

# An effective automatic system deployed in agricultural Internet of Things using Multi-Context Fusion Network towards crop disease recognition in the wild



Yushan Zhao<sup>a</sup>, Liu Liu<sup>b,c,\*</sup>, Chengjun Xie<sup>b,\*\*</sup>, Rujing Wang<sup>b</sup>, Fangyuan Wang<sup>b,c</sup>, Yingqiao Bu<sup>d</sup>, Shunxiang Zhang<sup>a</sup>

<sup>a</sup> College of Mathematics & Big Data, Anhui University of Science and Technology, Huainan 232001, China

<sup>b</sup> Institute of Intelligent Machines, and Hefei Institute of Physical Science, Chinese Academy of Sciences, Hefei 230031, China

<sup>c</sup> University of Science and Technology of China, Hefei 230026, China

<sup>d</sup> College of Electronic Countermeasures, National University of Defense Technology, Hefei, China

## ARTICLE INFO

### Article history:

Received 23 August 2019

Received in revised form 22 December 2019

Accepted 22 January 2020

Available online 28 January 2020

### Keywords:

Multiclass crop disease recognition  
Convolutional Neural Network  
Internet of Things  
Multi-Context Fusion Network  
ContextNet

## ABSTRACT

Automatic crop disease recognition in the wild is a challenging topic in modern intelligent agriculture due to the appearance variances and cluttered background among crop diseases. To overcome these obstacles, the popular methods are to design a Convolutional Neural Network (CNN) model that extracts visual features and identifies crop disease images based on these features. These methods work well on laboratory environment under simple background but achieve low accuracy and poor robustness in processing the raw images captured from practical fields that contain inevitable noises. In this case, Internet of Things (IoT) is attracting increasing attention, with many alternatives to collect high-level contextual information that helps modern recognition system to effectively identify crop diseases in the wild. Motivated by the usefulness of agricultural IoT, a deep learning system using a novel approach named Multi-Context Fusion Network (MCFN), is developed to be deployed in agricultural IoT towards practical crop disease recognition in the wild. Our MCFN firstly adopts a standard CNN backbone to extract highly discriminative and robust visual features from over 50,000 in-field crop disease samples. Next, we exploit contextual features collected from image acquisition sensors as prior information to assist crop disease classification and reduce false positives in our presented ContextNet. Finally, a deep fully connected network is designed to fuse visual features as well as contextual features and output the crop disease prediction. Experimental results on 77 common crop diseases captured in our newly built domain specific dataset show that MCFN with the deep fusion model outperforms the state-of-the-art methods in wild crop disease recognition, and achieves a good identification accuracy of 97.5%.

© 2020 Elsevier B.V. All rights reserved.

## 1. Introduction

The crop disease is known to be a major issue for modern agriculture, which always cause large losses to crop production. Thus, the automatic and precise diagnosis and identification of crop diseases play a crucial role in ensuring the high yield and high quality of crops. Besides, it helps avoid the labor-intensive manual crop disease identification in the fields. Currently, automatic crop disease identification and analysis using static image processing techniques [1,2] is a challenging topic, and has been actively researched for applications such as the timely diagnosis

of crop diseases [3], forecasting disease decisions [4], precision pesticide [5], etc. Most of current imaging techniques are based on the fluorescence, multispectral and hyperspectral [6,7], as well as digital images [8,9]. Among these techniques, digital images are usually captured by digital cameras, which could photograph images with a high resolution.

The past few years have witnessed the advancement in the efficiency of computing power and the availability of large training datasets. Some researchers focused on deep learning, especially Convolutional Neural Network (CNN) to solve the crop disease identification problem, which has been proven to hold the capability of outperforming previous works in various automatic identification fields, such as visual recognition [10,11] and natural language processing [12]. The most influential breakthrough in the field of computer vision occurred when deep convolutional

\* Corresponding author.

\*\* Corresponding author.

E-mail addresses: [liuliu66@mail.ustc.edu.cn](mailto:liuliu66@mail.ustc.edu.cn) (L. Liu), [cjxie@iim.ac.cn](mailto:cjxie@iim.ac.cn) (C. Xie).

neural networks are chosen [13] to win the competition of ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [10]. Since then, deep learning has become one of the most popular approaches for identification task [14–16]. However, despite the effectiveness of CNN in handling generic object recognition tasks, using deep learning as ready-to-use recipe in crop disease identification task is still suffering the problem of unsatisfied performance due to the influence of practical wild scenario. Despite that there appear lots of deep learning based approaches for recognizing crop diseases on various crops, such as banana leaf disease [17], tomato plant disease [18] and rice disease [19], few researchers focus on identifying multi-class crop disease simultaneously in the wild environment. In general, this could be attributed to the explanation that wild scenarios likely affect the classifier in distinguishing different kinds of diseases. In this paper, we target at solving the problem of multi-class automatic crop disease identification whose images are collected from wild circumstance using deep learning techniques, which requires that the identification algorithms are highly robust to various challenges of disease appearances.

An observation in current CNN based crop disease identification approaches is that most of them try to extract visual feature representations from 2D static images for prediction by designing powerful enough models [20,21]. However, the bottleneck of these methods is low-level visual features are not insufficient for highly precise crop disease identification because the lack of high-level semantic information provided. In addition, wild circumstance might bring a serious difficulty with various challenges (large illumination variation or background clutter) into the captured raw images. In this case, relying on the sample low-level visual features may not be a satisfactory solution. In this case, Internet of Things (IoT) is attracting more and more attention recently, with many alternatives to collect high-level contextual information that could be potential to be extracted as highly useful features to help modern recognition system to effectively identify crop diseases in the wild. Therefore, the key motivation for developing the deep CNN model for automatic and precise multi-class crop disease identification system in the wild is to exploit extra the contextual information captured from image acquisition sensors in agriculture IoT as prior information to assist crop disease classification and reduce false positives.

In this paper, motivated by usefulness of contextual information from IoT sensors, a deep learning system using Multi-Context Fusion Network (MCFN), is developed to be deployed in agricultural IoT for solving these problems, whose architecture is shown in Fig. 1. Our MCFN firstly adopts a CNN backbone to extract highly discriminative and robust visual features from over 50,000 in-field disease samples. Next, we exploit contextual features (temperature, humidity and so on) collected from sensors as prior information to assist crop disease classification in our presented ContextNet. Finally, a deep fully connected network is designed to fuse visual features as well as contextual features and output the crop disease prediction. Such this system could an effective solution to precisely identify the crop diseases in the wild environment and be used in practical IoT applications.

The main contributions of this paper are as follows:

- (1) A novel CNN based architecture Multi-Context Fusion Network (MCFN) is proposed and developed in agricultural IoT to achieve precise crop disease recognition in the wild. This approach exploits contextual information to improve the recognition performance.
- (2) Targeting at the wild crop disease identification task, a new domain specific dataset is published in this paper. This dataset contains over 50,000 crop disease samples with 77 categories, which covers almost all the common crop diseases.
- (3) A comprehensive and in-depth experimental evaluation on our dataset is provided for validating the correctness and usefulness of our proposed MCFN. The results show that MCFN could achieve superior performance over the state-of-the-art approaches in crop disease identification task.

## 2. Related work

### 2.1. Machine learning for crop disease identification

Conventional crop disease identification methods [22–24] focused on recognizing single disease or a small number of diseases and usually worked well only under controlled conditions in the laboratory rather than wild scenarios, which contain many influence factors such as uniform illumination, consistent scale or position of the disease in the image. Although progress has been made with respect to the disease identification problem, most of the traditional methods are only capable of identifying some specific types of diseases in the wild [25,26], but might not be transferred to a large range of crop disease categories with powerful ability of generalization. In general, this could be attributed to the explanation that wild scenarios likely affect the classifier in distinguishing different kinds of diseases. In this paper, we focus on the automatic crop disease identification whose images are collected from wild circumstance, which requires that the identification algorithms are highly robust to various challenges of disease appearances, such as complex backgrounds, illumination changes as shown in Fig. 2.

Generally, there are three separate steps to deal with this task in conventional computer vision techniques: object segmentation, feature extraction and pattern recognition. The goal of segmentation is to filter the background from the leaf image containing the symptoms [27–29], but segmenting the symptoms from the disease leaf images is particularly difficult when the background is affected by significant amount of variation elements, for example as can be seen in Fig. 1. Feature extraction, the second step, is required to manually design features for describing the relevant information of disease symptoms and have made drastic progresses in many disease recognition tasks. The common hand-crafted features contain SIFT [30] and HOG [31]. Thirdly, a lot of classifiers, such as Support Vector Machine (SVM) [32] and K-Nearest Neighbors (KNN) [33], are studied to predict the categories of an unknown crop disease image represented by hand-crafted features. Although these methods could achieve a passable performance on crop disease identification task [6,7], it is still difficult to obtain good performance from disease images in wild scenarios containing various challenges e.g. large illumination change and resolution variations.

### 2.2. Deep learning for agriculture insect pest and crop disease recognition

In recent years, deep learning approaches, which is a novel breakthrough in computer vision field, is promising for image recognition and classification. Within convolutional neural network adopted in modern computer vision and pattern recognition tasks, the labor-intensive and hand-crafted feature engineering is avoided. Based on these works, deep learning has been widely employed in intelligent agriculture area for insect pest identification and fine-grained crop disease recognition. For insect pest classification task, [34] proposed a publicly available dataset that contained 24 types of common pests for identification, which provided a benchmark for insect pest research. Then, to expand this work for much generalized, [20] investigated some hyperparameters in the state-of-the-art CNN based image classification

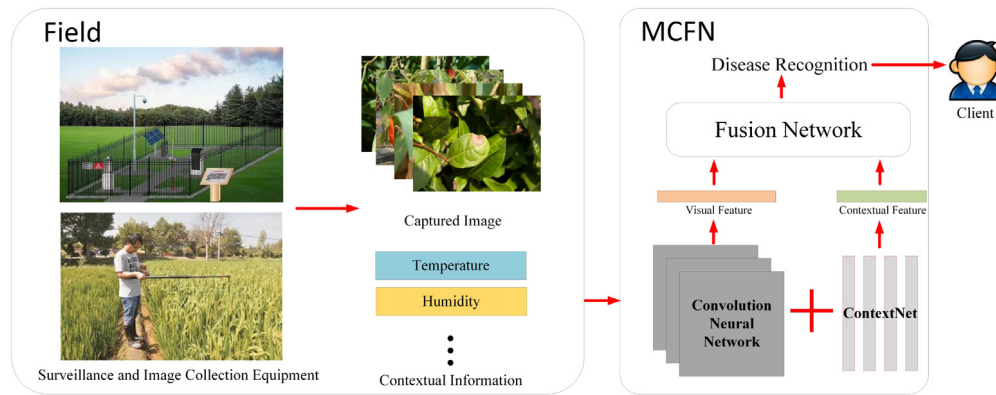


Fig. 1. System Architecture. Surveillance and image collection equipment collects in-field crop images as well as contextual information.

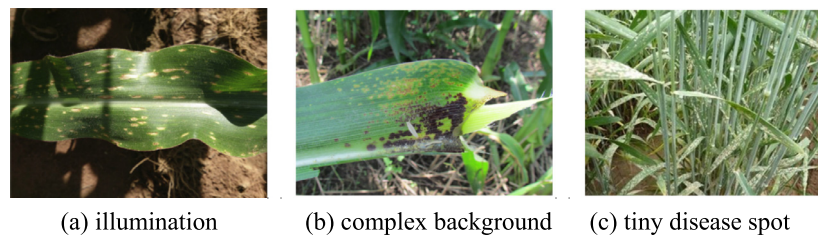


Fig. 2. Examples of various disease symptoms in the wild.

approaches and used them to successfully identify more than 40 classes of insect pests with 96.75% accuracy. In terms of crop disease recognition task, [19] built a dataset of 500 natural images of rice leaves and stems that were diseased or healthy, which were collected from rice experimental field. In this work, 10 common rice diseases were identified using deep CNN architecture and the accuracy reached 95.48%. In order to address the issue of small training samples in this dataset and insufficient disease categories, the methods in [21] and [35] attempted to recognize much more types of crop diseases rather than rice and finally could identify 13 and 26 crop diseases respectively. In addition, disease spot detection was also an effective way for precise crop disease classification. [18] introduced the state-of-the-art generic object detection approaches Faster RCNN and SSD for localizing crop leaves and disease spot with good performance (83% detection accuracy). Despite these approaches showed a great success in current crop disease recognition task, the advances in this area slowed down in the last few years. Furthermore, most of these methods might not achieve satisfactory performance in practical crop disease recognition applications because of various challenges in wild scenes during test image inference such as illumination and complicated background shown in Fig. 2. Therefore, inspired by break-through of deep learning in agriculture field, we design our MCFN to solve the problems on practical crop disease identification task by augmenting contextual information to improve the performance.

### 3. Materials and methods

#### 3.1. Data collection

There exist lots of open public dataset such as PlantVillage that is a common dataset for crop disease identification task [36,37]. However, most samples in PlantVillage are photographed under laboratory condition so the model trained on it might not perform well in the wild environment. In order to address the issue of crop disease identification in the wild, we build our own dataset to verify the performance of our approach, in which the images

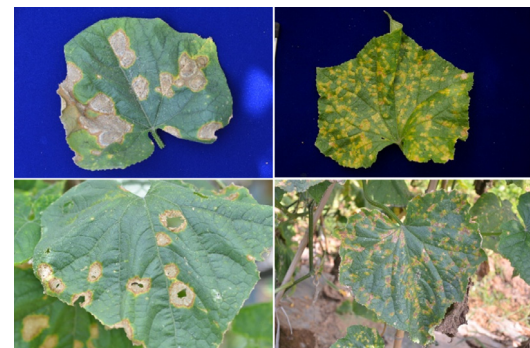


Fig. 3. Images captured in the laboratory and in the field. The two images on the top are captured in the laboratory environment, in which the background is pure blue. The bottom two images were samples that we captured, it had very complex background noise in the images.

are captured in the practical fields. Compared with the images collected in the laboratory environment, where noises are eliminated manually, the images in our dataset hold more complicated backgrounds as shown in Fig. 3. Thus, our built dataset is much more suitable for training our model to recognize crop diseases in the wild scenarios.

The hierarchical structure of our dataset is in accordance with the plant species designated by agricultural experts. Particularly, there are 19 different types of crops distributed in three main categories: grain, fruiter and vegetable in our dataset. Totally, these crops contain diseases of 77 categories of these crops, where most of samples are captured with CCD camera. In order to ensure the enough visual information for some certain diseases with small spot, the resolution of images is  $1920 \times 1440$ . After image collection, we manually filter out the noises among our data including faded, blurry and duplicated images to maintain the consistency of crop disease data and handle invalid information. With these pre-processing operations, the generalization and identification ability of our model trained on the dataset could be effectively



Fig. 4. Sample distribution in our dataset.

improved. Currently, our dataset contains over 50,000 crop disease images. In order to make sure the authority and authenticity of the image annotations, all of the captured crop disease images are screened by agricultural experts carefully. In terms of annotation, we invite more than 10 agriculture experts to label these images. To guarantee the professionalism and correctness of these labels, each image is labeled by at least 5 experts and the final label is obtained by voting. The sample distribution of our dataset is visualized in Fig. 4. After image collection, we randomly split all of the images into 10 folds, where nine-tenths are selected as training set and the rest part is used for validating our model with cross-validation strategy. During training phase, we adopt ‘mirror’ data augmentation method to horizontally flipped the original images to expand the training data volume, which could further improve the generalization ability of our model.

### 3.2. Multi-Context Fusion Network (MCFN)

In this paper, we propose a novel CNN based approach named Multi-Context Fusion Network (MCFN) whose pipeline is shown in Fig. 5(b), which exploits contextual features to explore the crop disease interaction between vision and prior knowledge of the crop disease. Compared to the general CNN architecture shown in Fig. 5(a), we introduce a novel branch to extract contextual features and encode to fuse them with visual features together. The contextual information, referred as temporal and spatial information in our work, consists of season, geographic location, temperature and humidity. Potentially, the contextual information is largely related to categories of crop diseases and could be taken advantage to improve the identification performance. In this case, we design our MCFN to make full use of the contextual information to achieve a better classification accuracy. MCFN contains three major parts: CNN backbone, ContextNet and fully connected network. The CNN backbone is used for visual feature extraction, which could automatically learn the appropriate features from the training set rather than hand-crafted. In terms of ContextNet, it is designed to fuse contextual information of crop diseases. Finally, a deep fully connected network is designed to fuse visual features as well as contextual features and output the crop disease prediction. In MCFN, visual and contextual features

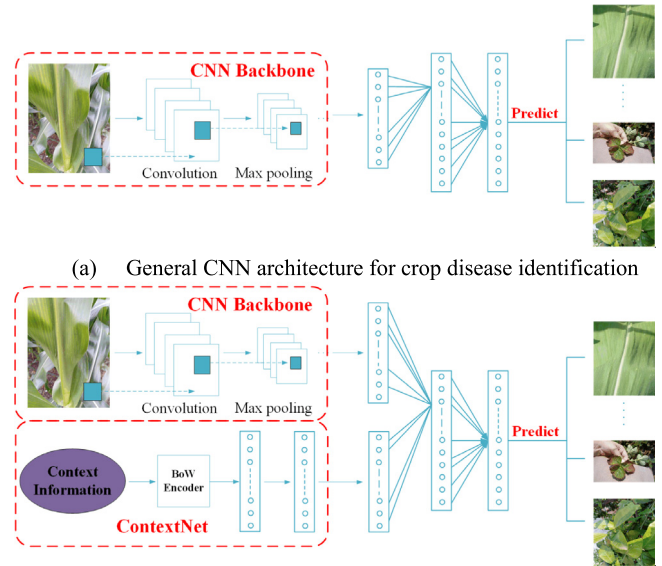


Fig. 5. General CNN and our proposed MCFN architecture.

could be learned jointly to capture the shared high-level semantic information between images and prior knowledge of the crop diseases. Eventually, MCFN could learn a joint embedding space which facilitates efficient crop disease matching. Therefore, MCFN closely connect visual and contextual information mechanisms into a unified deep learning framework, achieving the state-of-the-art performance in the crop disease recognition task under our newly built domain specific crop disease dataset.

### 3.3. Convolutional Neural Network (CNN)

A Convolutional Neural Network (CNN) is a type of feed-forward neural network in processing data with grid-like topology. In the standard CNN, images are input into convolutional layers computed with ‘convolutional kernel’ and pooling layers for down-sampling. In our work, we apply different combinations of these layers for extracting 3D image features which is called ‘feature map’ rather than 2D feature vectors. In general, a CNN consists of three operations: convolutional layer, activation function and pooling layer.

**Convolutional layer:** Convolutional layer is the main component in a standard CNN that could automatically extract feature maps from input 2D image by introducing a filterbank containing a series of convolution kernels. Generally, for each kernel  $k$ , the forward propagation process of convolution in layer  $l$  could be represented by:

$$a_k^l = \sigma(z_k^l) = \sigma(a^{l-1} * W_k^l + b^l) \quad (1)$$

Where the  $a_k^l$  and  $a^{l-1}$  are output of kernel  $k$  from layer  $l$  and  $l-1$ .  $\sigma(z_k^l)$  is the activation function.  $*$  indicate the convolution operation.  $W_k^l$  and  $b^l$  represent the convolution kernel and bias in layer  $l$  respectively. Therefore, the output convolutional layer could be computed as the sum of outputs from the filterbank:

$$a^l = \sigma(z^l) = \sigma\left(\sum_{k=1}^M z_k^l\right) = \sigma\left(\sum_{k=1}^M (a_k^l * W_k^l) + b^l\right) \quad (2)$$

**Activation function:** The activation function aims to introduce non-linear decision boundaries to the network. The Rectified Linear Unit (ReLU) that is defined as  $\sigma(z) = \max(0, z)$  is an often-used activation function in deep learning applications, as it is

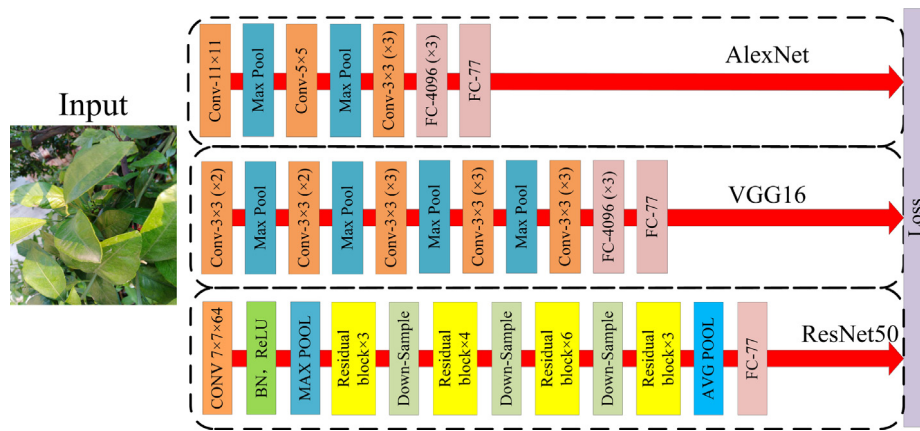


Fig. 6. AlexNet, VGG16 and ResNet50 training for crop disease identification.

considerably faster to train the model because of larger gradient in  $(0, \infty)$  rather than alternatives such as the sigmoid function that might cause gradient vanishing problem.

**Pooling layer:** A pooling function replaces the output of the net at a certain location with a summary statistic of the nearby outputs. Usually, convolutional layer is usually followed by a pooling layer for feature dimension reduction. Besides, spatial translational invariance is another benefit of pooling layer. The widely used pooling operation is the max pooling operation, which reports the maximum output in a receptive field and discards other values. Max pooling helps to make the representation become approximately invariant to small translations of the input.

### 3.4. Visual feature extraction: CNN backbone

In our research, CNN backbones are utilized to extract the features of the crop image and identify the corresponding diseases. From the perspective of feature extraction, we adopt AlexNet [13], VGG16 [38] and ResNet50 [11] as our visual feature extraction networks. The AlexNet was proposed by [13], and won the championship of the 2012 ILSVRC competition, achieved a winning top-5 accuracy of 84.7%, compared to 73.8% achieved by the second-best entry. The VGG network was proposed by the Oxford Vision Group and it won the championship in the 2014 ILSVRC competition. Furthermore, ResNet is the state-of-the-art CNN backbone for image recognition task, which adopts residual module that skips connection between convolution operations. Therefore, the current researches have proved the effectiveness of these two CNN backbones.

The structures of AlexNet and VGG16 used in our method are visualized in Fig. 6. As it can be seen, AlexNet is comprised of 8 layers of network, containing 5 convolutional layers and 3 max pooling layers. ReLU is followed by each convolutional layer. AlexNet selects large kernel size for enlarging the respective field to make sure the noises could be filter better at the beginning of training. Differently, VGG16 consists of 16 layers of network, in which 13 convolutional layers and 3 max pooling layers are augmented for feature extraction. The network uses nothing but small convolution kernel at  $3 \times 3$ , which is the smallest size capable of capturing all directions and central concept. Multiple convolution layers at  $3 \times 3$  display greater nonlinearity than one convolutional layer of large size, making the function more deterministic and greatly reducing the number of parameters. Due to network depth and small size of convolution kernel that helps realize implicit regularization, VGG network could start to converge when there is just a few of iteration times. For ResNet50, there are 50 layers that contains several residual modules. Due

the large depth, ResNet50 has been becoming one of the popular networks for image recognition task. Therefore, we select AlexNet, VGG16 as well as ResNet50 as CNN backbones in our system to accomplish the visual feature extraction of crop disease images.

### 3.5. Contextual feature extraction: Contextnet

The CNN backbones could predict the category of crop disease based on visual features from raw input image. But in the wild application scenarios, it is hard to extract powerful enough features for precise crop disease prediction. To address this issue, we design a parallel network ContextNet to extract the extra information: contextual features. During data collection, there are amount of contextual information that could be obtained, such as spatial, temporal and weather. The information could describe the crop disease from a high-level semantic aspect. So intuitively, the contextual information might be helpful to obtain a more correct category prediction of diseases. In order to extract and encode the contextual information it to integrate together with visual features, we design ContextNet for this specific purpose whose architecture is shown in Fig. 7.

At the beginning of ContextNet, we need to encode the collected context in-formation including spatial and temporal features to make them fit the network as the context collected as text form. In our method, we select a Bag-of-Word (BoW) [39] encoding algorithm to transform the contextual information into a 1D vector because each context text is relatively independent. In our BoW approach, we build a dictionary containing all of the contextual information (geometric locations, season). And then we add the information into the corresponding index in the dictionary to conduct a 1D context feature vector. Finally, we normalize the 1D feature vector into  $(0,1)$ . In this way, we could encode the contextual in-formation for training ContextNet.

In the next step, the context features and visual features are not clearly similar in nature and their scales are large different. Specifically, context features indicate a high-level semantic information while visual features are used to describe content information of images. In this case, we might not directly concatenate them together in the same level, so our ContextNet employs a feedforward subset to the end of CNN backbones, which consists of a series of fully connected layers. And then we fuse the extracted context features to visual features at the final fully connected layer. In this way, we could successfully augment the context information into our system.

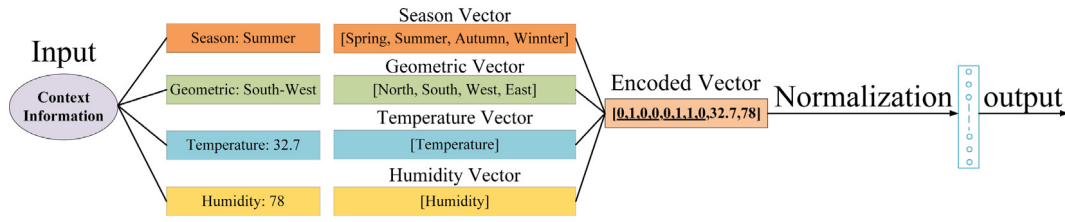


Fig. 7. ContextNet for contextual feature extraction.

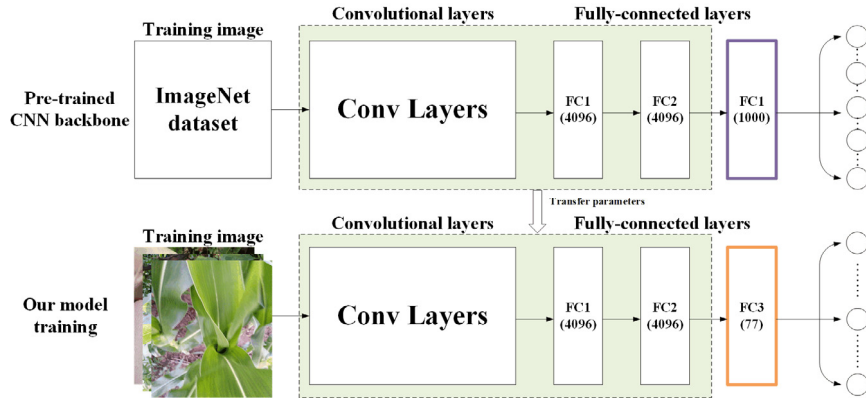


Fig. 8. Fine-tune strategy for crop disease identification.

### 3.6. Fine-tune

Fine-tune is a strategy of transfer learning [40], which utilize massive task weakly correlated data to obtain a pre-trained model, and couple with the task specific data, to train new model. The advantage of Fine-tune is that it could solve the problem of small amount data trained for neural network. Besides, Fine-tune strategy could also avoid the problem of gradient vanishing as it provides excellent weights for shallow layers in the new model.

First of all, the network goes through sufficient data on ImageNet dataset for pre-training. In ImageNet, the last full connected layer of network could output a 1000-Dimensional tensor for predicting possibility of 1000 categories of input image. Although this pre-trained model is not capable of identifying crop diseases for our task, it could provide initial values for our MCFN network which are good enough to extract shallow features. So, we modify the last fully connected layer and set its output dimensions to 77, corresponding to the number of crop disease categories in our task. Then, our MCFN model is re-trained using our wild crop disease dataset so the weights could be fine-tuned. In this way, our network could be trained and applied to the crop disease identification task with fewer iterations and better performance. The process of Fine-tune is shown in Fig. 8.

### 3.7. Optimization and evaluation

To optimize our model, loss function is the criterion for training process that guides the direction of model convergence. In our approach, Softmax function is selected to predict the final category of input sample and Cross-Entropy (CE) loss [41] is adopted as the loss function. Generally, Softmax is an extension of logistic regression used for multi-classification task without any learnable parameters, which is defined as:

$$\sigma(z_j) = \frac{e^{z_j}}{\sum_{k=1}^K e^{z_k}} \quad (3)$$

Where  $K$  represents the number of classes so the output  $\sigma(z_j)$  is the possibility that the input sample is predicted as  $j$ th class.

Based on the softmax function as predictor, the CE loss is computed by:

$$CE(y) = - \sum_{j=1}^K \hat{y}_j \log(y_j) \quad (4)$$

In which the  $\hat{y}_k$  and  $y_k$  are the ground truth and predicted possibility of the input sample.

## 4. Results and discussion

### 4.1. Comparative results

Table 1 present the accuracy of different identification methods, where the hand-crafted feature is SIFT (Giselsson, Midtiby and Jrgensen, 2013) that has been widely used in conventional machine learning methods. It can be observed that MCFN using ResNet50 as CNN backbone for extracting visual features achieves the best performance with the highest accuracy. Among these results, the lowest performance occurs in conventional machine learning method with hand-crafted features, which indicates that CNN could dramatically enhance the quality of features. In addition, after augmenting context features, we could obtain a large improvement compared with the methods without contextual information. Besides, within Fine-tune strategy, MCFN could also increase the identification accuracy by 0.8%, 4.3% and 3.9% using AlexNet, VGG16 and ResNet50 respectively compared to those using without fine-tune strategy.

### 4.2. Training analysis

Our method achieves the best performance in our targeted wild crop disease identification task. Additionally, we present the variation of loss during the training process using deep learning methods in Fig. 9. By comparing different loss curve, it could be observed that MCFN fits the distribution of our dataset better, which could obtain the lowest loss when reaching converge.

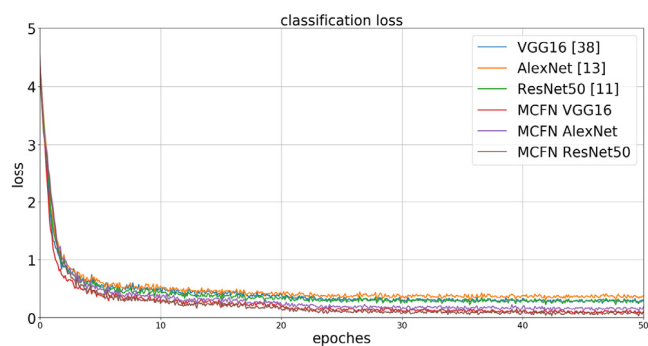


Fig. 9. Classification losses of the different models during training.

Table 1

Validation accuracy (%) of different methods. The column 'Accuracy' is the average classification accuracy with standard deviation over 10-folds cross validation.

Method	Feature extractor	Fine-tune	Accuracy
BP	Hand-Crafted	–	57.9 ± 2.45
SVM [32]	Hand-Crafted	–	50.8 ± 3.66
General CNN	AlexNet [13]		72.6 ± 2.34
General CNN	AlexNet [13]	✓	77.5 ± 2.08
MCFN (Ours)	AlexNet [13]		87.0 ± 1.77
MCFN (Ours)	AlexNet [13]	✓	87.8 ± 1.54
General CNN	VGG16 [38]		75.3 ± 2.14
General CNN	VGG16 [38]	✓	78.7 ± 1.98
MCFN (Ours)	VGG16 [38]		91.5 ± 1.43
MCFN (Ours)	VGG16 [38]	✓	95.8 ± 1.29
General CNN	ResNet50 [11]		82.6 ± 1.35
General CNN	ResNet50 [11]	✓	84.8 ± 1.21
MCFN (Ours)	ResNet50 [11]		93.6 ± 1.14
MCFN (Ours)	ResNet50 [11]	✓	<b>97.5 ± 0.69</b>

Fig. 10 shows the accuracy curves for each epoch during training. (a) shows the curves of training from scratch, (b) shows the curves of training by Fine-tune strategy. Compared with the models training from scratch, Obviously, both training modes show the similar trend, in which our MCFN gets a better accuracy during training. Furthermore, the accuracy curve of models using Fine-tune training mode could be trained dramatically faster than those training from scratch. Only when the training continues to iterate with more epochs, the gap between them start to decrease.

#### 4.3. Confusion matrix discussion

In order to analyze detailed crop disease identification results, we compare the confusion matrix of the category results predicted by the MCFN and other CNN models (AlexNet, VGG16 and ResNet50). Due to the fact that we aim to recognize a large number of categories of crop diseases, which are 77 types with 19 different crops, we may not visualize the confusion matrix of these diseases in one single matrix. So, we present the confusion matrixes for part of them. Fig. 11 shows the confusion matrixes for part of crops' diseases using our MCFN with Fine-tune strategy. As it can be seen, MCFN could get an excellent identification performance in almost every disease and the prediction accuracy of most crop diseases are above 90%. Therefore, our MCFN is capable of identifying various types of crop diseases in the wild.

#### 4.4. Result visualization and discussion

Finally, we visualize some of test samples in our dataset and their corresponding predicted diseases. As shown in Fig. 12, MCFN

could achieve crop disease identification under various wild scenarios, despite some intractable challenges such as illumination (Fig. 12(b) and (c)) and noisy background (Fig. 12(a) and (d)). Considering the comparative experimental results in Table 1, the major reason for this excellent performance is that contextual features, such as spatial and temporal information, are intuitively important that could be a constraint to reduce the possibility of prediction confusion.

#### 4.5. Limitations and future work

Despite that we implement a novel MCFN method for crop disease recognition in the field and achieve a successful performance in our dataset, there are two limitations for future study. Firstly, the problem of unbalanced data structure has not been well solved. Specifically, due to the seasonal and regional difficulty, it is almost impossible to collect a balanced dataset for various categories of crop diseases. Besides, the occurrence frequency of crop diseases might be also different in the practical applications. Thus, an excellent recognition system is expected to focus on identifying them based on practical crop disease occurrence regulation and introducing prior probability distribution for these categories. Secondly, our system might pay attention on identifying the class of crop disease rather than estimating the severity of the identified crop disease which is also significant for practical crop disease monitoring. Therefore, future work would target at solving the problem of unbalanced captured dataset and focus on more fine-grained identification of crop disease.

## 5. Conclusion

In this paper, we introduce an effective automatic crop disease recognition system deployed in agricultural IoT using Multi-Context Fusion Network (MCFN) which explores the potential of automatic crop disease identification in the wild through Convolutional Neural Network (CNN) augmenting contextual features as prior information. Specifically, MCFN encodes contextual information extracted during data collection, and adopts fully connected layers to concatenate them with high-level visual features. In order to validate the performance of our MCFN system in crop disease identification task, a new domain specific dataset is published in this paper, contains over 50,000 crop disease samples with 77 categories, which covers almost all the common diseases. Our designed comparative experiments have shown that MCFN is powerful enough approach for crop disease identification in the wild, which outperforms the popular CNN architectures AlexNet and VGG16 by a large margin.

#### Declaration of competing interest

No author associated with this paper has disclosed any potential or pertinent conflicts which may be perceived to have impending conflict with this work. For full disclosure statements refer to <https://doi.org/10.1016/j.asoc.2020.106128>.

#### CRediT authorship contribution statement

**Yushan Zhao:** Investigation, Data curation. **Liu Liu:** Conceptualization, Methodology, Writing - original draft. **Chengjun Xie:** Writing - review & editing. **Rujing Wang:** Supervision, Funding acquisition. **Fangyuan Wang:** Validation. **Yingqiao Bu:** Project administration. **Shunxiang Zhang:** Funding acquisition.

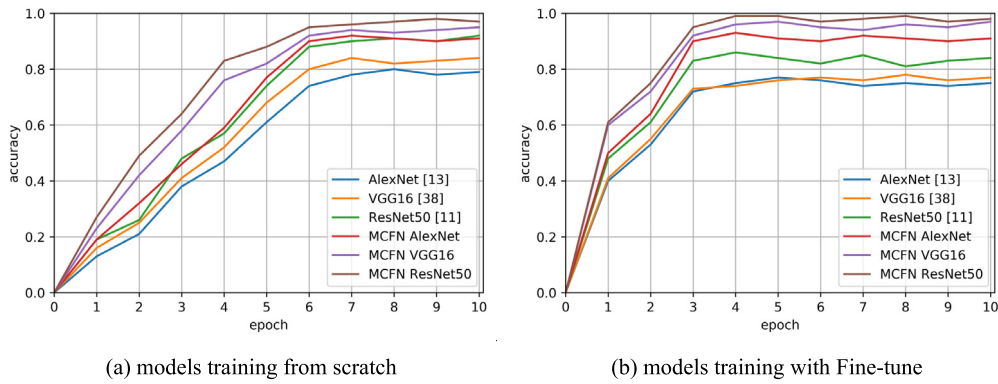


Fig. 10. Accuracy curves during training.

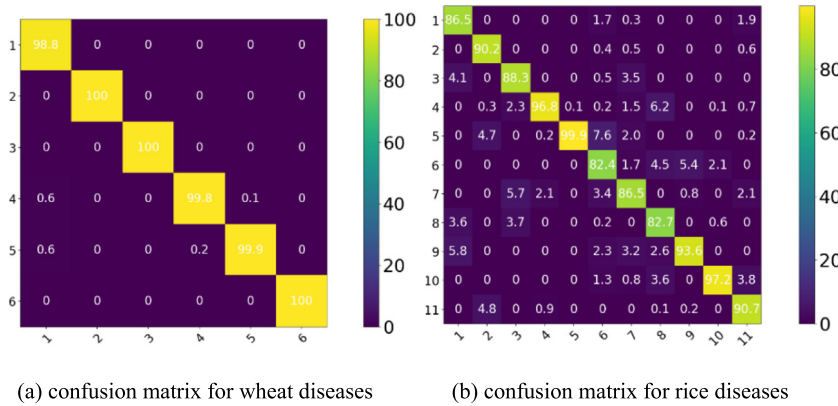


Fig. 11. Confusion matrixes for two different crops' diseases. Note that the x-axis represents the predicted disease ID and y-axis represents the actual disease ID.

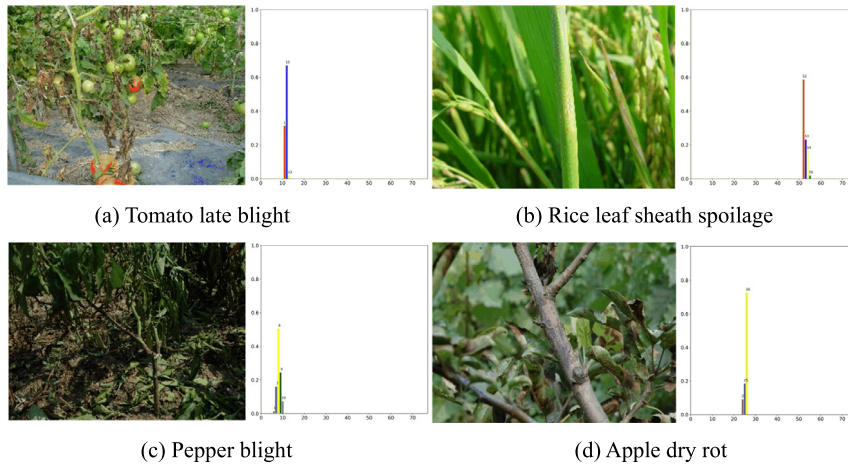


Fig. 12. Visualization of crop disease images identified by MCFN visualization.

**Acknowledgments**

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

**References**

[1] J. Barbedo, A review on the main challenges in automatic plant disease

identification based on visible range images, *Biosyst. Eng.* 144 (2016) 52–60.  
 [2] S. Sankaran, A. Mishra, R. Ehsani, C. Davis, A review of advanced techniques for detecting plant diseases, *Comput. Electron. Agric.* 72 (2010) 1–13.  
 [3] J. Lai, et al., An image-based diagnostic expert system for corn diseases, *Agric. Sci. China* 9 (8) (2005) 1221–1229.  
 [4] N. Hardwick, Disease forecasting, *The epidemiology of plant diseases* (2006) 239–267.  
 [5] J. Zheng, et al., Toward-target precision pesticide application and its system design, *Trans. Chin. Soc. Agric. Eng.* 11 (14) (2005).  
 [6] J. Barbedo, C. Tibola, J. Fernandes, Detecting Fusarium head blight in wheat kernels using hyperspectral imaging, *Biosyst. Eng.* 131 (2015) 65–76.  
 [7] J. Zhang, L. Yuan, R. Pu, R. Loraamm, G. Yang, J. Wang, Comparison between wavelet spectral features and conventional spectral features in detecting yellow rust for winter wheat, *Comput. Electron. Agric.* 100 (2014) 79–87.



- [8] R. Zhou, S. Kaneko, F. Tanaka, M. Kayamori, M. Shimizu, Image-based field monitoring of *Cercospora* leaf spot in sugar beet by robust template matching and pattern recognition, *Comput. Electron. Agric.* 116 (2015) 65–79.
- [9] C. Xie, J. Zhang, R. Li, J. Li, P. Hong, J. Xia, P. Chen, Automatic classification for field crop insects via multiple-task sparse representation and multiple-kernel learning, *Comput. Electron. Agric.* 119 (2015) 123–132.
- [10] O. Russakovsky, et al., Imagenet large scale visual recognition challenge, *Int. J. Comput. Vis.* 115 (3) (2015) 211–252.
- [11] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.
- [12] R. Hu, J. Xu, M. Rohrbach, J. Feng, K. Saenko, T. Darrell, Natural language object retrieval, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 4555–4564.
- [13] A. Krizhevsky, I. Sutskever, G. Hinton, Imagenet classification with deep convolutional neural networks, *Adv. Neural Inf. Process. Syst.* (2012) 1097–1105.
- [14] S. Wang, Y. Jiang, F.L. Chung, P. Qian, Feedforward kernel neural networks, generalized least learning machine, and its deep learning with application to image classification, *Appl. Soft Comput.* 37 (C) (2015) 125–141.
- [15] D. Gope, P.C. Gope, A. Thakur, A. Yadav, Application of artificial neural network for predicting crack growth direction in multiple cracks geometry, *Appl. Soft Comput.* 30 (C) (2015) 514–528.
- [16] E.I. Papageorgiou, A.T. Markinos, T.A. Gemtos, Fuzzy cognitive map based approach for predicting yield in cotton crop production as a basis for decision support system in precision agriculture application, *Appl. Soft Comput.* 11 (4) (2011) 3643–3657.
- [17] J. Amara, B. Bouaziz, A. Algergawy, A deep learning-based approach for banana leaf diseases classification, in: *BTW (Workshops)*, 2017, pp. 79–88.
- [18] A. Fuentes, S. Yoon, S. Kim, D. Park, A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition, *Sensors* 17 (9) (2017) 2022.
- [19] Y. Lu, S. Yi, N. Zeng, Y. Liu, Y. Zhang, Identification of rice diseases using deep convolutional neural networks, *Neurocomputing* 267 (2017) 378–384.
- [20] K. Thenmozhi, U.S. Reddy, Crop pest classification based on deep convolutional neural network and transfer learning, *Comput. Electron. Agric.* 164 (2019) 104906.
- [21] S. Sladojevic, M. Arsenovic, A. Anderla, D. Culibrk, D. Stefanovic, Deep neural networks based recognition of plant diseases by leaf image classification, *Comput. Intell. Neurosci.* 2016 (2016).
- [22] R. Oberti, M. Marchi, P. Tirelli, A. Calcante, M. Iriti, A. Borghese, Automatic detection of powdery mildew on grapevine leaves by image analysis: optimal viewangle range to increase the sensitivity, *Comput. Electron. Agric.* 104 (2014) 1–8.
- [23] G. Polder, G. Heijden, J. Doorn, T. Baltissen, Automatic detection of tulip breaking virus (TBV) in tulip fields using machine vision, *Biosyst. Eng.* 117 (2014) 35–42.
- [24] A. Pourreza, W. Lee, R. Ehsani, J. Schueller, E. Raveh, An optimum method for real-time in-field detection of Huanglongbing disease using a vision sensor, *Comput. Electron. Agric.* 110 (2015) 221–232.
- [25] S. Phadikar, J. Sil, Das, A, Rice diseases classification using feature selection and rule generation techniques, *Comput. Electron. Agric.* 90 (2013) 76–85.
- [26] P. Sanyal, S. Patel, Pattern recognition method to detect two diseases in rice plants, *Imaging Sci. J.* 56 (6) (2008) 319–325.
- [27] G. Jayme, A review on the main challenges in automatic plant disease identification based on visible range images, *Biosyst. Eng.* 144 (2016) 52–60.
- [28] G. Jayme, V. Luciano, T. Thiago, Identifying multiple plant diseases using digital image processing, *Biosyst. Eng.* 147 (2016) 104–116.
- [29] O. Kruse, J. Prats-Montalban, U. Indahl, K. Kvaal, A. Ferrer, C. Futsaether, Pixel classification methods for identifying and quantifying leaf surface injury from digital images, *Comput. Electron. Agric.* 108 (2014) 155–165.
- [30] T. Giselsson, H. Midtby, R. Jrgensen, Seedling discrimination with shape features derived from a distance transform, *Sensors* 13 (5) (2013) 5585–5602.
- [31] C. Xia, et al., Plant leaf detection using modified active shape models, *Biosyst. Eng.* 116 (2013) 23–35.
- [32] N. Larios, B. Soran, L. Shapiro, G. Martinez-Munoz, J. Lin, T. Dietterich, Haar random forest features and SVM spatial matching kernel for stonefly species identification, in: *International Conference on Pattern Recognition, ICPR*, 2010, pp. 2624–2627.
- [33] C. Wen, D. Guyer, W. Li, Local feature-based identification and classification for orchard insects, *Biosyst. Eng.* 104 (3) (2009) 299–307.
- [34] C. Xie, J. Zhang, R. Li, J. Li, P. Hong, J. Xia, P. Chen, Automatic classification for field crop insects via multiple-task sparse representation and multiple-kernel learning, *Comput. Electron. Agric.* 119 (2015) 123–132.
- [35] P. Mohanty, Sharada, D. Hughes, M. Salathe, Using deep learning for image-based plant disease detection, 2016, arXiv preprint [arXiv:1604.03169](https://arxiv.org/abs/1604.03169).
- [36] G. Wang, Y. Sun, J. Wang, Automatic image-based plant disease severity estimation using deep learning, *Comput. Intell. Neurosci.* 2017 (2017) 1–8.
- [37] C.S. Hlaing, S.M.M. Zaw, Model-based statistical features for mobile phone image of tomato plant disease classification, in: *International Conference on Parallel & Distributed Computing*, 2018.
- [38] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, 2014, arXiv preprint [arXiv:1409.1556](https://arxiv.org/abs/1409.1556).
- [39] R. Johnson, T. Zhang, Effective use of word order for text categorization with convolutional neural networks, 2014, arXiv preprint [arXiv:1412.1058](https://arxiv.org/abs/1412.1058).
- [40] S. Pan, Q. Yang, A survey on transfer learning, *IEEE Trans. Knowl. Data Eng.* 22 (10) (2010) 1345–1359.
- [41] P.T.D. Boer, D.P. Kroese, S. Mannor, R.Y. Rubinstein, A tutorial on the cross-entropy method, *Ann. Oper. Res.* 134 (1) (2005) 19–67.