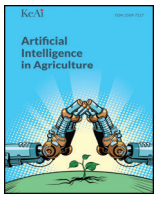
Artificial Intelligence in Agriculture 5 (2021) 13–23

Contents lists available at ScienceDirect

Artificial Intelligence in Agriculture

journal homepage: http://www.keaipublishing.com/en/journals/artificial-intelligence-in-agriculture/



Proximal detecting invertebrate pests on crops using a deep residual convolutional neural network trained by virtual images

Huajian Liu [a,*], Javaan Singh Chahl [b,c]

[a] The Plant Accelerator, Australian Plant Phenomics Facility, School of Agriculture, Food and Wine, University of Adelaide, Waite Campus, Building WT 40, Hartley Grove, Urrbrae, SA 5064, Australia
[b] School of Engineering, University of South Australia, Mawson Lakes 5095, Australia
[c] Joint and Operations Analysis Division, Defence Science and Technology Group, Australia

ARTICLE INFO

Article history:
Received 9 November 2020
Received in revised form 7 January 2021
Accepted 8 January 2021
Available online 10 January 2021

Keywords:
Insect detection
Deep convolutional neural network
Precision agriculture

ABSTRACT

Detecting invertebrate pests on crops at early stages is essential for pest management. Traditionally, traps were used to sample pests and then human experts undertook classification and counting to estimate the levels of infestation, which is subjective, error-prone and labour intensive. Recently, semi-automatic pest detection is possible by using computer vision technologies to classify and count pest samples in laboratories or insect traps, however, the decision made by the laboratory-based or trap-based approaches are still too late for more optimised pest management decisions. Today, precision agriculture needs detection of pests on crops so that real-time actions can be taken or optimised decision can be made based on accurate information of time and location pest occurs. In this study, we used computer vision and machine learning technologies to detect invertebrates on crops in the field. We first evaluated the performances of the state-of-art convolutional neural networks (CNNs) and proposed a standard training pipeline. Facing the challenge of rapidly developing comprehensive training data, we used a novel method to generate a virtual database which was successfully used to train a deep residual CNN with an accuracy of 97.8% in detecting four species of pests in farming environments. The proposed method can be applied to a robotic system for proximal detection of invertebrate pests on crops in real-time.

1. Introduction

Invertebrate pests are difficult to control and the losses caused by them are huge (Oerke 2006). Invertebrate pests can feed on leaves, affect photosynthesis and infect diseases (Nalam et al. 2019). In six major Australian grain crops, the estimated annual loss from invertebrate pests is $359.8 million (Murray et al. 2013). Although there are biological and chemical methods for pest control, biological control has to be carefully used for specific species and heavy doses and uniform pesticide applications have caused serious problems of pesticide resistance, environmental pollution and killing of beneficial species, such as bees (Liu et al. 2016b).

Increasing challenges for pest control and the need for better on-farm efficiencies are driving the implementation of integrated pest management (IPM) which involves pest detection, application of appropriate management methods and recording the result of the management action applied (Boissarda et al. 2008; Kogan and Hilton 2009). Early pest detection and estimation of potential infestation and yield loss are essential for a successful IPM program. Traps, sweep net or beet sheets (GRDC 2014) are widely adopted for systematic pest monitoring. If implemented properly, these sampling methods can successfully estimate populations over the entire area of interest (Yen et al. 2013). However, farmers have to manually collect and count samples and evaluate infestation visually, introducing a degree of subjectivity. Further, manual counting is time-consuming, labour-intensive and error-prone. Computer vision technologies for insect classification were initialized by entomologists (Martineau et al. 2017), and recently have been applied as semi-automatic methods for pest management by counting pests in sample containers or traps (Sun et al. 2017; Zhu et al. 2017). However, using the approach of "collect samples in traps and then estimate", decisions tend to be late, leading to reduced efficiency. Due to their short life cycle, pests might have finished reproduction and the population could have exceeded the threshold for minimal intervention management before adults are detected (Baker and Jennings 2015; Miles 2015). Also, traps cannot provide the exact time and location that pests first occur.

Early detection of pests on crops can ensure appropriate and timely management decisions. First, it can provide an accurate time when crops are attacked; second, it can provide the exact location where crops are infested; third, based on the time, location and species, a robotic system can take actions in real-time, such as selective spraying, or the information could be analysed in an IPM system to make more optimal decisions (Liu et al. 2016b). However, detecting pests in natural

* Corresponding author.
E-mail address: huajian.liu@adelaide.edu.au (H. Liu).

https://doi.org/10.1016/j.aiia.2021.01.003

environments is challenging. Natural sunlight is unstable and changes from time to time. Even using artificial illumination, the leaves or branches could block the light, resulting in uneven illumination (Liu et al. 2017c). The backgrounds of natural farming fields are complex and unpredictable with different types of soil and plant residues. To survive, many pests are camouflaged and they have the same colour and morphological features as their living environments (Liu et al. 2016a; Liu et al. 2017a), which raises challenges for detection. There are very limited studies that have been reported for detecting invertebrate pests in natural farming fields. The recent studies of Chahl and Liu (2018) demonstrated the possibility of detecting pests in natural farming environments using computer vision technologies. They could detect common invertebrate pests on green leaves using multispectral images of ultraviolet, blue, green, red and near-infrared with the limitation of further identifying the species of the pests. Liu et al. (2017a) reported that three-dimensional (3D) vision could be used to detect relatively large invertebrate pests on broad leaves. Xia et al. (2018) used the combination of a region proposal network (Ren et al. 2017) and a pre-trained VGG19 network (Simonyan and Zisserman 2014) for insect classification and achieved an average precision of 89.22% on 24 common insect species in crop fields.

This study aimed to use computer vision and machine learning technologies to detect common pests on crops in South Australia. First, we created a ten-class database to evaluate the performance of the state-of-art CNNs for pest classification and developed a standard training pipeline. Then we used a novel method to generate virtual images to create an optimised five-class database. The model trained using the virtual database had a high classification accuracy and can be applied to proximally detect pests on crops.

2. Materials and methods

2.1. Create databases

To compare the effects of the number of classes and size of databases for training CNNs models, two databases, 10C-database and 5C-database, were created to train and validate the models. The 10C-database included ten classes of background, bee and eight invertebrate species hard to control in Australia. The database included a total of 6073 images in which 5573 images were used for training and 500 images for validation. The class names and quantity of images are listed in Table 1 and the examples images are shown in Fig. A1. The class of background included images of crop leaves, fresh grasses, flowers, dry plant residuals and bare soil in different colours. The numbers of images in different classes were roughly balanced in the range of 450 to 627 except the background. Because the class of background had large intra-class variety, it included a relatively large number of images of 984. Ideally, pests should be detected as early as possible, however, when their sizes are rather small, e.g., at egg stage, macro-lenses have to be used for imaging, which is not practical in field applications. Invertebrate animals present different and complex morphological, colour and textural features at their different living stages and it would require detection across orders of magnitude of scale to detect them in all of their living stages. To simplify the problem, the sizes of the invertebrates for observation were larger than 5 mm (except aphid) and they were at the same living stage in each class. A part of the image set was collected in multiple fields using smartphones. When collecting the images, the working distances of the cameras was adjusted in the range of 20 to 50 cm so that the fields of view were as large as possible and at the same time the pests presented enough pixels for human identification. Another part of the image set was collected from either ImageNet (StandfordVisionLab 2018) or InsectImages (InsectImages 2018). In the original images, the invertebrates might distribute randomly and occupied only a small portion of pixels and were not suitable for training CNN models. The regions of interests of the targets were cropped out manually from the original images. The cropped images have sizes from 128 × 128 pixels to 300 × 300 pixels and the objects can be observed clearly in the images. In some of the cropped images, only parts of the invertebrates were visible, which was intentionally arranged to simulate the real scenarios in natural farming environments.

The 5C-database was created with fewer classes and more images in each class than the 10C-database. It included five classes of background, snail, slug, earwig and worm and composed of a total of 29,507 images in which 22,773 images were used for training and 6743 images for validation. The class names and quantity of images are listed in Table 2 and the example images are shown in Fig. A2. In contrast to the 10C-database, the background was mainly composed of green leaves with moderate soil and branches of trees and did not include flowers and grasses. Besides that, the intra-class variety of the class of worm was enlarged by including worms in different colour and textures. The images in the same classes of the 10C-database were copied to the 5C-database to create a part of the images. As the resources of the images of pests in the field environments were limited, to further increase the size of the database was challenging. Since the images of plants without pests can be easily collected, the database of background including 5554 images was first created. However, collecting the images of pests in the fields is time-consuming because pests in the fields are difficult to find and the data collection is limited by season. We used a novel method to create virtual images of pests to obtain a large number of training images while significantly reduced labour. First, we collected an image of a pest (or several pests). Then we used the Photoshop software (Adobe Photoshop, California, USA) to manually segment the pest from the background. Third, the image including only the pest was processed to generate random transformation, including random rotation, random crop, random resize and random contrast adjustment. Lastly, the transformed image of the pest was positioned to an image of background which was randomly selected from the class of background to create an image in the corresponding class of pest. Each original image of pest can be used to generate 200 to 300 virtual images in the database so that it significantly reduced the time and workload.

2.2. Evaluate the performances of CNN models

Seven deep CNN models achieved outstanding performances in the ImageNet classification were evaluated using the 10C-database and

Table 1
Class name and quantity of images in 10C-database.

No.	Class name	Train	Validation
0	Background (BK)	984	50
1	Aphid (AP)	530	50
2	Portuguese Millipede (PM)	465	50
3	Earwig (EA)	454	50
4	Vineyard snail and white garden snail (SN)	575	50
5	Slug (SL)	522	50
6	Honeybee (HB)	627	50
7	Locust (LO)	512	50
8	Orange striped oakworm (OO)	454	50
9	Redlegged earthmites (RE)	450	50
	Sub-total	**5573**	**500**
	Total	**6073**	

Table 2
Class name and quantity of images in 5C-database.

No.	Class name	Train	Validation
0	Background (BK)	5554	1578
1	Vineyard snail and white garden snail (SN)	3718	1135
2	Slug (SL)	4763	1652
3	Earwig (EA)	4223	1274
4	Worm (WO)	4515	1095
	Sub-total	**22,773**	**6734**
	Total	**29,507**	

Table 3
The statistics of training from scratch. a_t represents classification accuracy of training, a_v for classification accuracy of validation, t for time (hour) used for training and e for epoch used.

Model	AlexNet	VGG16	VGG19	ResNet18_v2	ResNet34_v2	ResNet50_v2	ResNet152_v2
a_t	100.00%	100.00%	99.51%	100.00%	100.00%	99.92%	99.91%
a_v	70.20%	64.00%	63.20%	71.80%	73.80%	70.40%	**74.00%**
t	0.70	1.30	1.50	0.40	0.41	1.20	2.50
e	160	50	50	40	40	60	50

Table 4
The statistics of training from scratch using data argument. a_t represents classification accuracy of training, a_v for classification accuracy of validation, t for time (hour) used for training and e for epoch used.

Model	AlexNet	VGG16	VGG19	ResNet18_v2	ResNet34_v2	ResNet50_v2	ResNet152_v2
a_t	94.23%	94.93%	94.21%	95.00%	99.61%	99.03%	98.88%
a_v	82.80%	87.00%	88.40%	89.80%	90.60%	**91.20%**	89.20%
t	3	13.5	15.8	2.1	2.4	7.6	9.2
e	500	500	500	300	200	400	200

they were AlexNet, VGG16, VGG19, ResNet18_v2, ResNet34_v2, ResNet50_v2 and Resent152_v2. AlexNet architecture is one of the first deep CNNs to push ImageNet classification accuracy by a significant stride in comparison to traditional methodologies. It is composed of five convolutional layers followed by three fully connected layers (Krizhevsky 2014). The VGGNets made improvements over AlexNet by replacing large kernel-sized filters (11 and 5 in the first and second convolutional layer, respectively) with multiple 3 × 3 kernel-sized filters one after another (Simonyan and Zisserman 2014). A deep CNN could suffer from problems of vanishing gradient and degradation. The first means that earlier layers are almost negligibly learned when a CNN is deep and the second refers to adding more layers leading to higher training error. The residual network (ResNet) successfully solved these problems by constructing the network through modules called residual modules. It achieved better accuracy than VGGNet and GoogLeNet (Szegedy et al. 2015) while being computationally more efficient (He et al. 2016).

The modes were trained in four different training pipelines: (1) training the models from scratch using the original images in the training dataset; (2) training the models from scratch with data argument; (3) training the models using transfer learning without data argument; (4) training the models using both transfer learning and data argument. The purpose of data argument is to increase the number of images in a dataset and it has proven to be useful to improve model accuracy. First, the input images were cropped with random size and aspect ratio. The areas of the cropped images were 0.5 to 1 times of that of the original images and the aspect ratio was from 0.5 to 2. Second, the images were randomly flipped from left to right and top to bottom. Lastly, the brightness of the images was adjusted randomly. The final input images were resized to 224 × 224 for input. Previous studies have proven that transfer learning is effective in many remote sensing applications (Hu et al. 2015; Fu et al. 2017; Maggiori et al. 2017). Generally, transfer learning method discards the last layer of a pre-trained model and appends a fully connected layer where the neurons correspond to the number of predicted classes. During the training stage, the final layer is trained from scratch, while the others are initialized from the pre-trained model and updated by back-propagation rule or kept fixed (Huang et al. 2018). The models of 1000 classes fully trained on ImageNet including 14,197,122 images were used for transfer learning. The training and validation pipeline was realised using the MXNet machine learning library written in Python 3.6 (MXNet 2018). The modes were trained with batch size 10 and learn rate 0.002 on a computer with a GeForce GTX1080 GPU.

Confusion matrices were calculated to provide a direct observation of the classification results. Four statistics were used to evaluate the performance of the models. The accuracy of training a_t is the percent of correctly classified samples in the training dataset, and similarly, a_v represents the accuracy of validation in the validation dataset. t and e are the time (hours) used and epochs run when a_v start converging.

2.3. Test model accuracy

After evaluating the performances of the CNN models, the trained model on the 10C-database had the highest a_v was considered as the best model on this database and it was named M10C. The training pipeline returning the M10C was considered as the standard training pipeline. The standard training pipeline was applied to the 5C-database to train another model M5C. M10C and M5C were tested using independent images collected by smartphones in real farming environments. The images were either collected in day time using solar illumination or at night using artificial illumination and the background included multiple crops and fruit trees.

The pests distributed randomly in the images and only occupy a small portion of the pixels. The locations and class names of the pests needed to be detected automatically in the images. The region proposal network (Ren et al. 2017) has been widely used to detect the locations of targets in images. However, this method needs extra training of bounding box, involving a large amount of manual labelling work of bounding boxes. Besides this, aphids and snails could cluster together or distribute randomly in an image, making it hard to define the borders of bounding boxes; thus the bounding box technique was not suitable for this study. We used a sliding window technique which is simple and efficient. The images were divided into consequent square windows and the classification model was applied to the windows one by one. The size of the windows needed to be adjusted so that the objects had a similar scale as those in the training data. Besides confusion matrices and classification accuracy a, true positive rate r_{tp} and true negative rate r_{tn} were used for evaluating the performances of the models. True positive rate r_{tp} is the number of windows correctly classified as the corresponding classes of pest to the number of windows have pests. True negative rate r_{tn} is the number of windows correctly classified as background to the number of windows have background only.

3. Results

3.1. Performances of the CNN models and training pipelines

The seven CNN models were trained using the four different pipelines introduced in section 2.2, resulting in 28 models. The statistics of the training and validation are summarised in Tables 3 to 6, in which

Table 5
The statistics of training using transfer learning. a_t represents classification accuracy of training, a_v for classification accuracy of validation, t for time (hour) used for training and e for epoch used.

Model	AlexNet	VGG16	VGG19	ResNet18_v2	ResNet34_v2	ResNet50_v2	Resen152_v2
a_t	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
a_v	92.20%	95.60%	96.60%	95.80%	96.40%	97.40%	**97.60%**
t	0.17	0.39	0.50	0.08	0.10	0.67	1.40
e	30	15	15	10	10	30	30

Table 6
The statistics of training using both transfer learning and data argument. a_t represents classification accuracy of training, a_v for classification accuracy of validation, t for time (hour) used for training and e for epoch used.

Model	AlexNet	VGG16	VGG19	ResNet18_v2	ResNet34_v2	ResNet50_v2	ResNet152_v2
a_t	98.31%	99.33%	99.58%	99.94%	99.71%	99.83%	99.80%
a_v	92.20%	95.60%	96.20%	96.00%	96.20%	97.00%	**97.80%**
t	0.17	0.68	1.29	0.17	0.23	0.37	0.86
e	40	25	40	20	20	20	20

the highest accuracies of validation are highlighted by bold font. In Figs. A3 to A6, the curves of accuracy-epoch corresponding to the highest values of a_v in Tables 3 to 6 are plotted on the left side and the corresponding confusion matrices of validation are on the right side. Tables 3 to 6 show that no matter what training pipeline was used, the ResNet models always returned the highest accuracy of validation; that is a_v 74.00% for ResNet152_v2, 91.20% for ResNet 50_v2, 97.60% for ResNet152_v2 and 97.80% for ResNet 152_v2. Further, the ResNet152_v2 are more accurate than ResNet34 and Resnet50. The values of a_v also show that data argument can increase the accuracy of validation and transfer learning can significantly improve the accuracy and reduce the time of training. The ResNet152_v2 converged in only 0.86 h and had the highest validation accuracy of 97.80%. Thus, using ResNet152_v2 with data argument and transfer learning was considered as the standard training pipeline.

3.2. Testing results

Ten images for each class of pests were randomly selected from the testing database for testing. The images were divided into small windows of size 400 × 400 and resized to 224 × 224 as input images, resulting in 1895 windows for testing M10C and 2133 windows for testing M5C. The overall test result of M10C is shown in Table 7 and the

Table 7
Test results of M10C. The 11 columns on the left side compose the confusion matrix, in which the first row and first column show the class name of background (BK), aphid (AP), portuguese millipede (PM), earwig (EA), vineyard snail and white garden snail (SN), slug (SL), honeybee (HB), locust (LO), orange striped oakworm (OO), redlegged earthmites (RE). The column of SUM is for the sum of the number of windows of each class. r_{tp} represents true positive rate, r_{tn} for true negative rate and a for classification accuracy.

	BK	AP	PM	EA	SN	SL	HB	LO	OO	RE	SUM	r_{tp}
BK	1126	45	78	103	38	56	17	157	67	13	1700	58.46%
AP	5	12	0	0	0	0	0	0	0	0	17	$\boldsymbol{r_{tn}}$
PM	9	0	17	6	0	0	0	0	0	0	32	66.24%
EA	7	0	9	15	0	0	0	0	0	0	31	$\boldsymbol{a}$
SN	2	0	0	0	16	5	0	0	0	0	23	65.44%
SL	6	0	2	1	4	17	0	0	0	0	30	
HB	4	0	0	0	0	0	9	2	0	0	15	
LO	5	2	0	0	0	0	2	8	0	0	17	
OO	4	0	0	0	0	0	0	3	9	0	16	
RE	3	0	0	0	0	0	0	0	0	11	14	

Table 8
Test results of M5C. It shows the confusion matrix and accuracy of classification of each class and overall classes. In each sub-table, the six columns on the left side compose the confusion matrix, in which the first row and first column show the class name of background (BK), snail (SN), slug (SL), earwig (EA) and worm (WO). The column of SUM is for the sum of the number of windows of each class. r_{tp} represents true positive rate, r_{tn} for true negative rate and a for classification accuracy.

Snail

	BK	SN	SL	EA	WO	SUM	r_{tp}
BK	483	0	0	0	2	485	95.83%
SN	1	23	0	0	0	24	$\boldsymbol{r_{tn}}$
SL	0	0	0	0	0	0	99.59%
EA	0	0	0	0	0	0	$\boldsymbol{a}$
WO	0	0	0	0	0	0	99.41%

Slug

	BK	SN	SL	EA	WO	SUM	r_{tp}
BK	333	1	0	4	4	342	56.25%
SN	0	0	0	0	0	0	$\boldsymbol{r_{tn}}$
SL	5	2	9	0	0	16	97.37%
EA	0	0	0	0	0	0	$\boldsymbol{a}$
WO	0	0	0	0	0	0	95.53%

Worm

	BK	SN	SL	EA	WO	SUM	r_{tp}
BK	641	1	4	1	2	649	73.33%
SN	0	0	0	0	0	0	$\boldsymbol{r_{tn}}$
SL	0	0	0	0	0	0	98.77%
EA	0	0	0	0	0	0	$\boldsymbol{a}$
WO	7	1	0	0	22	30	97.64%

Earwig

	BK	SN	SL	EA	WO	SUM	r_{tp}
BK	567	2	0	1	3	573	69.23%
SN	0	0	0	0	0	0	$\boldsymbol{r_{tn}}$
SL	0	0	0	0	0	0	98.95%
EA	3	0	1	9	0	13	$\boldsymbol{a}$
WO	0	0	0	0	0	0	98.29%

All classes

	BK	SN	SL	EA	WO	SUM	r_{tp}
BK	2025	4	4	6	11	2050	75.90%
SN	1	23	0	0	0	24	$\boldsymbol{r_{tn}}$
SL	5	2	9	0	0	16	98.78%
EA	3	0	1	9	0	13	$\boldsymbol{a}$
WO	7	1	0	0	22	30	97.89%

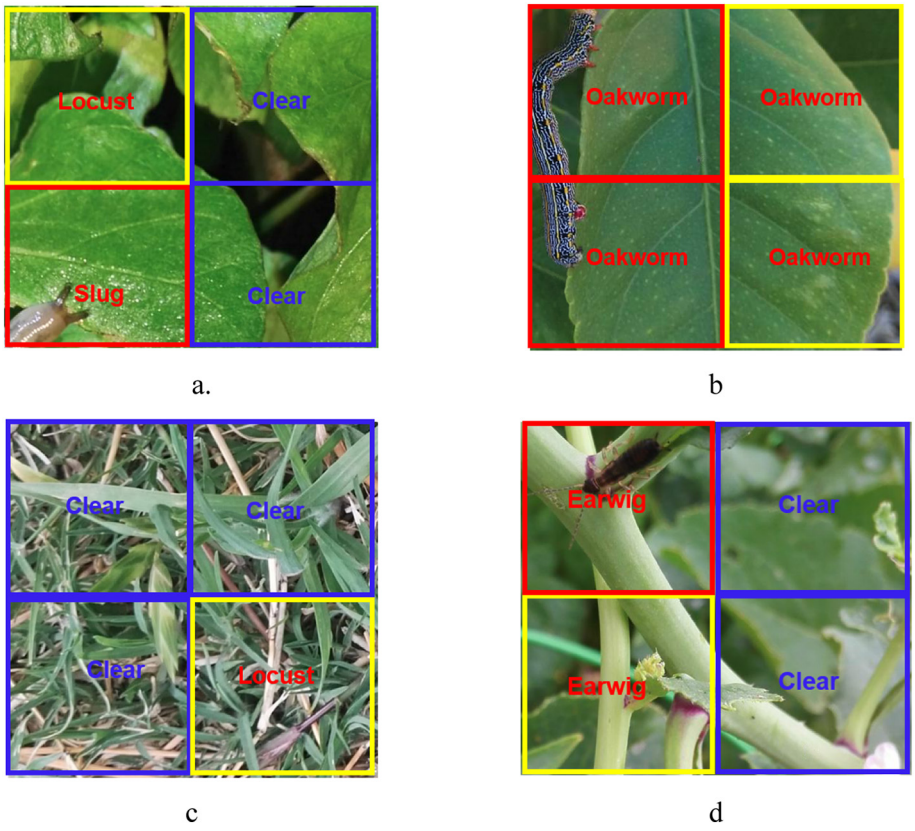


Fig. 1. Typical errors when testing the M10C model. The blue windows marked with "Clear" are correctly classified as background. The red windows with class name are correctly classified as the corresponding pests. The misclassified windows are highlighted with yellow colour.

results for each class of pests and all classes of M5C are illustrated in Table 8. M10C had a poor classification accuracy a of 65.44% with true positive rate r_{tp} 58.46% and true negative rate r_{tn} 66.24%. The confusion matrix shows that most of the errors are the misclassification of background to pests or vice versa. M5C had a high overall accuracy of 97.89% with the moderate true positive rate of 75.90% and an excellent true negative rate of 98.78%. For an individual class of pests, the classification accuracy and true negative rate are high, ranging from 95.53% to 99.59%. However, the true positive rates are relatively low for slug (56.25%), earwig (69.23%) and worm (73.33%).

4. Discussion

4.1. Deep CNN

Traditional methods for image classification need to first extract feature vectors in images and then input the vectors to machine learning algorithms for classification, such as support vector machine or logistic model tree (LMT) (Landwehr et al. 2005). For invertebrates, feature vectors could be composed of colour, textural, spectral (Liu et al. 2016a; Liu and Chahl 2018) and morphological features (Liu et al. 2017a) or abstract feature descriptors such as statistical moment (Han and He 2013), SIFT descriptor (Lowe 2004; Larios et al. 2008) or bag-of-region features (Csurka et al. 2004; Larios et al. 2008). Compared to the feature-based classification, CNN classification uses relatively less pre-processing. CNN classification combines feature extraction and classification modules into one integrated system without involving the hard work of feature extraction. Recently, the outstanding performances of deep CNNs have driven image classification technologies to a new milestone. Deep CNNs naturally integrate low-, middle- and high-level features and classifiers in an end-to-end multilayer fashion, and the levels of features can be enriched by the number of stacked layers (depth) (He et al. 2016).

When evaluating the performances of the CNN models using the 10C-database, the ResNets outperformed AlexNet and VGG nets, meanwhile, the accuracy improved when the nets were deeper. The leading CNN models that have been successful in the challenging ImageNet database all exploit "very deep" models, with a depth from sixteen to hundreds of layers (He et al. 2016), indicating that network depth is of crucial importance. However, deep CNNs could suffer the degradation problem: with the network depth increasing, accuracy gets saturated and then degrades rapidly (Simonyan and Zisserman 2014). He et al. (2016) successfully addressed this problem by using deep residual learning. In this study, ResNet152_v2 with 152 layers did not suffer from the degradation problem and it achieved the highest validation accuracy of 97.80%; thus, it is an ideal CNN for invertebrate classification.

4.2. Data argument and transfer learning

The 10C-database had limited training data, resulting in over-fitting when training the models from scratch (Table 3 and Fig. A3). The data argument suppressed the problem of overfitting by increasing the size of the training data. Table 4 shows that the data argument technique significantly improved the accuracy of validation from the range of 70.20% to 74.00% to the range of 82.80% to 89.20%. The transfer learning provided the surprisingly high accuracies of validation from 92.20% to 97.60% (Table 5). Another advantage of transfer learning is that it can significantly reduce training time. For example, Restnet152_v2 using data argument converged at 9.2 h, however, it converged at 1.4 h when using transfer learning (Table 5) and converged at 0.86 h when using the combination of transfer learning and data argument (Table 6).

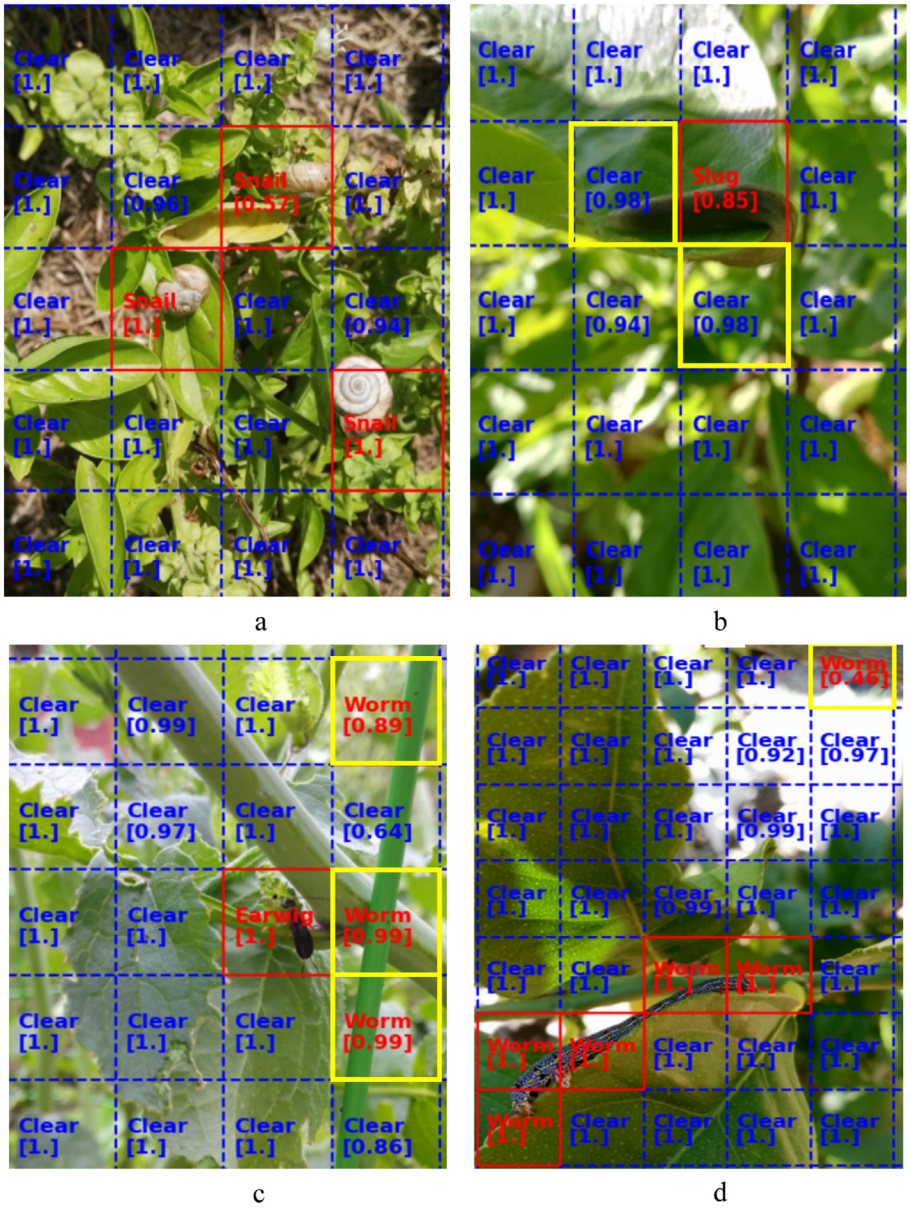


Fig. 2. Typical errors when testing the M5C model. The blue windows marked with "Clear" are correctly classified as background. The red windows with class name are correctly classified as the corresponding pests. The misclassified windows are highlighted with yellow colour. The numbers in the square brackets are the values of confidence of classification.

4.3. Database

Creating a high-quality database for training purpose is one of the most expensive and time-consuming tasks and it is a bottleneck of machine learning applications in the agricultural field (Barbedo 2020). Machine learning tools have evolved to a point in which the data fed to them has a much more prominent role in their success than their intrinsic characteristics, which explains the relatively similar performances yielded by different models when the data used to train and test them is the same (Kamilaris and Prenafeta-Boldú 2018; Liakos et al. 2018). The classification accuracy in Table 3 to Table 6 clearly show this trend. Conversely, image sets with different data distributions and variability can lead to significantly disparate performances even when the same machine learning model is employed, a fact that was reflected in the M10C and M5C database.

In testing, the M10C model got an unsatisfied classification accuracy of 65.44%. First, the 10C-database was not large enough to represent the true scenarios of natural farming environments, especially for the class of background. Second, the intra-class variance of the background was large because the background included soil, green leaves, dry plant residuals and flowers. Most of the errors are the misclassification of background to pests or pests to the background. For example, in Fig. 1a, the shadow of the leaf is misclassified as Locust. In Fig. 1b, the windows of the green leaf with yellow spots are identified as orange striped oakworm that has the similar texture to yellow spots. In Fig. 1c, the grass is misclassified as locust because locusts are camouflaged with the same colour and textural feature of the grass. In Fig. 1d, the stems are misclassified as earwigs because plant stems were not included in the training data.

The M5C model achieved a positive testing result with an overall accuracy of 97.89%, true positive rate of 75.90% and true negative rate of 98.78%. This achievement was obtained by cutting down the intra-class variance of background, reducing the number of classes and enlarging the size of the database. The class of background in the 5C-database was mainly composed of green leaves without flowers and dry plant residuals so to simplify the classification. In real

agricultural applications, although detecting a large scale of pest species on multiple host plants at different growing stages is attractive, a focus on detecting several pest species on a certain type of host plants is still valuable for pest management. Thus, simplifying the classification task to detect four types of pests on green leaves is reasonable and meets the requirements of most pest management tasks. The virtual images played a critical role in the high accuracy. By creating the virtual images, a large database with enough variance was created to represent the true scenario of the farming environments, at the same time, the cost and labour were significantly reduced. The true positive rate of M5C was relatively low. In the test images, the numbers of windows having background and pests were unbalanced, e.g., only a few windows had pests in a testing image. Thus, even a small number of pests misclassified as background could cause a significant drop in the true positive rate. However, for pest management, if a part of a pest is detected then the pest is considered as detected. Thus 75.90% true positive rate is acceptable for pest management applications.

Although the 5C-database included a large number of green worms without notable textural or colour features, the M5C had a limitation to detect green worms because they are well camouflaged in the green leaves. Green worms were often classified as background while stems or thin artificial materials were frequently classified as worm (Fig. 2). Using ultraviolet or near infrared images could improve the detection of some camouflaged invertebrates on crops (Liu et al. 2016a; Liu et al. 2017a; Liu et al. 2017b; Liu et al. 2017c; Chahl and Liu 2018). Hyperspectral images having the capability to reveal the biological and physiological characters of plants (Bruning et al. 2019; Bruning et al. 2020; Liu et al. 2020a, 2020b) have a potential to improve pest detection.

4.4. Feasibility of automation

We found that the wide-angle cameras of smartphones working at 20 cm to 50 cm can capture high-quality images for pest detection. This imaging approach can be automated by using a ground-based platform with a robotic arm equipped with a low-cost camera. The processing speed was tested on a computer with a 4.2G Hz CPU and the average time for processing a window was 0.21 s. By using parallel processing, this processing rate can support real-time pest detection in the field.

Appendix A

5. Conclusion

In precision agriculture, detection of invertebrate pests at an earlier stage on crops in the natural farming environments is a necessary prerequisite of IPM, however, this task is challenging and corresponding methods have not been well developed. We developed a computer vision method which can achieve proximal detection of invertebrate pests on crops. We found that ResNets outperformed AlexNet and VGG for pest classification and data argument and transfer learning can significantly improve classification accuracy and speed up convergence. We proposed a novel method to generate a virtual database which not only enabled training a model with high accuracy but also significantly saved the cost and time for collecting training data. Using three-dimensional modelling and gaming technologies to generate large-size training data would boost the applications of machine learning. In the future, using a large virtual database including millions of images and tens of classes to train a model with high accuracy for pest detection will be further investigated. Using multispectral or hyperspectral images to detect well-camouflaged pests would be another research direction. Developing a ground-based robotic system to conduct proximal detection of pests on crops in real-time for pest management will be considered.

Declaration of Competing Interest

This research is supported by Australian Plant Phenomics Facility, The Plant Accelerator and School of Engineering, University of South Australia. No potential conflict of interest was reported by the authors.

Acknowledgement

The authors would like to thank The Australian Plant Phenomics Facility funded by the Australian Government under the National Collaborative Infrastructure Strategy (NCRIS), School of Engineering, University of South Australia and Joint and Operations Analysis Division, Defence Science and Technology Group, Australia.

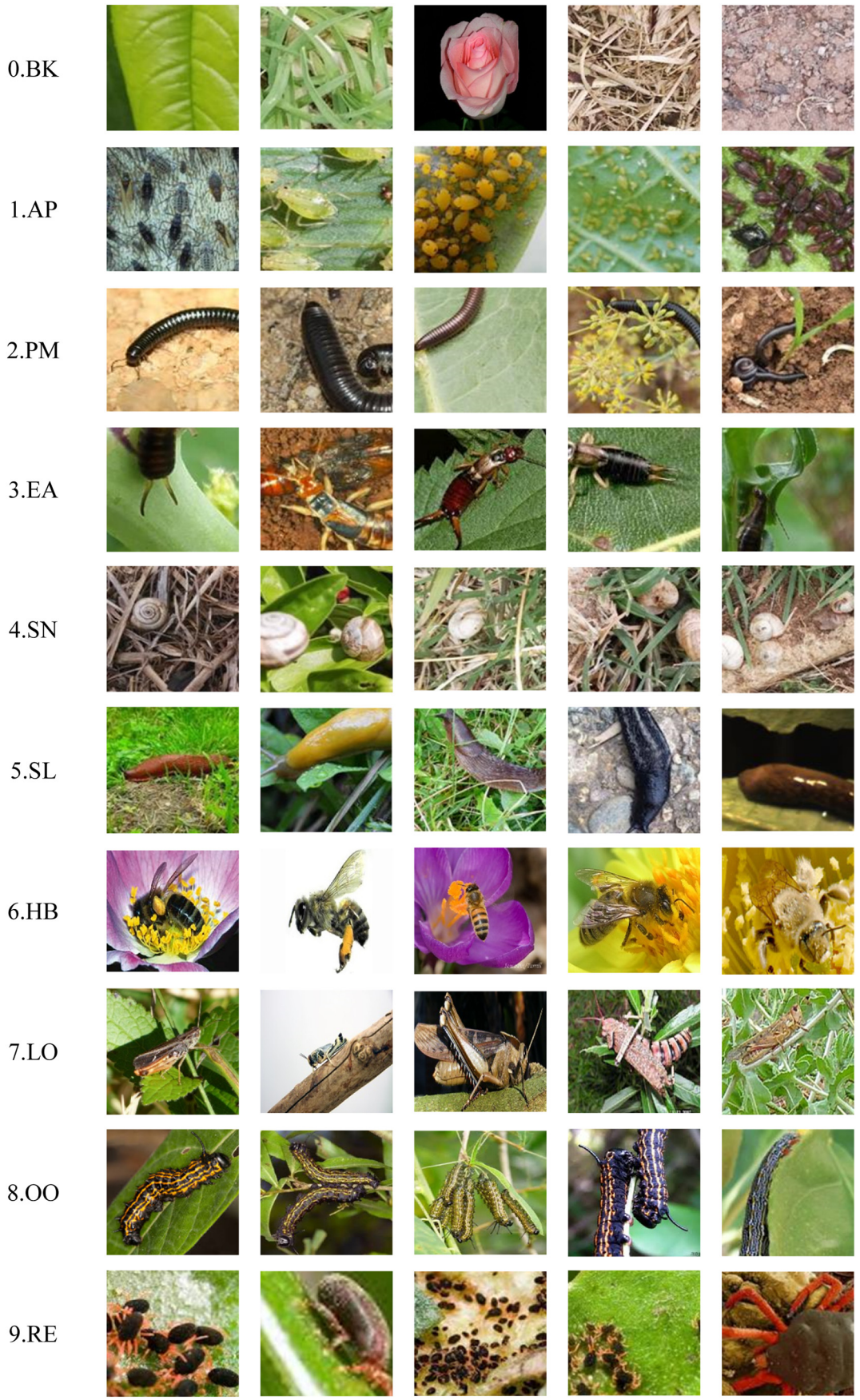
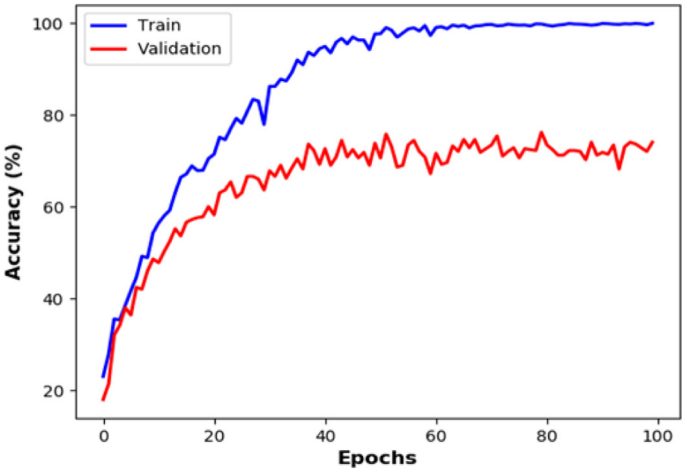


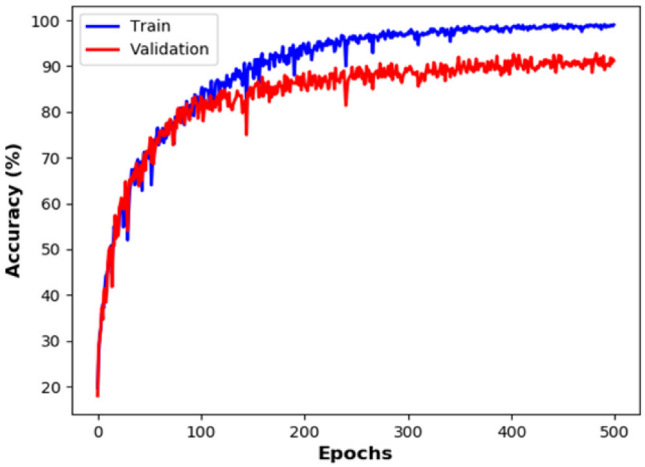
Fig. A1 Example images of 10C-database. The first column shows the class names of background (BK), aphid (AP), portuguese millipede (PM), earwig (EA), vineyard snail and white garden snail (SN), slug (SL), honeybee (HB), locust (LO), orange striped oakworm (OO), redlegged earthmites (RE).

Fig. A2 Example images of virtual 5C-database. The first column shows the class names of background (BK), vineyard snail and white garden snail (SN), slug (SL), earwig (EA), worm (WO).



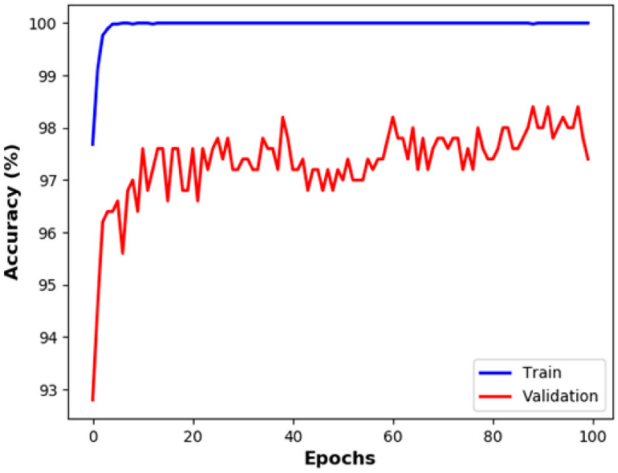
	0	1	2	3	4	5	6	7	8	9
0	38	2	0	2	3	1	2	1	0	1
1	1	32	1	1	1	5	0	2	3	4
2	2	1	39	0	3	3	0	0	1	1
3	0	3	3	39	2	1	0	0	0	2
4	0	0	2	0	46	0	0	1	1	0
5	0	3	4	1	2	32	0	2	3	3
6	7	1	1	0	0	1	34	6	0	0
7	1	0	0	3	1	1	11	28	4	1
8	0	0	2	1	2	1	2	4	37	1
9	4	0	0	1	0	0	0	0	0	45

Fig. A3 ResNet152_v2 trained from scratch. The numbers in the first row and column indicate classes (refer to Fig. A1).



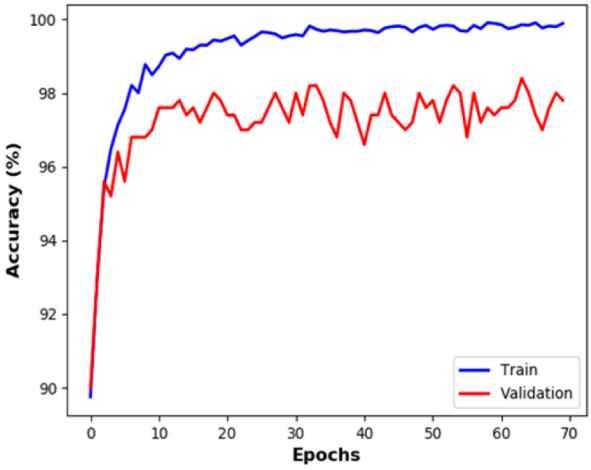
	0	1	2	3	4	5	6	7	8	9
0	48	0	0	0	0	0	2	0	0	0
1	0	45	0	1	0	2	2	0	0	0
2	0	1	45	0	1	1	1	0	1	0
3	1	0	0	46	1	0	1	0	1	0
4	2	0	0	0	48	0	0	0	0	0
5	0	1	2	0	1	44	0	2	0	0
6	5	0	0	0	0	0	42	3	0	0
7	3	0	0	0	2	0	1	44	0	0
8	0	0	0	0	0	3	2	1	44	0
9	0	0	0	0	0	0	0	0	0	50

Fig. A4 ResNet50_v2 trained from scratch using data argument. The numbers in the first row and column indicate classes (refer to Fig. A1).



	0	1	2	3	4	5	6	7	8	9
0	49	0	0	0	0	0	0	0	0	0
1	0	44	0	1	0	1	0	1	0	3
2	0	0	49	0	0	0	1	0	0	0
3	0	1	1	48	0	0	0	0	0	0
4	0	0	0	0	50	0	0	0	0	0
5	0	0	0	0	0	50	0	0	0	0
6	1	0	0	0	0	0	49	0	0	0
7	1	0	0	0	0	0	0	49	0	0
8	0	1	0	0	0	0	0	0	49	0
9	0	0	0	0	0	0	0	0	0	50

Fig. A5 ResNet152_v2 trained using transfer learning. The numbers in the first row and column indicate classes (refer to Fig. A1).



	0	1	2	3	4	5	6	7	8	9
0	47	0	0	0	1	0	2	0	0	0
1	1	46	0	2	0	1	0	0	0	0
2	0	0	50	0	0	0	0	0	0	0
3	0	0	0	50	0	0	0	0	0	0
4	0	0	0	0	50	0	0	0	0	0
5	0	0	0	0	0	50	0	0	0	0
6	2	0	0	0	0	0	47	1	0	0
7	0	0	0	0	0	0	0	50	0	0
8	0	0	0	0	0	0	0	0	49	0
9	0	0	0	0	0	0	0	0	0	50

Fig. A6 ResNet152_V2 trained using both transfer learning and data argument. The numbers in the first row and column indicate classes (refer to Fig. A1).

References

Baker, G., Jennings, R., 2015. Growers Chase Pest-Control Answers. Grains Research and Development Corporation. Adelaide, Australia https://grdc.com.au/Media-Centre/Ground-Cover/Ground-Cover-Issue-117-July-August-2015/Growers-chase-pest-control-answers Last accessed 8 Aug. 2016.

Barbedo, J., 2020. Detecting and classifying pests in crops using proximal images and machine learning: a review. Atificial Intelligen. 1, 312–328. https://doi.org/10.3390/ai1020021.

Boissarda, P., Martinb, V., Moisanb, S., 2008. A cognitive vision approach to early pest detection in greenhouse crops. Comput. Electron. Agric. 62, 81–93.

Bruning, B., Liu, H., Brien, C., Berger, B., Lewis, M., Garnett, T., 2019. The development of hyperspectral distribution maps to predict the content and distribution of nitrogen and water in wheat (Triticum aestivum). Front. Plant Sci. 10. https://doi.org/10.3389/fpls.2019.01380.

Bruning, B., Berger, B., Lewis, M., Liu, H., Garnett, T., 2020. Approaches, applications, and future directions for hyperspectral vegetation studies: an emphasis on yield-limiting factors in wheat. Plant Phenome J. https://doi.org/10.1002/ppj2.20007.

Chahl, J., Liu, H., 2018. Bioinspired invertebrate pest detection on standing crops. SPIE, Bioinspiration, Biomimetics, and Bioreplication VIII, Denver, Colorado, United States. SPIE, Colorado, United States, p. 105930B https://doi.org/10.1117/12.2296580.

Csurka, G., Dance, C.R., Fan, L., Willamowski, J., Bray, C., 2004. Visual categorization with bags of keypoints. ECCV International Workshop on Statistical Learning in Computer Vision, Prague, Czech Republic. CiteSeer, pp. 1–22 doi:10.1.1.72.604.

Fu, G., Liu, C., Zhou, R., Sun, T., Zhang, Q., 2017. Classification for high resolution remote sensing imagery using a fully convolutional network. Remote Sens. 9 (5), 498. https://doi.org/10.3390/rs9050498.

GRDC, 2014. Slugging slugs. Grains Research and Development Corporation, Barton, Canberra, Australia http://grdc.com.au/Media-Centre/Hot-Topics/Slugging-slugs Last accessed 18, July 2016.

Han, R., He, Y., 2013. Remote automatic identification system of field pests based on computer vision. Trans. Chin. Soc. Agric. Eng. 29 (3), 156–162.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE https://doi.org/10.1109/CVPR.2016.90.

Hu, F., Xia, G., Hu, J., Zhang, L., 2015. Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. Remote Sens. 7 (11), 14680–14707. https://doi.org/10.3390/rs71114680.

Huang, H., Deng, J., Lan, Y., Yang, A., Deng, X., Zhang, L., 2018. A fully convolutional network for weed mapping of unmanned aerial vehicle (UAV) imagery. PLoS One 13 (4), e0196302. https://doi.org/10.1371/journal.pone.0196302.

InsectImages, 2018. Insect Images. https://www.insectimages.org Last accessed Dec 1 2018.

Kamilaris, A., Prenafeta-Boldú, F., 2018. Deep learning in agriculture: a survey. Comput. Electron. Agric. 147, 70–90. https://doi.org/10.1016/j.compag.2018.02.016.

Kogan, M., Hilton, R.J., 2009. Conceptual framework for integrated pest management (IPM) of tree-fruit pests. Biorational Tree-fruit Pest Management. vol. 1. Centre for Agriculture and Biosciences International, Oxfordshire, UK. https://doi.org/10.1079/9781845934842.0001 ISBN:9781845934842.

Krizhevsky, A., 2014. One weird trick for parallelizing convolutional neural networks. arXiv 1404.5997. https://arxiv.org/abs/1404.5997v2 arXiv:1404.5997v2.

Landwehr, N., Hall, M., Frank, E., 2005. Logistic model trees. Mach. Learn. 59 (1–2), 161–205.

Larios, N., Deng, H., Zhang, W., Sarpola, M., 2008. Automated insect identification through concatenated histograms of local appearance features: feature vector generation and region detection for deformable objects. Mach. Vis. Appl. 19, 105–123.

Liakos, K., Busato, P., Moshou, D., Pearson, S., Bochtis, D., 2018. Machine learning in agriculture: a review. Sensors (Basel, Switzerland) 18 (8), 2674. https://doi.org/10.3390/s18082674.

Liu, H., Chahl, J., 2018. A multispectral machine vision system for invertebrate detection on green leaves. Comput. Elecron. Agric. 150, 279–288. https://doi.org/10.1016/j.compag.2018.05.002.

Liu, H., Lee, S.H., Chah, J.S., 2016a. An evaluation of the contribution of ultraviolet in fused multispectral images for invertebrate detection on green leaves. Precis. Agric. 17 (4). https://doi.org/10.1007/s11119-016-9472-7.

Liu, H., Lee, S.H., Chahl, J.S., 2016b. A review of recent sensing technologies to detect invertebrates on crops. Precis. Agric. 17 (4). https://doi.org/10.1007/s11119-016-9473-6.

Liu, H., Lee, S.H., Chahl, J.S., 2017a. A multispectral 3D vision system for invertebrate detection on crops. IEEE Sens., 1–14 https://doi.org/10.1109/JSEN.2017.2757049.

Liu, H., Lee, S.H., Chahl, J.S., 2017b. Registration of multispectral 3D points for plant inspection. Precis. Agric. https://doi.org/10.1007/s11119-017-9536-3.

Liu, H., Lee, S.H., Chahl, J.S., 2017c. Transformation of a high-dimensional color space for material classification. J. Opt. Soc. Am. A 34 (4), 523–532. https://doi.org/10.1364/josaa.34.000523.

Liu, H., Bruning, B., Garnett, T., Berger, B., 2020a. Hyperspectral imaging and 3D technologies for plant phenotyping: from satellite to close-range sensing. Comput. Electron. Agric. 175. https://doi.org/10.1016/j.compag.2020.105621.

Liu, H., Bruning, B., Garnett, T., Berger, B., 2020b. The performances of hyperspectral sensors for proximal sensing of nitrogen levels in wheat. Sensors 20 (16), 4550. https://doi.org/10.3390/s20164550.

Lowe, D., 2004. Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. 60, 91–110.

Maggiori, E., Tarabalka, Y., Charpiat, G., Alliez, P., 2017. Convolutional neural networks for large-scale remote-sensing image classification. IEEE Trans. Geosci. Remote Sens. 55 (2), 645–657. https://doi.org/10.1109/TGRS.2016.2612821.

Martineau, M., Conte, D., Raveaux, R., Arnault, I., Munier, D., Venturini, G., 2017. A survey on image-based insect classification. Pattern Recogn. 65 (C), 273–284. https://doi.org/10.1016/j.patcog.2016.12.020.

Miles, M., 2015. Insect Pest Management in Faba Beans. Grains Research and Development Corporation. QLD, Australia https://grdc.com.au/Media-Centre/Ground-Cover/Ground-Cover-Issue-117-July-August-2015/Insect-pest-management-in-faba-beans Last accessed 7 Oct 2015.

Murray, D., Clarke, M., Ronning, D., 2013. Estimating invertebrate pest losses in six major Australian grain crops. Aust. J. Entomol. 52 (3), 227–241. https://doi.org/10.1111/aen.12017.

MXNet, 2018. A Flexible and Efficient Library for Deep Learning MXNet Scientific Community. https://mxnet.apache.org/versions/1.7.0/ Last accessed August 2019.

Nalam, V., Louis, J., Shah, J., 2019. Plant defense against aphids, the pest extraordinaire. Plant Sci. (Limerick) 279, 96–107. https://doi.org/10.1016/j.plantsci.2018.04.027.

Oerke, E., 2006. Crop losses to pests. J. Agric. Sci. 144 (1), 31–43. https://doi.org/10.1017/S0021859605005708.

Ren, S., He, K., Ross, G., Sun, J., 2017. Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Mach. Intell. 39 (6), 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031.

Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv 1409.1556 arXiv:1409.1556.

StandfordVisionLab, 2018. ImageNet. http://www.image-net.org Last accessed Jan 1 2019.

Sun, Y., Cheng, H., Cheng, Q., Zhou, H., Li, M., Fan, Y., et al., 2017. A smart-vision algorithm for counting whiteflies and thrips on sticky traps using two-dimensional Fourier transform spectrum. Biosyst. Eng. 153, 82–88. https://doi.org/10.1016/j.biosystemseng.2016.11.001.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., et al., 2015. Going deeper with convolutions. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Boston, MA, USA. IEEE, pp. 1–9 https://doi.org/10.1109/CVPR.2015.7298594 ISBN:10636919.

Xia, D., Chen, P., Wang, B., Zhang, J., Xie, C., 2018. Insect detection and classification based on an improved convolutional neural network. Sensors 18 (12), 4169. https://doi.org/10.3390/s18124169.

Yen, A., Madge, D., Berry, N., Yen, J., 2013. Evaluating the effectiveness of five sampling methods for detection of the tomato potato psyllid, Bactericera cockerelli (Šulc) (Hemiptera: Psylloidea: Triozidae). Aust. J. Entomol. 52 (2), 168–174. https://doi.org/10.1111/aen.12006.

Zhu, L., Ma, M., Zhang, Z., Zhang, P., Wu, W., Wang, D., et al., 2017. Hybrid deep learning for automated lepidopteran insect image classification. Orient. Insects 51 (2), 79–91. https://doi.org/10.1080/00305316.2016.1252805.