Article:

Wang, F., Wang, R., Xie, C. et al. (2 more authors) (2020) Fusing multi-scale context-aware information representation for automatic in-field pest detection and recognition. Computers and Electronics in Agriculture, 169. 105222. ISSN 0168-1699

https://doi.org/10.1016/j.compag.2020.105222

Fusing multi-scale context-aware information representation for automatic in-field pest detection and recognition

Fangyuan Wang[a,b], Rujing Wang[a,*], Chengjun Xie[a,*], Po Yang[c,*], Liu Liu[a,b]

[a] *Institute of Intelligent MACHINES, Hefei Institute of PHYSICAL Science, Chinese ACADEMY of Sciences, Hefei 230031 CHINA*
[b] *University of Science AND Technology of CHINA, Hefei 230026 CHINA*
[c] *Sheffield University, Sheffield, United Kingdom*

A R T I C L E I N F O

Keywords:
Convolutional neural network
Context-aware attention network
Multi-projection pest detection model
In-field pest in food crop

A B S T R A C T

Automatic in-field pest detection and recognition using mobile vision technique is a hot topic in modern intelligent agriculture, but suffers from serious challenges including complexity of wild environment, detection of tiny size pest and classification of multiple classes of pests. While recent deep learning based mobile vision techniques have shown some success in overcoming above issues, one key problem is that towards large-scale multiple species of pest data, imbalanced classes significantly reduce their detection and recognition accuracy. In this paper, we propose a novel two-stages mobile vision based cascading pest detection approach (DeepPest) towards large-scale multiple species of pest data. This approach firstly extracts multi-scale contextual information of the images as prior knowledge to build up a context-aware attention network for initial classification of pest images into crop categories. Then, a multi-projection pest detection model (MDM) is proposed and trained by crop-related pest images. The role of MDM can combine pest contextual information from low-level convolutional layers with these in high-level convolutional layers for generating the super-resolved feature. Finally, we utilize the attention mechanism and data augmentation to improve the effectiveness of in-field pest detection. We evaluate our method on our newly established large-scale dataset In-Field Pest in Food Crop (IPFC) and sufficient experimental results show that DeepPest proposed in this paper outperforms state-of-the-art object detection methods in detecting in-field pest.

1. Introduction

Specialized pest control and prevention is always a highly-priority agricultural issue over all the world (Santangelo, 2018; Liu et al., 2017; Berenstein and Edan, 2018). Due to cost-effectiveness and efficient automation, pest monitoring approaches are widely utilized in practical crop monitoring systems. Its applications need to use either fixed stationary camera or mobile camera to observe the in-field pest images, then employ some advanced image processing algorithms (Ding and Taylor, 2016; Wang et al., 2013; Wu et al., 2014; Yan et al., 2017) to identify and analysis pest associated data for decision-making and prediction. While the application of above advanced techniques (Ding and Taylor, 2016; Wang et al., 2013) enables great success in effective recognition and classification of certain type of insect, one key problem appears that most researchers focus on increasing the recognition accuracy of certain type of insect by either introducing new features or employing new machine learning algorithms, yet paying more attentions on developing practical useful automatic pest monitoring systems towards Large-scale Multi-class Pest dataset in the Wild (LMPW).

As shown in Fig. 1, LMPW contains many challenging issues affecting the accuracy of pest detection approaches, such as shadow and sky influence, tiny size of pest objects, imbalanced data of multi-class pest, dense or sparse distribution of pests, etc. Traditional pest detection methods work well in lab-based small-scale pest dataset with few types, but achieve low accuracy and poor robustness in processing practical large-scale multi-class pest dataset. Their performance is strongly limited by many issues, like lighting illuminates, dense or overlapping distribution of tiny objects, etc.

Recently, the advances in deep learning techniques (Bulat and Tzimiropoulos, 2018; Zhou et al., 2018; Sermanet et al., 2013) have led to significantly promising progress in the field of object detection, with the majority of study focuses on designing more complex object detection networks for improved accuracy, such as Super-FAN (Bulat and Tzimiropoulos, 2018), Scale- Transferrable Object Detection (Zhou et al., 2018), unsupervised multi-stage feature learning (Sermanet et al., 2013) and other extended variants of these networks (He et al., 2016; Li

* Corresponding authors.
E-MAIL ADDRESSES: rjwang@iim.ac.cn (R. Wang), cjxie@iim.ac.cn (C. Xie), po.yang@sheffield.ac.uk (P. Yang).

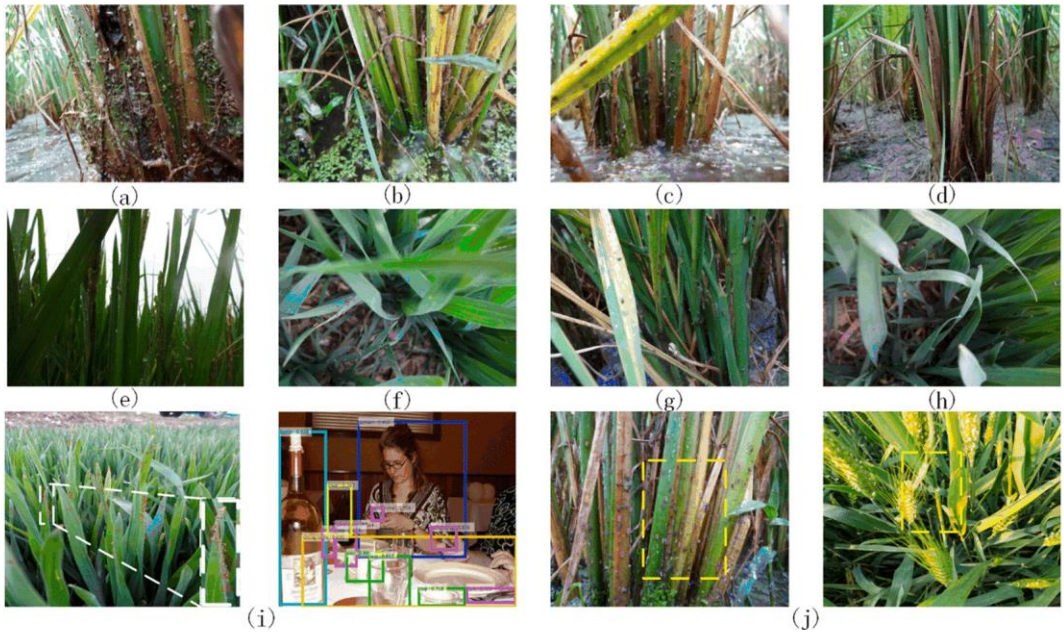


Fig. 1. Some typical challenges in the in-field pest detection. (a) shadow influence; (b) crop disease; (c) branch barrier; (d) leaf barrier; (e) sky interference; (f) water interference; (g) leaf specular reflection; (h) uneven distribution; (i) object size comparison between MS COCO dataset and in-field pest dataset; (j) demonstration of imbalanced in-field pest data.

et al., 2018; Lin et al., 2017). Despite the fact that above CNN based object detection networks have showed great accuracies in general object detection applications, their applications in LMPW still suffer from two key limitations: (1) Large variance of density distribution and sizes of tiny pests make the activation of some objects even smaller and insensitive with each pooling layer through a deep learning architecture. (2) The data volume of different types of pests is imbalanced, which make the deep learning approach difficult to achieve high accuracy over all types of pests.

If we train a specific detection model for each pest, and then inference different pests successively in the practical application, it will undoubtedly take several times as long as the single model. Compared with the state-of-the-art methods, it is more time-consuming and affect the user experience. On the other hand, if one single model is used to detect all of the pests, its accuracy cannot meet the requirements of actual use. Actually, certain categories of pests only appear on specific crops, it is necessary to use a two-stage mobile vision based deep learning approach to address this issue.

In this paper, we attempt to explore one new two-stage mobile vision based deep learning approach with fusion of multi-scale context-aware information for improving the recognition accuracy over large-scale multi-class pest in the wild. As shown in Fig. 2, the foundation of our idea is that most of mobile devices are able to record large-scale contextual information of pest images in the wild, like temporal information, geographic information and ambient information, also including exact time, longitude, latitude, air temperature, air humidity, soil temperature as well as soil humidity. We extract different contextual information of the images as prior knowledge to build up a context-aware attention network for initial classification of these pest images into crop categories. The imbalanced data challenge in the crop images is addressed as different crop images are separated according to crop species.

Then, a multi-projection pest detection model (MDM) is designed for training these crop-associated pest images. The role of MDM is to combine small-scale contextual information from low-level convolutional layers with these in high-level convolutional layers for generating the super-resolved feature. Note that the multi-projection detection model is trained on model pretrained on the context-aware attention network, i.e., the crop classification model is used as the pre-trained model to train the pest detection model. Finally, we utilize the attention mechanism and data augmentation to improve their effectiveness of in-field pest detection.

Inspired by ResNet (He et al., 2016) and DetNet (Li et al., 2018), we optimize the structure and parameters of the in-field pest detection model by short-cut and projection convolution so as to overcome the difficulty of pest feature vanishing under the high-level convolutional layer. In addition, we fine-tune pest detection models on a single crop using other in-field pest datasets to prevent the serious consequences of classification errors in certain cases. Through the synthesis of different scale features embedded in Feature Pyramid Networks (Lin et al., 2017), the network is effectively trained to detect the super-resolved pest feature representation, which enhance the performance of in-field pest detection.

We evaluate the proposed method on newly established large-scale dataset In-field Pest in Food Crop (IPFC), containing over 17 K photographed and labeled independently images over five years. The experimental results demonstrate our method achieves better performance comparable to the other state-of-the-art methods on the IPFC.

The major contributions of this paper are as follows:

(1) A novel two-stages mobile vision based cascading pest detection approach (DeepPest) towards large-scale multiple species of pest data is proposed, which is feasible to apply for practical tiny pest detection in the field.
(2) This approach has explored the possibility of fusing extensive multi-scale contextual information into CNN for extracting super-resolved feature representation. It shows some success to overcome the difficulty of pest feature vanishing in the high-level convolutional layer.
(3) A comprehensive and in-depth experimental evaluation on practical industry level large-scale in-field pest dataset is provided for verifying the usefulness and robustness of proposed approaches. The results show that our proposed method could deliver better performance comparable to the state-of-the-art methods.

2. Related work

2.1. *Small object detection*

Small object detection is a hot topic in computer vision community.

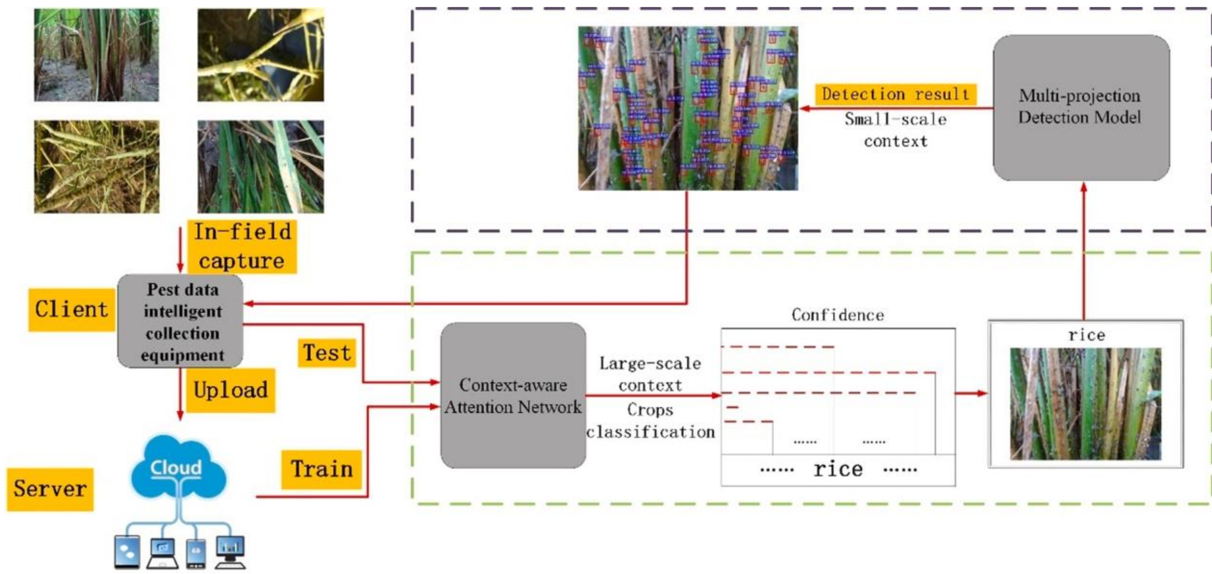


Fig. 2. Technical pipeline of our system architecture. The left part indicates the data acquisition process in which in-field pests can be photographed and uploaded by the pest data intelligent collection equipment. The part in green dotted box shows that the classification result is given by comparing the confidence in different crops. All of the pests are detected by multi-projection detection model which is embed in the top purple dotted box. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Many deep learning based approaches have been working in this field. Keshari et al. (2018) optimize the structure of the filter by a dictionary-based filter learning algorithm and provides the weight representation in convolutional neural network via small samples, which obtains great mean average precision even with small samples. Zhu et al. (2018) indicate the biggest challenge in face recognition is recent anchor boxes cannot well match with tiny faces and present a novel evaluation metrics Expected Max Overlapping score (EMO) for low overlapping issue. Bai et al. (2018) propose a generative adversarial network (GAN) to estimate the confidence of a high-resolution face generated from corresponding blurry tiny face respectively. Li et al. (2017) attempt to solve the small object detection problem by present a Perceptual Generative Adversarial Network (PGAN), which generate high resolution image by low resolution ones and obtain similar features and hence improve the detection rate. While above deep learning approaches achieve some great access in general object detection, they suffer from some typical inherited problems of deep learning, such as imbalanced data. Towards large-scale multi-class dataset, this issue will significantly reduce the performance of deep learning approaches.

2.2. Object detection with IMBALANCED DATA

Some scientists solve the problem by re-sampling (Chawla et al., 2002; Maciejewski and Stefanowski, 2011; Oquab et al., 2014; Dong et al., 2017) and others believe that this question could be work out via cost-sensitive weighting (Tang et al., 2009; Ting, 2000; Huang et al., 2016). Maciejewski and Stefanowski (2011) focus on the resampling techniques and exploits more precisely information about the local neighborhood of the considered examples. Dong et al. (2017) formulate a novel pipeline for batch incremental hard example mining of minority attribute classes from imbalanced large-scale training data. Ting (2000) also tries re-sampling as well as cost-sensitive weighting in multiple SVM variations and then presents novel granular SVMs-repetitive down-sampling algorithm as the appropriate method. Huang et al. (2016) indicate the restriction of class-level triplet loss and replace with cluster-level and class-level quintuple loss, which effectively mitigates the risk of class imbalance inherent in the dataset. While these re-sampling methods achieve some success in imbalanced data, their performance in practical application is usually constrained by the quality and diversity of collected data.

2.3. Context AWARE scheme

In mobile vision field, the mobile device can retrieve and record some contextual information associated images, which will be potentially used and improve the tasks of applications. For instance, semantic context achieves a great deal in 3D point matching, high-speed visual tracking, video frame interpolation, generative image inpainting, scene segmentation and object detection (Deng et al., 2018; Choi et al., 2018; Niklaus and Liu, 2018; Yu et al., 2018; Ding et al., 2018; Huang et al., 2018). Local descriptors learned (Deng et al., 2018) is highly aware of the global context, which enhance the local feature representation and improve the robustness and invariance in 3D descriptor extraction performance. Context-aware scheme with numerous auto-encoders make a great progress on deep feature compression, which provide high computational speed for visual tracking (Choi et al., 2018). Niklaus and Liu (2018) present a context-aware synthesis method that warps the input frames as well as their pixel-wise contextual information so as to interpolate a high-quality intermediate frame. Yu et al. (2018) propose a novel architecture which employ image features around the gap as the references to generative inpainting. Ding et al. (2018) come up with a dramatic context contrasted local feature that not only leverages the informative context but also spotlights the local information in contrast to the context. Huang et al. (2018) combine visual cues with instance-dependent weights and achieves state-of-the-art performance on face recognition task. Hu and Ramanan (2017) present multiple scale-specific detector for images of different scale and utilizes the features involving contextual information to address the small object detection problem, which reduce error of face recognition and show the importance of the context. In this paper, we demonstrate that the context is useful for solving small object detection issue.

3. Proposed approach: Deeppest

3.1. MOTIVATED IDEA of Deeppest ARCHITECTURE

Due to the occurrence time and damaged crops of different pests vary greatly, not all of the in-field pest data are easy to collected, which leads to the imbalanced data in most pest dataset. The current object detection method is difficult to deal with imbalanced data problem and always ignore the category that don't have enough pest data. By collecting and analyzing our dataset IPFC, we found different pests appear on different crops, different schedule as well as a variety of environment and location (Dyrmann et al., 2016; Ebrahimi et al., 2017).

It is important to introduce the geographic information, environmental parameters and temporal information of pest images into pest detection. A straightforward method is to classify the vulnerable crops in the pest images according to the contextual information, and then train the corresponding pest detection models for different crops. A context-aware attention network (CAN) could classify in-field pest images automatically according to the prior knowledge about plant category. Thus, the imbalanced pest data are classified into the several balanced pest data by the damage crop category.

Also, we observe that the in-field pests on images are mostly small and individual. Using state-of-the-art object detection approaches into these images will make in-field pest features are prone to lose after high-level convolution, and it is difficult to extract the in-field pest features in shallow network. Hence, a novel multi-projection detection model (MDM) can combine the delicate features in high-level convolutional layer and integral structure of pest come from low-level convolutional layer. Then we could fuse the contextual information around pests from low-level convolutional layer and address the issue of feature vanishing of small object in the deep convolution layer.

In the later section, we will present the alternative optimization for in-field pest detection from internal structure of convolutional neural network, and give details of the DeepPest, i.e., context-aware attention network and multi-projection detection model.

3.2. Context-aware Attention Network (CAN)

The overview of context-aware attention network is shown in Fig. 3, in-field pest images which photographed by CCD have some contextual information including geographic information, temporal information as well as ambient information. It is always unsatisfactory to use raw image alone for crop classification while some of the supportive information of raw images is exactly helpful for image classification. Motivated by this, we construct a multi-task learning paradigm to extract different contextual feature, then concatenate all of the contextual information and the final classification result is obtained by the decision net. In this paper, given the trade-off between efficiency and accuracy, ResNet-50 is employed as the backbone for extracting contextual information and two fully connected layers are used to output the crop category to which the pest image belongs. In this way, we have better used the supplementary information and accurately separated the different crop images.

Our dataset was collected in Anhui and Hunan provinces in China, the longitude and latitude of collection areas range from 110°E to 120°E and from 25°N to 35°N, respectively. Obviously, the geographic information can be divided into 10 categories according to longitude or latitude. Thus, 20 labels are used in one-hot coding to describe the geographic information. In terms of temporal information, each month corresponds to a temporal category, and the whole year is divided into 12 categories. From the perspective of ambient information, most of the pest images in IPFC are collected with the temperature range of 20 °C to 35 °C and the humidity range of 0–100%. The temperature range is divided every 3°as well as contributes 5 classes, while the humidity information contains 10 categories. Similarly, the ambient information contributes 15 labels to the decision net. The aforementioned contextual information is outputted and concatenated by multi-task learning paradigm in CAN and indicates the crop category by decision net.

Training: The training process consists of both contextual information extraction and decision net optimization. Firstly, different contextual information is encoded into different category labels in each task, thus the contextual information extraction can be parsed into numerous classification task. Obviously, the contextual information should have a basic correlation between them, therefore, the shared feature extraction layer can effectively improve the effect of the classification of different contextual information. Then, the decision net is trained by a given crop category with specific contextual information. In particular, the output of the contextual information extraction network and the input of the decision net are the same paradigm, thus we can integrate them into one model in the inference stage. By this means, the internal relations between contextual information and crop category are learned by decision net, which constrained false positive and false negative misclassifications.

Inference: Other than the training process, the contextual information extraction network and decision net are integrated to one model in the inference stage. Therefore, the input of the CAN is the pest image and the output is the category of crop suffered from pests, which subject to the contextual information extracted by CNN. Hence, we can select the specific pest detection model to detect the pest species and locations in the pest images.

3.3. Multi-projection Detection Model (MDM)

Architecture: After determining the crop category in the in-field pest images, we need calculate the location and the number of pests. Unfortunately, for most of the CNN models such as ResNet and Inception-v4, due to the resolution of feature map is reduced to 1/32 or

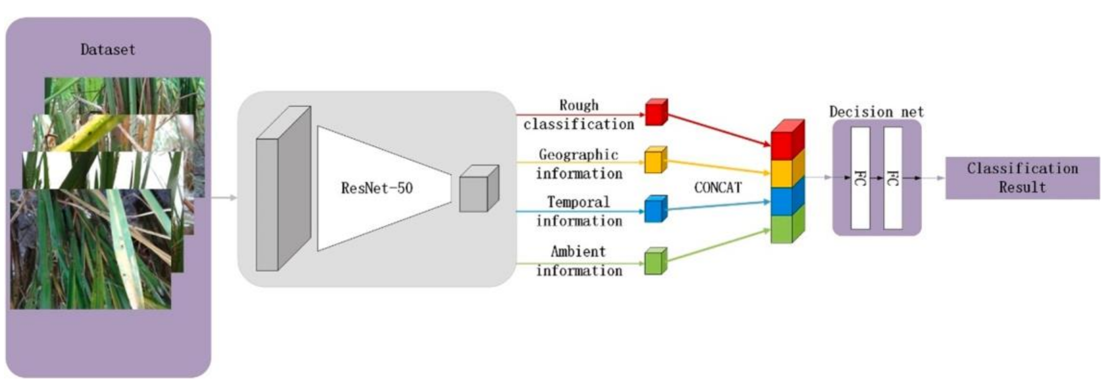


Fig. 3. Details of the proposed context-aware attention network. We encode different contextual information into image annotation. Different contextual codes are labelled for one image, therefore we employ corresponding multi-task CNN model to extract the specific contextual information in the training stage. The red branch is used to roughly classify pest images and the other branches in different color are utilized to recognize the geographic information, temporal information as well as ambient information, respectively. The final classification result is obtained by concatenating different contextual information extracted from each CNN model into feature vector and putting them into decision net. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

1/64 of the original image in the high-level convolutional layer, small objects are not visible on it (32*32 or 64*64 object takes up only one pixel in the feature map). In fact, the scene around the pest can also serve as contextual information for the pest detection task. (Pests, for example, tend to stick to food crop instead of appearing in the sky for no reason). Obviously, it is a good idea to integrate the small-scale context information in the shallow layers into the semantic information in the deep layers, which restricts the influence of the features vanishing in deep layers and ensures the size of the receptive field. Inspired by the research of ResNet and DetNet, we propose multi-projection detection model to increase the performance of the in-field pest detection which decrease the consequence of features vanishing. Different from conventional method like ResNet, Inception-v4 (Szegedy et al., 2017) and VGG (Simonyan and Zisserman, 2015), in our case, we introduce the multiple projection convolution block to increase the weight of low-level convolution which is illustrated in Fig. 4. We introduce the several projection convolutional layers to extract the small-scale contextual information such as texture, color and shape. Noted that small-scale contextual information is relative to the large-scale contextual information extracted from CAN.

In this paper, ResNet-50 is used as the backbone for extracting pest feature. Different from the residual block, each projection convolution block has only one downsampling operation directly by convolutional layers that have a 3*3 kernel and stride of 2, instead of the convolutional layer with stride of 1 which used to extract the image semantic information. This approach maximizes the retention of small-scale contextual information such as color, shape, and texture of the pest image. Considering the feature in projection convolution blocks need to be merged with that in residual blocks, therefore, the projection convolution blocks and residual blocks have the same amount of convolutional layer that have a stride of 2. Basically, batch normalization and ReLU are used to greatly accelerate the convergence process as well as avoid the risk of exploding gradient. The final layer of projection convolution blocks will have the same scale as the last layer of the backbone from which the image information is extracted. Thus, the output of the projection convolution blocks and backbone are concatenated and fed into FPN, which combines the small-scale context-aware information from shallow layers. After that, RPN (Ren et al., 2015) generates a number of region proposals for in-field pest detection. The accurate classification and regression results of the region proposals can be obtained by the finetuning of RoI Pooling and fully connected layer.

Fine-tuning: Generally, multi-projection detection model is used to detect in-field pest on the species-specific crop which means there are greatly different between in-field pest detection models applied to different crops. Notice that context-aware attention network cannot make sure that each in-field pest image can be classified accurately. We need to fine-tune specific multi-projection detection model on all of the in-field pest images in order to reduce the impact of image misclassification and improve the system performance and robustness.

3.4. *Alternative optimization*

Attentional Mechanism: DeepPest also integrates an attentional mechanism to modify feature map in low-level convolutional layer by image information in high-level convolutional layer which prevent the background noises and refine the in-field pest features. We introduce channel attention mechanism to obtain the weights for each channel and multiply with the raw feature map. Different from SENet (Hu et al., 2018) which put attentional mechanism behind each convolutional layer, in this paper, we set this module at each projection convolution block and residual block so as to regulate the feature map which fed into next block.

Data Augmentation: Before training our model, we utilize some data augmentation methods to expand data amount. Firstly, considering the variability of the shooting angle and the rotation invariability of the in-field pest, we rotate and flip the original images with unchanging image resolution. We horizontal and vertical flip the raw images so that obtain the other 2-fold images as well as the use of image rotation results in 7-fold increase at data amount (image is rotated by 45° at each time). Therefore, a total of 24-fold pest data can be obtained. Additionally, all of the pest images should be cropped into different scale which enlarge the data amount as well as enrich the pest knowledge in a single image.

4. Dataset collection and experiment setting

4.1. *In-field Pest in Food Crop (IPFC)*

While there are some available insect datasets containing images captured in the laboratory environment such as butterfly dataset (Kang et al., 2014) and bee dataset (Bozek et al., 2018). Intuitively, the models trained by these single pest image in laboratory cannot effectively recognize the pest.

Based on the cascading model proposed in this paper, we establish a task-specific in-field pest dataset In-Field Pest in Food Crop (IPFC), which is used to address the challenge of in-field pest detection and counting. IPFC is generated by data acquisition from the past three years, which contains 17,192 in-field pest images with 76,595 pest annotations. Note the fact that IPFC has been randomly divided into 10 folds, where nine-tenths are used as training set while the residual part is used as validation set. The statistical for IPFC dataset are shown in Table 1. The in-field pest images analyzed in this paper are collected in

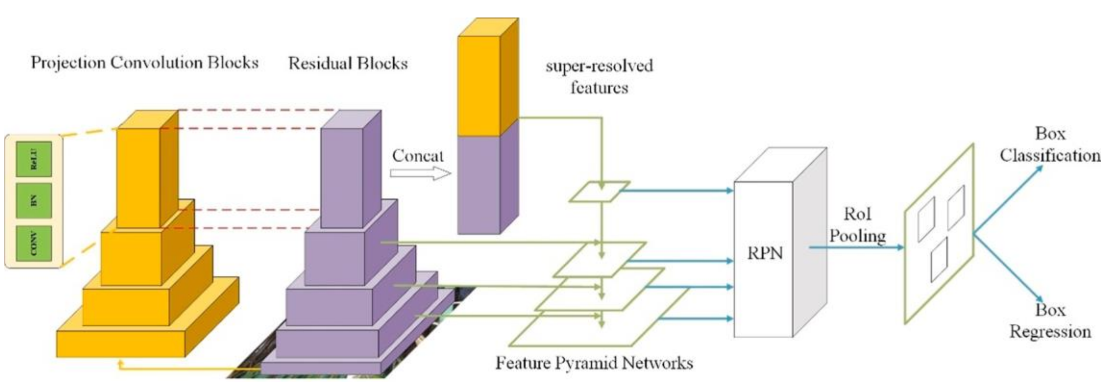


Fig. 4. Details of the proposed multi-projection detection model. The purple network is ResNet-50 including plentiful convolutional layers and shortcut connections. We introduce the shortcut connections with several projection detection layers which is yellow block in the figure. Results of multi-projection convolution are combined with the Res5 in ResNet-50 to generate super-resolved feature, which trained by FPN together with that of other residual blocks. RPN extracts the region proposals from the feature map, and the classification and regression results of the bounding boxes can be obtained after ROI pooling and finetune. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Details of in-field pest data in IPFC. For a class-specific pest, the endangered crop, the number of pests in training set and the number of pests in validation set are shown in this table. Some crops suffer from multiple pests at the same time (such as wheat), so that multiple species of pests maybe appear in one pest image, therefore the total number of pest images is not always equal to the sum of the various pest images. The column 'size' represents the percentage of the area of pest in the whole image.

Insect Name	Endangered crop	Size (%)	Training			Validation			All		
			Images	Pests	Avg.	Images	Pests	Avg.	Images	Pests	Avg.
Wheat mite	Wheat	0.089	11,505	54,423	4.73	1278	6095	4.76	12,783	60,518	4.73
Sticky worm	Wheat	1.512	2901	2980	1.02	303	320	1.05	3204	3300	1.02
Rice planthopper	Rice	0.148	1084	11,352	10.47	121	1425	11.77	1205	12,777	10.6
Total	/	/	15,490	68,755	4.44	1702	7840	4.61	17,192	76,595	4.46

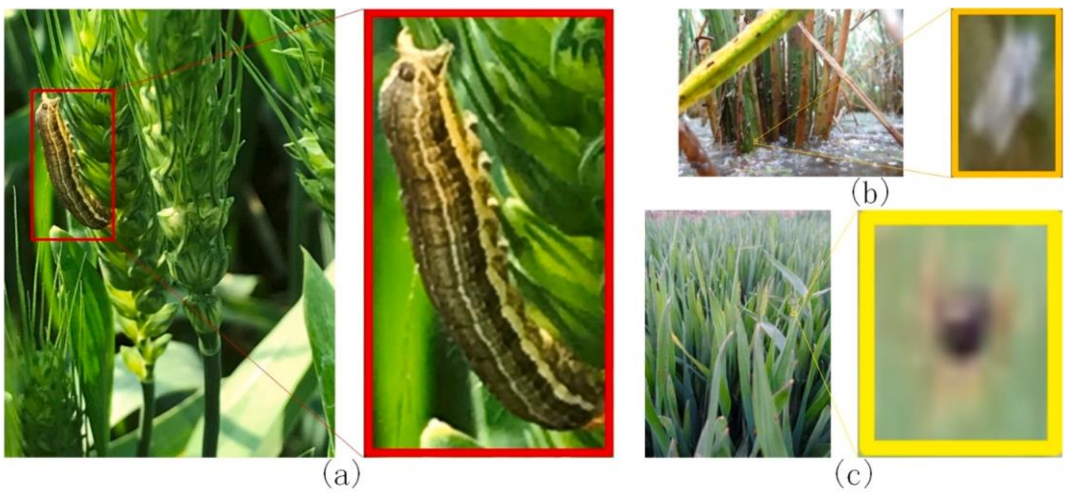


Fig. 5. Some pest image samples in IPFC. The pest in each sample are zoomed in on the right. (a) sticky worm on wheat crops in eastern China (117°4′N, 32°86′E), photographed at April 2018. The air temperature is 27 °C and air humidity is 55%. (b) rice planthopper on rice crops in southern China (111°47′N, 27°25′E), photographed at July 2017. The air temperature is 36 °C and air humidity is 76%. (c) wheat mite on wheat crops in northern China (114°77′N, 34°56′E), photographed at March 2018. The air temperature is 13 °C and air humidity is 34%.

Table 2
Experimental results of DeepPest with state-of-the-art architectures. The bold values indicate the best results.

Methods	Sticky worm	Rice planthopper	Wheat mite	mAP
Scale-specific detection	90.6	66.8	59.9	72.4
DeepPest	**90.7**	**69.2**	**61.7**	**73.9**
VGG-16 + FPN	49.3	54.5	43.8	49.2
ResNet-50 + FPN	90.2	67.3	59.4	72.3

China. All the images were captured by independent research and development device called pest intelligent collection equipment. As for image acquisition, Sony CX-10 CCD camera whose parameters are set to 4 mm focal length with an aperture of f/3.3. It should be noted that only one RGB color image (1440*1080) from each time series is labelled and used in this paper, therefore the labelled pest image is unique. Through the utilization of pest intelligent collection equipment, we not only collect numerous pest images, but also record the temperature, humidity and geographic information of pest images. The pest image samples are shown in Fig. 5. Particularly, the geographic information, ambient information, temporal information and food crop background of each pest image are described in detail.

Note that we employ few human labels (approximately 1 k-2 k) first in the process of labelling data, and then train these images by multi-scale context-aware information representation method. Next, we automatically label more pest images using the trained model and artificially justified the results of automatic annotation. So that the constantly iteration and correction improve the performance of the model

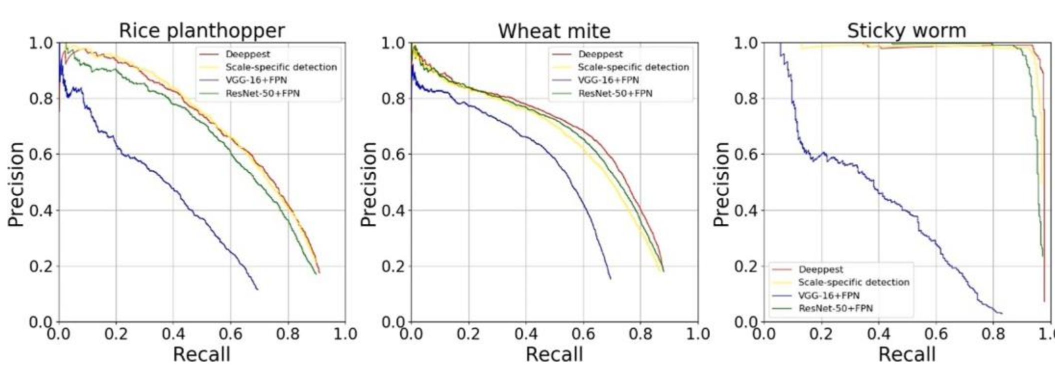


Fig. 6. Comparisons of pest detection performance with the state-of-the-arts methods on the IPFC.

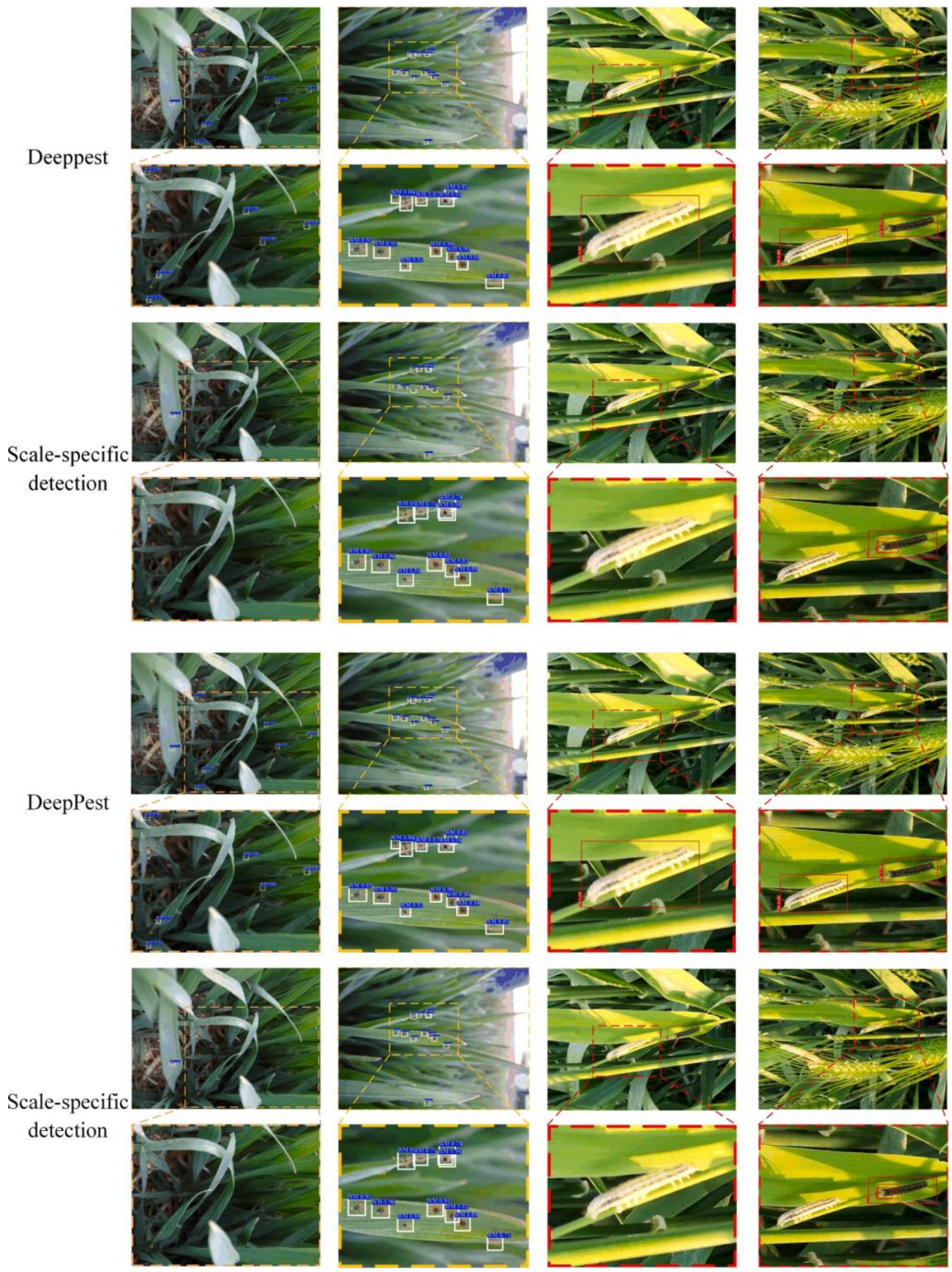


Fig. 7. Qualitative results on different challenging environment in IPFC.

proposed in this paper and also save the human resources and cost compared to completely human label.

4.2. Experiment setting

In order to verify that our method could be used in IPFC for detecting pests, we establish some experiments to compare the performance of DeepPest with that of other state-of-the-art CNN architectures on IPFC. Our codes are based on Caffe2 (Jia et al., 2014) with Python API and run on 12 GB Tesla P40 GPU. Some experiment details are given in this section. ResNet-50 is used for extracting contextual information from in-field pest images in CAN. The RMSprop is chosen as our optimizer with momentum equals to 0.9, which updates parameters based on one mini-batch at each iteration. This optimizer could partly keep the update gradient at previous iteration and fine-tune the final gradient considering the current mini-batch. In order to avoid overfitting problem, we utilize dropout method as well as early-stopping strategy to select the best training iteration.

As to learning rate policy, 'step' strategy is applied in gradient descent, in which we initialize learning rate to 0.001 and the learning rate will be divided by 10 per 15,000 iterations. In addition, mini-batch size is set to 2 and the number of region proposals of every training example is at least 128. Code and high-resolution demo are available at: http://t.cn/E2qcTYZ.

5. Experimental results

5.1. Performance comparison

Table 2 presents the comparison of DeepPest with other state-of-the-

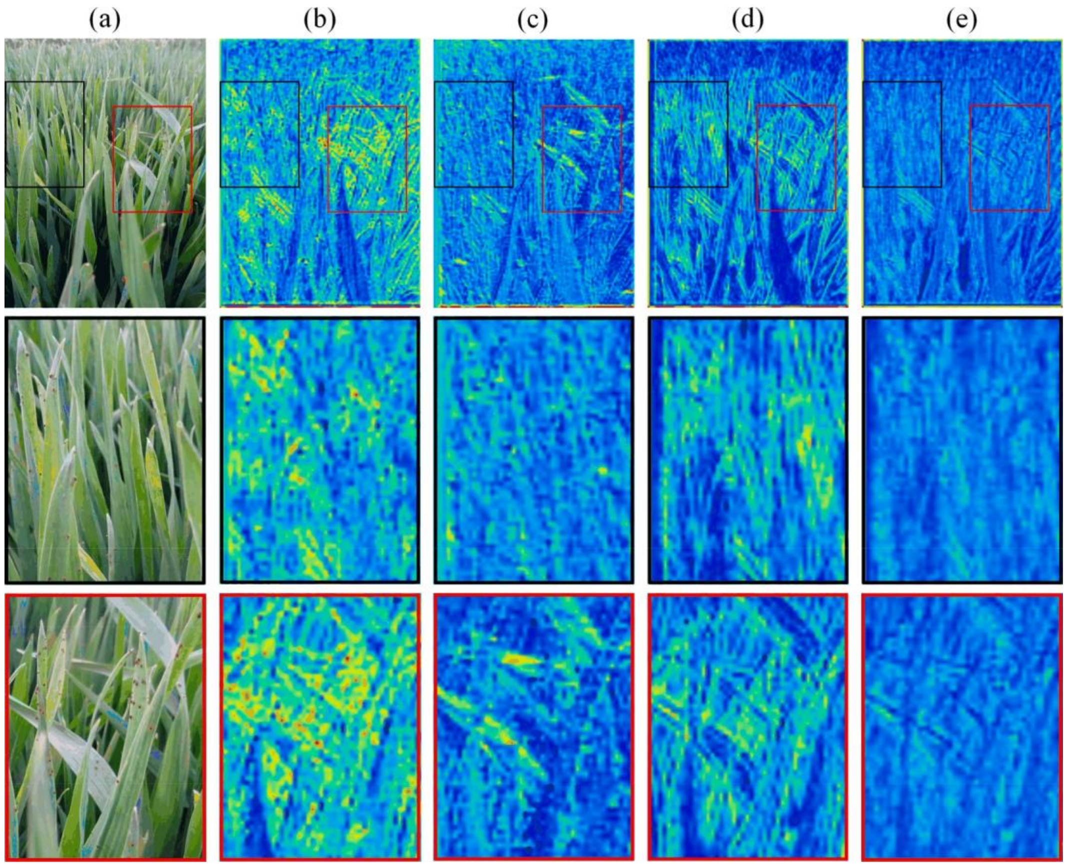


Fig. 8. Feature visualization. (a) raw image; (b) super-resolved feature from DeepPest; (c) feature from Scale-specific detection; (d) feature from ResNet-50 + FPN; (e) feature from VGG-16 + FPN.

Table 3
Experimental results of CAN on IPFC. The bold values indicate the best results.

Model	CAN	Top-1 error
ResNet-50	√	**0.01%**
		0.56%
Inception-v4	√	0.13%
		0.82%
VGG-16	√	0.19%
		1.52%

Table 4
Ablation studies of contextual information in CAN. The bold values indicate the best results.

Model	CAN			Top-1 error
	Geographic information	Temporal information	Ambient information	
ResNet-50				0.56%
	√			0.41%
		√		0.29%
			√	0.34%
	√	√		0.06%
	√		√	0.11%
		√	√	0.08%
	√	√	√	**0.01%**

Table 5
Comparisons of detection performance with or without CAN based on DeepPest and the state-of-the-art architectures. The bold values indicate the best results.

Methods	CAN	Sticky worm	Rice planthopper	Wheat mite	mAP
Scale-specific detection	√	90.6	66.8	59.9	72.4
		90	63.9	58.3	70.7
DeepPest	√	**90.7**	**69.2**	**61.7**	**73.9**
		90.4	65.4	61.1	72.3
VGG-16 + FPN	√	49.3	54.5	43.8	49.2
		39.1	37	44.7	40.3
ResNet-50 + FPN	√	90.2	67.3	59.4	72.3
		90.7	60.3	59.7	70.2

Table 6
Comparisons of detection performance with contextual information extracted from different residual blocks. The bold values indicate the best results.

Methods	Different sources	Sticky worm	Rice planthopper	Wheat mite	mAP
Deep-Pest	Res1	**90.7**	**69.2**	**61.7**	**73.9**
	Res2	90.6	68.7	60.8	73.4
	Res3	90.5	67.7	60.5	72.9
	Res4	90.2	67.3	60.6	72.7
ResNet-50 + FPN	/	90	63.9	58.3	70.7

arts CNN models in terms of mean average precision on each pest category. We firstly observe the results of DeepPest and other conventional CNN models. Our method could improve the mAP by 1.5%, 24.7% as well as 1.6% respectively compared to Scale-specific detection, VGG-16 with FPN and ResNet-50 with FPN on the IPFC, which illustrate the advantage of DeepPest in automatically and accurately detecting pests. Our method achieves the best performance in all of the pest categories. DeepPest clearly improves the detection performance of rice planthopper and wheat mite. However, in terms of larger size pests such as sticky worm, since state-of-the-art methods can obtain sufficient and effective pest features after multiple convolutional layers, the

Table 7
Comparisons of detection performance with or without fine-tuning strategy. The bold values indicate the best results.

Methods	Fine-tuning	Sticky worm	Rice planthopper	Wheat mite	mAP
Scale-specific detection	√	90.6	66.8	59.9	72.4
		90.3	64.5	58.7	71.2
DeepPest	√	**90.7**	**69.2**	**61.7**	**73.9**
		90.4	65.8	61	72.4
ResNet-50 + FPN	√	90.6	67.4	58.5	72.2
		90.2	63.3	58.1	70.5

Table 8
Comparisons of detection performance with or without alternative optimization. AM and DA represent attentional mechanism and data augmentation, respectively. The bold values indicate the best results.

Methods	Alternative Optimization	Sticky worm	Rice plant-hopper	Wheat mite	mAP
DeepPest	AM and DA	**90.8**	**70.0**	**62.1**	**74.3**
	\	90.4	65.4	59.2	71.7
	AM	90.6	65.3	61.8	72.6
	DA	90.7	69.2	61.7	73.9

Table 9
Comparisons of efficiency of different modules proposed in this paper with the state-of-the-arts methods.

Methods	CAN	MDM	Time cost on one image (ms)
Scale-specific detection			65
	√		93
		√	74
	√	√	114
DeepPest			60
	√	√	78
			67
	√	√	98
VGG-16 + FPN			68
	√		105
		√	77
	√	√	116

approach employing contextual information has little effect on the detection performance. More comparisons of precision-recall curves in terms of different categories are provided in Fig. 6, which can further illustrate the feasibility and effectiveness of DeepPest.

Several instances of the DeepPest for in-field pests are visualized in Fig. 7. We compare our visual results with those from scale-specific detection. Different from other state-of-the-art methods, CAN and MDM applied in DeepPest address the issue of in-field pest detection. As shown in Fig. 7, obviously, DeepPest can effectively detect tiny-size pests in small scales, which outperforms scale-specific detection at precision and misdetection due to small object detection challenge.

We further present visualization of super-resolved feature generated by combining contextual information. The learned general representation and the multi-scale context-aware information representation by the DeepPest for in-field pests are shown in Fig. 8. The second and the third row respectively demonstrate the details of local features from the convolutional layer. Pest features caught by DeepPest are shown in red or yellow in the feature map, while the background is shown in blue or green. We can observe that our architecture successfully extracts the in-field pest features easily lost in the deep convolutional layers, proving the effectiveness and feasibility of DeepPest.

5.2. Ablation experiments

5.2.1. Context-aware Attention Network (CAN)

Table 3 indicate the top-1 error of the state-of-the-art deep learning classification methods. It can be seen that ResNet-50 with CAN achieves the best performance with the lowest top-1 error. CAN could reduce the top-1 error by 0.55%, 0.69% as well as 1.33% respectively compared to ResNet-50, Inception-v4 and VGG-16 on the IPFC, which illustrate the effectiveness of our method in automatic and accurate pest image classification.

Besides, an ablation study is attempted to further recognize the influence of each part of the contextual information in CAN. The effect of different contextual information on food crop classification is shown in Table 4. One can observe that the more contextual information adopted, the better classification accuracy of CAN will be. In contrast with other contextual information, the temporal information has a great impact on the classification result due to the specific temperature condition of pest activity (Such as wheat spider damages wheat between 10 °C and 20 °C).

In order to verify the superiority of the crop recognition with CAN on in-field pest detection, we apply our method to some state-of-the-art CNN models. As shown in Table 5, CNN models with CAN outperforms the Scale-specific detection, VGG-16 with FPN and ResNet-50 with FPN by 1.6%, 8.9%, 2.1% in mAP respectively and also improve the accuracy of DeepPest with 1.6%, which verifies the effectiveness of CAN.

We find that the IPFC has a relatively small number of rice image in Table 1. If we directly detect in-field pest with no crop recognition (without CAN), rice images will be submerged in a flood of wheat images, so that not all rice planthoppers could be recognized. Multi-scale contextual information contained in the pest image can reasonably separate the small-amount rice images from the wheat images and alleviate the issue of class imbalance problem between the rice images and the wheat images, thus improving the robustness of CNN model.

5.2.2. Contextual information from different residual block

The MDM proposed in this paper learns contextual information about pest from the low-level convolutional layer of ResNet-50. Aiming to verify the effectiveness of contextual information in different convolutional layers, we establish a set of parallel experiments in which introduce MDM from four different residual blocks, since ResNet-50 has only five residual blocks.

As shown in Table 6, MDM at different scales can improve its

Fig. 9. Wild pest image acquisition equipment and its usage. (1) CCD camera; (2) carbon fiber telescopic rod; (3) mobile client. The parameters of CCD camera are set to 4 mm focal length with an aperture of f/3.3. CCD camera is used to collect pest images and is controlled by mobile client. The position of field would be located by global positioning system.

performance in detecting in-field pest like sticky worm, rice planthopper and wheat mite, which demonstrate that recursively employ contextual information in low-level convolutional layer is of great significance to detect in-field pests. Another issue is that the contextual information from shallower convolutional layer is more useful than that from deeper convolutional layer and the pest feature has been destroyed by several residual blocks if MDM is introduced at the third or fourth residual block. Besides, when ResNet-50 with FPN algorithm is used to calculate the category and location of pests which means we do not use any small-scale contextual information to assist pest detection, experimental result shown that its performance is lower than any approaches using MDM, which proves that the algorithm proposed in this paper is feasible and effective.

5.2.3. Fine-tuning

Particularly, in order to verify the necessity of fine-tuning strategy, we present the experimental results of DeepPest, Scale-specific detection and ResNet-50 with FPN in Table 7. “Fine-tuning” indicates the network is fine-tuned by the other crop images after training the model on the given crop, otherwise the pest detection model only trained on the specific crop outputted by CAN.

As shown in Table 7, DeepPest with fine-tuning performs better than that without fine-tuning, and the similar result emerge from the other two methods. By comparing the results of fine-tuning with non-fine-tuning, our strategy shows a clear advantage over the approach without any justification. Specifically, fine-tuning performs well on the category with fewer pictures like rice, which can be attributed the ability of fine-tuning in minimizing the impact of misclassification improving the robustness of the system.

5.2.4. Alternative optimization

DeepPest optimizes the training process of CAN and MDM with alternative optimization. In order to illustrate the necessity of the alternative optimization, we present the experimental results of DeepPest with or without optimization method during CAN and MDM. We simultaneously attempt attentional mechanism block and data augmentation strategy.

Comparing the method with or without attentional mechanism block in Table 8, we can observe that DeepPest with attentional mechanism outperforms that with non-attentional mechanism by 0.9% in mAP, and data augmentation enhance the effect of DeepPest for about 2.1%. This shows remarkable improvements in mAP on in-field pests can be obtained when employ these two alternatives.

5.2.5. Inference time

Since the method proposed in this paper operates in the mobile camera or other mobile phone, the inference efficiency (Time cost on one image) of our approaches is shown in Table 9. It should be noted that we use iPhone X as the mobile device.

One can observe that our method decreases the inference efficiency of each state-of-the-art detection algorithm. However, the whole algorithm is able to output the better pest detection result less than 120 ms for the slowest model (VGG-16 + FPN + CAN + MDM), which meet the requirement of actual use.

5.3. Application in real-world conditions

We have deployed the algorithm proposed in this paper to mobile devices, which can classify crop disease and estimate the disease severity in real orchard/field. The structure and usage of whole equipment and system are illustrated in Fig. 9. Specifically, whenever the CCD camera captures the plant images, mobile client employs proposed method and output the plant disease as well as disease severity in real-time. According to the classification result of our method, agricultural personnel is able to decide whether to spray or not in this region and provide the guidance for variable dosing/ curative treatment. Furthermore, they can analyze the current crop growth and forecast the crop yield in this region, which is of great significance to the development of agriculture.

6. Conclusion

We propose DeepPest, a new cascading convolutional neural network architecture in this paper, to solve the problem of small object detection and imbalance data. DeepPest employ CAN to integrate prior knowledge as contextual information on the basis of results outputted by image rough classification. Additionally, MDM introduces multiple projection convolution blocks and learns the in-field pest context from the low-level convolutional layer, which generates the super-resolved feature for in-field pest. Furthermore, we have shown that appropriate attentional mechanism and data augmentation strategy are advantageous in detecting imbalanced data. Our experiments have shown the superiority of DeepPest in this paper compared with other state-of-the-art methods.

CRediT authorship contribution statement

Fangyuan Wang: Conceptualization, Methodology, Software, Visualization, Writing - original draft, Writing - review & editing. **Rujing Wang:** Project administration, Funding acquisition, Supervision. **Chengjun Xie:** Supervision, Writing - review & editing. **Po Yang:** Funding acquisition, Supervision. **Liu Liu:** Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was supported by The National Key Technology R&D Program of China under Grant 2018YFD0200300, National Natural Science Foundation of China under grant 31401293, 31671586, 61773360, Major Special Science and Technology Project of Anhui Province under grant 201903a06020006.

References

Bai, Y., Zhang, Y., Ding, M., Bernard, G., 2018. Finding tiny faces in the wild with generative adversarial network. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 21–30. https://doi.org/10.1109/CVPR.2018.00010.

Berenstein, R., Edan, Y., 2018. Automatic adjustable spraying device for site-specific agriculture application. IEEE Trans. Autom. Sci. Eng. 15, 641–650. https://doi.org/10.1109/TASE.2017.2656143.

Bozek, K., Hebert, L., Mikheyev, A., 2018. Towards dense object tracking in a 2D honeybee hive. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4185–4193. https://doi.org/10.1109/CVPR.2018.00440.

Bulat, A., Tzimiropoulos, G., 2018. Super-FAN: integrated facial landmark localization and super-resolution of real-world low-resolution faces in arbitrary poses with GANs. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 109–117. https://doi.org/10.1109/CVPR.2018.00019.

Chawla, N., Bowyer, K., Hall, L., Kegelmeyer, W., 2002. Smote: synthetic minority oversampling technique. J. Artif. Intell. Res. 16, 321–357. https://doi.org/10.1613/jair.953.

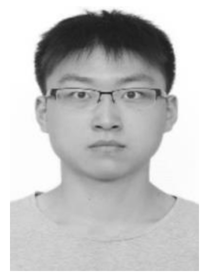
Choi, J., Chang, H., Fischer, T., Yun, S., Lee, K., Jeong, J., Demiris, Y., Choi, J., 2018. Context-aware deep feature compression for high-speed visual tracking. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 479–488. https://doi.org/10.1109/CVPR.2018.00057.

Deng, H., Birdal, T., Ilic, S., 2018. PPFNet: Global context aware local features for robust 3D point matching. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 195–205. https://doi.org/10.1109/CVPR.2018.00028.

Ding, H., Jiang, X., Shuai, B., Liu, A., Wang, G., 2018. Context contrasted feature and gated multi-scale aggregation for scene segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2393–2402. https://doi.org/10.1109/CVPR.2018.00254.

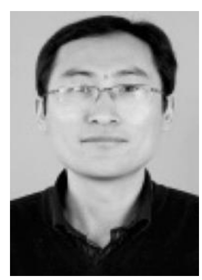
Ding, W., Taylor, G., 2016. Automatic moth detection from trap images for pest

management. Comput. Electron. Agric. 123, 17–28. https://doi.org/10.1016/j.compag.2016.02.003.
Dong, Q., Gong, S., Zhu, X., 2017. Class rectification hard mining for imbalanced deep learning. In: IEEE International Conference on Computer Vision (ICCV), pp. 1869–1878. https://doi.org/10.1109/ICCV.2017.205.
Dyrmann, M., Karstoft, H., Midtiby, H., 2016. Plant species classification using deep convolutional neural network. Biosyst. Eng. 151, 72–80. https://doi.org/10.1016/j.biosystemseng.2016.08.024.
Ebrahimi, M., Khoshtaghaza, M., Minaei, H., Jamshidi, B., 2017. Vision-based pest detection based on svm classification method. Comput. Electron. Agric. 137, 52–58. https://doi.org/10.1016/j.compag.2017.03.016.
He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778. https://doi.org/10.1109/CVPR.2016.90.
Hu, P., Ramanan, D., 2017. Finding tiny faces. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1522–1530. https://doi.org/10.1109/CVPR.2017.166.
Huang, C., Li, Y., Chen, C., Tang, X., 2016. Learning deep representation for imbalanced classification. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5375–5384. https://doi.org/10.1109/CVPR.2016.580.
Huang, Q., Xiong, Y., Lin, D., 2018. Unifying identification and context learning for person recognition. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2217–2225. https://doi.org/10.1109/CVPR.2018.00236.
Hu, J., Shen, L., Sun, G., 2018. Squeeze-and-excitation networks. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 7132-7141.
Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T., 2014. Caffe: Convolutional architecture for fast feature embedding. In: ACM international conference on. Multimedia (ACM MM, pp. 675–678.
Kang, S., Cho, J., Lee, S., 2014. Identification of butterfly based on their shapes when viewed from different angles using an artificial neural network. J. Asia-Pac. Entomol. 17, 143–149. https://doi.org/10.1016/j.aspen.2013.12.004.
Keshari, R., Vatsa, M., Singh, R., Noore, A., 2018. Learning structure and strength of CNN filters for small sample size training. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 9349–9358. https://doi.org/10.1109/CVPR.2018.00974.
Li, J., Liang, X., Wei, Y., Xu, T., Feng, J., Yan, S., 2017. Perceptual generative adversarial networks for small object detection. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1951–1959. https://doi.org/10.1109/CVPR.2017.211.
Li, Z., Peng, C., Yu, G., Zhang, X., Deng, Y., Sun, J., 2018. Detnet: Design backbone for object detection. In: European Conference on Computer Vision (ECCV). pp. 334-350.
Lin, T., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S., 2017. Feature pyramid networks for object detection. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 936–944. https://doi.org/10.1109/CVPR.2017.106.
Liu, H., Lee, S., Chahl, J., 2017. A multispectral 3D vision system for invertebrate detection on crops. IEEE Sens. J. 17, 7502–7515. https://doi.org/10.1109/JSEN.2017.2757049.
Maciejewski, T., Stefanowski, J., 2011. Local neighborhood extension of smote for mining imbalanced data. In: IEEE International Conference on Data Mining, pp. 104–111. https://doi.org/10.1109/CIDM.2011.5949434.
Niklaus, S., Liu, F., 2018. Context-aware synthesis for video frame interpolation. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1701–1710. https://doi.org/10.1109/CVPR.2018.00183.
Oquab, M., Bottou, L., Laptev, I., Sivic, J., 2014. Learning and transferring mid-level image representations using convolutional neural networks. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1717–1724. https://doi.org/10.1109/CVPR.2014.222.
Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. In: Neural Information Processing Systems (NIPS). pp. 91-99.
Santangelo, G., 2018. The impact of FDI in land in agriculture in developing countries on host country food security. J. World Bus. 53, 75–84. https://doi.org/10.1016/j.jwb.2017.07.006.
Sermanet, P., Kavukcuoglu, K., Chintala, S., LeCun, Y., 2013. Pedestrian detection with unsupervised multi-stage feature learning. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3626–3633. https://doi.org/10.1109/CVPR.2013.465.
Simonyan, K., Zisserman, A., 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition. In: International Conference on Learning Representations (ICLR).
Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A., 2017. Inception-v4, inception-resnet and the impact of residual connections on learning. In: Thirty-First AAAI Conference on Artificial Intelligence (AAAI). pp. 4278-4284.
Tang, Y., Zhang, Y., Chawla, N., Krasser, S., 2009. Svms modeling for highly imbalanced classification. IEEE Trans. Syst. Man Cybernet. Cybernet. 39, 281–288. https://doi.org/10.1109/TSMCB.2008.2002909.
Ting, K., 2000. A comparative study of cost-sensitive boosting algorithms. In: International Conference on Machine Learning, pp. 983–990.
Wang, K., Zhang, S., Wang, Z., Liu, Z., Yang, F., 2013. Mobile smart device-based vegetable disease and insect pest recognition method. Intell. Automat. Soft Comput. 19, 263–273. https://doi.org/10.1080/10798587.2013.823783.
Wu, C., Chan, Y., Fu, L., Hsiao, P., Huang, S., Chen, H., 2014. Combining multiple complementary features for pedestrian and motorbike detection. In: International IEEE Conference on Intelligent Transportation Systems, pp. 1358–1363. https://doi.org/10.1109/ITSC.2013.6728420.
Yan, C., Luo, M., Liu, H., Li, Z., Zheng, Q., 2017. Top-k multi-class SVM using multiple features. Inf. Sci. 432. https://doi.org/10.1016/j.ins.2017.08.004.
Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X., Huang, T., 2018. Generative image inpainting with contextual attention. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5505–5514. https://doi.org/10.1109/CVPR.2018.00577.
Zhou, P., Ni, B., Geng, C., Hu, J., Xu, Y., 2018. Scale-transferrable object detection. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 528–537. https://doi.org/10.1109/CVPR.2018.00062.
Zhu, C., Tao, R., Luu, K., Savvides, M., 2018. Seeing small faces from robust anchor's perspective. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5127–5136. https://doi.org/10.1109/CVPR.2018.00538.



Fangyuan Wang received the B.E. and M.S. degrees in electrical engineering and automation from Hefei University of Technology, Hefei, China, in 2014 and 2017, respectively. He is currently pursuing the Ph.D. degree in pattern recognition and intelligent system with University of science and technology of China, Hefei. His current interest is deep learning and computer vision.

Rujing Wang received the B.E. degree in computer science from Huazhong University of Science and Technology, Wuhan, China, in 1987, and M.S. degree in electronic engineering from Dalian University of Technology, Dalian, China, in 1990, and Ph.D. degree in pattern recognition and intelligent system from University of Science and Technology of China, Hefei, China, in 2005. He is currently with the Institute of Intelligent Machinery of the Chinese Academy of Sciences as Professor and Researcher. His main research interests include intelligent agriculture, agricultural internet of things, Agricultural knowledge engineering.



Chengjun Xie received the M.S. degree in software engineering from the Hefei University of Technology, Hefei, China, in 2008, and Ph.D. degree in image processing from in the Hefei University of Technology, Anhui, China, in 2011. He is currently working in the Institute of Intelligent Machinery of the Chinese Academy of Sciences as Associate Researcher. His research interests include crop disease and pest image recognition, agricultural big data, agricultural IOT.

Po Yang (M'11) received the B.E. degree in computer science from Wuhan University, Wuhan, China, in 2004, the M.S. degree in computer science from Bristol University, Bristol, U.K., in 2006, and the Ph.D. degree in electronic engineering from the University of Staffordshire, Stafford, U.K., in 2011. He is currently with the Department of Computer Science, Sheffield University, Sheffield, U.K. He has been a Postdoctoral Research Fellow with the Department of Computing, Bedfordshire University, Luton, U.K. Before he joined Bedfordshire University, he was a Research Assistant with the University of Salford, Salford, U.K. His main research interests include radio-frequency identification and sensor networking, document image processing, computer vision, GPU, and parallel computing.

Liu Liu received the B.E. degree in information engineering aerospace information applications from Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2015, and M.S. degree in advanced computer science from University of Manchester, Manchester, United Kingdom, in 2016. He is currently pursuing the Ph.D. degree in pattern recognition and intelligent system with University of Science and Technology of China, Hefei. His current interest is deep learning and computer vision.