Article

Monitoring the Damage of *Armyworm* as a Pest in Summer Corn by Unmanned Aerial Vehicle Imaging

Wancheng Tao [1, 2], Xinsheng Wang [1,2], Jing-Hao Xue [3], Wei Su [1, 2]*, Mingzheng Zhang [1, 2], Dongqin Yin[1, 2], Dehai Zhu [1, 2], Zixuan Xie[1, 2], Ying Zhang[1, 2]

[1] College of Land Science and Technology, China Agricultural University, Beijing 100083, China
[2] Key Laboratory of Remote Sensing for Agri-Hazards, Ministry of Agriculture, Beijing 100083, China
[3] Department of Statistical Science, University College London, London, WC1E 6BT, UK
* Correspondence: suwei@cau.edu.cn; Tel.: +86 010-6273-7855

Received: date; Accepted: date; Published: date

Abstract:
BACKGROUND: The timely, rapid, and accurate near real-time observations are urgent to monitor the damage of corn *armyworm*, because the rapid expansion of *armyworm* would lead to severe yield losses. Therefore, the potential of machine learning algorithms for identifying the *armyworm* infected areas automatically and accurately by multispectral Unmanned Aerial Vehicle (UAV) dataset is explored in this study. And the study area is in Beicuizhuang Village, Langfang City, Hebei Province, which is the main corn-producing area in the North China Plain.

RESULTS: Firstly, we identified the optimal combination of image features by *Gini*-importance and the comparation of four kinds of machine learning methods including Random Forest (RF), Multilayer Perceptron (MLP), Naive Bayes Classifier (NB) and Support Vector Machine (SVM) was done. And RF was proved to be the most potential with the highest *Kappa* and *OA* of 0.9709 and 0.9850, respectively. Secondly, the *armyworm* infected areas and healthy corn areas were predicted by an optimized RF model in the UAV dataset, and the *armyworm* incidence levels were classified subsequently. Thirdly, the relationship between the spectral characteristics of different bands and pest incidence levels within the Sentinel-2 and UAV images were analyzed, and the B3 in UAV images and the B6 in Sentinel-2 image were less sensitive for *armyworm* incidence levels. So the Sentinel-2 image was used to monitor *armyworm* in two towns.

CONCLUSIONS: The optimized dataset and RF model are effective and reliable, which can be used for identifying the corn damage by *armyworm* using UAV images accurately and automatically in field-scale.

Keywords: *armyworm*; summer corn; Unmanned Aerial Vehicle; Random Forest; Sentinel-2

1. Introduction

The incidence of crop insect pests is increasing in China due to the frequent occurrence of extreme weather and the increase of insect pests' resistance to insecticides in recent years[1]. Curtis et al. proven that crop pests will increase with global temperature. When average global surface temperature increases by 2°C, corn production will be reduced by 23% in China and 32%in the United States[2]. *Armyworm* (*Mythimna separata* Walker) is one of the most serious insect pests on cereal crops, resulting in great crop production loss every year[3]. Summer corn grows in hot rainy season, as the high temperature and humidity create a perfect environment for *Armyworm* to develop. The harm of *armyworm* to corn is huge, especially the second and third-generation larvae with the characteristics of rapid outbreak[4]. In 2012, the *armyworm* broke out in most corn planted areas of China, resulting in serious damage on corn production[5].

The chemical pesticide is the primary method to control the *armyworm* presently. However, over usage of pesticides will pollute the environment, and pesticide residues do harm to soil and human health. Thus, it is of great significance to determine the occurrence area, severity and spread pattern of *armyworm* pest rapidly and timely, for controlling the pest infestation by employing precise spraying of pesticides in the field[6]. The traditional methods of monitoring *armyworm* are mainly depended on the field campaign. It is time-consuming and laborious, and the field campaign can be done only in finite sampling areas. Therefore, the laborious investigation method in the field campaign cannot meet the requirements of timely and rational pesticide usage[7]. Therefore, many studies explore the easy way to monitor the damage of *armyworm* pests. The *armyworm* mainly feeds on the corn leaves and ears, especially in the tasseling-filling stage, which results in plant dwarfing and corn canopy spectral signature changing. Considering these facts, remote sensing has emerged as a promising way to identify crop diseases and insect pests non-destructively, quickly, and automatically on a regional scale.

Many studies[8-10] have revealed the effectiveness of satellite images in monitoring crop diseases and insect pests. Unfortunately, the long revisit period, low spatial resolution, and the requirements for clear weather of satellite images hampered monitoring the *armyworm* damage because the *armyworm* would break out in a very short time[11]. Recently, unmanned aerial vehicles (UAV) with different carrying sensors act as a rapid and ongoing tool in precision agriculture. Many researches[12-14] have made lots of effort on crop monitoring by UAV images. With the advances in flexible operation and high spatial, spectral, and temporal resolution, the UAV images can identify the damaged plants and monitor damaged severity accurately[15]. Dobbels and Lorenz[16] used a UAV system to monitor soybean chlorosis caused by iron deficiency. They achieved better results than traditional ground-based visual assessments. Yue et al.[17] extracted pest information from UAV images with an improved scale-invariant feature transform algorithm. Huang et al.[18] explored the potential using a photochemical reflectance index for quantifying the yellow rust in wheat from airborne hyperspectral images. Due to the UAV can observe Stereo Image pairs, and the images can be used to produce digital surface models (DSM) through photogrammetry using image-based modeling (IBM) algorithm, which offers a cost-effective way to collect crop canopy structure information[19]. As a co-product of UAV optical imaging, the DSM can be used to classify crop planted area[20], monitor crop growth condition[21], identify crop disasters, and estimate crop yield. However, the potential of jointing the UAV spectral images, DSM features and spectral features to monitor the *armyworm* infestation has few studies yet. Therefore, we search the optimal combination of the UAV spectral images, DSM features and image features by *Gini*-importance for monitoring the damage of *armyworm* for summer corn in this study.

Remote sensing data are massive and high-dimensional big data. Therefore, efficient and automated approaches are critical to image analysis and interpretation[22]. The Random Forest (RF) is one of the popular supervised classification methods with an ensemble learning algorithm, which uses multiple decision trees to predict samplings[23]. Moreover, RF can produce the importance of all features in the classification, which helps reduce data dimension and improve efficiency[24] . Consequently, RF has been used in pest damage monitoring by UAV images. Adelabu et al.[25] used RF and Support Vector Machine classification algorithms to identify different levels of insect defoliation in an African savanna based on RapidEye imagery. Aparecido et al.[26] developed a disease and pest warning system for coffee to predict the incidence of various diseases and insect pests using the RF Regressor and Artificial Neural Networks. In summary, the pixel-based RF method has attracted wide attention because it has good classification performance and processing speed for crop disaster classification. Therefore, we use the RF to monitor the *armyworm* damage of summer corn in this study based on the UAV dataset.

In this study, the RF algorithm is used to identify *armyworm* damage levels of summer corn by pest incidence based on the UAV dataset. Moreover, the spectral characteristics of different pest incidences in Sentinel-2 images are analyzed for exploring the potential of the satellite images. The contributions of this paper lie in:

i) Identifying the important image features in the UAV dataset for identifying the *armyworm* pests that occurred in corn planted areas.

ii) Exploring the potential of the RF model for monitoring the damage of *armyworm* pests of summer corn, compared with three kinds of machine learning algorithms including MLP, NB and SVM.

iii) Classifying the *armyworm* infected areas and healthy corn area by optimized RF model in the UAV dataset. And the pest Levels are divided into five levels using Natural Breaks (Jenks) and weighted average.

iv) Analyzing the relationship between the spectral characteristics of different bands and different pest incidence levels in Sentinel-2 and UAV images. Combining UAV classification results with Sentinel-2A images to monitor pest incidence levels in a large range.

2. Materials and Methods

2.1 Study area

The study area was located in Beicuizhuang Village, Langfang City, Hebei Province, China, ranging from 116°44′E 39°21′N to 116°46′E 39°22′N, and covering an area of about 70 ha (Figure 1 (a)). This area was in the central-eastern part of the North China Plain. And there were various crops were planted, including summer corn, soybean, peanut, apple trees, and other crops. This was a typical small-holding farming area with irrigation by pumping water, and the crop management (e.g., sowing, irrigation, fertilization and weeding) was varied. Conventionally, summer corn was sowed at the end of June and harvested at the beginning of October in the study area. There were some summer corn areas that were attacked seriously by *armyworm* in the middle of August 2019.

For monitoring the damage of *armyworm* pests, the UAV imaging and field campaign were done simultaneously on 18 August, 2019. The collected UAV images (B2/ B3/ B1) and field data are shown in Figure1 (b). According to the health condition of corn leaves, there were 24 representative field plots that were collected for monitoring the damage of *armyworm* pests. And there were 5 samplings in each plot which were distributed in the four corners and center of each plot. The size of each sampling was 1m×1m. All measured samplings were located using Huace i80 real-time kinematic (RTK) GPS receiver (Huace Ltd., Shanghai, China). The corn planted area in the study area could be divided into two main groups through field campaign: (1) the summer corn planted area attacked by *armyworm*, (2) the healthy summer corn planted area. In Figure1 (c), the left of each group was UAV sub-images, and the right of each group was the field sample photos. Figure1 (c) showed that the summer corn leaves attacked by *armyworm* are smaller and thinner, and the healthy corn planted areas were more evenly distributed. Considering the significant influence of the sample quantity and quality on model training, there were 1740 samples of healthy corn planted area, 1763 samples of corn damaged area by *armyworm* were selected manually from UAV images according to the measured data in field campaign and visual interpretation, and each sample is a pixel of UAV image.

2.2 Image collection and feature dataset construction

2.2.1 UAV image collection

The UAV images were collected between 10:50 and 11:25 AM, 18 August 2019, when it was sunny and windless. The Parrot Disco-Pro AG, a fixed-wing UAV system, was used to collect UAV images. This system carried an automated multispectral sensor (Parrot Sequoia camera), which was developed for agricultural applications. The multispectral bands include green, red, Red-edge and near-infrared (NIR), and the spectral characteristics of the UAV image collected using the Parrot Sequoia camera was shown in Table 1. The camera was connected to an irradiance sensor, which had the same spectral bands as the multispectral sensor that could record the light conditions. In this way, the image data was done radiometric calibration in real-time for illumination changing automatically.

Before the flight, the image of the calibration board was captured by the Parrot Sequoia camera for radiation correction (Figure 2). It was noted that the calibration board must face the sun to make no shadows covering the radiation calibration board. Figure 2 (a) – (d) are the green, red, Red-edge, and NIR bands captured for radiometric calibration. The height of flight above ground was 120 meters. The longitudinal overlap and side overlap were all 80%, with the ground sample distance (GSD) being 0.135 m/pixel. The georeferencing was achieved by Global Positioning System (GPS) built into the Parrot Sequoia camera and five ground control points (GCPs) were provided by using a Huace i80 real-time kinematic (RTK) GPS receiver (Huace Ltd., Shanghai, China) with 2.0 cm of positioning accuracy. The UAV photos, the calibration photos, and GCPs were input into the Pix4Dmapper Pro 4.1 software. The image splicing and point cloud modeling were done using "Ag Multispectral" template automaticly[27]. And the spliced multispectral image with 3,315 × 5,176 (Figure 1(b)) was generated with a spatial resolution of 0.135 m/pixel.

2.2.2 Feature dataset construction

The Normalized Difference Vegetation Index (NDVI) is the most commonly used remote sensing vegetation index, which is calculated from the reflectance measurements in the near infrared (790 nm) and red (660 nm) portion of the spectrum in UAV images[28]. The NDVI can be expressed by the following formula:

$$\mathrm{NDVI} = \frac{\rho_{NIR} - \rho_{red}}{\rho_{NIR} + \rho_{red}} \quad (1)$$

where, ρ_{NIR} and ρ_{red} are the NIR band and red band respectively. For the green crops, the reflectance of red band is small and the reflectance of NIR band is big. So the NDVI approaches 1 when the crops are dense, and the NDVI approaches 0 when the crops are small. Therefore, NDVI is used to depict if the corn leaves are attacked by *armyworm*.

From the photos taken in the field campaign, it is observed that the leaves of the summer corn affected by *armyworm* are generally blade yellowing. The corn leaves waned by *armyworm* attack resulting in the low vegetation cover and spectral change. So, the Red-edge Normalized Vegetation Index (RENDVI) is calculated based on two bands including the red band and the Red-edge band.

$$\mathrm{RENDVI} = \frac{\rho_{RE} - \rho_{red}}{\rho_{RE} + \rho_{red}} \quad (2)$$

where, ρ_{RE} is the Red-edge band. Similar to the spectral index NDVI, RENDVI approaches 1 when the crops are dense and the NDVI approaches 0 when the crops are thin, which can be used to depict if the corn leaves are attacked by *armyworm*.

Considering the height of the corn affected by *armyworm* is generally lower, the corn canopy DSM is built up to depicting the canopy height difference between corn plants attacked by *armyworm* and the healthy corn plants. The corn canopy DSM is generated with the spatial resolution of 0.131 m by Pix4Dmapper Pro 4.1 software. Therefore, the features of the UAV image are composed of four spectral bands (green, Red, Red-edge, and NIR), DSM, and the spectral index features including NDVI and RENDVI.0

2.3 Machine learning classification methods for dentifying corn pests areas

2.3.1 Multilayer Perceptron

The most important feature of multilayer perceptron (MLP) is that it has multiple neuron layers, so it is also called deep neural network. The first layer is the input layer, the last layer is the output layer, and the middle layer is the hidden layers. The optimal number of hidden layers and output neurons can be determined according to the specific application[29]. Because this method has good generalization ability, it is popular in classifying the pest areas of planted corns.

2.3.2 Naive Bayesian

Naive Bayesian (NB) based on Bayesian theory is a widely used classification algorithm in machine learning and data mining, with the assumption that the variables predicted are the

Gaussian distributed and independent of each other. The classification is based on the conditional probability that each sample belongs to various classes[30]. Compared with other methods, the NB method is applicable with few or no input parameters, which is uncomplicated and cost-effective in dealing with classification problems. In this study, the NB method is selected and used to identify the *armyworm* areas of planted corns.

2.3.3 Support vector machine

Support vector machine (SVM) classification via spatial features is a learning method based on statistical learning theory, which classifies the input sample features by solving the optimal hyperplane among many different classes[31]. The core of SVM is to solve the problem of dichotomy. For multi-classification problems, "one-to-many" classification method is usually adopted. For n classes, there are n hyperplanes that need to be solved. And there are n results will be produced after the prediction sample passes the discrimination of n optimal hyperplanes. And the optimal class will be produced finally. The SVM is a small sample learning method with good robustness, which is used to classify the pest areas of planted corns for the UAV dataset in this study.

2.3.4 Random Forest classification

Random forest classification algorithm is an integrated learning method based on the combination of multiple CART decision trees[32]. In order to divide the variable space completely, self-extraction of input samples and node random splitting techniques are used to construct multiple decision trees. The prediction results can be obtained through the majority voting strategy of the decision tree, and the error caused by a single parameter group can be avoided effectively. The RF can generate the importance of each dimension, which helps the feature selection and improve efficiency. Therefore, the RF algorithm can be adapted to classify the *armyworm* areas of planted corns for the UAV dataset.

2.4 Feature selection for classifying corn pests areas

The dimensions importance can be expressed with the *Gini*-importance[33] for high-dimensional UAV dataset. The *Gini* index is used to measure the impurity (degree of uncertainty) of the sampling set, which is the probability that a random sample is misclassified. The *Gini* impurity of the initial set is as the following formula:

$$Gini(K) = \sum_{i=1}^{n} P(k_i) \times \left(1 - P(k_i)\right) \quad (3)$$

where, $K = \{k_i;\ i = 1,2,3, \ldots, n\}$ is the collection of all classes, k_i is the classes i, n is total number of classes. The $P(k_i)$ is the probability of the k_i class. When the initial set is divided into multiple subsets, the *Gini* impurity is as the following formula:

$$Gini_s = \sum_{j=1}^{N} P(x_j) \times Gini(x_j) \quad (4)$$

where, $x_j = \{j = 1,2,3, \ldots, N\}$ is the j-th set in the initial set, N is the total number of subsets. $P(x_j)$ is the probability of x_j. $Gini_s$ is the *Gini* internal impurity of x_j, which is the *Gini*-importance and is used to calculate the contribution of each feature.

2.5 Accuracy assessment

There are two measurements used to assess the classification performance for the *armyworm* pest infested corn planted area, including the overall accuracy (*OA*) and Kappa coefficient. These measurements are calculated from a confusion matrix[34] of classification results. The confusion matrix includes TP (True Positive), FN (False Negative), FP (False Positive) and TN (Ture Negative). A 5-fold cross-validation[35] is used to evaluate the accuracy of the classification. All the selected samplings are divided into 5 groups randomly, including 4 groups used for training and 1 group used for testing. The configurations of hardware used in this study are Intel(R) Core (TM) i7-8700 CPU 3.20 GHz, 32 G RAM.

The *OA* indicates the proportion of correct pixels predicted, including pixels both healthy and attacked by insect pests. The *OA* can be as the following formula:

$$OA = \frac{\mathrm{TP+TN}}{\mathrm{TP+FN+FP+TN}} \quad (5)$$

where, TP and FN are the numbers of correct and wrong classifications of samples in the healthy corn category; FP and TN are the number of wrong and correct classifications of samples in the *armyworm* pest category.

The *Kappa* coefficient is a statistical indicator of interrater reliability, which is calculated by the *OA* and the probability of random agreement. The *Kappa* coefficient can be as the following formula,

$$Kappa = \frac{OA - P_e}{1 - P_e} \quad (6)$$

where, P_e is the probability of random agreement. The *Kappa* is an index to measure the spatial consistency of classification results, which reveals the spatial changes of classification results clearly. Then the P_e is as the following formula:

$$P_e = \frac{(\mathrm{TP+FN}) \times (\mathrm{TP+FP}) + (\mathrm{FP+FN}) \times (\mathrm{FN+TN})}{\mathrm{TP+FN+FP+TN}} \quad (7)$$

3. Results and Analysis

3.1 Optimize UAV dataset by *Gini*-importance and Random Forest method

The RF method is used to classify the summer corn planted area infested by *armyworm* for the UAV dataset. The zoomed sub-plots of the UAV dataset including pest infested area and healthy area are as shown in Figure 3. Figure 3 (a1) - (g1) are the image features of *armyworm* pest infested corn planted area, (a2) – (g2) are the features images of healthy corn planted area with the image features of green band, red band, red-edge band, NIR band, DSM, NDVI and RENDVI, respectively. From Figure 3(a1) - (g1), it can be seen that the tone difference is obvious and the image features are more heterogenous and blurry for *armyworm* infested areas than healthy area, especially in (a1), (b1), (f1) and (g1). Comparatively, the tone difference is relatively small and homogeneity exists in healthy corn planted areas in Figure 3 (a2) - (g2).

In order to determine the importance of each image feature in the UAV dataset, *Gini*-importance (Eq. 4) is used to measure and calculate the importance of image features in this study. Moreover, the parameters of RF are default, and the importance and the evaluation measures take the mean value of ten times experiment. The importance order of the UAV dataset image features for the identification of *armyworm* infested area is that NDVI (0.2933) > RENDVI (0.2061) > B2-Red (0.1677) > B4-NIR (0.1461) > DSM (0.0885) > B1-Green (0.0672) > B3-Red-edge (0.0308), as shown in Figure 4. And Figure 4 revealed that the importance of all features is greater than 0.03 and the top three important features for the classification model are the NDVI, RENDVI and B2-Red. It is clear that the vegetation index features (NDVI and RENDVI) bring the greatest contribution to classification.

With the same training samples, different features combinations are used to identify *armyworm* infested areas. The features in the UAV dataset are sorted by importance and the construction of each model in line with the importance of all features, and the performance of models is compared using the accuracy for selecting the optimal feature combination. The *Kappa* and *OA* are used to evaluate the performance of the different models quantitatively. And the experimental results are shown in Table 2. According to Table 2, the Model7 with all the features has the best quantitative evaluation results with *Kappa* and *OA* are 0.9709 and 0.9850, respectively. Compared to Model1 with a single image feature (RENDVI) only, the Model2, Model3, Model4, Model5, Model6 and Model 7 improved 0.1860/0.0906, 0.1933/0.0993, 0.1925/0.0990, 0.2019/0.1037, 0.2072/0.1064 and 0.2098/0.1078 in *Kappa* and *OA*, respectively. We also found that the model performance is constantly improved with the increase of features numbers. Compared with Model8 which only uses four features (Green, Red, Red-edge and NIR) and Model9 which uses five features (Green, Red, Red-edge, NIR and DSM), the Model7 improved 0.0301/0.0209 and 0.0195/0.0115 in *Kappa*/*OA*. This revealed that the joint of the DSM, NDVI and RENDVI could improve the performance of the classifier effectively. It is further proof that the features of the designed dataset in this study are reasonable and effective.

3.2 Comparison of performance for different machinelearning algorithms

The performance of different machine learning methods is compared to verify the superiority of the RF. And there are three machine learning methods including MLP, NB and SVM are compared with RF. For all the methods in these experiments, this comparison is done using the same cross-validation method and the evaluation measures, which are 5-fold cross-validation and *Kappa*/*OA*, respectively. In addition, all training and testing datasets of the RF experiments (Model7) are used to the other three classification (MLP, NB and SVM) experiments for ensuring a fair comparison. The good model parameters of MLP, NB and SVM are selected according to experience, and the parameters of RF are default. For MLP classification method[29], the initial learning the activation functions are 0.001 and "RELU", there are 2 hidden layers with 100 neurons per hidden layer. For NB classification method[30], there is only one major parameter, the prior probability, which is set as default without giving the prior probability. For SVM classification method[31], the main parameters are as follows: the kernel type is radial basis function, gamma value is "auto" in the kernel function, and cost or slack parameter is 1.0. The quantitative evaluation results for different machine learning methods are shown in Table 3.

According to Table 3, the classification performance of four machine learning methods is greater than 0.91/0.95 (*Kappa*/*OA*). Compared with MLP, NB and SVM, the RF method improved 0.0513, 0.0308 and 0.0285 in *Kappa*, and improved 0.0258, 0.0213 and 0.0205 in *OA*, respectively. It revealed that the RF classifier had the highest accuracy in distinguishing healthy and *armyworm* pests on corn. At the same time, the superiority of the RF classifier is also proved.

3.3 Mapping of *armyworm* infested area

Based on the comparison of the performance of different methods, the RF method is more applicable to mapping *armyworm* areas. The parameters of the RF method (Model7) can be optimized by random search and grid search furtherly. Firstly, the random searching method is used to estimate roughly the range of parameters, the number of decision trees is (80, 100), the maximum depth of decision trees is (90, 120). Secondly, the grid search method is used to optimize the parameters furtherly, the search step is 1. Finally, the optimal combination of parameters is obtained, the number of decision trees is 91, the maximum depth of decision trees is 107. The *Kappa* and *OA* of retrained RF (Model 7) are 0.9735 and 0.9864 based on optimized parameters. The mapping of *armyworm* infested areas is done using optimized RF (Model 7) subsequently, and the results as shown in Figure 5.

According to Figure 5, the pink areas are *armyworm* infested areas, and the green areas are healthy corn planted areas. A, B, C and D are the measured field plots in field campaigns where *armyworm* occurs seriously. The armyworm infestation occurs in the southwest of field plot A, the top center of field plot B, and nearly the whole of field plot C and D mainly. In addition, the reliability of the classification results is verified by field survey points. Four representative field survey photos are shown at four measured points from (p1) to (p4). The (p1) and (p2) are located in plots A and B, which are in the area with serious *armyworm* pests. It can be seen from the photos that most of the corn leaves are affected by *armyworm*, and the leaves are relatively few and small, resulting in the serious changes in corn canopy and morphology. On the other hand, (p3) and (p4) are in the healthy area, which is consistent with the obtained field survey photos.

3.4 Mapping of pest incidence level

To determine the severity of *armyworm* pests, and give some hints for pesticide spraying, the pest incidence level mapping is done. The NDVI can describe the vegetation growing condition, and the *Gini*-importance of NDVI features in Section 3.1 is the most prominent to UAV image classification. Natural Breaks (Jenks) method can best group similar values and maximize the differences between classes[36], which is a good method for grade classification. According to corn damage scales[37], the *armyworm* pest areas (the pink areas) in Figure 5 are divided into four levels based on NDVI value using Natural Breaks (Jenks). The levels include Level 1(0.23<=NDVI<=0.522,

little damage), Level 2(0.12<=NDVI<0.23, medium damage), Level 3(0.01<=NDVI<0.12, heavy damage) and Level 4(-0.22<=NDVI<0.01, very heavy damage), and healthy corn is Level 0(No damage). The result of pest incidence levels is shown in Figure 6(a), the green area is Level 0, and the red area indicates the most serious pest incidence (Level 4).

Infesting levels of insect pests can be extended to 10m × 10 m grid. First, assign values 0, 1, 2, 3, and 4 to Level 0, Level 1, Level 2, Level 3, and Level 4 in Figure 6(a). Then, the mean value of pixels in 10m × 10m grid is calculated as the grid value. The method Natural Breaks (Jenks) is used to classify damaged corn of the grids. The levels include Level 0[0, 0.172), Level 1[0.172, 0.58), Level 2[0.58, 1.21), Level 3[1.21, 2.04) and Level 4[2.04, 4]. The result of pest levels based on grids is shown in Figure 6(b).

It can be found that the red areas are the largest in A, B, C and D field plots, indicating the highest pest levels, which is consistent with field observations and Figure 5. In terms of pest level, plots A and B, C and D are adjacent to each other, and the infested areas are connected. In the early stage of the pest incidence, the farmers in field plots A, B, C, and D did not find out the *armyworm* infection timely and did not take appropriate management for preventing the pest from spreading. This could be the reason for the local *armyworm* outbreak in the cornfield. Furthermore, we also found that the height of weeds is higher than the corn plants in damaged corn plots, and result in a humid and airless environment, which might also be one of the important reasons for the outbreak of *armyworm*.

3.5 Spectral reasoning for mapping of pest incidence level

The Sentinel-2 image[38] has more abundant spectral information than the UAV image, which is likely to be beneficial for pest monitoring. The spectral characteristics of different pest levels are analyzed for the Sentinel-2 image and the UAV images. For unifying the analyzing unit of spectral difference, the obtained Sentinel-2 image on 18 August 2019 in the study area is reconstructed with super resolution using SupReME[39], and each band is unified to 10m spatial resolution. After that, the spectral characteristics of Sentinel-2 image before and after super resolution reconstruction are analyzed, mainly in the building and corn planted area, which is as shown in Figure 7. In Figure 7, the orange line (predictive value) is the reflectance spectral curve after SupReME reconstruction, the red line (truth value) is the reflectance spectral curve of Sentinel -2. Figure 7 (a) is the spectral contrast of corn planted area, (b) is spectral contrast of buildings, the RMSE (Root Mean Square Error) are 0.00169 (a) and 0.008085 (b), respectively. The RMSE is close to zero, and the spectrum is basically unchanged.

The Sentinel-2 image with 10m spatial resolution and UAV images with a resolution of 0.135m are used to analyze the spectral characteristics of different pest levels. There are 600 sampling points and 250 sampling points are selected based on Figure 6(a) and Figure 6(b) respectively to calculate the mean values of each band of UAV and Sentinel-2 images. The statistical results as shown in Figure 8. The X-axis are band numbers and the Y-axis is spectral reflectance, different colors indicate different levels of pests. In Figure 8(a), with the pest level declines, the spectral reflectance of B1, B2, B3, B4, B5, B11 and B12 decreases, the spectral reflectance of B7, B8 and B8A increases. The spectral reflectance of B9 can distinguish between healthy and unhealthy corn areas, B6 is not clear. In Figure 8(b), B1 and B2 reveal that the spectral reflectance decreases with the pest level declines. B4 can distinguish between healthy corn and damaged corn. B3 is similar to B6 in the Sentinel-2 image, the relationship between reflectivity and pest level is not very clear. Sentinel-2 and UAV images have similar characteristics in different pest levels in corresponding wavelength bands.

4. Discussions

The corn area damaged by *armyworm* resulted in the changing of corn canopy spectral reflectance since the changed morphological and chemical characteristics of leaves[40]. In Figure 8(b), the bands involved in the vegetation index calculation are sensitive to the *armyworm* pest infestation, the NDVI and RENDVI features bring the greatest contribution to the classification. The height of

corn plants infested by *armyworm* is lower than healthy plants, which brings the DSM is also sensitive to the pest infestation of summer corn. It shows that the addition of vegetation index and DSM is beneficial to accurate classification.

As is known, the red-edge band (B5, B6 and B7 of Sentinel-2 and B3 of UAV) was one of the most sensitive bands to vegetation diseases[41]. In Figure 8 (a), with the change of corn *armyworm* incidence levels, the spectral characteristics of the red-edge bands change as well. With the decrease of corn *armyworm* incidence, the reflectance of B5 decreases, and the reflectance of B7 increases. The B6 of Sentinel-2 and the B3 of UAV are between B5 and B7, and the spectral intervals were only about 40 nm, which makes the sensitivity of the band (B6 of Sentinel-2 or the B3 of UAV) to the pest change lower than other bands[42]. B1, B2, B3 and B4 of Sentinel-2 and the B1 and B2 of UAV were the visible spectrum band where the leaves could rely on various pigments (i.e., chlorophyll, carotenoids) to absorb this energy. with corn canopy pigment decrease, visible light absorption decreases, which resulted in the increase of visible reflectance. B7, B8, B8A and B9 of Sentinel-2 and the B4 of UAV were the near-infrared bands and were related closely to the cell structure. The canopy reflectance was lower than healthy corn, indicating that the *armyworm* had seriously damaged the cell structure of the corn. Moreover, B11 and B12 were related to the water content of vegetation. In the area damaged by *armyworm*, the leaves of the plants gradually withered and the water content decreased, which led to the increase in spectral reflectance compared with normal vegetation.

The result of Figure 7 and Figure 8 revealed the performance of abundant spectral information of Sentinel-2 image had a great advantage in monitoring the damage of *armyworm* in the regional area. The Towns (An ci) in the UAV flight area and adjacent towns (Yong qing) are used as experimental areas (Figure 9(a)), the obtained Sentinel-2 images (18 August 2019) after SupReME reconstruction were attempted to classify pest levels. There are selected 250 sampling points based on Figure 6(b), and the training dataset and test dataset are divided into 7:3 for the RF model. The classification results as shown in Figure 9(b), the *OA* of train and validation are 0.76 and 0.69. Different colors correspond to different pest levels, the 'Level 5' in legend was the non-corn area. The non-corn area was obtained by Google Earth Engine platform classification using RF based on Sentinel-2 image (18 August 2019), which overall validation accuracy is 0.92. Figure 9(c) is the sub-image of Figure 9 (b), and the black box is the UAV flight area. The UAV flight area in Figure 9(c) was consistent with the classification results (Figure 6(b)) of UAV basically, which revealed the effectiveness of classifying *armyworm* levels based on Sentinel-2 images. The areas of Level 3 and Level 4 were very small and the proportion of insect pest areas was low, indicating that the planted corns were generally healthy. The damaged levels by *armyworm* are classified from Sentinel-2 images, which could provide data support for pesticide spraying and pest prevention in a large area.

5. Conclusions

The *armyworm* is one of the most serious insect pests of corn. We proposed a method to monitor the damaged scale by pest in summer corn based on UAV images. The conclusions can be drawn as follows.

(1) The importance of image features in the UAV dataset is determined by *Gini*-importance. The importance of images features is sorted as NDVI (0.2933) > RENDVI (0.2061) > B2-Red (0.1677) > B4-NIR (0.1461) > DSM (0.0885) > B1-Green (0.0672) > B3-Red-edge (0.0308). The NDVI is the most sensitive, while the B3 of UAV has the lowest sensitivity to *armyworm* pests.

(2) The RF model has the best performance for the classification of the *armyworm* pest and healthy corn compared to different machine learning methods (MLP, NB and SVM). The parameters of the RF can be optimized by random search and grid search furtherly, which are used to retrain the model. The classification results of the UAV dataset are predicted subsequently.

(3) The spectral characteristics of sentinel-2 and UAV images with different pest levels are analyzed. With the pest level declines, the spectral reflectance of B1 and B2 decreases, B3 is not very clear, B4 can distinguish healthy corn and damaged corn in UAV images. With the pest level

increases, the spectral reflectance of B1-B5 and B11-B12 increased, B7, B8 and B8A decreased, B6 and B9 have no clear characteristics in the Sentinel-2 image. Furthermore, the classification of pest levels based on Sentinel-2 image using the RF model has achieved good results.

Due to the small geographical coverage of UAV images, the experimental area is small and restricted relatively. In the future, we will expand the research to explore the *armyworm* infestation of other similar crops (e.g., wheat and rice) and monitor the damage of pests on a larger scale. Then, coupling the crop growth models (e.g., WOFOST, DSSAT) or radiation transfer models (e.g., PROSAIL, DART) with machine learning techniques will be tried, which maybe have potential in estimating the damage of *armyworm* pests. Furthermore, the multi-temporal UAV images and satellite images can be combined tentatively to detect insect pests in the field dynamically, which can support more informed farming decisions.

Author Contributions: This work was a cooperation of our research team, and the contributions were as follows: Conceptualization, Mingzheng Zhang, Jing-Hao Xue and Wei Su; Formal analysis and Writing – review & editing, Jing-Hao Xue and Dongqin Yin; Investigation, Xinsheng Wang, Wei Su and Dehai Zhu; Methodology and Writing – original draft, Mingzheng Zhang; Software, Mingzheng Zhang and Wancheng Tao; Validation, Zixuan Xie.

Funding This study was funded by the National Natural Science Foundation of China under the projects (No. 42171331), the 2115 Talent development Program of China Agricultural University.

Availability of data and materials

The datasets used and/or analyzed during the current study available from the corresponding author on reasonable request.

Acknowledgments: We thank the farmers for allowing us to measure their fields.

Competing interests:

The authors declare that they have no competing interests.

References

1. Xu HX, Yang YJ, Lu YH, Zheng XS, Tian JC, Lai FX, et al., Sustainable Management of Rice Insect Pests by Non-Chemical-Insecticide Technologies in China. *Rice Sci* **24**: 61-72 (2017).
2. Deutsch CA, Tewksbury JJ, Tigchelaar M, Battisti DS, Merrill SC, Huey RB, et al., Increase in crop losses to insect pests in a warming climate. *Science* **361**: 916-919 (2018).
3. Yasoob H, Abbas N, Li YF and Zhang YL, Selection for resistance, life history traits and the biochemical mechanism of resistance to thiamethoxam in the maize armyworm, Mythimna separata (Lepidoptera: Noctuidae). *Phytoparasitica* **46**: 627-634 (2018).
4. Serbesoff-King K, Melaleuca in Florida: A literature review on the taxonomy, distribution, biology, ecology, economic importance and control measures. *J Aquat Plant Manage* **41**: 98-112 (2003).
5. Liu J and Jiang Y, Characteristic of Corn Diseases and Pest in 2012 and Cause Analysis. *Chinese Agricultural Science Bulletin* **30**: 270-279 (2014).
6. Godfray HCJ, Beddington JR, Crute IR, Haddad L, Lawrence D, Muir JF, et al., Food Security: The Challenge of Feeding 9 Billion People. *Science* **327**: 812-818 (2010).
7. Klauser D, Challenges in monitoring and managing plant diseases in developing countries. *J Plant Dis Protect* **125**: 235-237 (2018).
8. Yuan L, Bao ZY, Zhang HB, Zhang YT and Liang X, Habitat monitoring to evaluate crop disease and pest distributions based on multi-source satellite remote sensing imagery. *Optik* **145**: 66-73 (2017).

9. Chemura A, Mutanga O and Dube T, Separability of coffee leaf rust infection levels with machine learning methods at Sentinel-2 MSI spectral resolutions. *Precis Agric* **18**: 859-881 (2017).
10. Meigs GW, Kennedy RE, Gray AN and Gregory MJ, Spatiotemporal dynamics of recent mountain pine beetle and western spruce budworm outbreaks across the Pacific Northwest Region, USA. *Forest Ecol Manag* **339**: 71-86 (2015).
11. Eklundh L, Johansson T and Solberg S, Mapping insect defoliation in Scots pine with MODIS time-series data. *Remote Sens Environ* **113**: 1566-1573 (2009).
12. Su W, Zhang MZ, Bian DH, Liu Z, Huang JX, Wang W, et al., Phenotyping of Corn Plants Using Unmanned Aerial Vehicle (UAV) Images. *Remote Sens-Basel* **11**2019).
13. Li X, Giles DK, Andaloro JT, Long R, Lang EB, Watson LJ, et al., Comparison of UAV and fixed-wing aerial application for alfalfa insect pest control: evaluating efficacy, residues, and spray quality. **n/a**.
14. Roosjen PP, Kellenberger B, Kooistra L, Green DR and Fahrentrapp J, Deep learning for automated detection of Drosophila suzukii: potential for UAV-based monitoring. **76**: 2994-3002 (2020).
15. Zarco-Tejada PJ, Gonzalez-Dugo V and Berni JAJ, Fluorescence, temperature and narrow-band indices acquired from a UAV platform for water stress detection using a micro-hyperspectral imager and a thermal camera. *Remote Sens Environ* **117**: 322-337 (2012).
16. Dobbels AA and Lorenz AJ, Soybean iron deficiency chlorosis high throughput phenotyping using an unmanned aircraft system (vol 15, 97, 2019). *Plant Methods* **15**2019).
17. Yue JW, Lei TJ, Li CC and Zhu JQ, The Application of Unmanned Aerial Vehicle Remote Sensing in Quickly Monitoring Crop Pests. *Intell Autom Soft Co* **18**: 1043-1052 (2012).
18. Huang W, Lamb DW, Niu Z, Zhang Y, Liu L and Wang J, Identification of yellow rust in wheat using in-situ spectral reflectance measurements and airborne hyperspectral imaging. *Precis Agric* **8**: 187-197 (2007).
19. Zarco-Tejada PJ, Diaz-Varela R, Angileri V and Loudjani P, Tree height quantification using very high resolution imagery acquired from an unmanned aerial vehicle (UAV) and automatic 3D photo-reconstruction methods. *Eur J Agron* **55**: 89-99 (2014).
20. Diaz-Varela RA, Zarco-Tejada PJ, Angileri V and Loudjani P, Automatic identification of agricultural terraces through object-oriented analysis of very high resolution DSMs and multispectral imagery obtained from an unmanned aerial vehicle. *J Environ Manage* **134**: 117-126 (2014).
21. Yang MD, Huang KS, Kuo YH, Tsai HP and Lin LM, Spatial and Spectral Hybrid Image Classification for Rice Lodging Assessment through UAV Imagery. *Remote Sens-Basel* **9**2017).
22. Troya-Galvis A, Gancarski P and Berti-Equille L, Remote sensing image analysis by aggregation of segmentation-classification collaborative agents. *Pattern Recogn* **73**: 259-274 (2018).
23. Svetnik V, Liaw A, Tong C, Culberson JC, Sheridan RP and Feuston BP, Random forest: A classification and regression tool for compound classification and QSAR modeling. *J Chem Inf Comp Sci* **43**: 1947-1958 (2003).
24. Belgiu M and Dragut L, Random forest in remote sensing: A review of applications and future directions. *Isprs J Photogramm* **114**: 24-31 (2016).
25. Adelabu S, Mutanga O and Adam E, Evaluating the impact of red-edge band from Rapideye image for classifying insect defoliation levels. *Isprs J Photogramm* **95**: 34-41 (2014).
26. Aparecido LED, Rolim GD, De Moraes JRDC, Costa CTS and de Souza PS, Machine learning algorithms for forecasting the incidence of Coffea arabica pests and diseases. *Int J Biometeorol* **64**: 671-688 (2020).

27. Fernandez-Guisuraga JM, Sanz-Ablanedo E, Suarez-Seoane S and Calvo L, Using Unmanned Aerial Vehicles in Postfire Vegetation Survey Campaigns through Large and Heterogeneous Areas: Opportunities and Challenges. *Sensors-Basel* **18**2018).
28. Hassan MA, Yang MJ, Rasheed A, Yang GJ, Reynolds M, Xia XC, et al., A rapid monitoring of NDVI across the wheat growth cycle for grain yield prediction using a multi-spectral UAV platform. *Plant Sci* **282**: 95-103 (2019).
29. Zhao CH, Gao YS, He J and Lian J, Recognition of driving postures by multiwavelet transform and multilayer perceptron classifier. *Eng Appl Artif Intel* **25**: 1677-1686 (2012).
30. Wang Q, Garrity GM, Tiedje JM and Cole JR, Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microb* **73**: 5261-5267 (2007).
31. Koda S, Zeggada A, Melgani F and Nishii R, Spatial and Structured SVM for Multilabel Image Classification. *Ieee Transactions on Geoscience and Remote Sensing* **56**: 5948-5960 (2018).
32. Zhang J, Huang Y, Reddy KN and Wang BJPMS, Assessing crop damage from dicamba on non-dicamba-tolerant soybean by hyperspectral imaging through machine learning. **75**: 3260-3272 (2019).
33. Menze BH, Kelm BM, Masuch R, Himmelreich U, Bachert P, Petrich W, et al., A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. *Bmc Bioinformatics* **10**2009).
34. Goutte C and Gaussier E, A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. *Advances in Information Retrieval* **3408**: 345-359 (2005).
35. Rodriguez JD, Perez A and Lozano JA, Sensitivity Analysis of k-Fold Cross Validation in Prediction Error Estimation. *Ieee T Pattern Anal* **32**: 569-575 (2010).
36. Chen J, Yang S, Li H, Zhang B and Lv JJIAPRSSIS, Research on geographical environment unit division based on the method of natural breaks (Jenks). **3**: 47-50 (2013).
37. Toepfer S, Fallet P, Kajuga J, Bazagwira D, Mukundwa IP, Szalai M, et al., Streamlining leaf damage rating scales for the fall armyworm on maize. *J Pest Sci* **94**: 1075-1089 (2021).
38. Sims NC, De Barro P, Newnham GJ, Kalyebi A, Macfadyen S and Malthus TJJPms, Spectral separability and mapping potential of cassava leaf damage symptoms caused by whiteflies (Bemisia tabaci). **74**: 246-255 (2018).
39. Zhang M, Su W, Fu Y, Zhu D, Xue J-H, Huang J, et al., Super-resolution enhancement of Sentinel-2 image for retrieving LAI and chlorophyll content of summer corn. **111**: 125938 (2019).
40. Gilioli G, Schrader G, Baker RHA, Ceglarska E, Kertesz VK, Lovei G, et al., Environmental risk assessment for plant pests: A procedure to evaluate their impacts on ecosystem services. *Sci Total Environ* **468**: 475-486 (2014).
41. Ramoelo A, Skidmore AK, Cho MA, Schlerf M, Mathieu R and Heitkonig IMA, Regional estimation of savanna grass nitrogen using the red-edge band of the spaceborne RapidEye sensor. *Int J Appl Earth Obs* **19**: 151-162 (2012).
42. Zheng Q, Huang WJ, Cui XM, Shi Y and Liu LY, New Spectral Index for Detecting Wheat Yellow Rust Using Sentinel-2 Multispectral Imagery. *Sensors-Basel* **18**2018).

Tables

Table 1. The spectral characteristics of the Parrot Sequoia sensor.

Band	Central wavelength (nm)	Bandwidth(nm)
B1- Green	550	40
B2- Red	660	40
B3- Red-edge	735	10
B4-NIR	790	40

Table2. Random Forest model performance over features

Model	B1	B2	B3	B4	DSM	NDVI	RENDVI	OA	Kappa
Model1	-	-	-	-	-	-	+	0.8772	0.7611
Model2	-	-	-	-	-	+	+	0.9678	0.9471
Model3	-	+	-	-	-	+	+	0.9765	0.9544
Model4	-	+	-	+	-	+	+	0.9762	0.9536
Model5	-	+	-	+	+	+	+	0.9809	0.9630
Model6	+	+	-	+	+	+	+	0.9836	0.9683
Model7	+	+	+	+	+	+	+	**0.9850**	**0.9709**
Model8	+	+	+	+	-	-	-	0.9641	0.9408
Model9	+	+	+	+	+	-	-	0.9735	0.9514

Notes: "+" represents the added modeling features. "-" represents the removed modeling feature.

Table 3. Quantitative evaluation results different method

Method	MLP	NB	SVM	RF
Kappa	0.9196	0.9401	0.9424	**0.9709**
OA	0.9592	0.9637	0.9645	**0.9850**

Figures

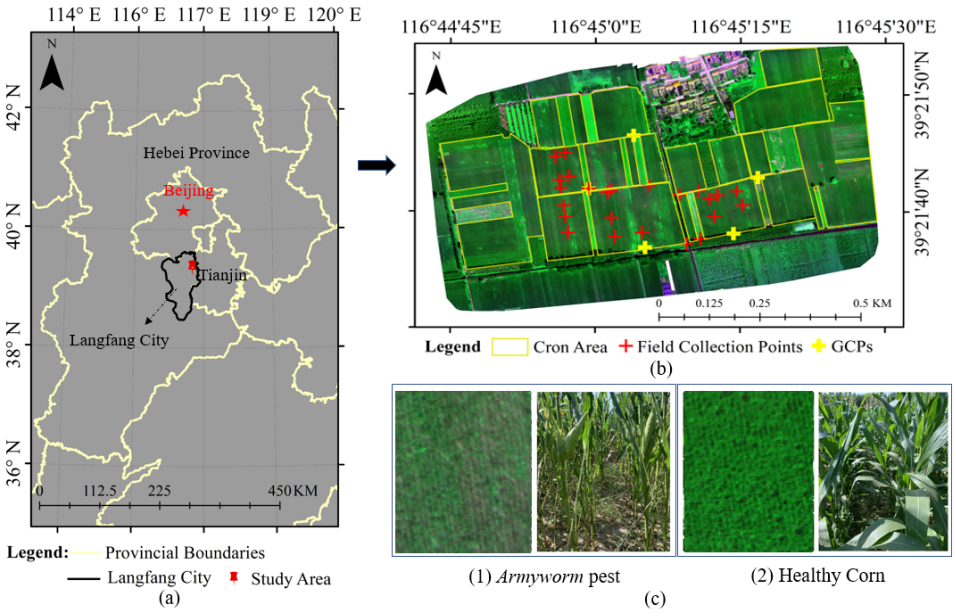


Figure 1. The location of study area (a). (b) UAV multispectral image (R (B2, red band)/G (B3, Red-edge band)/B (B1, green band)) and the spatial distribution of the field collection points and the GCPs. (c) The image pairs with UAV sub-image (R (B2, red band)/G (B3, Red-edge band)/B (B1, green band)) and corresponding photos taken in field campaign.

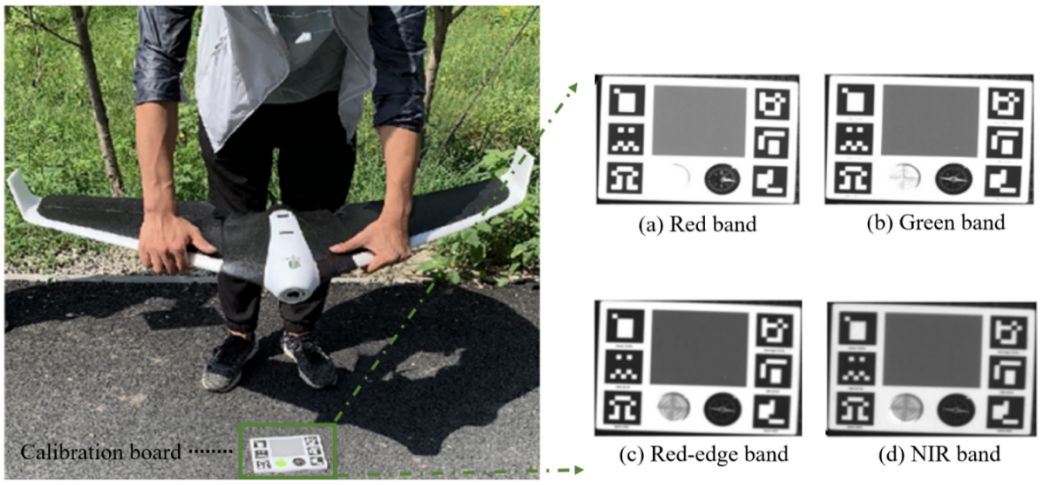


Figure 2. The fixed-wing UAV (left) and the captured images of calibration board with four spectral bands (right).

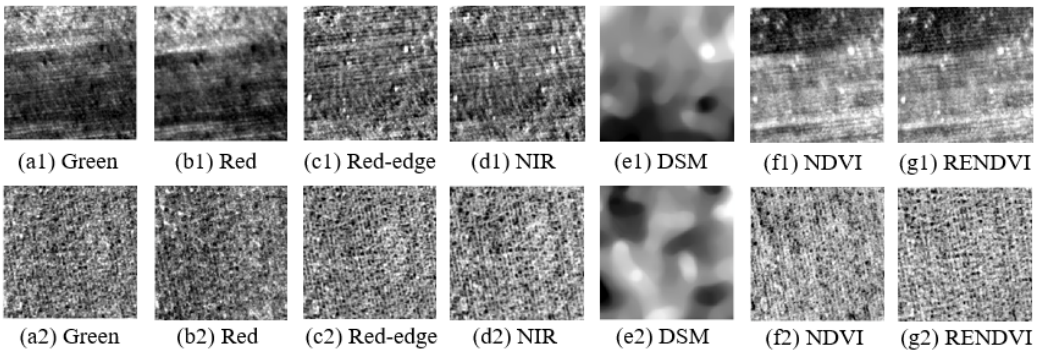


Figure 3. Visualization of image features for *armyworm* infested area (a1-g1) and healthy corn planted area (a2-g2).

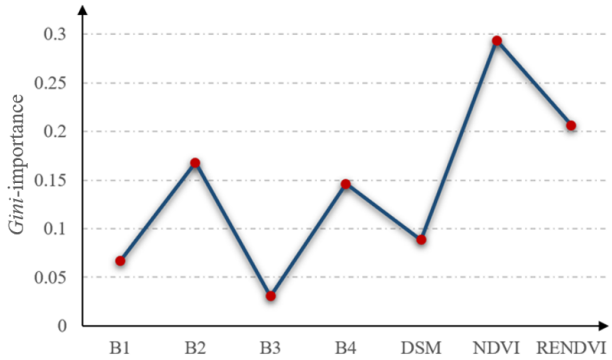
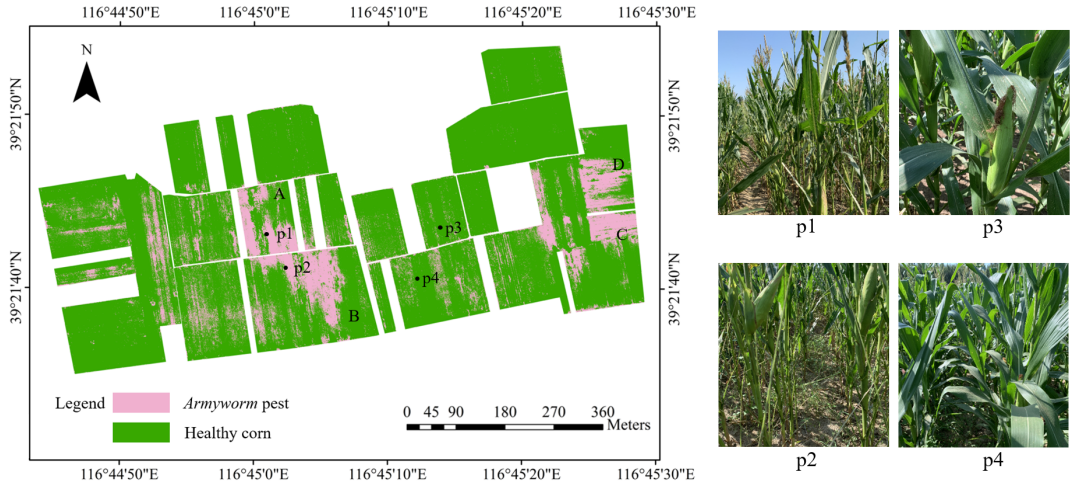


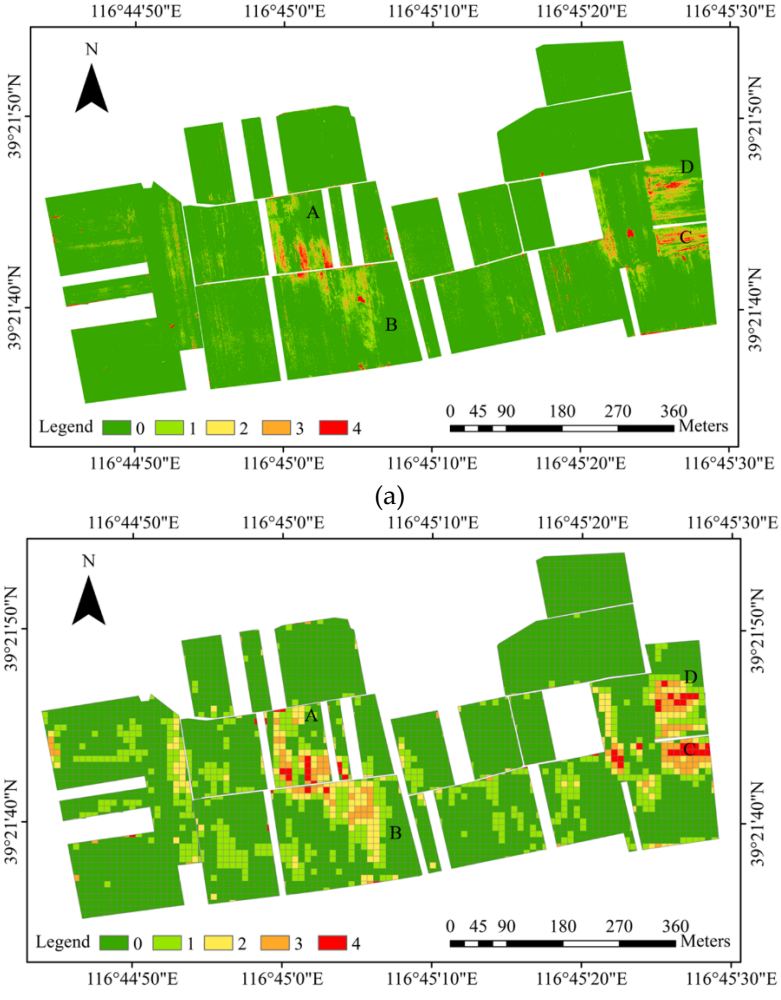
Figure 4. The *Gini*-importance of the image features.



581

582 **Figure 5.** Mapping result of corn planted area infested by *armyworm* pest by optimized RF model.
583 The mark A, B, C and D are the investigated field plots with heavy *armyworm* pest. The photos of
584 (p1) and (p2) are taken in the area infested by *armyworm*, and the photos of (p3) and (p4) are taken in
585 the healthy corn planted area.

586



587
588
589
590

591 **Figure 6.** *Armyworm* pest incidence levels of corn planted area. (a) Infesting levels of insect pests
592 based on UAV images. (b) Infesting levels resulting from (a) using statistical results of 10m × 10m

grid. The mark A, B, C and D are the investigated cornfield plots with serious *armyworm* pest infested area.

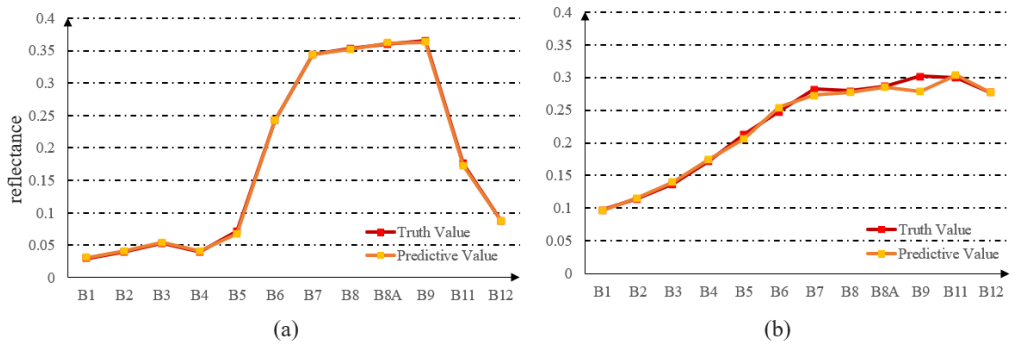


Figure 7. Spectral changes of corn planted area (a) and buildings (b) before and after SupReME reconstruction.

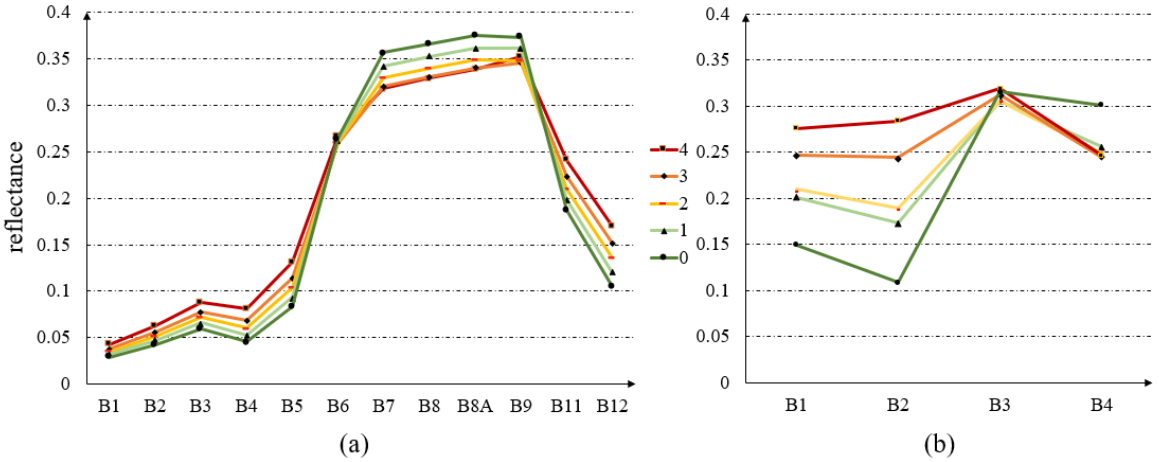


Figure 8. Spectral characteristics of different pest levels in Sentinel-2 image (a) and UAV image (b).

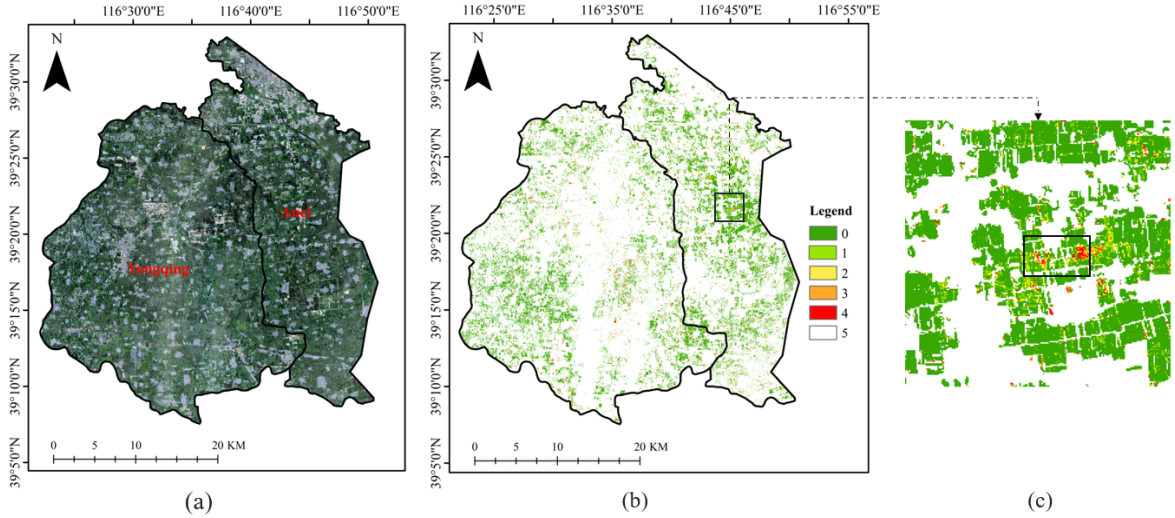


Figure 9. Classification of armyworm pests based on Sentinel-2 image. (a) is sentinel-2 true color map. (b) Classification result of armyworm pest levels. (c) sub-image of (b).