IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received September 26, 2019, accepted October 20, 2019, date of publication October 28, 2019, date of current version November 14, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2949852

An Effective Data Augmentation Strategy for CNN-Based Pest Localization and Recognition in the Field

RUI LI[1,2], RUJING WANG[1], JIE ZHANG[1], CHENGJUN XIE[1], LIU LIU[1,2], FANGYUAN WANG[1,2], HONGBO CHEN[1], TIANJIAO CHEN[1], HAIYING HU[1], XIUFANG JIA[1], MIN HU[3], MAN ZHOU[1,2], DENGSHAN LI[1,2], AND WANCAI LIU[4]

[1]Institute of Intelligent Machines, and Hefei Institute of Physical Science, Chinese Academy Sciences, Hefei 230031, China
[2]Department of Automation, University of Science and Technology of China, Hefei 230026, China
[3]Anhui Province Key Laboratory of Affective Computing and Advanced Intelligent Machine, School of Computer and Information, Hefei University of Technology, Hefei 230009, China
[4]National Agro-Tech Extension and Service Center, Beijing 100125, China.

Corresponding authors: Rujing Wang (rjwang@iim.ac.cn), Jie Zhang (76609080@qq.com), and Chengjun Xie (cjxie@iim.ac.cn)

This work was supported in part by the National Key Technology Research and Development Program of China under Grant 2018YFD0200300, in part by the National Natural Science Foundation of China under Grant 31401293, Grant 31671586, and Grant 61773360, in part by the Chinese Academy of Science and Technology Service Network Planning under Grant KFJ-STS-ZDTP-048-02, and in part by the Fundamental Research Funds for the Central Universities of China under Grant ACAIM190101.

ABSTRACT In agriculture, pest always causes the major damage in fields and results in significant crop yield losses. Currently, manual pest classification and counting are very time-consuming and many subjective factors can affect the population counting accuracy. In addition, the existing pest localization and recognition methods based on Convolutional Neural Network (CNN) are not satisfactory for practical pest prevention in fields because of pests' different scales and attitudes. In order to address these problems, an effective data augmentation strategy for CNN-based method is proposed in this paper. In training phase, we adopt data augmentation through rotating images by various degrees followed by cropping into different grids. In this way, we could obtain a large number of extra multi-scale examples that could be adopted to train a multi-scale pest detection model. In terms of test phase, we utilize the test time augmentation (TTA) strategy that separately inferences input images with various resolutions using the trained multi-scale model. Finally, we fuse these detection results from different image scales by non-maximum suppression (NMS) for the final result. Experimental results on wheat sawfly, wheat aphid, wheat mite and rice planthopper in our domain specific dataset, show that our proposed data augmentation strategy achieves the pest detection performance of 81.4% mean Average Precision (mAP), which improves 11.63%, 7.93% ,4.73% compared to three state-of-the-art approaches.

INDEX TERMS Pest localization, pest recognition, convolutional neural network, multi-scale, data augmentation.

I. INTRODUCTION

In modern agricultural production, pest predicting and forecasting are very important for Integrated Pest Management (IPM) in field. However, the manual recognition and counting are very time-consuming, labor intensive and inefficient. Automatic pest localization and recognition in field can dramatically reduce labor intensity and improve work efficiency.

With the rapid development of computer vision technology, automatic disease and pest localization and recognition has become a research hotspot in recent years [1], [6], [11], [12]. All the above methods could achieve satisfactory performances, but they aim to detect pests under simple background conditions rather than with the complex backgrounds typical of the field environment.

The associate editor coordinating the review of this manuscript and approving it for publication was Tossapon Boongoen.

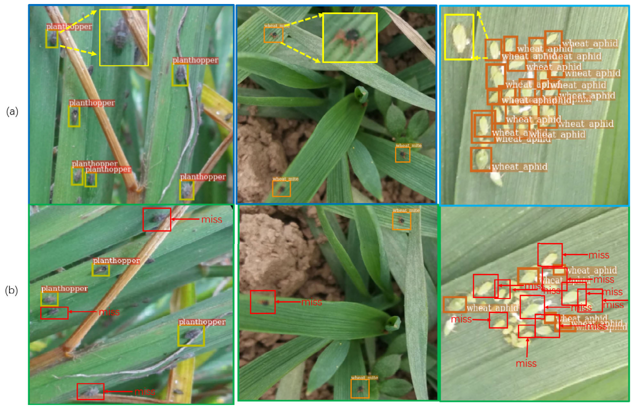


FIGURE 1. **(a) The detection results before rotation. (b) The detection results after 60° rotation. The numbers of pest are much less than before rotation.**

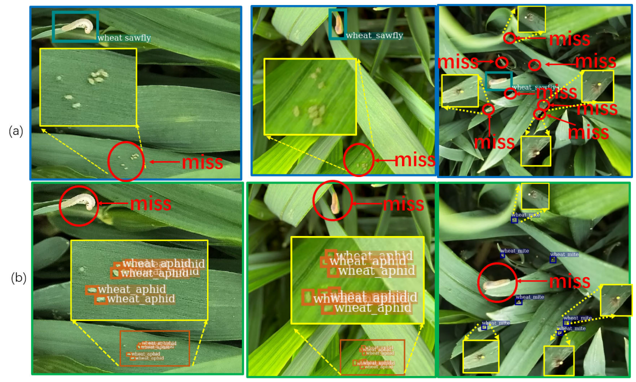


FIGURE 2. **(a) When image size is 800*600, the larger wheat sawfly is detected, but the smaller wheat aphids, wheat mite are miss. (b) When image size is 1400*1200, there is a better detection result for wheat aphids, wheat mite but the wheat sawfly is miss.**

In practical application, the scales, different attitudes, complexity of image background and illumination are the major challenges in pest localization and recognition. These traditional approaches always have low accuracy and poor robustness in practical pest monitoring. Recent advances in deep learning techniques based on convolutional neural network (CNN) have led to significantly promising progress in object localization and recognition under natural conditions like R-CNN [19], SPP-net [20], Fast-RCNN [21], Faster-RCNN [22], SSD [23], YOLO [24], R-FCN [27], and other extended variants of these networks [25]–[26]. Faster RCNN, SSD, YOLO are effective object detectors towards general object detection in a specific field, but intractable for use in practical real-world multi-scale object detection. Feature Pyramid Network (FPN) [28] and DSSD [29] could achieve satisfied results in multi-scale object detection.

CNN has achieved good performance in object localization and recognition, but in wild field, pest localization and recognition in the field have many constraints: 1) The rotation invariance is not specially processed by CNN [30], which only compensates through maximum pooling. The detection result may not be satisfactory, when the image rotation angle changes too large. As FIGURE1(b) shows the numbers of detected pests (planthopper, wheat mite, wheat aphid) are much less than those before rotating (as FIGURE1 (a) shown). 2) Different scales (changes of image resolution) may result lower detection rate. As FIGURE2 shows the scale of wheat sawfly is dozen times larger than that of wheat mite, wheat aphid and rice planthopper in field. FIGURE2 shows that it is difficult to detect tiny and large size pests at the same time. Therefore, in this paper, we developed an effective data augmentation strategy for CNN-based method, which can improve the localization and recognition accuracy of pests in field. To address rotation invariant problem, in training phase, we pre-process training samples through rotating them by five degrees, so that the different attitudes of pests would be obtained. To tackle multi-scale problem, the trainset after rotating is cropped by different grids. In the way of dividing blocks, the proportion of each pest in the image is different so that the multi-scale of the training samples could be obtained. Finally, a multi-scale and multi-rotation pest localization and recognition model is trained by these augmented images. In order to improve the robustness of multi-resolution, in terms of test phase, we utilize the test time augmentation (TTA) strategy to split the input image into four different resolutions so that they could be separately detected by the trained multi-scale model. Finally, we fuse the four detection results from different image resolutions for the final one.

The major contributions of multi-scale CNN are as follows:

1. An effective data augmentation in training phase is proposed, which is feasible to apply for practical pest prevention. Experimental results demonstrate the advantages of the proposed algorithm over the state-of-the-art approach.

2. A test time augmentation (TTA) strategy in test phase is proposed in this paper. Compared with the method without multi-resolutions test, our method can improve 2.1%.

II. RELATED WORK

There are many advanced techniques developed and applied in agricultural field. To monitor the variety and quantity of cucumber pests in greenhouse and predict the development trend of pest, Yang *et al.* [3] proposed an image recognition algorithm based on Prewitt, Canny edge detection operator segmentation and support vector machine (SVM). However, their monitoring equipment is fixed and can only detect specific pests in a small region. To tackle these issues, Wang *et al.* [5] developed a mobile smart device-based vegetable disease and insect pest recognition method. Compared with fixed monitoring equipment, the smart mobile is more convenient for agricultural technicians to monitor pests and diseases, but this method has low accuracy and poor robustness in complex background. In order to improve the accuracy, Xie *et al.* [7] developed an insect recognition system using advanced multiple-task sparse representation and multiple-kernel learning (MKL) techniques. Although this method achieved a satisfactory performance, it takes longer time to train the dictionary for sparse coding and it focuses on pest classification rather than localization. Furthermore, some researchers have focused on the methods for detection of specific pests and obtained satisfactory results. Wen and Guyer [4] developed a feature fusion method for orchard insect localization and recognition with an accuracy of 96%. Dey *et al.* [9] presented an automated method based on Gray Level Run Length Matrix and Gray Level Co-occurrence Matrix operator feature extraction for detecting white fly pests. All above methods had achieved good performances, but their applications are limited to specific pests and cannot be widely used. Thus, some pest localization and recognition methods for crops are proposed in recent years. To monitor the variety pests in vegetable field and predict the development trend of pest, Xiao *et al.* [2] developed a visual perception method for counting vegetable pests. In cotton ecosystems, a neural network method is proposed by Gassoumi *et al.* [8] for detecting cotton pest. Roldan-Serrato *et al.* [10] studied a recognition system for the Colorado potato beetle with a recognition rate of 85% by using a special neural network and the random subspace classifier. However, all above methods might mainly focus on hand-crafted feature extraction and classifiers, which suffer from a number of limitations. In the practical application, the complexity of image background, illumination, scales and different attitudes are the major challenges in feature extraction and classifiers.

To tackle these issues, some researchers have focused on deep learning. Wang *et al.* [16] used deep CNN to classify 82 types of crop pest images in the field with the accuracy of 91%. However, this method only concentrates on classification rather than quantitative counting. To monitor the quantity of adult red turpentine beetles, Sun *et al.* [13] proposed a deep learning detection method, which could show the promising performance both qualitatively and quantitatively. Furthermore, there are many researchers focus on detecting the pests in fixed monitoring equipment. Shen *et al.* [15] developed a detection and identification method with 88% mAP for six species of common stored-grain insects. Ding and Taylor [18] developed an automatic detection pipeline based on deep learning for identifying and counting pest insects in images taken inside field traps. Liu *et al.* [14] proposed a region-based end-to-end approach named PestNet for large-scale multi-class pest localization and recognition based on deep learning and achieved 75.46% mAP for 16 species of pests. These methods all had good performances, but they were constrained by the limitation that fixed regions could not be widely applied in wild field. Pest predicting and forecasting on large-scale areas are very important for IPM. They can help agricultural technicians understand the occurrence of pests more comprehensively and formulate control programs more quickly. Thus, the detection of pest on large-scale areas in the wild field had become the research hotspots in recent years. Liu *et al.* [17] presented a pipeline for the visual localization and classification of agricultural pest insects by computing a saliency map and applying deep convolutional neural network learning. However, the pests might be incorrectly located, due to the saliency maps method's poor robustness. For pest detection task, CNN has become a widely used method to extract pest features. However, the rotation invariance is not specially processed by CNN and the existing object detection methods based on convolutional neural network are not satisfactory for multi-scale pest localization and recognition in the field. Therefore, we propose an effective data augmentation strategy for CNN-based pest localization and recognition in field, which improves performance by the following aspects. Firstly, in training step, a data augmentation method is proposed for generating multi-scale detection model. Secondly, for pest localization and recognition, a test time augmentation strategy is used. We split the input image into four different scales for detecting pests of different sizes. The four detection results are fused for the final result. Compared to three state-of-the-art approaches, our method achieved satisfactory results.

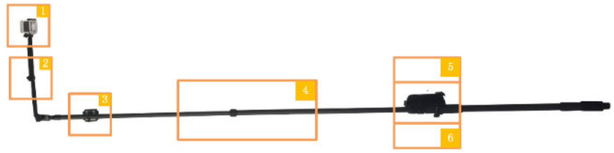


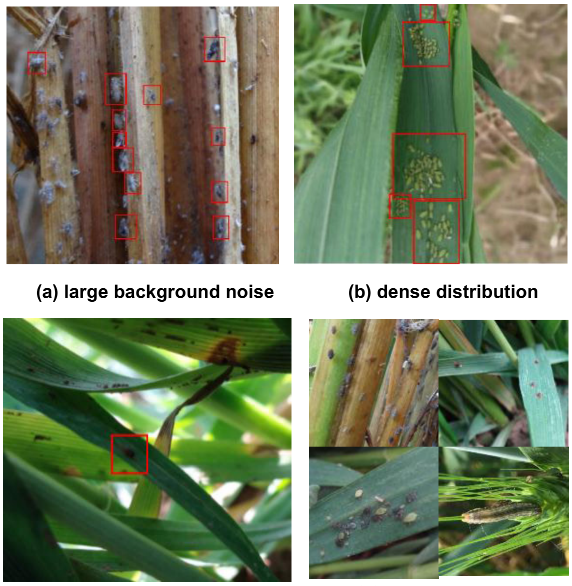
FIGURE 3. pest intelligent collection equipment. (1) CCD camera; (2) soil moisture sensor; (3) ambient temperature and humidity sensor; (4) carbon fiber telescopic rod; (5) mobile client; (6) global positioning system. CCD camera is used to collect pest images and controlled by mobile client. Soil moisture sensor, ambient temperature and humidity sensor are used to obtain the soil moisture, temperature and humidity in field, they are all transferred to cloud server by using mobile client. The position of field would be located by global positioning system.

FIGURE 4. The usage of pest intelligent collection equipment in the field.

III. MATERIALS AND METHODS

A. DATASET

All the images analyzed in this paper were collected in the Anhui province of China. These images were captured by independent research and development device called pest intelligent collection equipment, whose structure and usage are shown in FIGURE 3 as well as FIGURE 4. The resolution



(a) large background noise (b) dense distribution

(c) sunlight illumination influence (d) different attitudes

FIGURE 5. Dataset characteristics.

TABLE 1. Statistics on two subsets for our dataset with training subset and validation subset. For each pest, the number of images and objects are shown in this table. The totals shown in the '#images' columns are not simply the sum of the corresponding columns because one image might contain several types of pests.

pest name	Training		Validation	
	#images	#objects	#images	#objects
wheat sawfly	1600	2080	129	172
wheat mite	1450	8257	136	754
wheat aphid	1630	9674	147	921
planthopper	1000	7362	100	752
Total	4000	27373	400	2599

of these images taken by CCD camera with 4 mm focal length with an aperture of f/3.3 is 1440 × 1080 pixels. Our dataset covers four kinds of pest (wheat mite, wheat aphid, wheat sawfly and rice planthopper) and various attributes in field: large background noise, dense distribution, sunlight illumination influence that are shown in FIGURE 5. Then we select 4400 images and randomly split the dataset into training subset and validation subset at ratio of 9:1. The statistics of these two subsets are illustrated in Table 1.

After image acquisition, our dataset is annotated with several labels and bounding boxes by agricultural experts to guarantee the professionalism of these annotations. The LabelImg (https://tzutalin.github.io/labelImg/) is used as our annotation tool. The annotation files are saved in XML format as PASCAL VOC-style.

B. MODEL ARCHITECTURE

The state-of-the-art object detection methods contained CNN backbone and region proposal network (RPN) as shown in FIGURE 6. Convolutional feature maps are extracted by

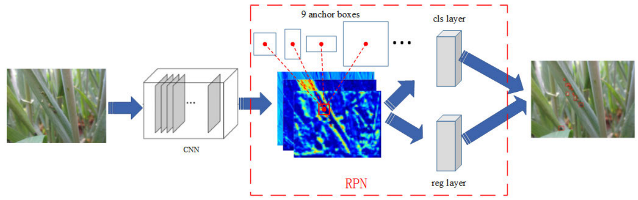


FIGURE 6. Convolutional neural network based pest detection architecture.

CNN and the object's location and class are obtained by using RPN.

1) CONVOLUTIONAL NEURAL NETWORK (CNN)

Convolutional neural network (CNN) is a kind of artificial neural network (ANN) based on the deep learning theory, and it can automatically extract spatial features from 2D grid-style images, so it has been widely used in the field of object recognition and detection. CNN is mainly composed of convolutional layer, activation function, pooling layer and fully connected layer.

a: CONVOLUTIONAL LAYER

Convolutional layer is used to extract image features, and it uses the convolution operation to replace the matrix multiplication operation in the traditional neural network for learning the mapping between input and output layers. The parameter sharing in the convolution operation allows the network to learn only one parameter set, which greatly reduces the number of parameters and significantly improves the computational efficiency. Convolution operation is defined as:

$$a_{i,j} = \sum_{m=0}^{s} \sum_{n=0}^{s} w_{m,n} x_{i+m,j+n} \tag{1}$$

where $w_{m,n}$ is the weight of convolutional kernel at $mandn$, $x_{i,j}$ is the pixel value of image at i and j, s is the height and width of convolutional kernel.

b: ACTIVATION FUNCTION

Rectified linear unit (ReLU) is usually used as activation function in CNN for avoiding vanishing gradient and improving training speed. The function of ReLU is defined as:

$$ReLU\,(x) = \begin{cases} x & x > 0 \\ 0 & x \le 0 \end{cases} \tag{2}$$

c: POOLING LAYER

The pooling layer can make the information in feature maps more concentrated and the computational complexity of the network simplified. Max pooling is usually used as pooling layer:

$$MaxPool(h_o, w_o) = \begin{cases} h_o = floor\left(\dfrac{(h_i + 2p - k)}{s} + 1\right) \\ w_o = floor\left(\dfrac{(w_i + 2p - k)}{s} + 1\right) \end{cases} \tag{3}$$

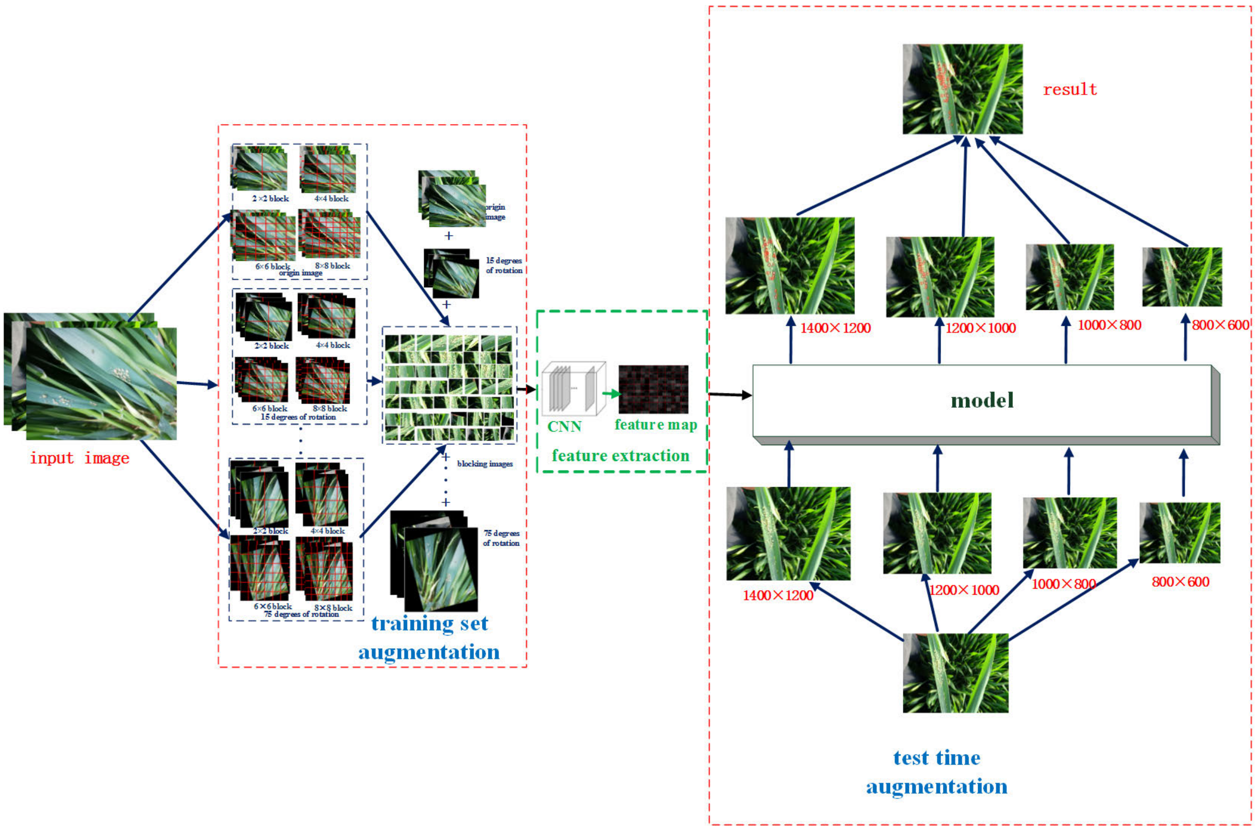


FIGURE 7. Framework of automatic pest recognizing and detecting based on the multi-scale CNN.

where *floor* (x) is the function of round up the number, h_o is the output height of feature map, w_o is the output width of feature map, h_i is the input height of feature maps, w_i is the input width of feature maps, p is the padding of feature maps, k is the kernel size of max pooling, s is the kernel stride of max pooling.

2) REGION PROPOSAL NETWORK (RPN)

The feature maps from the CNN are fed into RPN and generate many candidate bounding boxes with their corresponding probability scores. Specifically, all the boxes are obtained via a sliding window to traverse each position in feature maps, whose size is set to 6×6 in this paper. Then a classification layer and a box-regression layer are designed to predict the scores and offsets of bounding boxes respectively. Finally, the pests are localized by the RPN architecture.

C. DATA AUGMENTATION

In order to improve the poor pest localization and recognition performance caused by the limitations of single-scale and angle-invariance, a data augmentation strategy is proposed in this paper, including trainset augmentation and test time augmentation as shown in FIGURE7.

1) TRAINSET AUGMENTATION

There are two methods used in trainset augmentation step: multi-scale and rotation augmentation.

a: ROTATION AUGMENTATION

The specific steps of rotation augmentation are as follows:

In training step, a training image would be rotated vertically for three consecutive times. Thus, the training images (4000 pest images) are only rotated by five degrees (15, 30, 45, 60 and 75) as shown FIGURE 7, thus 20,000 new images are obtained after rotating. Finally, a total of 24,000 new training subset are generated. The new annotation of rotated image can be figured out as FIGURE 8 shown and the process are calculated as follows:

a) Firstly, the coordinates of the vertices of bounding box are calculated:

$$\begin{cases} \mathrm{x}' = (x - x_c)\cos\theta - (y - y_c)\sin\theta + x_c \\ \mathrm{y}' = (x - x_c)\sin\theta + (y - y_c)\sin\theta + x_c \end{cases} \tag{4}$$

where x_c, y_c are the central coordinates of image, $x_c = w/2$, $y_c = h/2$; x, y are the coordinates of bounding box before rotation, x', y' are the coordinates of bounding box after rotation , θ is the rotation angle.

b) Then, the top-left point of new bounding box is:

$$\begin{aligned} \mathrm{x}_{t,l} &= \min(x'_1, x'_2, x'_3, x'_4) \\ \mathrm{y}_{t,l} &= \min(y'_1, y'_2, y'_3, y'_4) \end{aligned} \tag{5}$$

The down-right point of new bounding box is:

$$\begin{aligned} \mathrm{x}_{d,r} &= \max(x'_1, x'_2, x'_3, x'_4) \\ \mathrm{y}_{d,r} &= \max\left(y'_1, y'_2, y'_3, y'_4\right) \end{aligned} \tag{6}$$

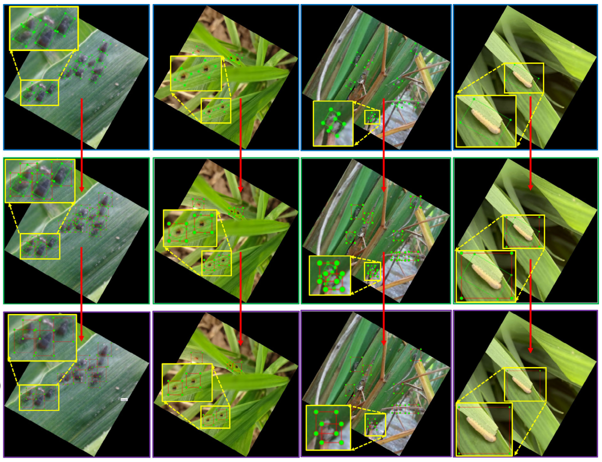


FIGURE 8. The calculating process of new annotation coordinate after image rotating. (a) The rotated coordinates are not a horizontal rectangle, these annotation files cannot be used in CNN. (b) We take the circumscribed rectangle of the original rectangle as the new annotation rectangle. (c) The final annotation file.

where $x_{t.l}$ is the x-coordinate of top-left point, $y_{t.l}$ is the y-coordinate of top-left point, $x_{d.r}$ is the x-coordinate of down-right point, $y_{d.r}$ is the y-coordinate of down-right point; x_i' is the new x-coordinate of bounding box, y_i' is the new y-coordinate of bounding box.

b: MULTI-SCALE AUGMENTATION

The specific steps of multi-scale augmentation are as follows:

a) The images in trainset are cut with blocks of $2 \times 2, 4 \times 4, 6 \times 6$ and 8×8 as shown FIGURE 7, in which the block size is set by considering image resolution and pest sizes Then we filter these cropped images by removing the sub-images that do not contain any pest objects.
b) All the training images are normalized to 1200×1000. Finally, a total of 219,752 images generated by rotating and cropping are used as the final training subset to train our pest detection model.

After image cutting, the scales of pests in sub-image are significantly enlarged.

2) TEST TIME AUGMENTATION

The multi-scale trainsets are obtained in training step, but the pests in trainsets are all labeled in a fixed-size image, so some pests are not located and recognized in one resolution. Thus, in order to detect different sizes of pest, test image is detected with four different resolution as shown FIGURE 7. The specific steps are as follows:

a) The test image is respectively resized to 1400×1200,1200×1000,1000×800 and 800×600. The size of 1400×1200 is used to detect the tiny pests such as wheat mites, nymph of wheat aphids, nymph of rice planthoppers. The moderate size of wheat aphids and rice planthoppers would be detected by 1200×1000. The sizes of 1000×800 and 800×600 are used to detect the larger wheat aphids, rice planthoppers and wheat sawfly.
b) The pest coordinates in the four detection images are unified the size of 1200×1000.
c) There are many overlapping boxes after merging coordinates, to solve this issue, the non-maximum suppression (NMS) is adopted in this paper.

IV. EXPERIMENTS

A. EXPERIMENTAL SETTINGS

Some experiments are built to evaluate the performance of our method. Caffe2 [34] with Python API 2.7 is used in our experiment and run on 12GB Tesla P40 GPU. We preprocessed the trainset by horizontal and vertical flips. All models were trained with mini-batch stochastic gradient descent (SGD) with a mini-batch size of 2. And the SGD was set to 0.9, this optimizer could partly keep the update gradient from previous iteration and fine-tune the gradient considering the current mini-batch. We found the suggested learning rate of 0.02 to be too high, using 0.01 instead. Dropout method [35] was utilized for avoiding over-fitting problem as well as early-stopping strategy [36] to select the best training iteration. Furthermore, the iterations of training are 500,000 times, the Resnet-50 is used as backbone of CNN.

B. EVLUATION METRICS

For validating the performance of our model in detecting pests, we select Precision-Recall (PR) curve and Average Precision (AP) as the evaluation metric in this task.

Precision-Recall (PR) curve represents the balance between false positive reduction and misdetections. We used the area under precision-recall curve to precisely measure the performance of the method, The Precision-Recall (PR) is calculated by:

$$\begin{aligned} Precision\,(c) &= \frac{TP\,(c)}{TP\,(c)+FP\,(c)} \\ Recall(c) &= \frac{TP(c)}{TP(c)+FN(c)} \end{aligned} \tag{7}$$

where $\boldsymbol{c}$ denotes the class, in which TP, FP and FN represent True Positive, False Positive and False Negative samples respectively, so the Precision measures the samples that are incorrectly detected while Recall measures those misdetection samples.

AP is updated for combining localization and recognition task together. Given an IoU thresh, the AP is defined as the area under Precision-Recall:

$$AP = \int \text{Precision } d \text{ Recall} \tag{8}$$

in which the Precision measures the samples that are incorrectly detected and Recall measures those misdetection samples. Mean Average Precision (mAP) is the mean of Average Precision (AP) value among classes and is obtained by taking mean:

$$mAP = \frac{1}{|C|} \sum\nolimits_{c \in C} AP(c) \tag{9}$$

TABLE 2. Result of different methods.

method	wheat sawfly (AP)	wheat aphid (AP)	wheat mite (AP)	rice planthopper (AP)	mAP
DAG-CNN	88.6%	78.9%	63.47%	64.12%	73.76%
HR	89.31%	84.8%	65.83%	67.15%	76.77%
FPN	87.39%	75.77%	57.3%	54.51%	68.74%
Ours	**90.88%**	**88.76%**	**70.2%**	**75.77%**	**83.23%**

(a)
(b)
(c)
(d)

FIGURE 9. Results of our method. (a) and (c) The results without augmentation were used. (b) and (d) The results with augmentation. The blue boxes are the result of wheat mites, the chocolate boxes are the result of wheat aphids, the red boxes are the result of rice planthopper and the pink boxes are the result of wheat sawfly. The results show that data augmentation method could improve the accuracy and robustness of pest detection.

V. RESULTS AND DISCUSSION

A. TRAIN AND TEST TIME AUGMENTATION RESULTS

Four state-of-the-art methods are experimented and the results are shown in Table 2, the best result is marked with the bold font. We firstly observe that our data augmentation method has the highest detection accuracy. DAG-CNN [31]and HR [32] both used the multi-scale methods for the object detection. DAG-CNN realized the multi-scale object detection by extracting features from multiple layers. HR used one model to detect different resolutions. Generative Adversarial Network (GAN) [33] is used to enhance the pest samples for ensuring the same training numbers as our method. The experimental results show that among all of the approaches, the best detection performance occurs in our data augmentation method, which achieves mAP with 83.23%. Compared with DAG-CNN and HR, rotation invariance is processed in our method, so it could improve the detection rate of pests in field. The experimental results show that the mAP values of the method proposed in this paper increase by 12.44%, 7.93%, and 4.72% than those of previous three methods, respectively, as FIGURE 9 shown, it has a higher accuracy and robustness in the automatic identification and counting of pests than without data augmentation.

B. ROTATION ANALYSIS

The rotation invariance is not specially processed by CNN, when the angle greatly changes, there might generate some effects on the detection results. Therefore, the effects on image rotation is experimented by using FPN, and the results

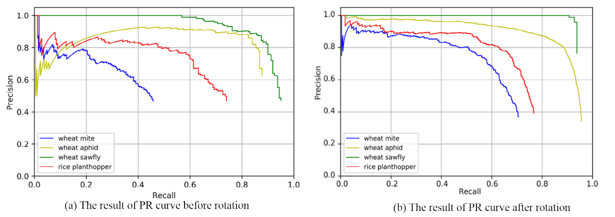


FIGURE 10. Rotation detection results. (a) The PR curve before the rotation images join the training set. (b) The PR curve after the rotation images joins the training set.

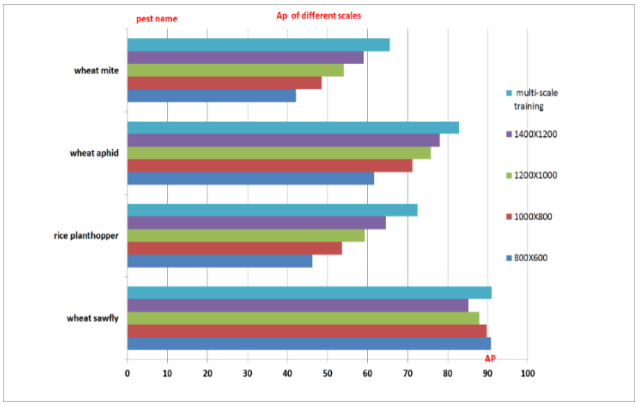


FIGURE 11. Result of multi-scale training. The experiment shows that compared with single-scale model, our multi-scale training has a good performance in both tiny and large pest detection.

are shown as Figure 10. The experiment shows that compared with the result before the rotation images joining the training set, the PR curve of four pest obviously rise after rotation. Rotating samples are added to the training set to compensate for the defects on rotation invariance of CNN. Thus, it is proved that the effect of localization and recognition could be improved by adding the rotating images into the training.

C. MULTI-SCALE TRAINING ANALYSIS

The scales of pest are very different, for instance, the scale of wheat sawfly is dozen times larger than that of wheat mite. Thus, it cannot have good detection results in both tiny and large size pests at the same time, which brings great difficulties in the pest localization and recognition. In order address this problem, multi-scale training is experimented by using FPN in this paper. Firstly, the scales of training images are normalized to 800×600, 1000×800, 1200×1000, and 1400×1200. As shown in Figure 11, average precision (Ap) of wheat aphid, wheat mite and rice planthopper all increase with the scale increasing except for wheat sawfly. The reason is that wheat aphid, wheat mite and rice planthopper are all tiny pests, and their scale ratio is less than 0.5% of the whole image scale. However, the proportion of wheat sawfly is much larger, it is insensitive to detect large-scale objects by using a tiny-scale model. Hence, it is important for the multi-scale processing on training samples so as to detect both large objects and small objects at the same time.

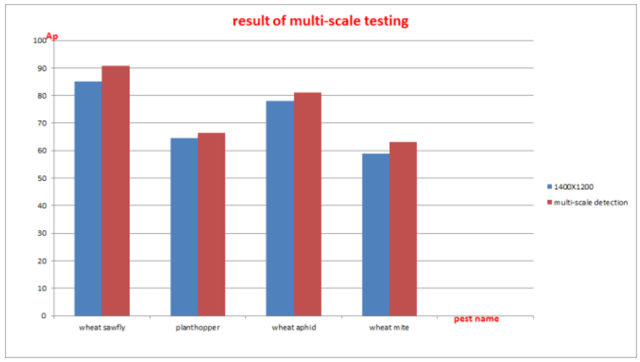


FIGURE 12. Result of test time augmentation strategy.

As shown in Figure 11, Ap of pests all increased by using multi-scale training, the experiment shows that multi-scale training could improve the detection.

D. TEST TIME AUGMENTATION ANALYSIS

In order to detect more pests in field, in addition to multi-scale training, we also carried out experiments of test time augmentation. Figure 12 shows the detection results with a single scale of 1400*1200 by using multi-scale training model and the four-scales fusing detection results by using multi-scale training model. We firstly observe that multi-scale testing has the highest detection accuracy. Fixed templates with different scales are used in multi-scale training, so multi-scale training cannot guarantee to detect all different scales of pest, but multi-scale testing could address this problem. Thus, the findings suggest that the multi-scale tests can improve the effects of recognition and detection.

VI. CONCLUTION

In order to improve the detection accuracy of pests with different scales and different attitudes in field, an effective data augmentation method is proposed in this paper. The experiment results show that our method could improve the accuracy of pest localization and recognition. Specifically:

[1] In consideration of CNN without the characteristics of rotation invariance, the rotating images are added into training set, compared with training set without rotation, the Ap of pests all improves.

[2] Based on the image feature extracted by CNN without scale invariance, the training sample can be divided into different blocks, and the images with the pests are selected as the training sets. The proportion of pests in each image is different after blocking, and multi-scale training set is obtained. Compared with other methods our method makes the AP of four kinds of pests increase by 6.1%.

[3] Due to the defect of detection for single-scale image after multi-scale training, the testing image is reset to four scales. Compared with the single scale detection, our test time augmentation could improve the effects of detection.

In this paper, the proposed method could realize the localization and recognition of pests in field environment, which provides a new way for intelligent prediction in the field of crop protection.

REFERENCES

[1] X. Zhang, Y. Qiao, F. Meng, C. Fan, and M. Zhang, "Identification of maize leaf diseases using improved deep convolutional neural networks," *IEEE Access.*, vol. 6, pp. 30370–30377, 2018.

[2] D. Xiao, Y. Zhang, M. Fan, C. Pan, W. Yao, and J. Cai, "Vegetable pest counting algorithm based on visual perception," *Trans. Chin. Soc. Agricult. Mach.*, vol. 49, no. 03, pp. 51–58, Mar. 2018.

[3] X. Yang, M. Liu, J. Xu, L. Zhao, S. Wei, W. Li, M. Chen, and M. Li, "Image segmentation and recognition algorithm of greenhouse whitefly and thrip adults for automatic monitoring device," *Trans. Chin. Soc. Agricult. Eng.*, vol. 34, no. 1, pp. 164–170, Jan. 2018.

[4] C. Wen and D. Guyer, "Image-based orchard insect automated identification and classification method," *Comput. Electron. Agricult.*, vol. 89, no. 3, pp. 110–115, Nov. 2012.

[5] K. Wang, S. Zhang, Z. Wang, Z. Liu, and F. Yang, "Mobile smart device-based vegetable disease and insect pest recognition method," *Intell. Automat. Soft Comput.*, vol. 19, no. 3, pp. 263–273, 2013.

[6] R. Li, R. Wang, C. Xie, L. Liu, J. Zhang, F. Wang, and W. Liu, "A coarse-to-fine network for aphid recognition and detection in the field," *Biosyst. Eng.*, vol. 187, pp. 39–52, Nov. 2019.

[7] C. Xie, J. Zhang, R. Li, J. Li, P. Hong, and J. Xia, "Automatic classification for field crop insects via multiple-task sparse representation and multiple-kernel learning," *Comput. Electron. Agricult.*, vol. 119, pp. 123–132, Nov. 2015.

[8] H. Gassoumi, N. R. Prasad, and J. J. Ellington, "Neural network-based approach for insect classification in cotton ecosystems," in *Proc. Int. Conf. Intell. Technol.*, Dec. 2000, pp. 13–15.

[9] A. Dey, D. Bhoumik, and K. N. Dey, "Automatic detection of whitefly pest using statistical feature extraction and image classification methods," *Int. Res. J. Eng. Technol.*, vol. 3, no. 9, pp. 950–959, 2016.

[10] L. Roldan-Serrato, T. Baydyk, E. Kussul, A. Escalante-Estrada, and M. T. G. Rodriguez, "Recognition of pests on crops with a random subspace classifier," in *Proc. 4th Int. Work Conf. Bioinspired Intell. (IWOBI)*, Jun. 205, pp. 21–26.

[11] Y. Li, C. Xia, and J. Lee, "Vision-based pest detection and automatic spray of greenhouse plant," in *Proc. IEEE Int. Symp. Ind. Electron.*, Jul. 2009, pp. 920–925.

[12] Q. Yao, G.-T. Chen, Z. Wang, C. Zhang, B.-J. Yang, and J. Tang "Automated detection and identification of white-backed planthoppers in paddy fields using image processing," *J. Integrative Agricult.*, vol. 16, no. 7, pp. 1547–1557, Jul. 2017.

[13] Y. Sun, X. Liu, M. Yuan, L. Ren, J. Wang, and Z. Chen, "Automatic in-trap pest detection using deep learning for pheromone-based dendroctonus valens monitoring," *Biosyst. Eng.*, vol. 176, pp. 140–150, Dec. 2018.

[14] L. Liu, R. Wang, C. Xie, P. Yang, F. Wang, C. Xie, P. Yang, F. Wang, S. Sudirman, and W. Liu, "PestNet: An end-to-end deep learning approach for large-scale multi-class pest detection and classification," *IEEE Access*, vol. 7, pp. 45301–45312, 2019.

[15] Y. Shen, H. Zhou, J. Li, F. Jian, and D. S. Jayas, "Detection of stored-grain insects using deep learning," *Comput. Electron. Agricult.*, vol. 145, pp. 319–325, Feb. 2018.

[16] R. Wang, J. Zhang, W. Dong, C. Xie, and R. Li, "A crop pests image classification algorithm based on deep convolutional neural network," *Telkomnika*, vol. 15, no. 3, pp. 1239–1246, 2017.

[17] Z. Liu, J. Gao, G. Yang, H. Zhang, and Y. He, "Localization and classification of paddy field pests using a saliency map and deep convolutional neural network," *Sci. Rep.*, vol. 6, Feb. 2016, Art. no. 20410.

[18] W. Ding and G. Taylor, "Automatic moth detection from trap images for pest management," *Comput. Electron. Agricult.*, vol. 123, pp. 17–28, Apr. 2016.

[19] R. Girshick, J. Donahue, T. Darrelland, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 580–587.

[20] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Sep. 2014.

[21] R. Girshick, "Fast R-CNN," Apr. 2015, *arXiv:1504.08083*. [Online]. Available: https://arxiv.org/abs/1504.08083

[22] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[23] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2016, pp. 21–37.

[24] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.

[25] Z. Cai and N. Vasconcelos, "Cascade R-CNN: Delving into high quality object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 6154–6162.

[26] Y. Chen, W. Li, C. Sakaridis, D. Dai, and L. Van Gool, "Domain adaptive faster R-CNN for object detection in the wild," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2018. pp. 3339–3348.

[27] J. Dai, Y. Li, K. He, and J. Sun, "R-FCN: Object detection via region-based fully convolutional networks," in *Proc. Neural Inf. Process. Syst. Conf. (NIPS)*, 2016, pp. 379–387.

[28] T. Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 936–944.

[29] C. Y. Fu, W. Liu, A. Ranga, and A. Tyagi, "DSSD: Deconvolutional single shot detector," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017. [Online]. Available :https://arxiv.org/pdf/1701.06659

[30] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2017, pp. 3856–3866.

[31] S. Yang and D. Ramanan, "Multi-scale recognition with DAG-CNNs," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Dec. 2015, pp. 1215–1223.

[32] P. Hu and D. Ramanan, "Finding tiny faces," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1522–1530.

[33] D. Hu, L. Wang, W. Jiang, S. Zheng, and B. Li, "A novel image steganography method via deep convolutional generative adversarial networks," *IEEE Access*, vol. 6, pp. 38303–38314, 2018.

[34] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," in *Proc. ACM Conf. Multimedia*, Nov. 2014, pp. 675–678.

[35] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, pp. 1929–1958, Jan. 2014.

[36] R. Caruana, S. Lawrence, and C. L. Giles, "Overfitting in neural nets: Backpropagation, conjugate gradient, and early stopping," in *Proc. Adv. Neural Inf. Process. Syst.*, 2001, pp. 402–408.

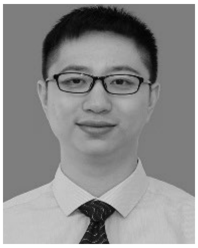
RUI LI received the B.E. degree from the Hebei University of Architecture, China, in 2010, and the M.S. degree in computer applied technology from the Hefei University of Technology, Hefei, China, in 2013. He is currently pursuing the Ph.D. degree in electronic information with the University of Science and Technology of China, Hefei. His current research interests include deep learning and computer vision.

RUJING WANG received the B.E. degree in computer science from the Huazhong University of Science and Technology, Wuhan, China, in 1987, the M.S. degree in electronic engineering from the Dalian University of Technology, Dalian, China, in 1990, and the Ph.D. degree in pattern recognition and intelligent system from the University of Science and Technology of China, Hefei, China, in 2005.

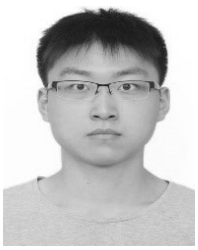
He is currently with the Institute of Intelligent Machines, Chinese Academy of Sciences, as a Professor and a Researcher. His current research interests include intelligent agriculture, the agricultural Internet of Things, and agricultural knowledge engineering.

JIE ZHANG received the M.S. degree from the Hefei University of Technology, in 2009, and the Ph.D. degree from the University of Science and Technology of China, in 2014. He was with the Institute of Intelligent Machines, Hefei Institutes of Physical Science, Chinese Academy of Sciences, where he is currently an Associate Professor. His current research interests include image processing, pattern recognition, and artificial intelligence.

CHENGJUN XIE received the M.S. degree in software engineering and the Ph.D. degree in image processing from the Hefei University of Technology, Hefei, Anhui, China, in 2008 and 2011, respectively. He is currently an Associate Researcher with the Institute of Intelligent Machines, Chinese Academy of Sciences. His current research interests include crop disease and pest image recognition, agricultural big data, and the agricultural Internet of Things.



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



FANGYUAN WANG received the B.E. and M.S. degrees in electrical engineering and automation from the 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 the University of Science and Technology of China, Hefei. His current research interests include deep learning and computer vision.

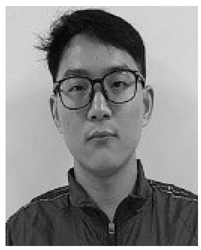
HONGBO CHEN received the B.S. degree from Hefei University, in 2012, and the M.S. degree from the Hefei University of Technology, in 2015. He was with the Institute of Intelligent Machines, Hefei Institutes of Physical Science, Chinese Academy of Sciences. His current research interests include image processing, pattern recognition, and artificial intelligence.

TIANJIAO CHEN received the B.S. and M.S. degrees from Anhui University, in 2011 and 2014, respectively. She was with the Institute of Intelligent Machines, Hefei Institutes of Physical Science, Chinese Academy of Sciences. Her current research interests include image processing, pattern recognition, and artificial intelligence.

HAIYING HU received the degree in radio physics from Sichuan University, in 1995, and the master's degree in control engineering from the Hefei University of Technology, in 2011. Her current research interests include pattern recognition and intelligent systems.

XIUFANG JIA received the M.S. degree in basic mathematics from Shanxi University, Taiyuan, China, in 2008, and the Ph.D. degree in pattern recognition and intelligent system from the University of Science and Technology of China, Hefei, China, in 2014. She is currently an Assistant Researcher with the Institute of Intelligent Machines, Chinese Academy of Sciences. Her current research interests include data mining, deep learning, and computer vision.

MIN HU received the M.S. and Ph.D. degrees in automation from the Hefei University of Technology, Hefei, China, in 1994 and 2004, respectively. He is currently a Professor and a Researcher with the School of Computer and Information, Hefei University. His current research interests include intelligent agriculture, the agricultural Internet of Things, and agricultural knowledge engineering.



MAN ZHOU received the B.E. degree in automation from the Wuhan University of Science and Technology, Wuhan, China, in 2013. He is currently pursuing the Ph.D. degree in pattern recognition and intelligent system with the University of Science and Technology of China, Hefei, China. His current research interests include deep reinforcement learning and computer vision.

DENGSHAN LI received the B.E. degree in electronic science and technology from the Hefei University of Technology, Hefei, China, in 2008, and the M.S. degree in computer science and technology from the Anhui University of Technology, Ma'anshan, China, in 2017. He is currently pursuing the Ph.D. degree in computer application technology with the University of Science and Technology of China, Hefei. His current research interests include deep learning and computer vision.

WANCAI LIU received the B.E. degree in plant protection from Northwest A&F University, Yangling, China, in 1991. He is currently a Researcher with the National Agro-Tech Extension and Service Center. His current research interests include technology for major crop pest and disease monitoring, and early warning.