Applied Entomology and Zoology (2021) 56:427–442
https://doi.org/10.1007/s13355-021-00732-y

ORIGINAL RESEARCH PAPER

Crop pest recognition using attention-embedded lightweight network under field conditions

Junde Chen[1] · Weirong Chen[2] · Adnan Zeb[1] · Defu Zhang[1] · Yaser Ahangari Nanehkaran[1]

Received: 14 October 2020 / Accepted: 7 March 2021 / Published online: 30 March 2021

Abstract
Plant pests have a negative effect on crop yields. If the various insect pests are not identified and controlled properly, they can spread quickly and cause a significant decline in agricultural production. To overcome the challenges, the convolutional neural network (CNN)-based methods have shown excellent performance as it performs automatic feature extraction in image identification and classification. In this study, to enhance the learning capability for pest images with cluttered backgrounds, the MobileNet-V2 pre-trained on ImageNet was chosen as the backbone network and the attention mechanism along with a classification activation map (CAM) were incorporated in our architecture to learn the significant pest information of input images. Moreover, the optimized loss function and two-stage transfer learning were adopted in model training. This kind of progressive learning first makes the model discover the large-scale structures, and then shifts its attention to delicate details step by step, improving the identification accuracy of plant pest images. The proposed procedure achieves an average accuracy of 99.14% on the publicly available dataset, and even in heterogeneous background conditions, the average accuracy also reaches 92.79%. Experimental results prove the efficacy of the proposed procedure, and it delivers outperformance compared with other state-of-the-art methods.

Keywords Crop pest recognition · Lightweight network · Classification activation map · Attention mechanism · Transfer learning

Introduction

World population is expected to rise to 9.7 billion in 2050 and may reach a peak of nearly 11 billion around 2100 (https://www.un.org/development/desa/en/news/population/world-population-prospects-2019.html), which requires substantial food supplies for the considerable growing population by minimizing crop damage. The occurrence of insect pests has a negative impact on crop growth, and if the dangerous insect pests are not recognized in time, they may cause a disastrous effect on food security (Deng et al. 2018). Early detection and warning can prevent the spread of insect pests and decrease the unnecessary usage of pesticides (Picon et al. 2019). It plays a crucial role to ensure the effective yields of food crops. Nevertheless, so far, the visual observations of the experienced insect pest experts or crop producers are still the primary approach for the identification of diverse plant pests in many places, especially in developing countries. This approach requires continuous monitoring of the crops and is undoubtedly time-consuming, labor-intensive, subjective, and expensive for large farms (Chen et al. 2020; Thenmozhi and Reddy 2019). Moreover, compared with the number of farmers, there are not adequate plant pest specialists in many areas, though some crop pest outbreaks can be prevented with this method. In particular, owing to economic and social reasons, few people are engaged in plant protection originally. As a consequence, there are great needs and important realistic significance to create a useful, reliable, automatic, and efficient tool to identify crop pests.

The new era of plant pest identification is being provided with the rapid development of digital cameras and computational capacity. Image processing and machine learning techniques have gained significant attention due to their outstanding ability to automate work (Shah et al. 2014), and specific

✉ Defu Zhang
dfzhang@xmu.edu.cn

[1] School of Informatics, Xiamen University, Xiamen 361005, China

[2] Department of Information and Electrical Engineering, Ningde Normal University, Ningde 352100, China

classifiers are usually employed to categorize plant pests into different types. For example, Gassoumi et al. (2000) proposed an artificial neural network (ANN)-based approach to recognize and categorize insects in cotton ecosystems, and they achieved an average identification accuracy of 90%. Wang et al. (2012) used the support vector machine (SVM) classifier to develop an insect identification system, which performed with good stability and accuracy reached 93%. Faithpraise et al. (2013) adopted k-means clustering along with relative filter to identify plant pests by manually extracting the features, and the correspondence filter achieved rotational invariance of pests up to angles of 360 degrees, indicating the effectiveness of their algorithm for identifying plant pests. Using six machine learning algorithms, Wen et al. (2009) proposed an effective local feature-based insect classification method for orchard insect identification and classification, and their maximum classification accuracy attained 89.5%, etc. In recent years, the novel machine learning technique named deep learning (DL), particularly convolutional neural network (CNN), has been widely used in image identification and classification and is becoming the standard methods for solving many image recognition tasks (Gadekallu et al. 2020; Hayashi et al. 2019; Kessentini et al. 2019). Based on VGG16 (Simonyan and Zisserman 2014) and transfer learning, Shijie et al. (2017) constructed a CNN model to detect tomato pests and diseases. A total of 10 common tomato diseases and pests were detected, and they achieved an average classification accuracy of 89%. Nanni et al. (2020) reported an ensemble CNN approach to perform the insect pest detection and recognition; they reached the state-of-the-art accuracy of 92.43%. Wen et al. (2015) proposed an improved pyramidal stacked de-noising auto-encoder (IpSDAE) architecture to build a deep neural network for moth identification, and they attained a good identification accuracy of 96.9%. Liu et al. (2016) developed an 8-layer CNN model to learn powerful local features for classifying paddy insects, and their architectures achieved a mean accuracy precision (mAP) of 0.951, a significant improvement over compared methods. Additionally, Nazki et al. (2020) introduced a data augmentation scheme using Activation Reconstruction-Generative Adversarial Networks (AR-GAN) for plant disease recognition and realized an accuracy improvement of 5.2% compared with the traditional method, etc. In summary, there are two varieties of popular approaches, such as the strong supervision methods and weakly supervised methods, for the identification of crop pests. Strong supervision scheme mainly includes object detection technology, which relies on more manual annotation information like the bounding box, key points, and coordinate information of the target object. In practice, it is time-consuming and labor-intensive to obtain a large amount of annotation information for model training. Conversely, the fine-grained image classification based on weak supervision information only needs the label data of the images, e.g., the images of insect pests belonging to the same species are stored in the same folder, and do not request additional annotation information. Thus, more and more researchers have paid much attention to the fine-grained image classification method based on weakly supervised information, and this approach is also adopted in our work. On the other hand, it is well known that there is no internet connection or the internet is very slow in remote rural areas, especially in developing countries. For this, we need to deploy the model offline and allocate memory for the CNN model applied on the mobile portable devices. Whereas, the classical deep CNN models with huge sizes require larger computing resources and are not suitable for such memory allocation. Therefore, in this study, to create a best-of-both-worlds with memory efficiency and classification accuracy, we utilized a lightweight CNN model to identify plant pest types. The MobileNet-V2 paired with attention mechanism and CAM were chosen in our approach. Based on the transfer learning (TL), we migrated the common knowledge of MobileNet-V2 learned from ImageNet and embedded the attention mechanism along with the CAM module to create a new network, namely Atten-MobNet, for identifying plant pest types. The top layer of the MobileNet-V2 was truncated and an additional convolutional layer was added for the extraction of high-dimensional characteristics. Then, the attention mechanism was incorporated in the network to learn the significant pest information, which is realized by creating a shortcut connection manner in the network. The output of the backbone layers is available as input to the attention module using a summation operation instead of being concatenated. Moreover, the optimized loss function and two-stage transfer learning were used in model training. This progressive training strategy helps the model discover the large-scale structures in the pest images first, and then shifts the attention to delicate details gradually, improving the identification accuracy of pest images in heterogeneous background conditions.

The remainder of this paper is structured below. In Sect. 2, following an introduction of data collection and overall process, this section primarily discusses the methodology to accomplish the task of plant pest identification. Later in Sect. 3, a series of experiments are conducted to probe the performance of the proposed procedure, and the experimental results are assessed by comparative analysis. Finally, Sect. 4 summarizes the writing and gives future work.

Materials and methods

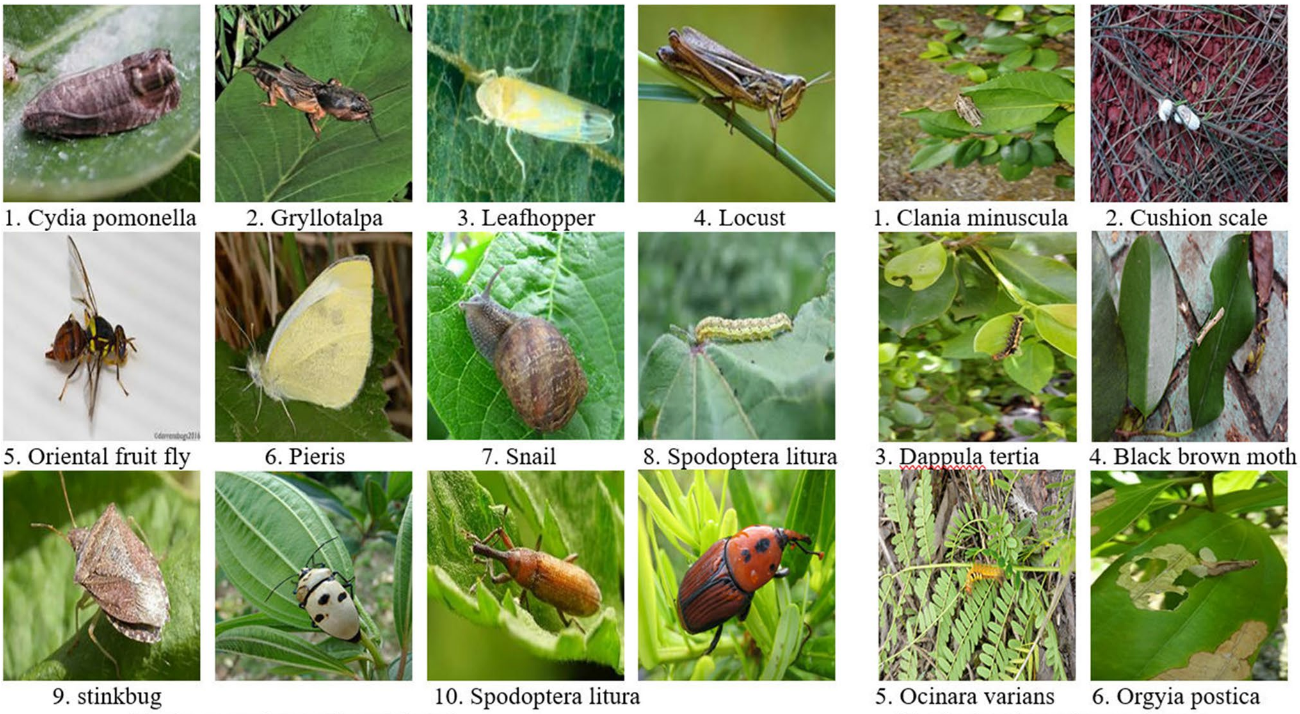
Image dataset

In this study, we have used two datasets including an open dataset from Li et al. (2020) and our collected local pest image dataset to perform the experiments together. There are a total of 5629 pest images in the open dataset which is composed of 10 common species of plant pests, namely cydia pomonella, gryllotalpa, leafhopper, locust, oriental fruit fly, pieris rapae linnaeus, snail, spodoptera litura, stinkbug, and weevil. This dataset collection was primarily performed by downloading images from the internet and partial images were photographed outdoors using the Apple 7 Plus mobile phone. Note that the number of samples in each class is not consistent and individual different types of insect pests are classified into the same category. In our experiments, the dimensions of all the images are uniformly re-sized as 224 × 224 pixels to fit the models. Also, the one-hot encoding of the categorical variable is first done for model training. The partial sample images are displayed in Fig. 1a.

Approximately 200 plant pest images were provided by the Xiamen Institute of Forest Pest Control and Quarantine, Fujian, China. These images were photographed from real-field wild scenarios with cluttered backdrop conditions and inconsistent lighting strengths. For example, some images were taken under the background conditions of heterogeneous grasses or leaves, and in some other images, the backgrounds were the surroundings of the field. The photographers' fingers or the ground with different colors might occur in the background of some plant images sometimes. Moreover, the illumination intensities were inconsistent for the images photographed at different times were with the varied weather conditions like sunny, cloudy, and overcast weather, etc. According to the knowledge of specialists in the field, these types of plant pests have been confirmed and they are primarily composed of 6 classes including clania minuscula butler, cushion scale, dappula tertia templeton, black brown moth, ocinara varians, and orgyia postica. The images are uniformly saved as JPG format. For subsequent computations, the Photoshop tool has been used to process the crop pest images into the RGB mode first, and then the dimensions of images are uniformly re-sized to 224 × 224 pixels to fit the models. Figure 1b shows the partial sample images.

Overall process

The overall process of our approach for the recognition of plant pests is described below. Initially, the plant pest images were captured from real-field wild scenarios and labelled based on the domain specialists' knowledge. The images of insect pests belonging to the same species were stored in the same folder, and the image pre-processing techniques, such as image sharpening, image filtering, and image re-sizing, were performed on the sample images. The image sharpening was executed for some blurred images to make them clearer, the image filtering was implemented to remove the noise of sample images, and the image re-sizing was conducted to uniformly adjust the sizes of images sizes as 224 × 224 pixels. Then, apart from remaining some original images to evaluate the performance of the model, the



(a) The open dataset from Li et al.

(b) The local pest image dataset.

Fig. 1 The sample images of open dataset and local pest images dataset. **a** The open dataset from Li et al. **b** The local pest image dataset

data augmentation scheme was used to synthesize new images to diversify the sample images and suppress the overfitting risks. The enhanced Generative Adversarial Networks (GAN) (Goodfellow et al. 2014) combined with the traditional methods were used in our data augmentation scheme to generate new synthetic images. After that, the training sample images were input to the proposed Atten-MobNet network, which fused the merits of the lightweight MobileNet-V2 as well as the spatial and channel attention mechanism, to train the model. Ultimately, the optimized loss function and two-stage transfer learning were implemented in model training, and the yielded optimum model was used for the identification of crop pest types. With this method, the finally identified results of plant pests are obtained and can also be used to update the expert sample library. Figure 2 depicts an overall flowchart, and the detailed descriptions of these processes are discussed in subsequent sections.

Related work

Lightweight networks

As stated previously, deep CNNs have shown great promise in image identification and classification. However, due to the large number of parameters that need to be trained, it is hard for a classical deep CNN to meet the application requirements of embedded systems. Therefore, the research and application of lightweight CNNs have obtained increasing attention in recent years. Mobile-nets are shaped up based on the streamlined structures that adopt depth-wise separable convolution (DWSC) to build lightweight CNNs (Shen et al. 2019). The simplified mobile-nets possess the characteristics of a small structure, low delay, and low power consumption to effectively maximize accuracy while considering the limited resources. In other terms, there is a trade-off between the calculation memory and identifying accuracy in CNN models.

Generally speaking, the standard convolution (SC) extracts features from all three dimensions of each image, including two spatial dimensions (width and height) and one-channel dimension (Kaiser et al. 2017). Therefore, the standard convolution slides over the image (y) using a filter with weights (W), as written in Eq. (1).

$$SC(W,y)_{(i,j)} = \sum_{h,l,m}^{H,L,M} W_{(h,l,m)} \times y_{(i+h,j+l,m)}, \tag{1}$$

where H and L separately represent the height and width of images, W is the weights of filters, and M denotes the number of filters. DWSC divides a standard convolution into a depthwise convolution (DWC) and a 1 × 1 convolution named pointwise convolution (PWC) (Sifre et al. 2014), as depicted in Fig. 3. Among them, DWC uses a convolution kernel to implement the convolution operation for each channel of the input feature map, and PWC conducts a 1 × 1 convolution for the results of DWC. The formulas are presented in Eqs. (2, 3).

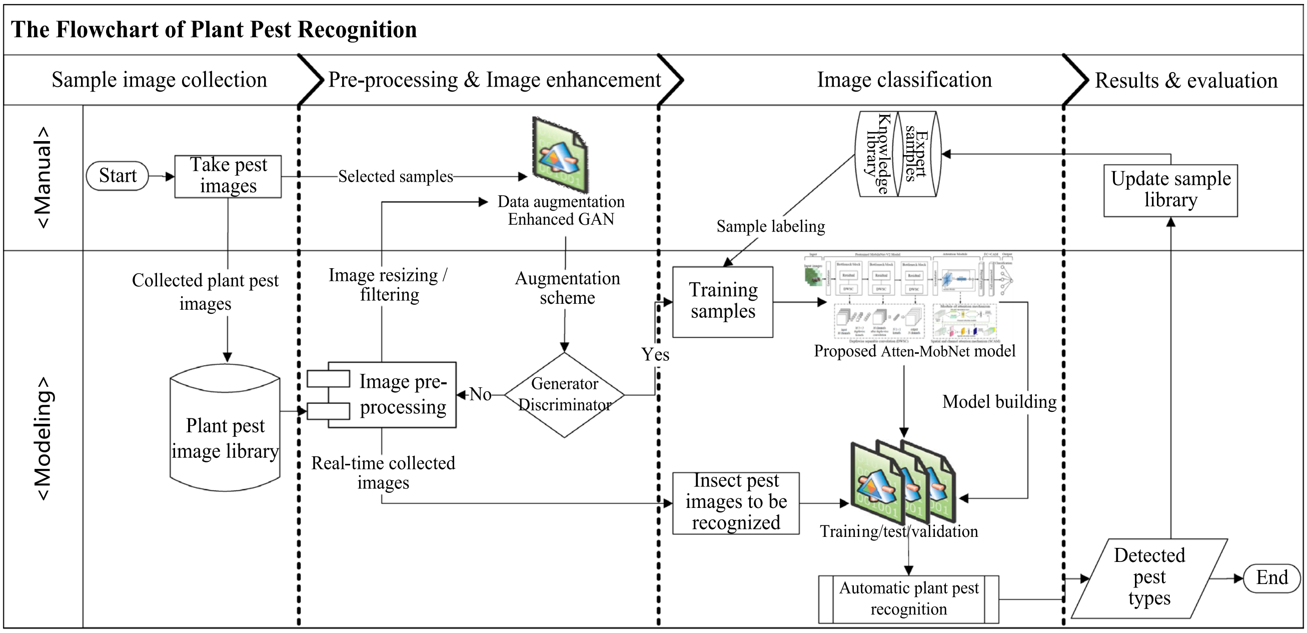


Fig. 2 The overall flowchart of plant pest identification

$$\mathrm{DWC}\left(W_d, y\right)_{(i,j)} = \sum_{h,l}^{H,L} W_{d(h,l)} \odot y_{(i+h,j+l)}, \tag{2}$$

$$\mathrm{PWC}\left(W_p, y\right)_{(i,j)} = \sum_{m}^{M} W_m \times y_{(i,j,m)}, \tag{3}$$

where (i, j) index position of the image, $\odot$ is the element-wise multiplication, W is the weights of convolution kernel, and y denotes the input image. Consequently, the DWSC can be calculated by

$$\mathrm{DWSC}\left(W_d, W_p, y\right)_{(i,j)} = \mathrm{PWC}_{(i,j)}\left(W_p, \mathrm{DWC}\left(W_d, y\right)_{(i,j)}\right). \tag{4}$$

MobileNet (Howard et al. 2017) is the benchmark network of mobile-nets. Among the mobile-nets, MobileNet-V2 addresses the vanishing gradient problem during the training process and has a certain improvement over V1. Instead of the ReLU activation function, MobileNet-V2 applies a linear bottleneck to avoid damage to features, and has a smaller model size compared to the other start-of-the-art methods, as shown in Table 1. It is the most optimum deep learning architecture till date (Sandler et al. 2018; Shen et al. 2019). As a consequence, the pre-trained MobileNet-V2 was chosen as the foundation network in our work to identify plant pest types.

Table 1 The sizes of commonly used CNN models

Types	Networks	Model size (MB)	Parameters (million)	Depth
Classical deep CNNs	VGG19	549	143	26
	Inception V3	92	23.8	159
	ResNet50	98	25.6	–
	DenseNet121	33	8.1	121
Lightweight CNNs	Xception	88	22.9	126
	MobileNet	16	4.2	88
	NASNetMobile	23	5.3	–
	MobileNet V2	14	3.5	88
	EfficieNet-B0	29	5.3	–

Spatial and channel attention mechanism (SCAM)

The attention mechanism is a selective mechanism that can focus on some significant information in the context while ignoring other unwanted information at the same time (Li et al. 2019). The goal of attention mechanism is to highlight the diverse influence of different input data on output data (Vaswani et al. 2017), and much recent research has been performed on attention mechanism for its excellent performance of feature extraction (Anderson et al. 2018; Wang et al. 2017; Woo et al. 2018). In general, spatial attention mechanism (SM) (Nan and Xi 2019) and channel attention mechanism (CM) are two kinds of most-used attention mechanisms in deep learning. Among the two attention mechanisms, SM is well at probing the location of the object in the feature map, while CM is prominent when searching for the desired target in multiple feature maps. Concretely, the calculations involving CM and SM are described below.

Suppose that an intermediate feature map $F \in R^{W \times H \times C}$ is input to the attention module, and thus the CM will infer a 1D channel attention map $M_C \in R^{C \times 1 \times 1}$ and SM will generate a 2D spatial attention map $M_s \in R^{1 \times H \times W}$. Mathematically, the calculation output of the attention module can be summarized by:

$$F_{\mathrm{att}} = \mathrm{CM}(F) + \mathrm{SM}(F) = M_C(F) \otimes F + F \otimes M_s(F), \tag{5}$$

where $\otimes$ represents the element-wise multiplication. The CM and SM both adopt the average pooling and maximum pooling to calculate the input feature F, and the CM adds the results of these two pooling operations to acquire the final feature, as expressed in Eq. (6).

$$M_C(F) = \delta(\mathrm{MLP}(\mathrm{avgpool}(F)) + MLP(\max\mathrm{pool}(F))), \tag{6}$$

where δ denotes the sigmoid activation function, MLP is a multilayer perceptron, and $F \in R^{c \times w \times h}$. The SM concatenates the final features obtained from CM and performs the convolution by a standard convolution layer, producing the spatial attention map. It is computed by:

$$M_S(x) = \delta(f^{n \times n}([\mathrm{avgpool}(F); \max\mathrm{pool}(F)])), \tag{7}$$

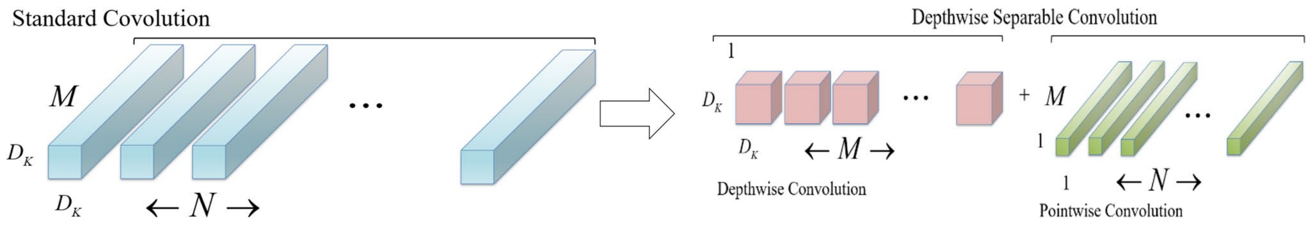


Fig. 3 Standard convolution and depthwise separable convolution

where $f^{n\times n}$ denotes a convolution operation with the kernel size of $n\times n$. Particularly, taking the parallel or sequential manner into account, the sequential arrangement gives a better result than a parallel arrangement in practice (Woo et al. 2018).

Classification activation map

Classification activation map (CAM) is introduced by Zhou et al. (2016) to visualize the regions that include distinctive object parts, and it is produced by the interaction among the final convolution layer, the global average pooling (GAP) layer, and the classification layer of CNNs. By projecting the weights of the output layer onto the convolutional feature map, CAM can learn the significance of the image regions (Zhou et al. 2016) and indicate the highest activated regions used by the CNNs for image classification.

Let $f_k(x, y)$ indicate the activation of unit k in the last convolutional layer at a spatial location (x, y) of an input image. Thus, the result of GAP for unit k is calculated by

$$F^k = \sum\nolimits_{x,y} f_k(x, y). \tag{8}$$

Consequently, for a given category c, the input to the Softmax is expressed in Eq. (9)

$$S_c = \sum\nolimits_k \omega_k^c F^k, \tag{9}$$

Among them, ω_k^c indicates the importance (weight) of F^k for the category c, and the Softmax output for the category c can be computed in Eq. (10).

$$P_c = \exp(S_c) / \sum\nolimits_c \exp(S_c). \tag{10}$$

Here, the bias term is ignored by explicitly setting the input bias of the Softmax to 0 because it has little to no influence on the classifying effect.

Substituting Eq. (8) into the class score S_c of Eq. (9), we can obtain:

$$S_c = \sum\nolimits_k (\omega_k^c \cdot \sum\nolimits_{x,y} f_k(x, y)) = \sum\nolimits_{x,y} \sum\nolimits_k w_k^c \cdot f_k(x, y). \tag{11}$$

Thus, M_c is defined as the CAM for class c, where each spatial element can be written in

$$M_c(x, y) = \sum\nolimits_k \omega_k^c f_k(x, y). \tag{12}$$

As a consequence, $M_c(x, y)$ directly reveals the importance of the activation at spatial grid (x, y), which results in the image being categorized into class c. By simply upsampling the CAM to the size of the input image, the image regions that are most related to a specific category can be identified accordingly.

Proposed approach

Atten-MobNet model

As described previously, MobileNet-V2 is a kind of lightweight CNNs, and it not only improves the calculation speed but also decreases the complexity of the model (Liu et al. 2019; Rabano et al. 2018). Although the accuracy may be no more than that of state-of-the-art deep CNNs, it has a smaller model volume and fewer parameters. On the other hand, the attention module can focus on significant information while ignoring the needless data received in the context, which is useful to extract the crucial features of plant pest images. In particular, it realizes the maximum re-use of channel interdependencies and makes good use of both the channel-wise attention and spatial attention to learn the significance of plant pest features, which helps recognize the plant pests under heterogeneous background conditions. To this end, the attention mechanism emphasizes the differential influence of different input data on output data. Moreover, CAM can ascertain the importance of the region by projecting the weight of the output layer onto the convolutional feature map. Therefore, motivated by the impressive performance, the MobileNet-V2 paired with the attention mechanism and CAM were selected in our approach. Using the method of transfer learning (TL) (Pan and Yang 2009), we transferred the common knowledge of MobileNet-V2 pre-trained from ImageNet (Russakovsky et al. 2015) and added the attention mechanism along with the CAM module in the pre-trained model to create a new network, which we termed the Atten-MobNet, for identifying the plant pest types. More concretely, to enhance the learning capability for minute plant pest features, we modified the network structure of conventional MobileNet-V2 and loaded the pre-trained weights from ImageNet (https://keras.io/api/applications/). The top layer of the MobileNet-V2 was truncated and the attention module was embedded in the pre-trained network, which was followed by an additional 1280 × 1 × 1 convolution layer for the extraction of high-dimensional features. The shortcut connection approach was used for the incorporation of the attention module to the MobileNet-V2 backbone network, and thus the output of the backbone layers is available as input to the attention module using a summation operation rather than being concatenated. In addition, one Batch Normalization (BN) layer was introduced to make the network converge faster and more stable. Subsequently, a global average pooling layer was substituted for the completely linked layer, and at the tail of the modified networks, a new fully connected (FC) Softmax layer with the practical number of classes was added as the new classification layer, where the CAM algorithm was employed for the activated region presentation. Last but not least, the L2 regularization with the *lambda* hyper-parameter of $l_2 = 10^{-2}$ was utilized in this FC layer to prevent the overfitting risk. In this manner, the newly formed network called Atten-MobNet is used to identify the plant pest types. Note that the parameters of newly extended layers were initialized by following (He et al. 2015), and the channel first and then the spatial order was employed in the attention module. Figure 4 depicts the network structure and relevant parameters are presented in Table 2.

Model training

As is well known, transfer learning is particularly important in deep CNNs because deep learning-based CNN algorithms require plenty of labelled data to train the model, while collecting a large number of labelled data in a domain is undoubtedly a challenging task. Hence, the scheme of transfer learning is naturally employed in practical application scenarios and has increasingly become the leading approach. In this work, the two-stage transfer learning approach was adopted in the process of model training. The first stage only trains the weight of newly extended layers while the bottom convolution layers are kept frozen with the weight pre-trained from ImageNet. The second stage retrained (fine-tuned) all the weight parameters using the target dataset by loading the model trained in the first stage. More specifically, the optimum model was trained by applying the following procedures.

Initially, the common image knowledge learned from ImageNet was transferred and the parameters of additional layers were trained in the network. Based on the transfer learning, the new auxiliary layers were trained using the target dataset and the common knowledge of images was transferred to the Atten-MobNet model by freezing all the weight of the bottom convolution layers. In this step, Adam optimizer (Kingma and Ba 2014) was utilized for updating the weight, as computed by

$$\theta_{c+1} = \theta_c - \eta \times \hat{b}_c / (\sqrt{\hat{s}_c} + \varepsilon), \tag{13}$$

where θ denotes the weight matrix, c is the index of categories, η is the learning rate, $\hat{b}_c$ and $\hat{s}_c$ represent the bias-corrected first and second moments, respectively.

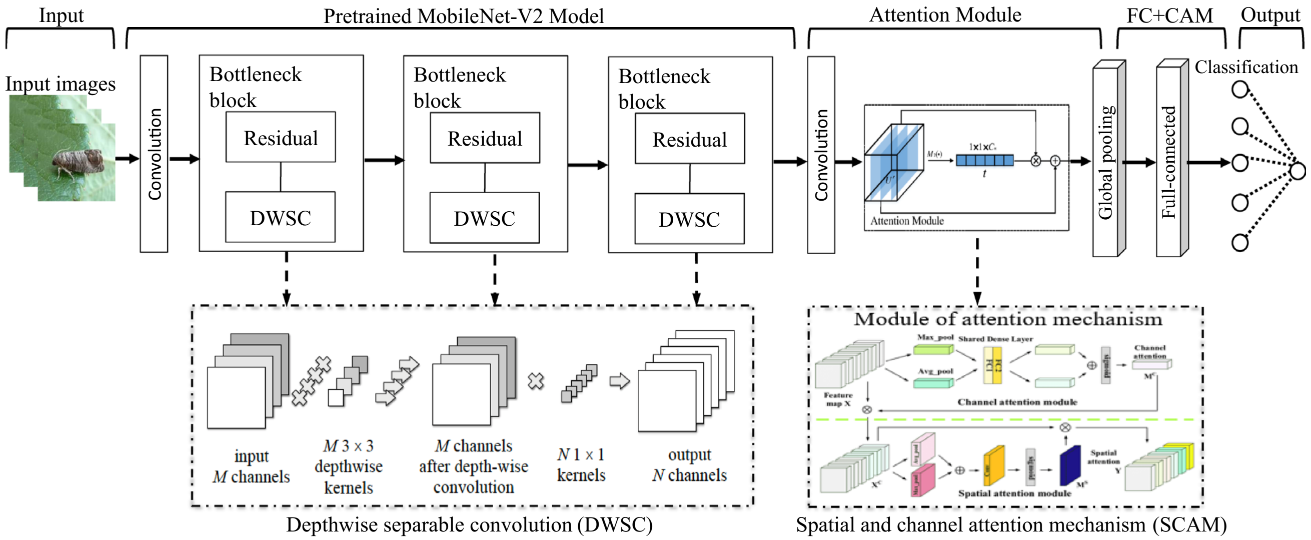
Then, by loading the model trained in the first stage, the weight parameters were injected into the network, and on this basis, the model was retrained using the target dataset. With this method, the optimum model of plant pest recognition was obtained eventually. Stochastic gradient descent (SGD) optimizer (Ghazi et al. 2017) was utilized to update the weight.

$$\theta_{c+1} = \theta_i - \eta \times (\partial L(\theta) / \partial \theta), \tag{14}$$

where $\partial L(\theta)/\partial\theta$ denotes the partial derivative of the loss function $L(\cdot)$ to the weight θ. In general, this two-stage progressive training approach can help the model discover the large-scale structures of plant pest images first, and then shift its focus on detailed features gradually, with no need to learn all scales meantime.

In addition, taking the fine-grained pest features and multi-classification tasks into account, the classical Focal-Loss function (Lin et al. 2017) was enhanced and employed in our model, as expressed in Eqs. (15, 16).

$$\mathrm{FL}(p_c) = -\sum_{c}^{C} \theta_c (1 - p(c|x))^{\gamma} y_c \log(p(c)), \tag{15}$$



Depthwise separable convolution (DWSC)
Spatial and channel attention mechanism (SCAM)

Fig. 4 The schematic diagram of the proposed Atten-MobNet architecture

Table 2 The key parameters of the network

Module (type)	Input shape	Expansion factor	Output shape	Repeated times	Stride
Input layer	224 × 224 × 3	–	224 × 224 × 3	1	–
ZeroPadding2D	224 × 224 × 3	–	225 × 225 × 3	1	–
Conv2d	225 × 225 × 3	–	112 × 112 × 32	1	2
Boottleneck	112 × 112 × 32	1	112 × 112 × 16	1	1
Boottleneck	112 × 112 × 16	6	56 × 56 × 24	2	2
Boottleneck	56 × 56 × 24	6	28 × 28 × 32	3	2
Boottleneck	28 × 28 × 32	6	14 × 14 × 64	4	2
Boottleneck	14 × 14 × 64	6	14 × 14 × 96	3	1
Boottleneck	14 × 14 × 96	6	7 × 7 × 160	3	2
Boottleneck	7 × 7 × 160	6	7 × 7 × 320	1	1
Conv2d 1 × 1	7 × 7 × 320	–	7 × 7 × 1280	1	1
BatchNormalization	7 × 7 × 1280	–	7 × 7 × 1280	1	1
Conv2d 1 × 1	7 × 7 × 1280	–	7 × 7 × 1280	2	1
GlobalAvgPool	7 × 7 × 1280	–	1280	2	1
Add 1:Add	1 × 1 × 1280, 1 × 1 × 1280	–	1 × 1 × 1280	1	–
Multiply1	7 × 7 × 1280, 1 × 1 × 1280	–	7 × 7 × 1280	1	–
Concatenate1	7 × 7 × 1, 7 × 7 × 1	–	7 × 7 × 2	1	–
Conv2d	7 × 7 × 2	–	7 × 7 × 1	1	1
Multiply2	7 × 7 × 1280, 7 × 7 × 1	–	7 × 7 × 1280	1	–
Add 2:Add	7 × 7 × 1280, 7 × 7 × 1280	–	7 × 7 × 1280	1	–
GlobalAvgPool	7 × 7 × 1280	–	1280	1	1
Softmax	1 × 1 × k	–	k	1	1

$$y_c = \begin{cases} 0, & c \neq \text{true\_type} \\ 1, & c = \text{true\_type} \end{cases}, \tag{16}$$

where p_c denotes the predicted probability distribution, and γ is the modulating parameter.

Experimental results and analysis

Different types of tools have been used to perform the experiments, and the image pre-processing work is primarily accomplished by the Photoshop software. The data augmentation and deep learning algorithms are conducted using the software of Anaconda3 (Python 3.6), where the OpenCV-python3, Tensorflow, Keras libraries, and others are employed and accelerated by GPU. The experimental hardware configurations consist of Intel® Xeon(R) E5-2620 CPU (2.10 GHz), 64-GB memory, and NVIDIA GeForce RTX 2080 graphics card, which are used for the software running.

Experiments on an open dataset

As mentioned previously, Li's dataset is an open pest image dataset and has some similarities with our collected plant pest images. Therefore, we first conducted experiments on this dataset to investigate the performance of the proposed approach. It is worth noting that the distribution of sample images is unbalanced, and the number of images in some categories is relatively small. Therefore, the data augmentation scheme was adopted to enrich sample images for the categories with small sample data. The enhanced DCGAN paired with the traditional methods, such as color jittering, random angle rotation, random horizontal or vertical flipping, shearing, and scale transformation, was used in our data augmentation scheme here.

To our best knowledge, many successful CNN models assign a relatively large image size, such as 224 × 224 pixels for the input to improve the performance of the model, while the architecture of classical DCGAN (Radford et al. 2015) is designed for generating 64 × 64 pixel images because training higher resolutions are unstable as one network becomes stronger than the other. Besides, other DCGAN variants, such as ProGAN (Karras et al. 2017), can generate higher-resolution insect pest images, but it does not synthesize images with a rich feature set representing the insect pest symptoms well enough. Hence, the classical DCGAN model is modified and the assignment of input size is set as 224 × 224 pixels instead of the original 64 × 64 pixels to enlarge the resolution of images and fit the model. Then, a 128 × 64 × 3 convolution block followed by a 32 × 3 convolution block was embedded in the generator module. Correspondingly, the 32 × 3 convolution block followed by a 64 × 128 × 3 convolution block was incorporated in the discriminator module. Further, the input shape of the discriminator module was set 224 × 224 × 3 as well. The hyper-parameters of the modified DCGAN were: a batch size of 16, a learning rate of 1×10^{-4}, an epoch of 5×10^{5}, and the Adam optimizer. With this method, the synthetic images were generated and no less than 1000 training samples were guaranteed for each category. Figures 5 and 6, respectively, present the sample images augmented by the traditional methods and the enhanced DCGAN approach.

Subsequently, using the proposed approach, we performed the model training and testing on the pest image dataset. The training and testing sets were divided with the proportion of 8:2. In particular, to know how the proposed approach would perform on new unseen data, a certain number of raw images were remained to validate the effectiveness of the model (about 1/10 of the original number in each category), as shown in Table 4. Also, to compare the proposed approach with other state-of-the-art methods, we chose five influential small CNNs, including MobileNet-V1, MobileNet-V2, NASNetMobile (Zoph et al. 2018), Efficient-Net-B0 (Tan and Le 2019), and DenseNet121 (Huang et al. 2017), as the baseline algorithms to perform the comparative analyses. Based on the transfer learning, the top layers of all the models were abandoned and a new completely linked Softmax layer with the practical number of categories was used as the classification layer. By this means, these CNN models were created and the parameters were initialized

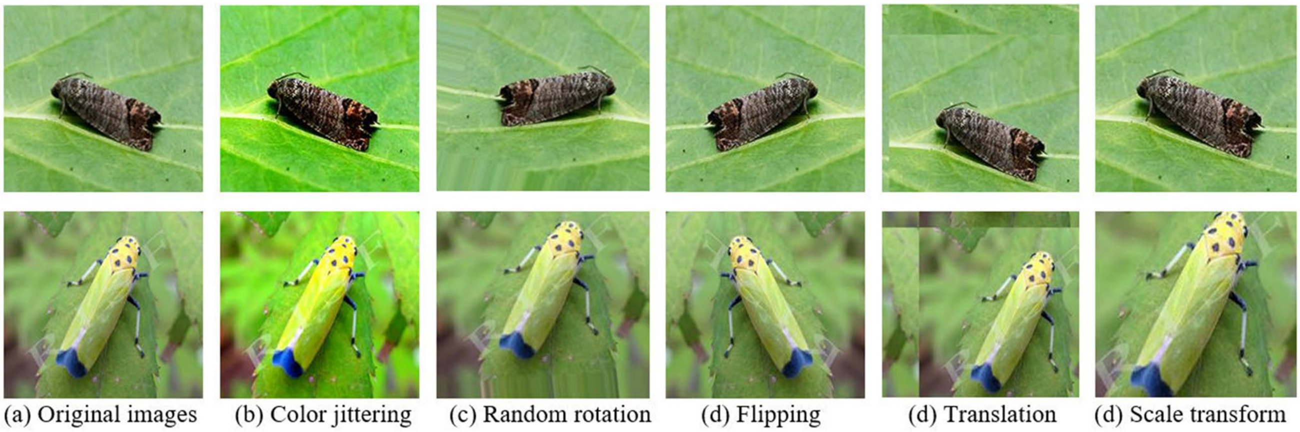


Fig. 5 The augmented pest images by the traditional geometric transformation

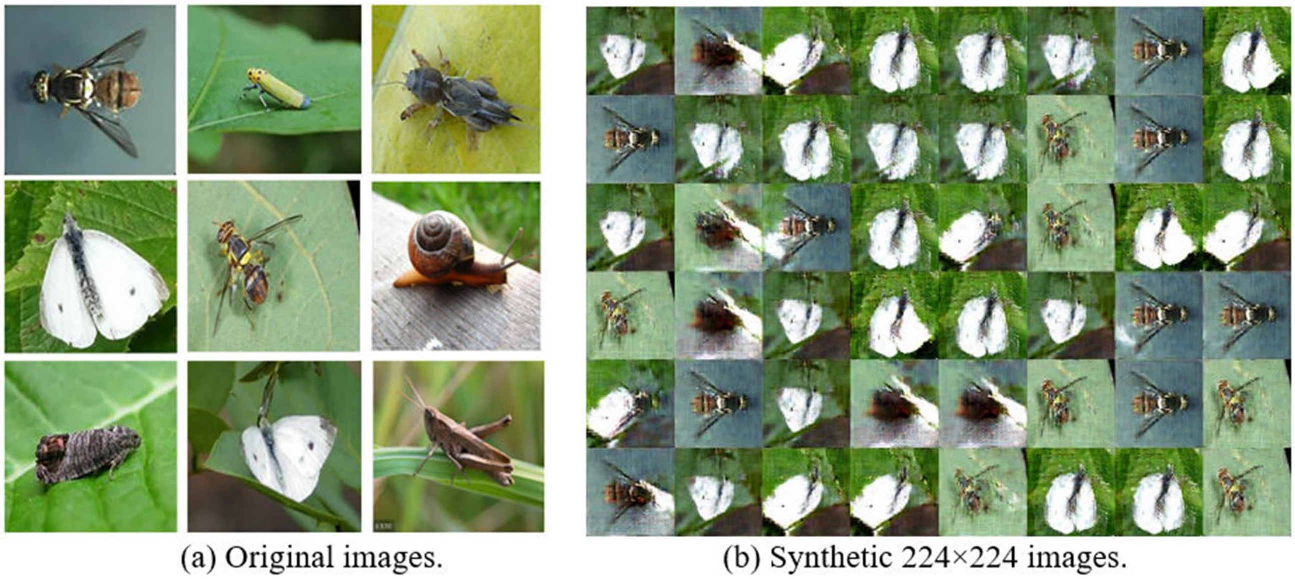


Fig. 6 The synthetic pest images by enhanced DCGAN

through injecting the weights pre-trained from ImageNet. The model was trained for 30 epochs, with a batch size of 64 and a learning rate of $\mathrm{lr} = 10^{-4}$. Particularly, to ensure fairness, all the CNN models were trained in the same manner. Thus, a series of experiments were performed on this open-source dataset, and Table 3 presented the training and test results.

Looking at Table 3, it is apparent that the proposed procedure delivers comparable effects compared to other state-of-the-art methods. After training for 10 epochs and 30 epochs, the training accuracy of the proposed procedure separately reaches 99.07% and 99.81%, and particularly the test accuracy attains 93.55% and 95.08%, respectively. These are the top performances of all the algorithms except for DenseNet-121, which is a deep CNN with relatively large volume and consumes more computational resources. In contrast, the proposed Atten-MobNet is memory-efficient, and the model volume is only about half of that of DenseNet-121. It achieves an increased efficacy and relatively high accuracy in the experiments, even though the optimum classifier is adopted. The main reason behind the outperformance of the proposed procedure is that the Atten-MobNet integrates the merits of the lightweight MobileNet-V2 as well as the spatial and channel attention mechanism to train the model. Also, the two-stage transfer learning and optimized loss function are performed in the model training. These measures enhance the capability of learning pest image features and make the optimum weight parameters be obtained for the model. By comparison, the other methods are individual networks; although the transfer learning approach is adopted and the model parameters were initialized with pre-trained weights rather than inferring from scratch, the optimum results are not achieved for these models. Hence, using the model trained by the proposed procedure, the new unseen plant pest images are chosen for the test of insect pest identification. Figure 7 depicts the confusion matrix and the receiver operating characteristic (ROC) curve of identification results.

Considering the accurate detections and misdetections, the performances of the proposed procedure and other state-of-the-art CNN methods discussed in this study are measured using different metrics including Accuracy, Sensitivity (Recall), Specificity, Precision, and F1-Score, as defined in Eqs. (17–21).

$$\mathrm{Accuracy} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}, \tag{17}$$

$$\mathrm{Sensitivity} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}, \tag{18}$$

$$\mathrm{Specificity} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}}, \tag{19}$$

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}, \tag{20}$$

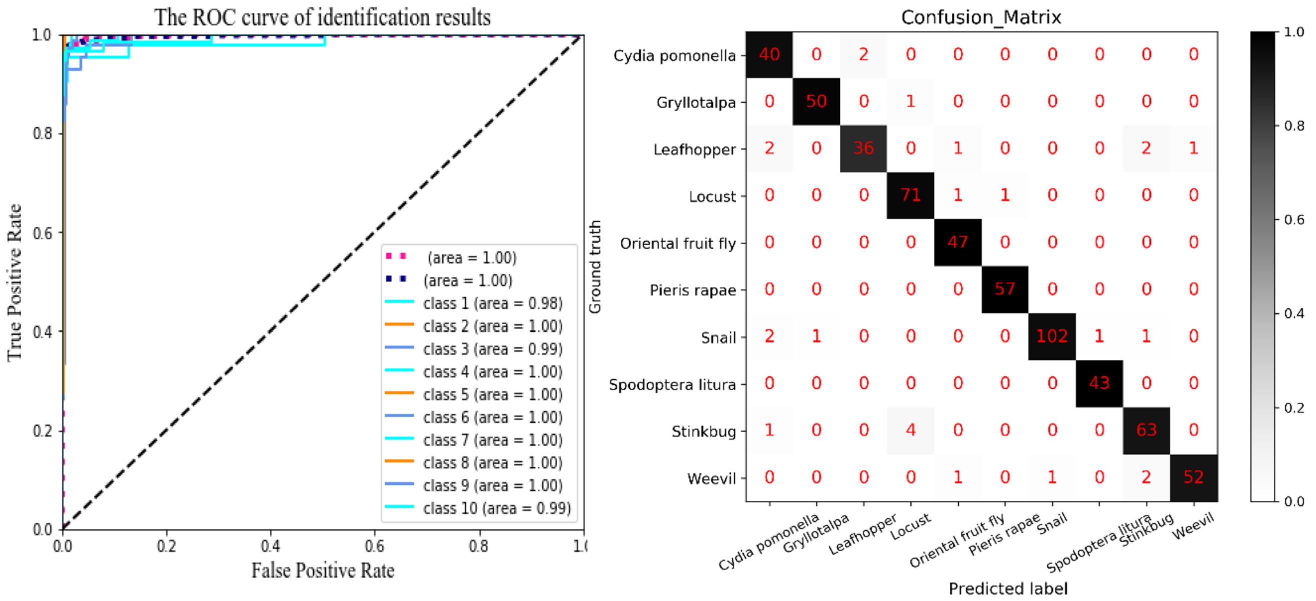
$$F1 - \mathrm{Score} = \frac{2\mathrm{TP}}{2\mathrm{TP} + \mathrm{FN} + \mathrm{FP}}, \tag{21}$$

where *TP*, *TN*, *FP*, and *FN* values are defined below. TP (true positive) denotes the number of positive samples that are accurately identified; FP (false positive) is the number of negative samples that are mistakenly identified; TN (true negative) is the number of negative samples that are identified correctly; FN (false negative) is the number of positive samples which are identified incorrectly. Table 4 presents the actual analysis of the results which are investigated with different metrics.

As can be visualized in Fig. 7a, the ROC curves of each class are closed to the upper left corner of the figure, which shows ideal operating characteristics and indicates the validity of the proposed procedure. Moreover, it can be reflected by the confusion matrix of Fig. 7b. Most of the pest samples in each class have been correctly recognized by the proposed procedure. For example, 40 samples are accurately identified in 42 "*Cydia pomonella*" pest images except for 2 misdetections, and the *Accuracy* realizes 98.90%. Similarly, the correct number is 50 for the recognition of 51 "Gryllotalpa" pest samples, and the

Table 3 The training accuracy and loss of different methods on the open dataset

Pre-trained models	10 epochs			30 epochs			
	Training accuracies %	Test accuracies %	Training losses	Training accuracies %	Test accuracies %	Training losses	Test losses
MobileNet-V1	98.95	91.80	0.6426	99.82	93.03	0.1617	4.2056
MobileNet-V2	98.80	92.58	0.6331	99.85	94.47	0.1047	3.9458
NASNetMobile	98.06	92.01	1.1761	99.52	92.77	0.4511	4.7411
EfficientNet-B0	94.23	91.02	3.3761	97.29	93.16	1.4766	3.5440
DenseNet-121	99.64	95.90	0.3555	99.92	96.48	0.0827	1.6645
Proposed procedure	99.07	93.55	0.5926	99.81	95.08	0.2824	4.2057



Ground truth \ Predicted label	Cydia pomonella	Gryllotalpa	Leafhopper	Locust	Oriental fruit fly	Pieris rapae	Snail	Spodoptera litura	Stinkbug	Weevil
Cydia pomonella	40	0	2	0	0	0	0	0	0	0
Gryllotalpa	0	50	0	1	0	0	0	0	0	0
Leafhopper	2	0	36	0	1	0	0	0	2	1
Locust	0	0	0	71	1	1	0	0	0	0
Oriental fruit fly	0	0	0	0	47	0	0	0	0	0
Pieris rapae	0	0	0	0	0	57	0	0	0	0
Snail	2	1	0	0	0	0	102	1	1	0
Spodoptera litura	0	0	0	0	0	0	0	43	0	0
Stinkbug	1	0	0	4	0	0	0	0	63	0
Weevil	0	0	0	0	1	0	1	0	2	52

(a) ROC curve

(b) Confusion matrix

Fig. 7 ROC curve and confusion matrix of identifying plant pests on open dataset

Table 4 The identification results of different insect pest types

No.	Category names	Identified samples	Correct samples	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1-Score (%)
1	Cydia pomonella	42	40	98.80	95.23	99.08	88.88	91.95
2	Gryllotalpa	51	50	99.65	98.03	99.81	98.03	98.03
3	Leafhopper	42	36	98.63	85.71	99.63	94.73	90.00
4	Locust	73	71	98.80	97.26	99.02	93.42	95.30
5	Oriental fruit fly	47	47	99.48	100.00	99.44	94.00	96.90
6	Pieris rapae linnaeus	57	57	99.82	100.00	99.81	98.27	99.13
7	Snail	107	102	98.97	95.32	99.79	99.02	97.14
8	Spodoptera litura	43	43	99.82	100.00	99.81	97.72	98.85
9	Stinkbug	68	63	98.29	92.64	99.03	92.64	92.64
10	Weevil	56	52	99.14	92.85	99.81	98.11	95.41
–	Average	–	–	99.14	95.73	99.52	95.73	95.73

Accuracy is 99.65%, as shown in Table 4. Consequently, a total of 561 instances are successfully recognized in 586 samples, and the average *Accuracy* reaches 99.14%, which demonstrates the proposed procedure has significant capability to recognize diverse plant pests. Furthermore, the performance investigation of our proposed procedure compared with the results reported in the literature (Li et al. 2020; Li and Yang 2020) is displayed in Table 5. The highest accuracy used models, and corresponding references are included in the table. Experimental findings show that the proposed procedure can effectively identify plant pest types in natural scenes.

Experiments on our local dataset

Similar to the above experiments conducted on the publicly available dataset, the proposed procedure is further tested on our collected local pest image dataset, which is captured in real-field wild scenarios with complex backdrop conditions and inconsistent lighting strengths. For example, photographs are taken in the background of various surroundings with cluttered leaves or grasses, soils of different colors, and sometimes the photographers' fingers, etc. Besides, the images photographed at different times are with varied weather conditions, such as overcast, cloudy, and sunny. To

Table 5 Comparison of our proposed procedure with other methods

ID	Researches	Models	Identification accuracy (%)
1	Li et al. (2020)	Fine-tuned GoogLeNet	98.91
2	Li and Yang (2020)	Few-shot model (CNN)	96.20
5	Proposed procedure	Atten-MobNet	99.14

diversify the images and prevent the overfitting problem, data augmentation techniques, including random angle rotation, random horizontal or vertical flipping, color jittering, image translation, and scale transformation, were applied to generate new synthetic images for enriching the dataset. The main parameters of data augmentation algorithms used in the experiment were: the rotation operation was conducted with an arbitrary angle in the interval of [0, 360°], the image translation was performed in the range of ± 20%, color jittering was changing the saturation, brightness, and contrast of color with a random adjustment factor in the interval (0, 3.1), and the scale transformation was implemented with the ratio of 0.8 to 1.2. By doing this, the new synthetic images were generated and at least 200 training samples were ensured for each category. Figure 8 displays the partial augmented sample images.

Based on the proposed procedure, we performed the model training and validation on the collected local pest image dataset, and all the newly synthesized images were uniformly re-sized to the fixed-dimension of 224 × 224 pixels to fit the models. Apart from remaining a certain number of raw images for verifying the validity of the models, the sample images were divided into a training set and a test set according to the ratio of 8:2. That is to say, the sample images were separately split into the training and test sets to train the models and determine whether the models were overfitting, while the new images outside modeling were used to evaluate the models. Referring to the experimental procedures implemented on the open-source dataset, the new unseen images were chosen to recognize the plant pests using the model trained by the proposed approach. Note that "unseen" images in this context indicate unknown insect pest images that have never been utilized by the CNN models during the training and test. Figure 9 depicts the final recognition results of local plant pests and an exhaustive measurement analysis of different metrics is presented in Table 6.

As seen in Fig. 9a, the recognition results of diverse insect pests are depicted in the ROC curves, and the AUC (area under the curve) of each class is no less than 0.93. The most of samples have been successfully recognized by the proposed procedure, as depicted in Fig. 9b. Except for 1 misdetection instance, all the plant pest samples are correctly recognized in the insect pest types of both "Clania minuscula butler" and "Cushion scale", and the recognition accuracy realizes 97.29% and 89.17%, respectively. Also, 9 instances are accurately identified in 11 "Dappula tertia templeton" samples and the recognition accuracy reaches 89.18%. In summary, a total of 29 instances are correctly recognized in 37 unseen pest samples. The average Accuracy, Sensitivity, and Specificity achieve 92.79%, 78.37%, and 95.67%, respectively, as shown in Table 6. The key explanation for the accurate recognition of insect pest images in heterogeneous backgrounds is that the proposed Atten-MobNet incorporates the merits of the attention module and MobileNet-V2 backbone network, which enhances the learning of significant information and suppresses the interference signals, realizes the maximum re-use of inter-channel relation and space-wise point features, and extracts the high-level features of crop pest images. Moreover, the two-stage transfer learning was implemented in the model training, which makes the model learn the general knowledge of insect pest images and obtain the optimum weight parameters. This

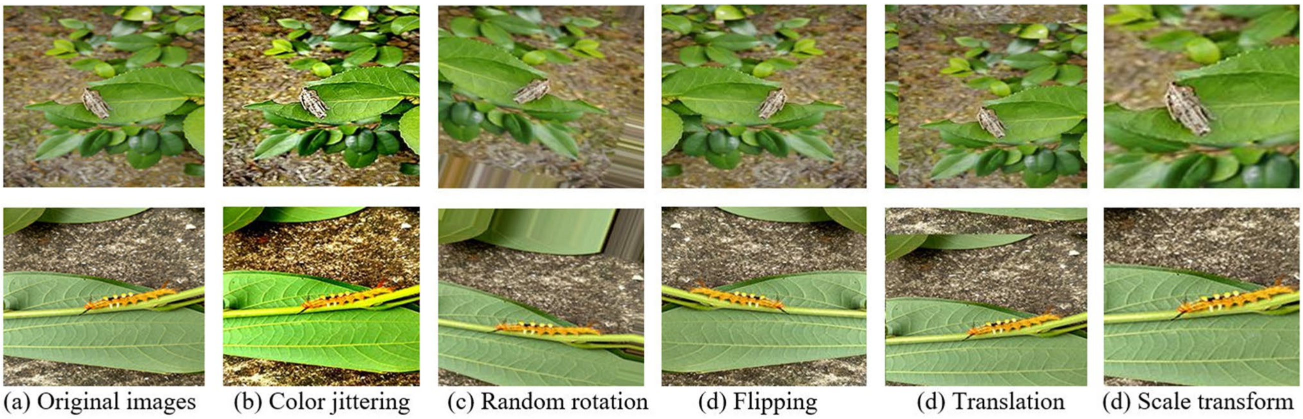
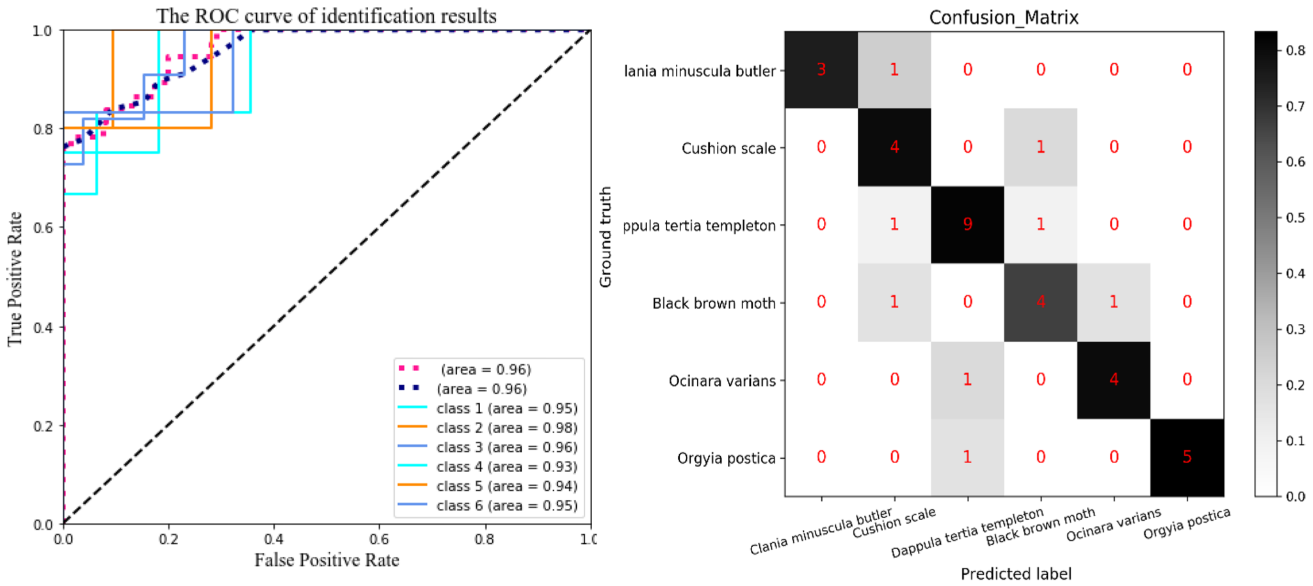


Fig. 8 The augmented samples of plant pest images



(a) ROC curve

(b) Confusion matrix

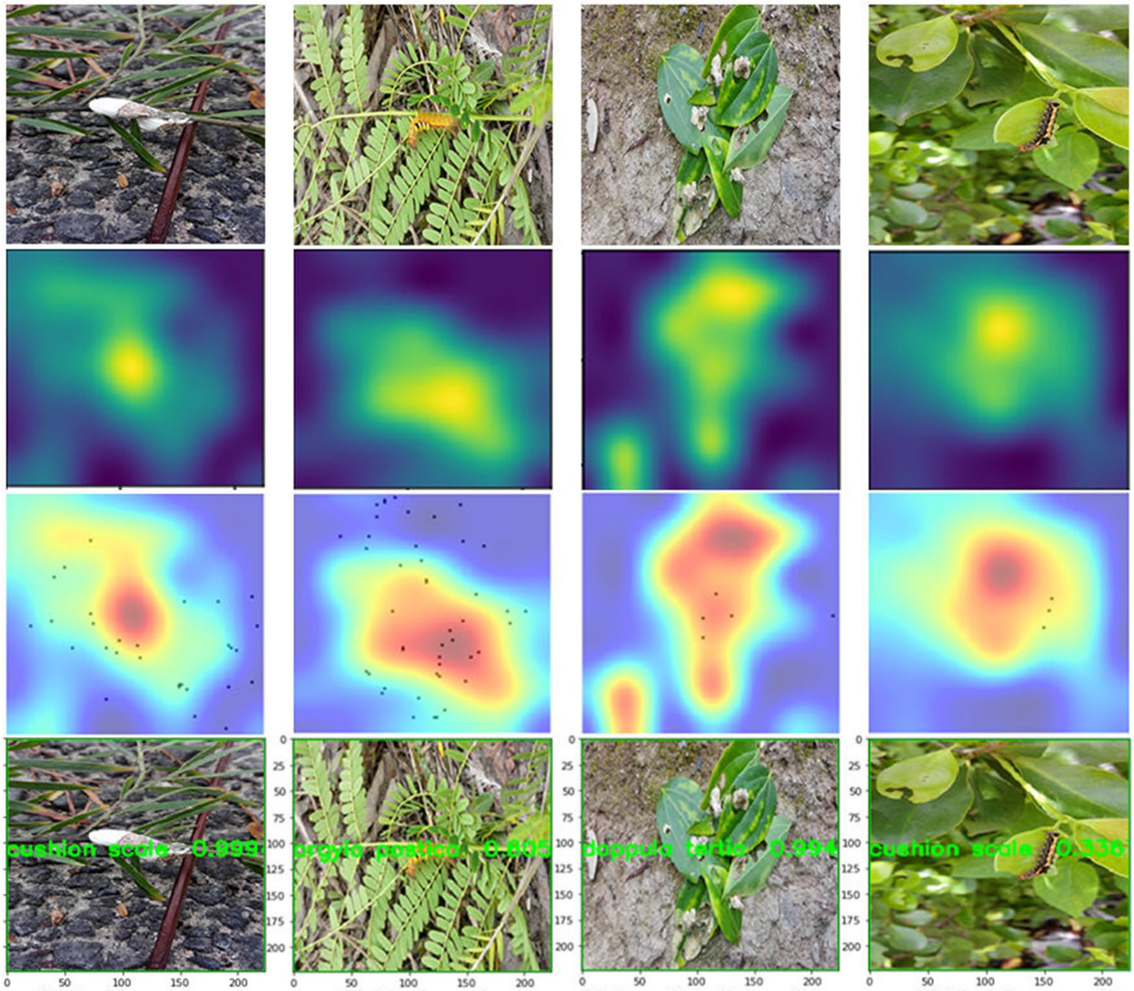
Fig. 9 ROC curve and confusion matrix of plant pest identification on local dataset

Table 6 The identified results of different plant pest classes

No	Insect pest types	Identified samples	Correct samples	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1-Score (%)
1	Clania minuscula butler	4	3	97.29	75.00	100.00	100.00	85.71
2	Cushion scale	5	4	89.18	80.00	90.62	57.14	66.67
3	Dappula tertia templeton	11	9	89.18	81.81	92.30	81.81	81.81
4	Black brown moth	6	4	89.18	66.67	93.54	66.67	66.67
5	Ocinara varians	5	4	94.59	80.00	96.87	80.00	80.00
6	Orgyia postica	6	5	97.29	83.33	100.00	100.00	90.91
–	Average	–	–	92.79	78.37	95.67	78.37	78.37

kind of progressive training helps the model discover the large-scale structures of insect pest images first, and then shifts its attention to detailed features of insect pest images gradually, improving the identification accuracy of plant pest images in heterogeneous background conditions. Conversely, there are also some misclassifications, such as two samples in the plant pest types of “Dappula tertia templeton”, which are caused by the extreme clutter field backdrops and inconsistent illumination intensities. Additionally, individual low-quality, blurred photographs may cause incorrect classification as well. Figure 10 displays the partially identified samples.

As seen in Fig. 10, the top images are the original insect image samples, the images in the second layer are the positioning images presented by CAM (heatmap), the third-layer images are the overlay images synthesized by heatmap and original images, and the bottom images are the results identified by the proposed procedure. From Fig. 10, we can see that the important positions of most sample images for classification are well visualized by the CAM module and most of the insect pest types have been successfully identified by the proposed procedure except for individual images. Such as Fig. 10a, the actual plant pest type of this sample is the “Cushion scale”, which is accurately identified with the probability of 0.99. Similarly, Fig. 10b, c are all correctly identified by the proposed procedure and the identification probabilities are equal to 0.81 and 0.99 separately. By contrast, as mentioned previously, individual sample images are incorrectly misidentified due to serious clutter backdrop conditions and uneven lighting strengths. For instance, Fig. 10d, which belongs to the "black brown moth", but is mistakenly classified as the "Cushion scale" type with a probability of 0.33. Despite individual misidentifications, the probability value of misclassification is low and most of the plant pests are accurately identified by the proposed procedure. The high accuracy has been attained on the unseen images for a



(a) Cushion scale (b) Orgyia postica (c) Dappula tertia templeton (d) Black brown moth

Fig. 10 The identified samples of different plant pests

number of experiments, which indicates that the proposed Atten-MobNet approach has a significant capability to identify plant insect pests. Based on the experimental findings, it can be concluded that the proposed procedure is successful in identifying plant pest types, and can also be extended to other related fields, such as online defect detection, computer-aided assessment, etc.

Conclusion

Identification and classification of various insect pest types by means of digital images are very important to obtain the high-quality yields of crop products. Deep learning techniques, particularly deep convolutional neural networks, can efficiently and effectively recognize most of the symptom features related to insect pests. In this study, considering both the memory efficiency and model accuracy, we proposed a lightweight network called Atten-MobNet to identify plant pest types. The MobileNet-V2 was chosen as the foundation network architecture of our proposal. Using the approach of transferring learning, the common knowledge of MobileNet-V2 pre-trained from ImageNet was transferred in our model and the spatial and channel attention mechanism as well as the CAM module were embedded in the pre-trained model to create a new network for identifying plant pest types. Further, the loss function was optimized and the two-stage transfer learning was performed for model training. The experimental results show the feasibility and effectiveness of the proposed procedure, and it is successfully accomplished to identify diverse insect pest types on both the publicly available dataset and the local dataset. On the other hand, the pest positioning images were located by the visualization technology of the classical CAM currently. Although the major areas that caused the classification made by the CNN model were revealed, there were still individual false positives of the attention map in the experiments. In future work, we would study the special crop pest activation map to improve the performance and the accuracy of the visual exhibition techniques of crop pest identification. Moreover, we would deploy the model on mobile devices for monitoring crop pests and transplant it to other fields for more real-world applications.

Acknowledgements The work is supported by the National Natural Science Foundation of China under grant No. 61672439 and the Fundamental Research Funds for the Central Universities No. 20720181004. The authors also would like to thank the editors and all the anonymous reviewers for their constructive advice.

Declarations

Conflict of interest The authors declare no conflicts of interest.

References

Anderson P, He X, Buehler C, Teney D, Johnson M, Gould S, Zhang L (2018) Bottom-up and top-down attention for image captioning and visual question answering. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 6077–6086

Chen J, Wang W, Zhang D, Zeb A, Nanehkaran YA (2020) Attention embedded lightweight network for maize disease recognition. Plant Pathol. https://doi.org/10.1111/ppa.13322

Deng L, Wang Y, Han Z, Yu R (2018) Research on insect pest image detection and recognition based on bio-inspired methods. Biosys Eng 169:139–148

Faithpraise F, Birch P, Young R, Obu J, Faithpraise B, Chatwin C (2013) Automatic plant pest detection and recognition using k-means clustering algorithm and correspondence filters. Int J Adv Biotechnol Res 4:189–199

Gadekallu TR, Rajput DS, Reddy MPK, Lakshmanna K, Bhattacharya S, Singh S, Alazab M (2020) A novel PCA–whale optimization-based deep neural network model for classification of tomato plant diseases using GPU. J Real-Time Image Proc 342:1–14

Gassoumi H, Prasad NR, Ellington JJ (2000) Neural network-based approach for insect classification in cotton ecosystems. In: International Conference on Intelligent Technologies, pp 13–15

Ghazi MM, Yanikoglu B, Aptoula E (2017) Plant identification using deep neural networks via optimization of transfer learning parameters. Neurocomputing 235:228–235

Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Bengio Y (2014) Generative adversarial nets. In: Advances in neural information processing systems, pp 2672–2680

Hayashi M, Tamai K, Owashi Y, Miura K (2019) Automated machine learning for identification of pest aphid species (Hemiptera: Aphididae). Appl Entomol Zool 54:487–490

He K, Zhang X, Ren S, Sun J (2015) Delving deep into rectifiers: surpassing human-level performance on imagenet classification. In: Proceedings of the IEEE international conference on computer vision, pp1026–1034

Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Adam H (2017) Mobilenets: efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:17040 4861

Huang G, Liu Z, Van Der Maaten L, Weinberger K Q (2017) Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4700–4708

Kaiser L, Gomez AN, Chollet F (2017) Depthwise separable convolutions for neural machine translation. arXiv preprint arXiv: 170603059

Karras T, Aila T, Laine S, Lehtinen J (2017) Progressive growing of gans for improved quality stability and variation. arXiv preprint arXiv:171010196

Kessentini Y, Besbes MD, Ammar S, Chabbouh A (2019) A two-stage deep neural network for multi-norm license plate detection and recognition. Expert Syst Appl 136:159–170

Kingma DP, Ba J (2014) Adam: a method for stochastic optimization. arXiv preprint arXiv:14126980

Li M, Zhang Y, Huang M, Chen J, Feng W (2019) Named entity recognition in chinese electronic medical record using attention mechanism. In: 2019 International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), pp 649–654, IEEE

Li Y, Yang J (2020) Few-shot cotton pest recognition and terminal realization. Comput Electron Agric 169:105240

Lin TY, Goyal P, Girshick R, He K, Dollár P (2017) Focal loss for dense object detection. In: Proceedings of the IEEE international conference on computer vision, pp 2980–2988

Liu Z, Gao J, Yang G, Zhang H, He Y (2016) Localization and classification of paddy field pests using a saliency map and deep convolutional neural network. Sci Rep 6:1–12

Liu X, Jia Z, Hou X, Fu M, Ma L, Sun Q (2019) Real-time marine animal images classification by embedded system based on mobilenet and transfer learning. In: OCEANS 2019-Marseille, pp 1–5

Nan Y, Xi W (2019) Classification of press plate image based on attention mechanism. In: 2019 2nd International conference on safety produce informatization (IICSPI), pp 129–132, IEEE

Nanni L, Maguolo G, Pancino F (2020) Insect pest image detection and recognition based on bio-inspired methods. Ecol Inform 57:101089

Nazki H, Yoon S, Fuentes A, Park DS (2020) Unsupervised image translation using adversarial networks for improved plant disease recognition. Comput Electron Agric 168:105117

Pan SJ, Yang Q (2009) A survey on transfer learning. IEEE Trans Knowl Data Eng 22:1345–1359

Picon A, Seitz M, Alvarez-Gila A, Mohnke P, Ortiz-Barredo A, Echazarra J (2019) Crop conditional convolutional neural networks for massive multi-crop plant disease classification over cell phone acquired images taken on real field conditions. Comput Electron Agric 167:105093

Rabano SL, Cabatuan MK, Sybingco E, Dadios EP, Calilung EJ (2018) Common garbage classification using mobilenet. In: 2018 IEEE 10th International Conference on Humanoid Nanotechnology Information Technology Communication and Control Environment and Management (HNICEM), pp 1–4, IEEE

Radford A, Metz L, Chintala S (2015) Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:151106434

Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Berg AC (2015) Imagenet large scale visual recognition challenge. Int J Comput Vision 115:211–252

Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC (2018) Mobilenetv2: inverted residuals and linear bottlenecks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4510–4520

Shah MA, Khan AA (2014) Imaging techniques for the detection of stored product pests. Appl Entomol Zool 49:201–212

Shen Y, Sun H, Xu X, Zhou J (2019) Detection and positioning of surface defects on galvanized sheet based on improved MobileNet v2. In: 2019 Chinese Control Conference (CCC), pp 8450–8454, IEEE

Shijie J, Peiyi J, Siping H (2017) Automatic detection of tomato diseases and pests based on leaf images. In: 2017 Chinese Automation Congress (CAC), pp 2537–2510, IEEE

Sifre L, Mallat S (2014) Rigid-motion scattering for image classification. PhD thesis

Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:14091556

Tan M, Le Q V (2019) Efficientnet: rethinking model scaling for convolutional neural networks. arXiv preprint arXiv:190511946

Thenmozhi K, Reddy US (2019) Crop pest classification based on deep convolutional neural network and transfer learning. Comput Electron Agric 164:104906

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Polosukhin I (2017) Attention is all you need. In: Advances in neural information processing systems, pp 5998–6008

Wang F, Jiang M, Qian C, Yang S, Li C, Zhang H, Tang X (2017) Residual attention network for image classification. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3156–3164

Wang J, Lin C, Ji L, Liang A (2012) A new automatic identification system of insect images at the order level. Knowl-Based Syst 33:102–110

Wen C, Guyer DE, Li W (2009) Local feature-based identification and classification for orchard insects. Biosys Eng 104:299–307

Wen C, Wu D, Hu H, Pan W (2015) Pose estimation-dependent identification method for field moth images using deep learning architecture. Biosyst Eng 136:117–128

Woo S, Park J, Lee JY, So Kweon I (2018) Cbam: convolutional block attention module. In: Proceedings of the European conference on computer vision (ECCV), pp 3–19

Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A (2016) Learning deep features for discriminative localization. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2921–2929

Zoph B, Vasudevan V, Shlens J, Le QV (2018) Learning transferable architectures for scalable image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 8697–8710