., Ampatzidis, Y.: Automated vision-based system for monitoring asian citrus psyllid in orchards utilizing artificial intelligence. Computers and Electronics in Agriculture, vol. 162, pp. 328–336 (2019)
14. Qing, Y., Jun, L., Liu, Q.-j., Diao, G.-q., Yang, B.-j., Chen, H.-m., Jian, T.: An insect imaging system to automate rice light-trap pest identification. Journal of Integrative Agriculture, vol. 11, no. 6, pp. 978–985 (2012)
15. Qing, Y., Xian, D.-x., Liu, Q.-j., Yang, B.-j., Diao, G.-q., Jian, T.: Automated counting of rice planthoppers in paddy fields based on image processing. Journal of Integrative Agriculture, vol. 13, no. 8, pp. 1736–1745 (2014)
16. Rajan, P., Radhakrishnan, B., Suresh, L. P.: Detection and classification of pests from crop images using support vector machine. In: 2016 international conference on emerging technological trends (ICETT). pp. 1–6. IEEE (2016)
17. Selvaraj, M. G., Vergara, A., Ruiz, H., Safari, N., Elayabalan, S., Ocimati, W., Blomme, G.: Ai-powered banana diseases and pest detection. Plant methods, vol. 15, pp. 1–11 (2019)
18. Yu, H., Liu, J., Chen, C., Heidari, A. A., Zhang, Q., Chen, H., Mafarja, M., Turabieh, H.: Corn leaf diseases diagnosis based on k-means clustering and deep learning. IEEE Access, vol. 9, pp. 143824–143835 (2021)