Remote Sensing of Environment 250 (2020) 112012

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

WHU-Hi: UAV-borne hyperspdectral with high spatial resolution (H^2) benchmark datasets and classifier for precise crop identification based on deep convolutional neural network with CRF

Yanfei Zhong[a,d], Xin Hu[a,*], Chang Luo[a], Xinyu Wang[b,*], Ji Zhao[c], Liangpei Zhang[a]

[a] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China
[b] School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China
[c] School of Computer Science, China University of Geosciences, Wuhan 430079, China
[d] Hubei Provincial Engineering Research Center of Natural Resources Remote Sensing Monitoring, Wuhan 430079, China

ARTICLE INFO

Keywords:
Precise crop classification
UAV-borne hyperspectral imagery
Convolutional neural network
Conditional random fields
WHU-Hi dataset

ABSTRACT

Unmanned aerial vehicle (UAV)-borne hyperspectral systems can acquire hyperspectral imagery with a high spatial resolution (which we refer to here as H^2 imagery). As a result of the low operating cost, high flexibility, and the ability to achieve real-time data acquisition, UAV-borne hyperspectral systems have become an important data source for remote sensing based agricultural monitoring. However, precise crop classification based on UAV-borne H^2 imagery is a challenging task when faced with a number of different crop classes. The traditional hyperspectral classification methods, such as the spectral-based and object-oriented classification methods, have difficulty in classifying H^2 imagery, faced with the problems of salt-and-pepper (SP) noise and scale selection. In this article, the deep convolutional neural network with a conditional random field classifier (CNNCRF) framework is proposed for precise crop classification with UAV-borne H^2 imagery. In the proposed method, a deep convolutional neural network (CNN) is designed to extract and fuse in-depth spectral and local spatial features, and the conditional random field (CRF) model further incorporates the spatial-contextual information to improve the problem of holes and isolated regions in the classification map. Meanwhile, virtual sample augmentation based on the hyperspectral imaging mechanism is used to lessen the issue of the limited labeled samples. To validate the results, a new dataset—the Wuhan UAV-borne hyperspectral image (WHU-Hi) dataset—has been built for precise crop classification. The experimental results obtained using the WHU-Hi dataset confirm the accuracy and visualization performance of the proposed CNNCRF classification method, which outperforms the previous methods. In addition, the WHU-Hi dataset could serve as a benchmark dataset for hyperspectral image classification studies.

1. Introduction

Precise crop classification and mapping are an important foundation for agricultural production management, agricultural policy setting, and food safety, and provide essential reference information for agricultural decision support (Löw et al., 2013; Wang et al., 2019). Hyperspectral remote sensing has become an important way to distinguish crop categories and achieve growth information monitoring for precision agriculture, based on the fine spectral response to the attributes of crops (Galvao et al., 2005; Nidamanuri and Zbell, 2011; Zhao et al., 2020). In the visible spectrum range of 0.4–0.7 um, the chlorophyll in crops strongly absorbs radiation at around 0.45 um (blue) and 0.65 um (red), which results in very low reflectance in the blue and green spectral regions. Furthermore, in the near-infrared spectral range of 0.7–1.0 um, the leaf area density and cellular structure of crops result in high reflectance around 0.8–1.0 um, and the reflectivity increases sharply in the range of 0.7–0.8 um (Jiang et al., 2006; Pinter Jr et al., 2003; Wardlow and Egbert, 2008). Recent years have witnessed the development of hyperspectral remote sensing technology, and rich hyperspectral data sources for precise crop classification have been provided by satellites, airplanes, and now unmanned aerial vehicle (UAV) observation platforms. The advantages of UAV-borne platforms over the other platform types is that they are fast, simple, macroscopic, and non-destructive (Honkavaara et al., 2013; Zhang and Kovacs, 2012b).

Hyperspectral remote sensing was first proposed by NASA's Jet

* Corresponding authors.
E-mail addresses: whu_huxin@whu.edu.cn (X. Hu), wangxinyu@whu.edu.cn (X. Wang).

https://doi.org/10.1016/j.rse.2020.112012
Received 11 December 2019; Received in revised form 12 June 2020; Accepted 20 July 2020

Propulsion Laboratory (JPL) in the early 1980s, and has since become an essential tool for Earth observation (Goetz et al., 1985; Tong et al., 2013). Every pixel in hyperspectral remote sensing imagery consists of dozens or hundreds of narrow spectral bands, which can be applied to high-precision detection and the identification of the attribute information of ground objects. Hence, with the fine spectral resolution of hyperspectral imagery, the crop classification ability can be greatly improved. Compared with the primary crop classification strategies that consider the phenological information of crops based on multi-temporal and multispectral remote sensing imagery (Cai et al., 2018; Son et al., 2014; Tatsumi et al., 2015), hyperspectral imagery can allow us to precisely identify multiple crop types by the spectral detail information, without requiring multi-temporal imagery. It is also difficult to obtain multi-temporal remote sensing images over the crop growth cycle in cloudy and rainy climate regions, such as the middle and lower reaches of the Yangtze River in China (Meng et al., 2020). Furthermore, space-borne hyperspectral imagery is difficult to apply in precise crop classification in smallholder agriculture regions with serious land fragmentation and heterogeneity, due to the low to medium spatial resolution (more than 30 m). For example, China, as a typical representative of smallholder agriculture regions, the average plot size was 0.087 ha in 2003 (Xiao et al., 2013), which means that the average plot size is smaller than the coverage area of one pixel in Earth Observing-1 (EO-1) and Gaofen-5 (GF-5) hyperspectral imagery with a 30-m spatial resolution. Therefore, it is a very difficult task to apply space-borne hyperspectral imagery in the precise classification of multiple crop types in smallholder regions. Despite the superiority over space-borne imagery in the spatial and spectral resolutions, the cost of manned aerial hyperspectral remote sensing imagery is high, and the data acquisition is greatly affected by the weather and the local airspace policies (Adão et al., 2017; Zhong et al., 2018a).

Over the last decade, with the continuing improvement of the payload and endurance time of UAVs, and the rapid development of lightweight hyperspectral imaging sensors, a new Earth observation technology—UAV-borne hyperspectral remote sensing—has become an essential supplement to the current space-borne and air-borne hyperspectral imaging systems (Adão et al., 2017; Suomalainen et al., 2014; Zhong et al., 2018a). As illustrated in Fig. 1, UAV-borne hyperspectral remote sensing systems can simultaneously acquire high spectral resolution (~nm level) and high spatial resolution (~cm level) remote sensing imagery, which we refer to here as H^2 imagery. As a result, UAV-borne hyperspectral systems have been widely applied in the agricultural field for their advantages of the flexible operation, low cost, and real-time data acquisition ability (Adão et al., 2017; Tu et al., 2018; Zhang and Kovacs, 2012a; Zhao et al., 2018).

Abundant research has been carried out on precise crop classification with hyperspectral imagery, adopting two primary classification strategies. The first strategy solely uses the spectral features, based on the rationale that different plants have distinctive spectral features which can be used for classification. For example, Rao et al. (2007) constructed a crop spectral library for the automatic discrimination of 12 crop categories, including rice, chili, sugarcane, and cotton. This study highlighted the possibility of classifying various crop categories with in-situ and space-borne hyperspectral data. Piiroinen et al. (2015) used the support vector machine (SVM) combined with minimum noise fraction (MNF) classification method to map crops based on the hyperspectral data acquired by an AisaEAGLE imaging spectrometer in a highly heterogeneous study area in southeastern Kenya. The second strategy is combining the spectral and spatial information. For example, Zhang et al. (2016) developed a combined spectral and spatial crop classification method by first building an optimal feature band set from the spectral indices feature, spectral feature, and spatial texture feature, and then combined an object-oriented classifier and SVM for the classification. The results obtained with two air-borne hyperspectral images confirmed that this method can achieve an excellent accuracy while maintaining the ground object boundary information.

However, UAV-borne hyperspectral imagery has abundant spatial information, and its spatial resolution can reach the centimeter level, but the very high spatial resolution causes severe spectral variability and spatial heterogeneity. Therefore, there are still some challenges when applying the previous crop classification strategies to UAV-borne H^2 imagery. Firstly, as illustrated in Fig. 1b, the spectral information of

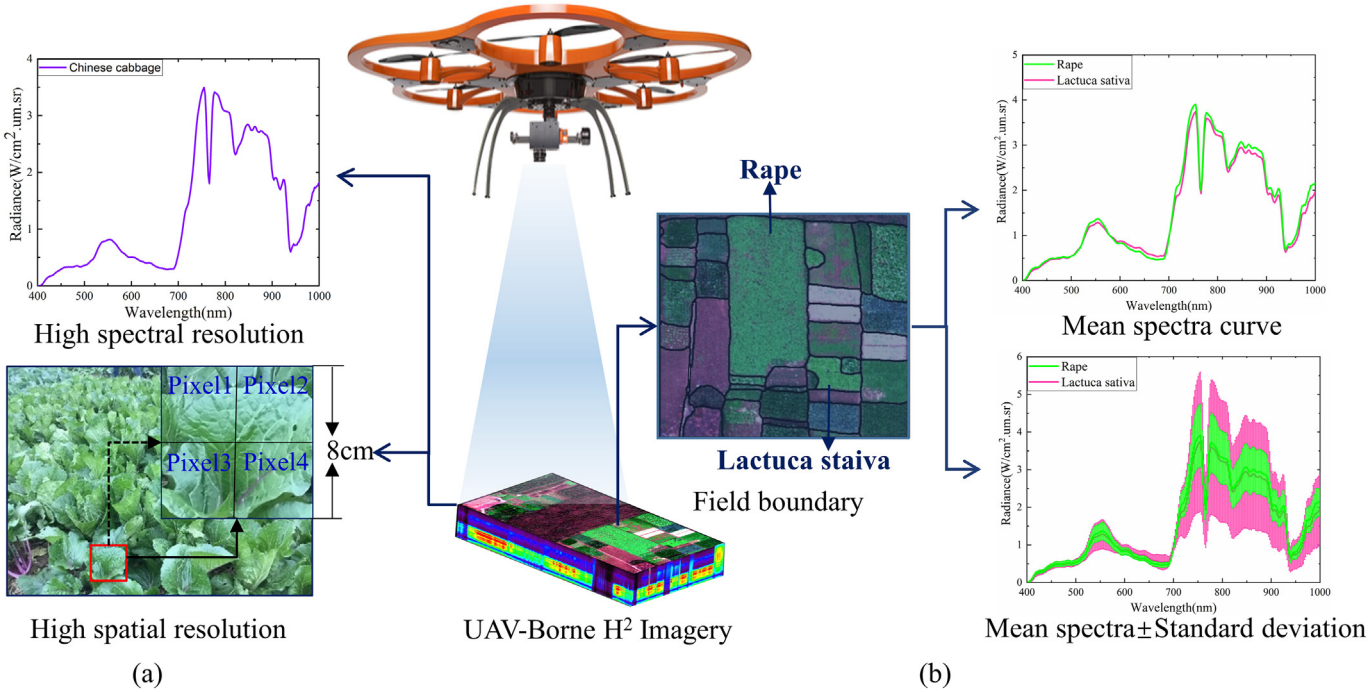

Fig. 1. UAV-borne H^2 imagery with a high spectral resolution (nm level) and a very high spatial resolution (cm level). (a) UAV-borne H^2 imagery (4-cm spatial resolution and 6-nm spectral resolution). (b) The intraclass pixels show serious spectral variability, and the interclass pixels show the spectral similarity of *Rape* and *Lactuca sativa*.

different crop classes can be similar, and there can also be a big difference in the spectra of the same crop class, which makes accurate separation of crop types difficult. However, the previous studies have been reliant on hand-crafted features based on domain knowledge and the experience of experts, and the hand-crafted features can be regarded as shallow features, which limits the application potential for the precise classification of crops based on complex UAV-borne H^2 imagery. Secondly, the plot sizes of crops can also vary in size in smallholder agriculture regions, which makes it a challenging task to select an optimal scale for the object-oriented classification methods.

Deep learning is now widely used in image processing due to its powerful feature learning capabilities. At present, the most popular deep learning based network architecture is the convolutional neural network (CNN). CNNs have the characteristics of sparse interaction, parameter sharing, and equivariant mapping, which can reduce the complexity and training parameter size of the network. These characteristics also allow the model to create some degree of invariance to shifting, distortion, and scaling, while also having strong robustness and fault tolerance (LeCun et al., 1998). As a result, CNNs have been widely used in hyperspectral imagery classification (Chen et al., 2016; Li et al., 2019; Yu et al., 2017; Zhao and Du, 2016). Although deep learning based classifiers have shown great potential in the field of hyperspectral image classification, we are still faced with several challenges when processing H^2 imagery with CNNs. Firstly, in UAV-borne H^2 imagery, the intraclass pixels can show serious spectral variability, while the interclass pixels can show spectral similarity. As a result, the CNNs, which only explore local spatial information, can easily fall into local optima, resulting in holes or isolated regions in the classification maps. In addition, the training samples in H^2 imagery are limited due to the high cost of field surveys, which makes it difficult to train a deep CNN classifier with massive network parameters.

In this article, to address the above-mentioned problems, the deep convolutional neural network with a conditional random field classifier (CNNCRF) framework is proposed to precisely identify crops and estimate their spatial distribution in UAV-borne H^2 imagery. Specifically, a deep CNN is designed to extract the in-depth spectral and spatial features within a local receptive field to preserve the richness of the land-cover details, and a Mahalanobis distance boundary constrained CRF model is proposed to further incorporate the spatial-contextual information and reduce the isolated regions in the classification maps. In addition, a virtual sample augmentation strategy based on the hyperspectral imaging mechanism is introduced to avoid overfitting of the deep CNN model and improve the robustness of the classification results. In addition to the proposed CNNCRF framework, the authors have also built the WHU-Hi dataset, which is an open-source UAV-borne H^2 dataset for precise crop classification. The WHU-Hi dataset consists of three individual UAV-borne H^2 datasets—WHU-Hi-LongKou, WHU-Hi-HanChuan, and WHU-Hi-HongHu—which were collected in Hubei province, China. The experimental results obtained with the WHU-Hi dataset demonstrate that the CNNCRF framework is an excellent method for precise crop classification based on UAV-borne H^2 imagery.

The rest of this article is composed of four sections. Section 2 describes the study regions and the WHU-Hi dataset. Section 3 provides the details of the proposed CNNCRF classification framework. The experimental results and an analysis are presented in Section 4, followed by a discussion in Section 5. Finally, the conclusions and future work are summarized in Section 6.

2. Materials

2.1. Study regions

In this article, as shown in Fig. 2, three experimental regions were selected in Hubei province in southern China, in the cities of Hanchuan and Honghu. Land fragmentation is common in the three study regions, where crops are planted on a small scale, which is representative of China's smallholder agriculture (Shu-hao et al., 2003; Xiao et al., 2013). A UAV hyperspectral observation platform was used to collect the H^2 imagery data. The hyperspectral sensor used was a Headwall Nano-Hyperspec imaging sensor, the specific parameters of which are listed in Table 1. A field survey in conjunction with global positioning system (GPS) technology was used as the reference for the crop planting distribution in the study regions.

2.2. WHU-Hi dataset

In this study, a new benchmark dataset, which is named the Wuhan UAV-borne hyperspectral image (WHU-Hi) dataset, was built for precise crop classification. The WHU-Hi dataset contains three individual UAV-borne hyperspectral datasets: WHU-Hi-LongKou, WHU-Hi-HanChuan, and WHU-Hi-HongHu. All the datasets were acquired in farming areas with various crop types in Hubei province, China, via a Headwall Nano-Hyperspec sensor mounted on a UAV platform.

The UAV hyperspectral data preprocessing included radiometric calibration and geometric correction, which were undertaken in the HyperSpec software provided by the instrument manufacturer. For the radiometric calibration, the raw digital number values were converted into radiance values by the laboratory calibration parameters of the sensor, which can be written as:

$$DN = (L_1 G_1 + L_2 G_2) \times ET + DF \quad (1)$$

where L_1 is the desired radiance of the first-order diffraction in radiometric units, $L_2$is the desired radiance of the second-order diffraction in radiometric units, $G_1$and $G_2$are the system gain of L_1 and L_2, ET is the exposure time of the sensor, and DF is the dark field measurement. The L_2 intensity is very weak and can be filtered through the use of a spectral filter. Under normal circumstances, the calibration parameters are measured by the manufacturer in a laboratory environment using an integrated sphere, and are recorded in the calibration software.

The geometric correction was undertaken based on a collinear equation and the position and attitude information recorded by the GPS and inertial measurement unit (IMU) modules. For the HyperSpec software, the input data are the raw hyperspectral data, the frame index, the timestamp, the digital surface model (DSM) data, and the GNSS/IMU files with latitude, longitude, latitude, roll, pitch, and yaw. Furthermore, ground control points (GCPs) can also be used to further promote the geometric correction accuracy.

2.2.1. WHU-Hi-LongKou dataset

The WHU-Hi-LongKou dataset was acquired from 13:49 to 14:37 on July 17, 2018, in Longkou Town, Hubei province, China, with an 8-mm focal length Headwall Nano-Hyperspec imaging sensor equipped on a DJI Matrice 600 Pro (DJI M600 Pro) UAV platform. The DJI M600 Pro UAV possesses a 6-kg maximum payload capacity, it has a flight time of approximately 30 min, and is equipped with GPS/IMU modules for centimeter-level navigation. During the data collection, the weather was clear and cloudless, the temperature was about 36 °C, and the relative air humidity was about 65%. The study area contains six crop species: corn, cotton, sesame, broad-leaf soybean, narrow-leaf soybean, and rice. The UAV flew at an altitude of 500 m, the size of the imagery is 550 × 400 pixels, there are 270 bands from 400 to 1000 nm, and the spatial resolution of the UAV-borne hyperspectral imagery is about 0.463 m. An overview of this dataset is provided in Fig. 3. The spectral means of the training samples of the nine classes of ground objects extracted from the WHU-Hi-LongKou dataset are presented in Fig. 4.

2.2.2. WHU-Hi-HanChuan dataset

The WHU-Hi-HanChuan dataset was acquired from 17:57 to 18:46 on June 17, 2016, in Hanchuan, Hubei province, China, with an 8-mm focal length Headwall Nano-Hyperspec imaging sensor equipped on a Leica Aibot X6 UAV V1 platform. The Leica Aibot X6 UAV V1 platform possesses a 2-kg maximum payload capacity, it has a 30-min flight time,

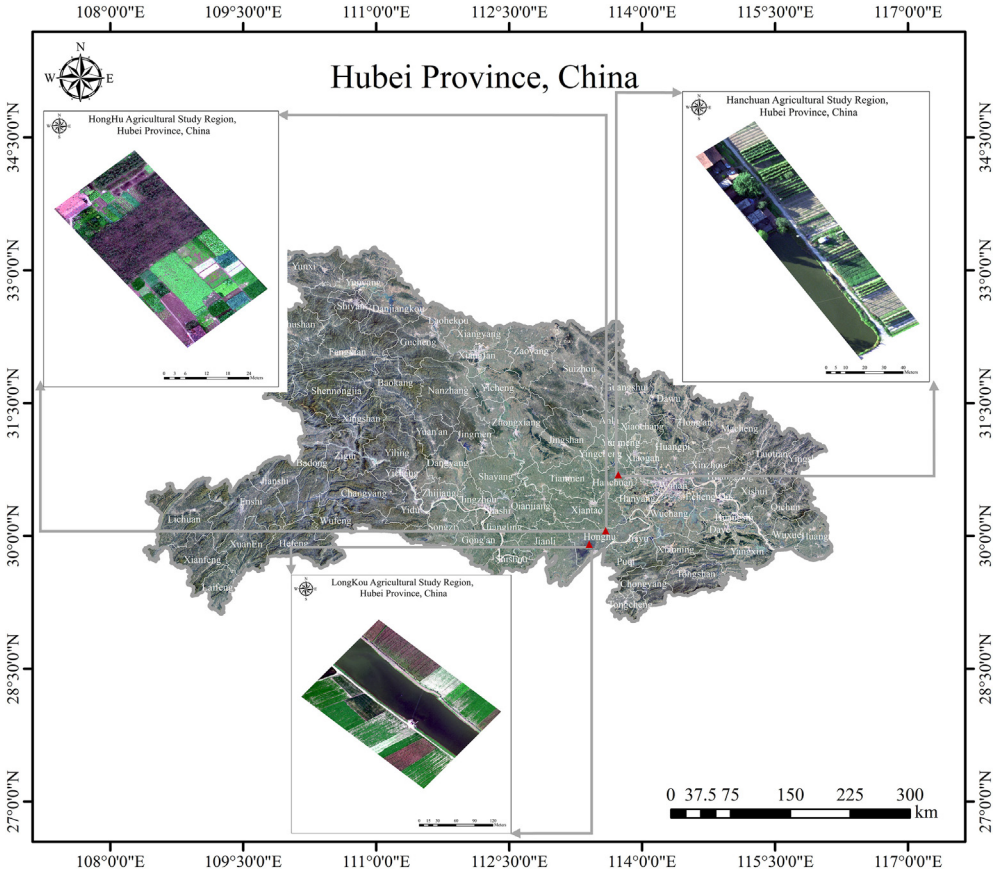

Fig. 2. Study regions.

Table 1
The parameters of the Headwall Nano-Hyperspec sensor.

Wavelength range:	400–1000 nm	Camera technology:	CMOS
Spatial bands:	640	Bit depth:	12-bit
Spectral bands:	270	Detector pixel pitch:	7.4 μm
Dispersion/pixel (nm/pixel):	2.2	Weight without lens, GPS:	0.5 kg
FWHM slit image:	6 nm	Focal length:	4.8 mm, 8 mm, 12 mm, 17 mm, 24 mm, 25 mm, 37 mm, 70 mm
Integrated 2nd order filter:	Yes		
Entrance slit width:	20 μm		

and is equipped with GPS/IMU modules for navigation. During the data collection, the weather was clear and cloudless, the temperature was about 30 °C, and the relative air humidity was about 70%. The study area contains seven crop species: strawberry, cowpea, soybean, sorghum, water spinach, watermelon, and greens. The UAV flew at an altitude of 250 m, the size of the imagery is 1217 × 303 pixels, there are 274 bands from 400 to 1000 nm, and the spatial resolution of the UAV-borne hyperspectral imagery is about 0.109 m. Notably, since the WHU-Hi-HanChuan dataset was acquired during the afternoon when the solar elevation angle was low, there are many shadow-covered areas in the image. An overview of this dataset is given in Fig. 5. The spectral means of the training samples of the 16 classes of ground objects extracted from the WHU-Hi-HanChuan dataset are presented in Fig. 6.

2.2.3. WHU-Hi-HongHu dataset

The WHU-Hi-HongHu dataset was acquired from 16:23 to 17:37 on November 20, 2017, in Honghu City, Hubei province, China, with a 17-mm focal length Headwall Nano-Hyperspec imaging sensor equipped on a DJI Matrice 600 Pro UAV platform. During the data collection, the weather was cloudy, the temperature was about 8 °C, and the relative air humidity was about 55%. The study area is typical of the regions affected by land fragmentation, and is planted with 17 crop types, including cotton, rape, and cabbage. Notably, the region is planted with different cultivars of the same crop type; for example, Chinese cabbage/cabbage and *Brassica chinensis*/small *Brassica chinensis*. The UAV flew at an altitude of 100 m, the size of the imagery is 940 × 475 pixels, there are 270 bands from 400 to 1000 nm, and the spatial resolution of the UAV-borne hyperspectral imagery is about 0.043 m. An overview of this dataset is provided in Fig. 7. The spectral means of the training samples of the 22 classes of ground objects extracted from the WHU-Hi-HongHu dataset are presented in Fig. 8.

3. The CNNCRF framework for precise crop classification

In this section, the CNNCRF framework for precise crop classification with UAV-borne H^2 imagery is described. As shown in Fig. 9, the acquired three-dimensional (3D) patch of the labeled pixel neighborhood is enhanced based on the hyperspectral imaging procedure. A deep benchmark CNN is then used to extract and fuse the in-depth spectral and local spatial features of the crops from the input 3D patch, and the output probability image is used as the input of the unary potential term in the CRF model. The pairwise potential term of the CRF model with the Mahalanobis distance boundary constraint is then applied to model the spatial-contextual relationship between the pixel and its neighborhood pixels, which encourages the adjacent pixels to be assigned the same label. In summary, this approach can filter the holes and isolated regions in the classification results while preserving the detailed boundary information of the land cover, and ultimately improves the classification accuracy.

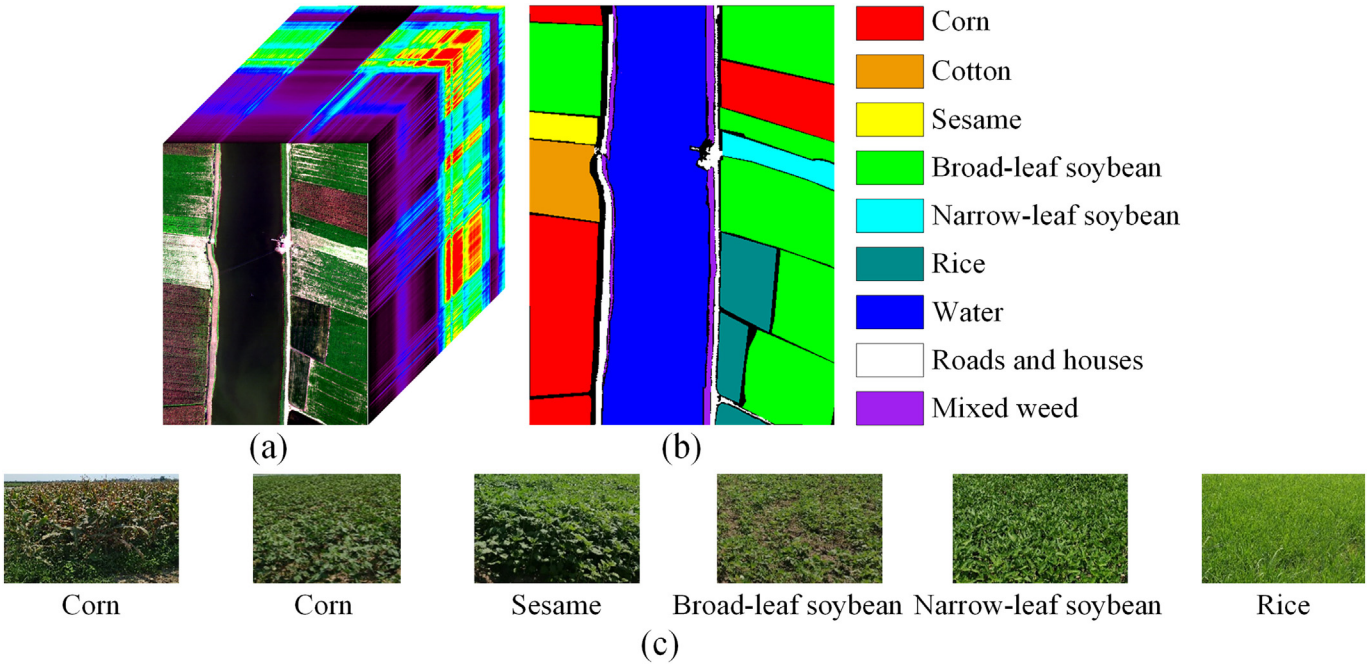

Fig. 3. The WHU-Hi-LongKou dataset. (a) Image cube. (b) Ground-truth image. (c) Typical crop photos in the study area.

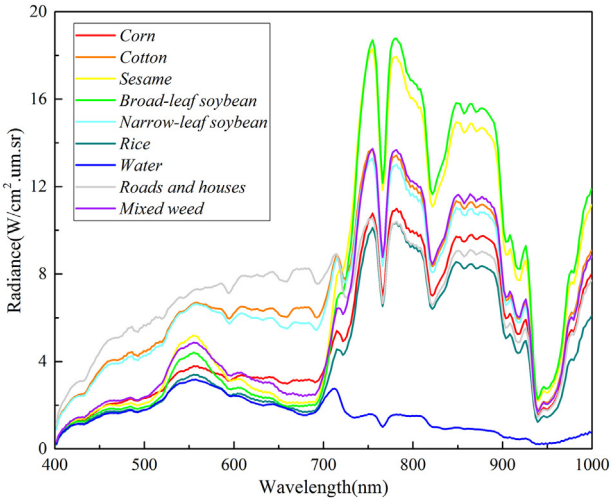

Fig. 4. Spectra of the WHU-Hi-LongKou dataset land-cover classes.

3.1. The deep benchmark CNN for precise crop classification

The deep benchmark CNN architecture extracts and fuses the in-depth spectral and local spatial features by convolutional layers and fully connected (FC) layers. To prevent the feature map size dropping too fast, we do not use a pooling layer when constructing the deep network. The input data are the 3D patch P selected centered on the labeled pixel and its corresponding label. As shown in Fig. 10, the deep benchmark CNN is made up of six layers, with four convolutional layers designed to extract the in-depth spectral and local spatial features, and two FC layers constructed to integrate the in-depth features. Finally, the output high-level semantic features of the FC layer are classified by a softmax classifier.

In the deep CNN, the core building block is the convolutional layer, which applies a convolution operation to the input data, with convolutional kernels to extract features. For the l-th convolutional layer, the i-th output feature map y_i^l can be expressed as:

$$y_i^l = f\left(\sum_j w_{i,j}^l * y_j^{l-1} + b_i^l\right) \tag{2}$$

where y_j^{l-1} represents the j-th output feature map in the $(l-1)$-th convolutional layer, $w_{i,j}^l$ is the j-th channel in the i-th convolutional kernel of the l-th convolutional layer, and b_i^l is the bias corresponding to the i-th convolutional kernel. $f(\cdot)$ represents the nonlinear activation function, for which rectified linear units (ReLUs) are commonly used, and * denotes the convolutional operator. The convolutional kernel slides on the input feature map to achieve weight sharing, which dramatically reduces the learning parameters and calculation amount. The FC layer is usually placed after the convolutional layer, and it constructs the connection between each output layer and the input layer neurons to achieve the transition from local features to global features. Therefore, the number of network parameters in the FC layer is generally the largest.

Softmax regression is a common multi-category classifier in CNNs, which is mainly used to output the probability of each pixel belonging to the different classes. For an n-class classification task, the softmax classifier outputs an n-dimensional vector $Y \in R^n$, and the value $p(Y_i = j | X_i, \theta)$ at the $j-th$ dimension indicates the probability of input X_i belonging to class j. The formula can be expressed as:

$$h_\theta(X_i) = \begin{bmatrix} p(Y_i = 1 \mid X_i, \theta) \\ p(Y_i = 2 \mid X_i, \theta) \\ \vdots \\ p(Y_i = n \mid X_i, \theta) \end{bmatrix} = \frac{1}{\sum_{j=1}^{n} e^{\theta_j^T X_i}} \begin{bmatrix} e^{\theta_1^T X_i} \\ e^{\theta_2^T X_i} \\ \vdots \\ e^{\theta_n^T X_i} \end{bmatrix} \tag{3}$$

where θ represent the model parameters, and $\sum_{j=1}^{n} p(Y_i = j \mid X_i, \theta) = 1$. The loss function $J(\theta)$ of the softmax classifier is defined as follows:

$$J(\theta) = -\frac{1}{m}\left[\sum_{i=1}^{m}\sum_{j=1}^{n} I\{Y_i = j\} \log \frac{e^{\theta_j^T X_i}}{\sum_{j=1}^{n} e^{\theta_j^T X_i}}\right] \tag{4}$$

$$I\{Y_i = j\} = \begin{cases} 0, & if\ Y_i \neq j \\ 1, & if\ Y_i = j \end{cases} \tag{5}$$

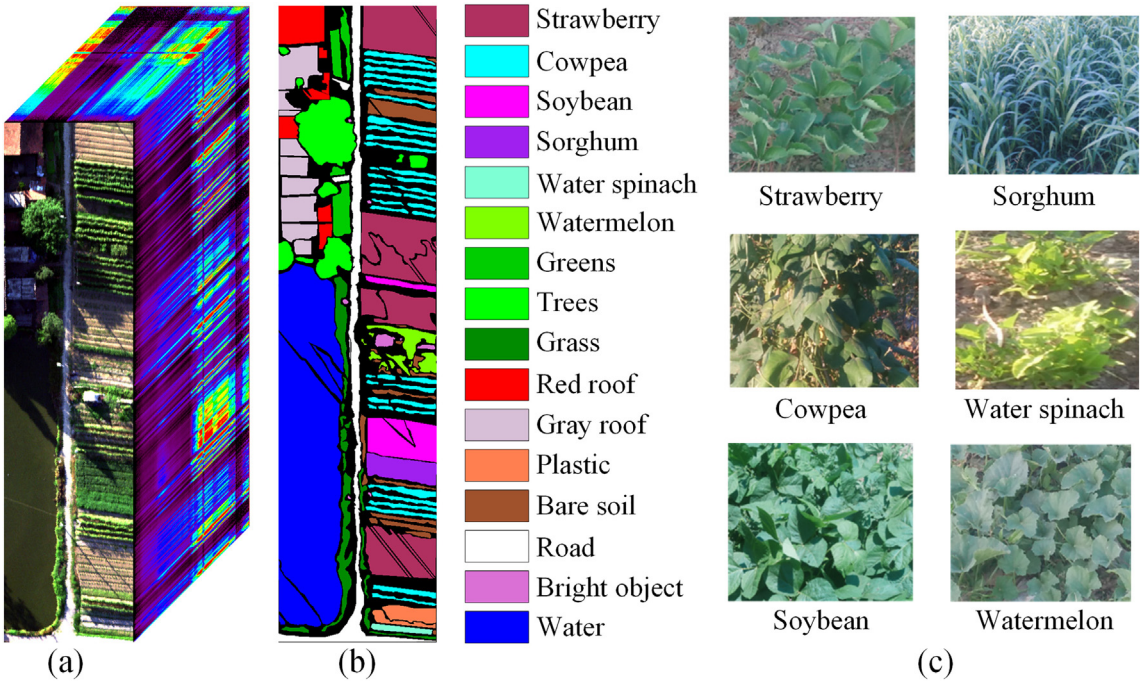

Fig. 5. The WHU-Hi-HanChuan dataset. (a) Image cube. (b) Ground-truth image. (c) Typical crop photos in the study area.

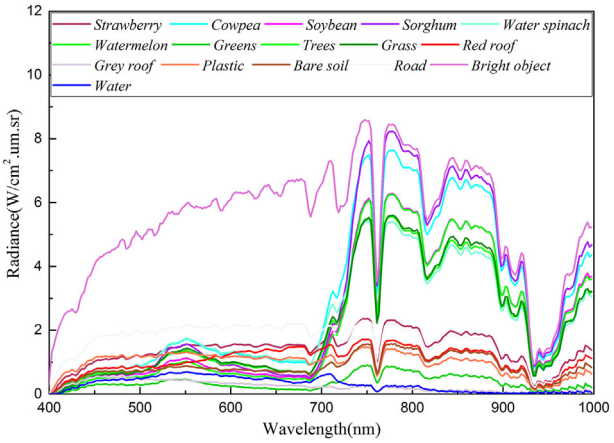

Fig. 6. Spectra of the WHU-Hi-HanChuan dataset land-cover classes.

where m represents the batch size and $I\{Y_i = j\}$ is an indicator function. If predicted label Y_i is equal to the actual label$j$, then$I\{Y_i = j\} = 1$; otherwise, $I\{Y_i = j\} = 0$.

3.2. The virtual sample augmentation strategy based on the hyperspectral imaging mechanism

A deep CNN requires abundant training samples to learn the massive parameters, but labeling the crop types of a hyperspectral image is a very expensive task, which is usually achieved by field survey. Accordingly, four hyperspectral data augmentation methods are used to alleviate the problem of the limited training samples: 1) flip and rotate based virtual samples; 2) mixture-based virtual samples; 3) changing radiation based virtual samples; and 4) denoising and adding noise based virtual samples. The flip and rotate based virtual samples were proposed by Ciregan et al. (2012). The mixture-based virtual samples and changing radiation based virtual samples were proposed by Chen et al. (2016), based on the imaging procedure perspective. The denoising and adding noise based virtual samples proposed in this article simulate the noise phenomena in hyperspectral imaging. In the following formulas, P represents the selected 3D patch centered on the labeled pixel of the training set.

The flip and rotate based virtual samples include up-down flip, left-right flip, and randomly rotating from -180° to 180° on the 3D patch spatial dimension.

The mixture-based virtual samples were inspired by the mixed pixel phenomenon in remote sensing. The virtual sample P_{ij}can be generated from two labeled samples $P_i$and P_j of the same class:

$$P_{ij} = \alpha_i P_i + \alpha_j P_j + \lambda n \quad s.\,t. \quad 0 < \alpha_i, \alpha_j < 1, \ \ \alpha_i + \alpha_j = 1 \tag{6}$$

where λ is the weight of the random Gaussian noise$n$, which simulates the error caused by the mixed pixel abundance.

The same class in different locations during data acquisition can be affected by different illumination conditions, resulting in differences in radiance. As a result, the changing radiation based virtual sample $P_r$is obtained by multiplying the light intensity factor α_m by the actual samplesP_m, and a random Gaussian noise term λn is then added to consider the errors from atmospheric changes.

$$P_r = \alpha_m P_m + \lambda n \quad s.\,t. 0.9 < \alpha_m < 1.1 \tag{7}$$

Because of the influence of weather, ground roughness, sensor noise, and other factors, the signals received by a single pixel will exhibit varying degrees of noise. However, the data augmentation method that directly adds noise to the original data further aggravates the noise interference. Therefore, we propose a data augmentation method of denoising and adding noise based virtual samples. Firstly, the spectral dimension of the original data is denoised by wavelet denoising to eliminate the interference of the other factors in the imaging process. Then, to increase the robustness of the data, random Gaussian noise is added to the denoised data:

$$P_d = f(P_n) + \lambda n \tag{8}$$

whereP_n represents the training samples; $f(\cdot)$is a wavelet denoising function, where the basis function selected is the discrete Meyer ("Dmey") wavelet; and λn represents the added Gaussian noise.

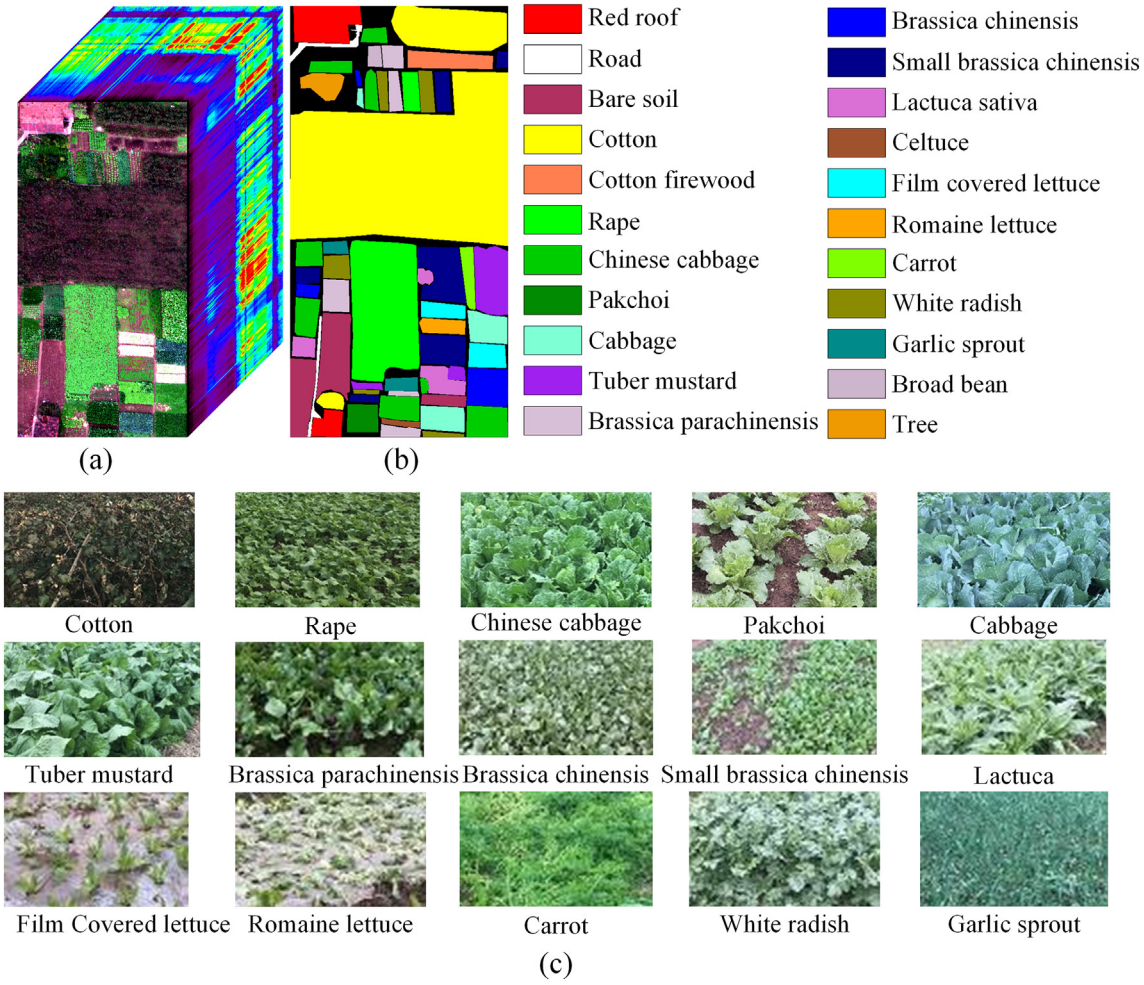

Fig. 7. The WHU-Hi-HongHu dataset. (a) Image cube. (b) Ground-truth image. (c) Typical crop photos in the study area.

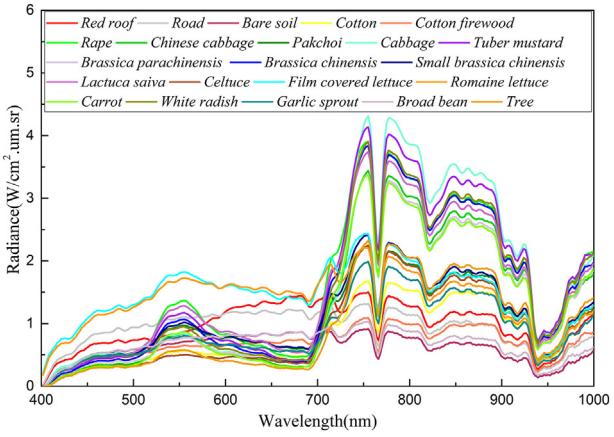

Fig. 8. Spectra of the WHU-Hi-HongHu dataset land-cover classes.

3.3. Incorporation of the spatial-contextual information based on the CRF model

UAV-borne H^2 imagery of a very high spatial resolution causes severe spatial heterogeneity and spectral variability. However, a deep CNN can only use the local spatial information, and it can easily fall into local optima during the model training, which can lead to holes or isolated regions in the classification maps. In this article, to address these problems of the deep CNNs, the CRF model is proposed to reduce the holes and isolated regions in the classification maps by incorporating the spatial-contextual information through modeling the spatial interaction among the image pixels.

CRF is a classical probabilistic discriminative model that was proposed by Lafferty (Lafferty et al., 2001), and is the statistical model that was initially applied to labeling serialized text data. In the two-dimensional data processing, to incorporate the spatial-contextual information, the CRF model considers the spatial interaction among image pixels by the edge connection between adjacent pixels in the probability graph model (Kumar, 2003). Based on its flexibility in context information modeling, the CRF model has been widely applied in remote sensing image processing, including optical image classification (Zhao et al., 2018; Zhao et al., 2015; Zhong et al., 2014) and change detection (Lv et al., 2016; Zhou et al., 2016). The most common CRF classification model is the pairwise CRF model, which considers the spatial interaction of the local neighborhoods based on the unary and pairwise potentials. The pairwise CRF model has been applied in high spatial resolution image classification (Zhao et al., 2015; Zhong et al., 2014) and air-borne hyperspectral image classification (Zhao et al., 2018; Zhong and Wang, 2011).

For hyperspectral remote sensing image classification, the CRF model directly models the potential energy function of the discriminative classifier to maximize the posterior probability of the classification label. For the observed data from a 3D hyperspectral image $Y = \{y_1, y_2, \cdots y_i \cdots y_m\}_{i \in S}$, where y_i is a spectral vector of pixel i in the input hyperspectral image, site$i \in S$, $S = \{1, 2, \cdots, m\}$ is the pixel index of the input data, and m denotes the number of pixels in the input data. The labeled data $X = \{x_1, x_2, \cdots x_i \cdots x_m\}_{i \in S}$ are the land-cover

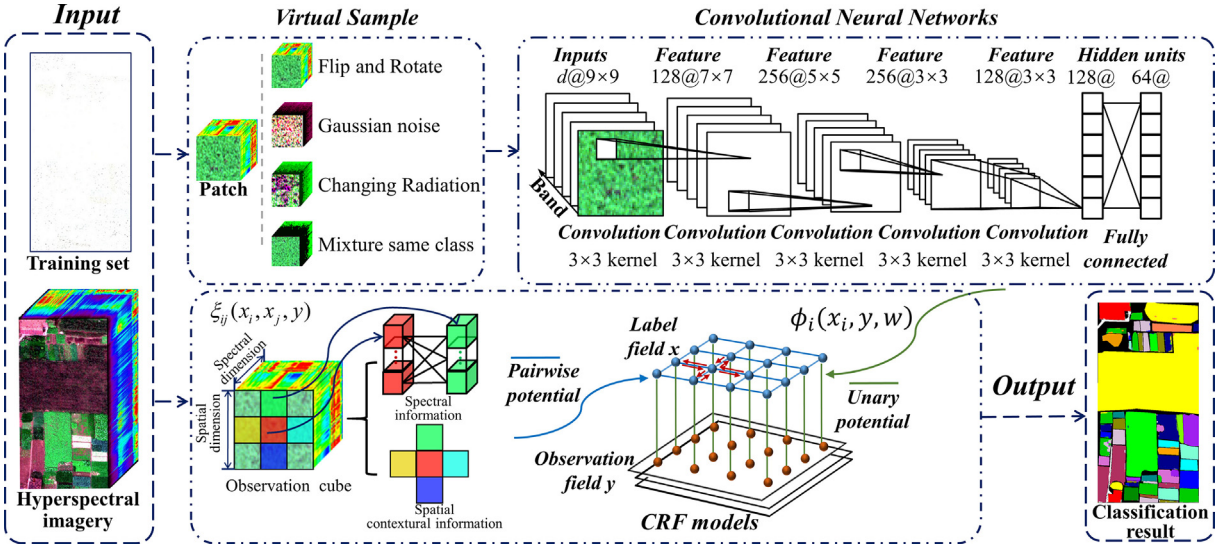

Fig. 9. Flowchart of the CNNCRF classification framework.

classes, where $x_i \in \{1, 2, \cdots, L\}$ in site$i \in \{1, 2, \cdots, m\}$, and L indicates the number of land-cover classes. For hyperspectral image classification, the pairwise CRF model can be represented as shown in Eq. (9):

$$P(x \mid y) = \frac{1}{Z(y)} \exp\left\{\sum_{i \in S} \psi_i(x_i, y_i) + \sum_{i \in S} \sum_{j \in \eta_i} \xi_{ij}\left(x_i, x_j, y\right)\right\} \tag{9}$$

where $Z(y)$ is the normalized function to constrain the sum of the class probabilities to 1, and η_i represents the neighborhood pixel index of pixel i. In CNNCRF, the unary potential energy function $\psi_i(x_i, y_i)$ models the relationship between pixel label x_i and spectral vector y_i. The pairwise potential energy function $\xi_{ij}(x_i, x_j, y)$ can model the spatial-contextual interaction relationship of pixel index i and its neighborhood pixel index η_i, which allows us to utilize the spatial-contextual information. As a result, the pairwise potential energy function acts as a smoothing prior and considers the constraint of the pixel class labels, to guarantee that the neighboring pixels with the same class in the homogeneous regions are classified into the same label, while preserving the boundaries of the ground objects, thus improving the smoothing and correcting misclassified pixels. In CNNCRF, η_i is set as a four-connected neighborhood. However, differing from the patch-based deep CNNs that only use the local spatial information, the CRF model has global spatial-contextual information integration capabilities, as it delivers the interaction information of neighboring pixels onto the entire image. A schematic drawing of the CRF model delivering the interaction information is shown in Fig. 11. The unary and pairwise potential energy functions of the CRF model are shown in Eq. (10) and Eq. (11).

In Eq. (10), the unary potential energy function $\psi_i(x_i, y_i)$ is the cost of pixel y_i obtaining the label l_k, based on the class probabilities P

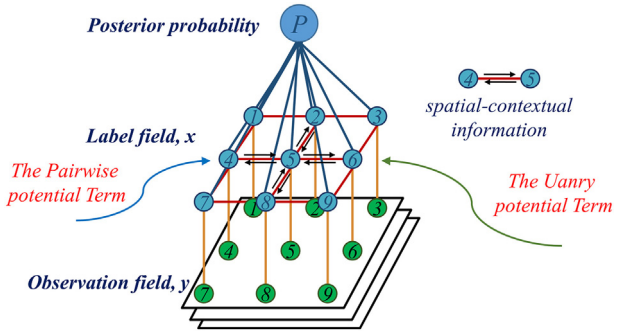

Fig. 11. Schematic drawing of the CRF model delivering the interaction information.

$$\psi_i(x_i, y_i) = P(x_i = l_k \mid y_i) \tag{10}$$

$$\xi_{i,j}(x_i, x_j, y) = \begin{cases} 1 & x_i = x_j \\ 1 - dist(i,j)^{-1} \exp(-\beta \|y_i - y_j\|) & x_i \neq x_j \end{cases} \tag{11}$$

$(x_i = l_k)$. For the CNNCRF framework, the class probability P is obtained by the benchmark CNN method. For the pairwise potential energy function $\xi_{i,\ j}(x_i, x_j, y)$ in Eq. (11), (i,j) is the neighborhood pixel pair, $y_i$and$y_j$represent the spectral vectors at positions$i$and$j$, and$dist(y_i, y_j)$ denotes the scalar distance between neighborhood pixel spectral vectors $y_i$andy_j. In this study, the Mahalanobis distance with the Pearson product-moment correlation coefficient (PPMCC) was used as the measure of similarity in the boundary constraint model, which adaptively adjusts the measurement criteria by considering the correlation

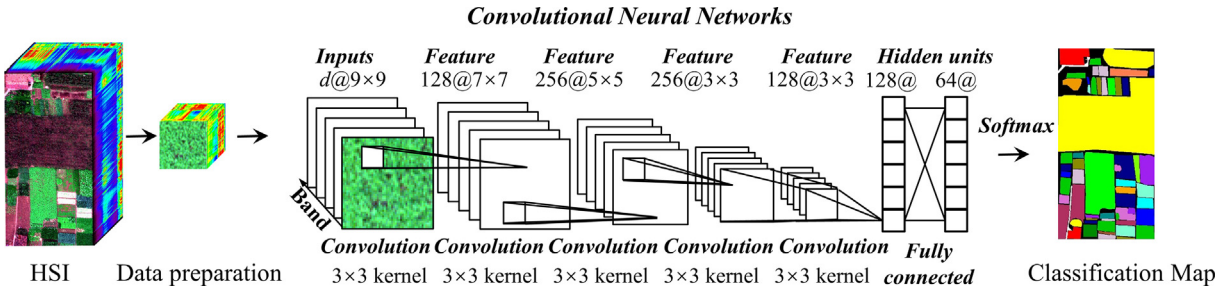

Fig. 10. The deep CNN network architecture, where d represents the number of channels, "@9 × 9" represents the size of the feature map, and "3 × 3 kernel" indicates that the size of the convolutional kernel is 3 × 3.

of the input data based on a multivariate normal distribution probability function (Kato et al., 1999). Therefore, the representation of $dist(y_i, y_j)$ is as follows:

$$d(y_i, y_j) = \frac{D_M(y_i, y_j)}{|Corr(y_i, y_j)|} \tag{12}$$

$$D_M(y_i, y_j) = \sqrt{(y_i - y_j)^T M (y_i - y_j)} \tag{13}$$

$$Corr(y_i, y_j) = \frac{\sum_{u=v=1}^{d} (y_{iu} - \overline{y_i})(y_{ju} - \overline{y_j})}{\sqrt{\sum_{u=1}^{d} (y_{iu} - \overline{y_i})^2} \sqrt{\sum_{v=1}^{d} (y_{jv} - \overline{y_j})^2}} \tag{14}$$

where $D_M(y_i, y_j)$, $Corr(y_i, y_j)$, and Σare the Mahalanobis distance, the PPMCC, and the covariance matrix between neighborhood pixel spectral vectors $y_i$andy_j, respectively. M denotes the Mahalanobis matrix, and $\overline{y_i}$ and $\overline{y_j}$ denote the average values of y_i and y_j, respectively. Parameter β is set to the average value of the reciprocal of twice the variance of pixel spectral vector y_i with regard to all the neighboring pixel spectral vectors y_{η_i}, expressed as$\beta = (2\langle \|y_i - y_{\eta_i}\|^2 \rangle)^{-1}$, where $\langle\cdot\rangle$ represents an averaging operation. Therefore, the intensity of the spatial interaction within the neighborhood is related to the imagery, supporting class consistency in similar spectral regions. If pixel spectral vectors y_i and y_j are similar, then $dist(i,j)^{-1} \exp(-\beta \|y_i - y_j\|)$ will be close to one, and $1 - dist(i,j)^{-1} \exp(-\beta \|y_i - y_j\|)$ will be close to zero. As a result, the pixel i and pixel j category similarity is very high. Conversely, if pixel spectral vectors y_i andy_j differ greatly, the value of $1 - dist(i,j)^{-1} \exp(-\beta \|y_i - y_j\|)$ will increase.

4. Experiments and analysis

4.1. Experimental parameter settings

4.1.1. The CNN model parameter settings

The CNN of the CNNCRF model is made up of four convolutional layers and two FC layers, as shown in Table 2. In the experiments, for the convolutional layers, the stride value was set to one, the activation function was ReLU, and we added batch normalization (BN) for faster convergence. Finally, all the neurons connected with FC layers were followed by a softmax classifier to produce the class probability vector of each pixel. The training epochs were 150, 200, and 200 for the WHU-Hi-LongKou, WHU-Hi-HanChuan, and WHU-Hi-HongHu datasets, respectively.

In the experiments, the patch sizes of the three datasets were set as $9 \times 9 \times d$, where d denotes the band number of the remote sensing image. The patch size was obtained empirically, so that each patch could basically cover a single crop. The data augmentation was conducted 20 times, which was also the empirical value obtained by considering the computing resources and accuracy.

4.1.2. Training sample settings

For each class, 100 labeled pixels were randomly selected for the model training, and the remaining pixels were used for the testing. To be more specific, the total number of training pixels was only 0.44%, 0.62%, and 0.57% of all the labeled pixels for the WHU-Hi-LongKou, WHU-Hi-HanChuan, and WHU-Hi-HongHu datasets, respectively. More class information details are provided in Tables 3–5.

4.2. Experimental results and analyses

4.2.1. Comparison methods

Four comparison methods were utilized in this study. The first method was a spectral-based pixel-wise SVM classification algorithm using the LIBSVM package (Chang and Lin, 2011). The kernel of SVM was the radial basis function (RBF), and penalty factor C was optimized by five-fold cross-validation. Simultaneously, to introduce the spatial information from the hyperspectral imagery, we also compared object-oriented and CRF-based classification methods. The object-oriented classification (OOC) method was the multi-resolution segmentation algorithm (Baatz and Schäpe, 2000) incorporated in eCognition 8.0 (the fractal net evolution approach (FNEA)). The OOC results were obtained by performing a majority vote on each object, based on the classification map of pixel-by-pixel SVM. In order to determine the optimal segmentation scale, the experimental data were segmented at a scale factor from 3 to 15, and the optimal classification result was selected as the comparative experimental result. This OOC method was named FNEA-OO. The CRF-based method was the support vector conditional random fields classifier with a Mahalanobis distance boundary constraint (SVRFMC) (Zhong et al., 2014). Finally, to prove the advantages of the CNNCRF classification framework in the use of spatial-contextual information, the benchmark CNN of the CNNCRF was also used for comparison. To quantitatively evaluate the experimental results, four evaluation indicators are used: the accuracy of each class, the overall accuracy (OA), the average accuracy (AA), and the kappa coefficient (kappa).

4.2.2. Results for the WHU-Hi-LongKou dataset

The visual representation of the imagery classification has always been important for classification methods. Fig. 12c–f show the qualitative classification results of the classification methods of CNNCRF and the comparison methods (i.e., SVM, FNEA-OO, SVRFMC, and the benchmark CNN) for the WHU-Hi-LongKou dataset. As shown in Fig. 12c, because of the similarity of the spectral information among the different crops, the classification results of SVM using only the spectral information show noticeable salt-and-pepper (SP) noise, as well as misclassification. For example, the spectral curves of sesame and broad-leaf soybean are similar, resulting in a large number of cotton pixels being misclassified as sesame in the classification map. Moreover, on the edge of the plot, part of the area is misclassified as mixed weeds and roads, due to the sparse nature of the crops. As displayed in Fig. 12d–f, compared with the pixel-wise SVM method using only spectral information, the classification methods considering the spatial neighborhood information (FNEA-OO, benchmark CNN, and CNNCRF) show a better visual performance. Therefore, spatial information should be considered in the classification process, to lessen the SP noise and improve the classification accuracy. Although FNEA-OO delivers a smooth classification map, it aggravates the misclassification between roads/

Table 2
The architecture of the CNN.

No.	Input	Output	Convolutional kernel		ReLU	BN	Stride
			Number	Size			
Conv1	$9 \times 9 \times d$	$7 \times 7 \times 128$	128	3×3	Yes	Yes	1
Conv2	$7 \times 7 \times 128$	$5 \times 5 \times 256$	256	3×3	Yes	Yes	1
Conv3	$5 \times 5 \times 256$	$3 \times 3 \times 256$	256	3×3	Yes	Yes	1
Conv4	$3 \times 3 \times 256$	$1 \times 1 \times 128$	128	3×3	Yes	Yes	1
FC5	$1 \times 1 \times 128$	$1 \times 1 \times 128$	128		Yes	Yes	
FC6	$1 \times 1 \times 128$	$1 \times 1 \times 64$	64		Yes	Yes	

Table 3
Class information for the WHU-Hi-LongKou dataset.

No.	Class name	Training samples	Verification samples	No.	Class name	Training samples	Verification samples
C1	Corn	100	34,411	C6	Rice	100	11,754
C2	Cotton	100	8274	C7	Water	100	66,956
C3	Sesame	100	2931	C8	Roads and houses	100	7024
C4	Broad-leaf soybean	100	63,112	C9	Mixed weed	100	5129
C5	Narrow-leaf soybean	100	4051				

Table 4
Class information for the WHU-Hi-HanChuan dataset.

No.	Class name	Training samples	Verification samples	No.	Class name	Training samples	Verification samples
C1	Strawberry	100	44,635	C9	Grass	100	9369
C2	Cowpea	100	22,653	C10	Red roof	100	10,416
C3	Soybean	100	10,187	C11	Gray roof	100	16,811
C4	Sorghum	100	5253	C12	Plastic	100	3579
C5	Water spinach	100	1100	C13	Bare soil	100	9016
C6	Watermelon	100	4433	C14	Road	100	18,460
C7	Greens	100	5803	C15	Bright object	100	1036
C8	Trees	100	17,878	C16	Water	100	75,301

Table 5
Class information for the WHU-Hi-HongHu dataset.

No.	Class name	Training samples	Verification samples	No.	Class name	Training samples	Verification samples
C1	Red roof	100	13,941	C12	*Brassica chinensis*	100	8854
C2	Road	100	3412	C13	Small Brassica chinensis	100	22,407
C3	Bare soil	100	21,721	C14	*Lactuca sativa*	100	7256
C4	Cotton	100	163,185	C15	Celtuce	100	902
C5	Cotton firewood	100	6118	C16	Film covered lettuce	100	7162
C6	Rape	100	44,457	C17	Romaine lettuce	100	2910
C7	Chinese cabbage	100	24,003	C18	Carrot	100	3117
C8	Pakchoi	100	3954	C19	White radish	100	8612
C9	Cabbage	100	10,719	C20	Garlic sprout	100	3386
C10	Tuber mustard	100	12,294	C21	Broad bean	100	1228
C11	*Brassica parachinensis*	100	10,915	C22	Tree	100	3940

houses/mixed weeds, for which the optimal segmentation scale is much less than for the crops and water classes. The SVRFMC method takes full advantage of the spatial-contextual information, which significantly improves the experimental results, i.e., a higher classification accuracy and less SP noise. However, as the unary potential generated by the SVM classifier is not accurate, there are still some misclassifications between broad-leaf soybean/narrow-leaf soybean/cotton. The benchmark CNN can only use local spatial information, and it easily falls into local optima during training, resulting in isolated misclassified regions in the classification map. Meanwhile, CNNCRF takes full advantage of the spatial-contextual information, achieving excellent results, and there are very few examples of holes or isolated misclassified regions.

From another perspective, the quantitative accuracy evaluation also reflects the performance of the classification methods. The test samples are obtained by field survey, and usually do not include uncertain mixed pixels. The quantitative evaluation results for the effectiveness evaluation of the various classification methods are listed in Table 6, where similar results to the visual performance can be seen. The OAs of the FNEA-OO, SVRFMC, benchmark CNN, and CNNCRF classification methods show improvements of 4.53%, 4.31%, 3.24%, and 4.85% over the pixel-wise SVM classifier, respectively, which shows that spatial information plays an important role in hyperspectral image classification. The CNNCRF method achieves the highest classification accuracies (OA, AA, and kappa). Except for the broad-leaf soybeans class, the classification accuracies for the other crops are all over 99%. In summary, it can be demonstrated from the experimental results that the CNNCRF method can achieve a competitive classification accuracy for the WHU-Hi-LongKou dataset.

4.2.3. Results for the WHU-Hi-HanChuan dataset

Fig. 13c–f show the classification maps of the different classification methods (SVM, FNEA-OO, SVRFMC, benchmark CNN, and CNNCRF) for the WHU-Hi-HanChuan dataset, and Table 7 lists the corresponding quantitative evaluation results. As in Experiment 1, there is severe SP noise in the SVM classification results. Furthermore, as shown in Fig. 14, since the energy reflected by the shadow-covered ground objects is low, the spectral values are low and show severe variability. Therefore, there is more misclassification of the ground objects in the shadow-covered areas. As shown in Fig. 15, in the sub-image of the SVM classification results, it can be found that there is serious misclassification in the shadow-covered areas. For example, cowpea is misclassified as greens, and strawberry, road, and water are misclassified as the gray roof class. The FNEA-OO, SVRFMC, benchmark CNN, and CNNCRF classification methods can deliver smoother classification results by taking the spatial neighborhood information into account, providing improvements in OA of 8.02%, 8.92%, 9.52%, and 16.34%, respectively, compared with the pixel-wise SVM classifier. For the object-oriented methods, the classification accuracy is connected with the quality of the segmentation. However, due to the complexity of the scene in the WHU-Hi-HanChuan dataset, the determination of the optimal segmentation scale is a challenging task. The classification map of FNEA-OO appears smoother, but it exacerbates the misclassification of some crops, such as cowpea and greens. Furthermore, the upper part of the road is severely misclassified as gray roof and plastic. The SVRFMC method shows a competitive visual effect, and the classification accuracy is greatly improved. However, there is still some serious misclassification in the shadow-covered areas, due to the inaccurate

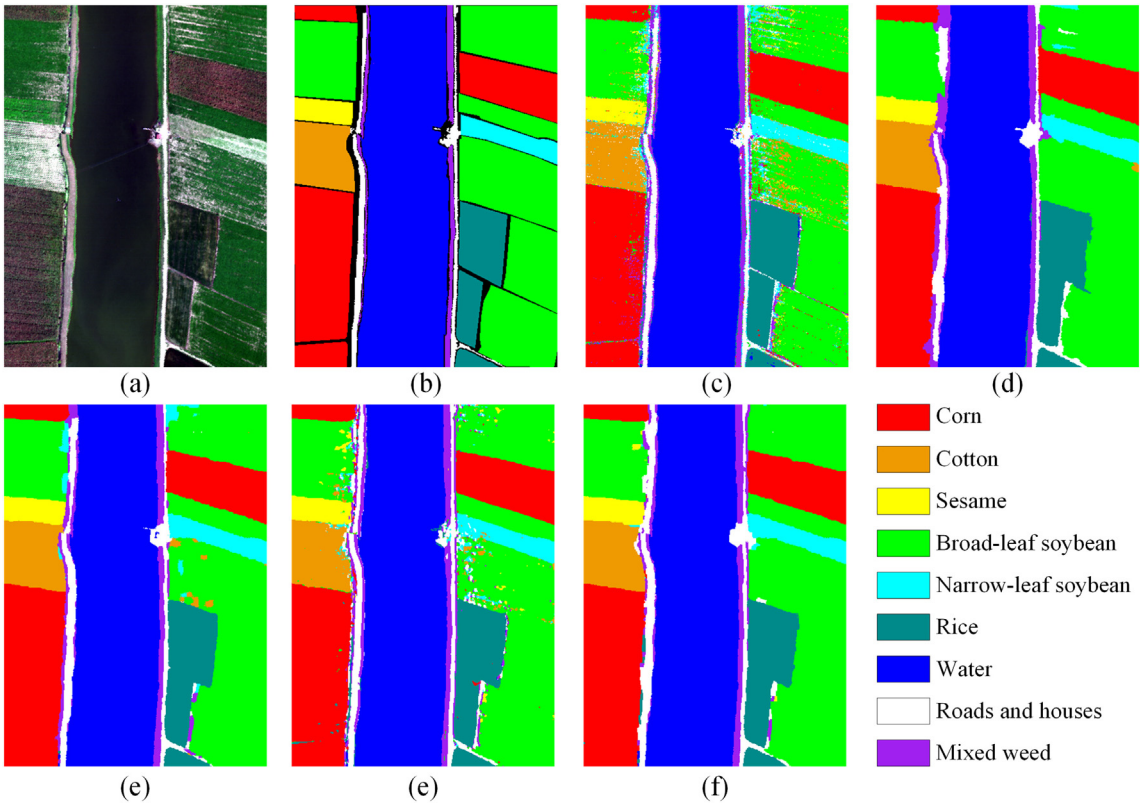

Fig. 12. The classification results for the WHU-Hi-LongKou dataset. (a) True-color image. (b) Ground-truth image. (c) SVM. (d) FNEA-OO. (e) SVRFMC. (f) Benchmark CNN. (g) CNNCRF.

unary potential provided by SVM. For this more complex scene, the benchmark CNN method embodies the ability to extract in-depth information, and the classification accuracy is better than that of SVRFMC. As shown in Fig. 15, it can be found that there is good distinction of ground objects in the shadow-covered area in the sub-image of the benchmark CNN classification results. However, some holes and isolated regions are found in the classification results, which are well relieved in the result of the CNNCRF method. Furthermore, the CNNCRF method achieves the highest accuracy for each crop type in the classification results.

4.2.4. Results for the WHU-Hi-HongHu dataset

Fig. 16c–f are the classification maps of the different classification methods (SVM, FNEA-OO, SVRFMC, benchmark CNN, and CNNCRF) for the WHU-Hi-HongHu dataset, and Table 8 lists the corresponding quantitative evaluation results. The WHU-Hi-HongHu dataset contains 18 kinds of crops, which makes it difficult to distinguish the species types using only spectral information. The OA value is 73.55% in the classification result of SVM, and there is a lot of SP noise in the classification map. The OA of the FNEA-OO method shows a 15.28% improvement over SVM, but there are still apparent misclassifications, such as the road showing a discontinuous shape and being seriously misclassified as bare soil and red roof. The OA of SVRFMC is more than 16% improved, compared to SVM, but the misclassification of isolated regions is still obvious, such as the bare soil in the bottom right of the classification map being misclassified as broad beans. The benchmark CNN method achieves an 11.88% accuracy improvement over the SVM classifier, but many isolated regions are still found in the classification results. CNNCRF achieves much better classification results than the other four methods, and the OA shows a 20% improvement over SVM. However, the cotton and cotton firewood classes still show serious misclassification, because the spectral and texture features are similar between withered cotton and cotton firewood. In summary, from these experimental results, it is shown that the CNNCRF framework exhibits a

Table 6
Classification accuracies for the WHU-Hi-LongKou dataset, the best results are highlighted in bold font.

Class		SVM	FNEA-OO	SVRFMC	Benchmark CNN	CNNCRF
Per-class accuracy (%)	Corn	98.34	99.51	99.93	99.62	**99.99**
	Cotton	93.62	**99.90**	99.67	96.23	99.08
	Sesame	96.93	99.32	99.90	98.74	**100.00**
	Broad-leaf soybean	87.46	97.72	96.56	93.30	**97.79**
	Narrow-leaf soybean	95.73	97.80	**99.43**	96.37	99.11
	Rice	99.23	99.67	**99.97**	98.01	99.80
	Water	99.97	99.93	**99.99**	99.92	**99.99**
	Roads and houses	92.95	88.77	94.55	96.64	**97.02**
	Mixed weed	92.42	94.74	86.66	**97.76**	91.07
OA (%)		94.96	98.59	98.37	97.30	**98.91**
KAPPA		0.9345	0.9815	0.9786	0.9647	**0.9857**
AA (%)		95.18	97.48	97.41	97.40	**98.21**

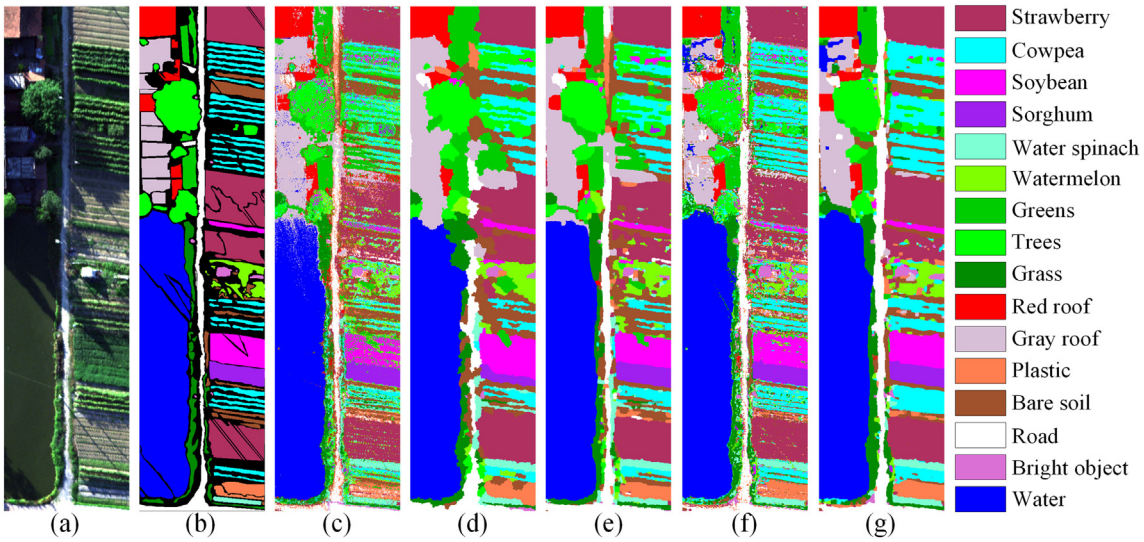

Fig. 13. The classification results for the WHU-Hi-HanChuan dataset. (a) True-color image. (b) Ground-truth image. (c) SVM. (d) FNEA-OO (e) SVRFMC. (f) Benchmark CNN. (g) CNNCRF.

Table 7
Classification accuracies for the WHU-Hi-HanChuan dataset, the best results are highlighted in bold font.

Class		SVM	FNEA-OO	SVRFMC	Benchmark CNN	CNNCRF
Per-class accuracy (%)	Strawberry	72.30	86.36	86.94	84.67	**95.01**
	Cowpea	50.77	63.37	61.59	84.44	**93.95**
	Soybean	72.66	91.01	93.46	90.25	**99.13**
	Sorghum	95.68	96.00	98.82	98.50	**99.28**
	Water spinach	82.91	98.91	96.64	98.45	**99.91**
	Watermelon	49.09	72.01	72.84	59.64	**83.28**
	Greens	90.38	95.81	98.60	94.00	**99.05**
	Trees	61.02	73.67	73.79	67.19	**77.60**
	Grass	63.19	70.04	79.82	76.08	**89.09**
	Red roof	89.50	88.87	92.34	95.29	**97.49**
	Gray roof	93.61	98.81	**98.98**	74.16	84.16
	Plastic	63.17	83.74	89.49	90.64	**98.83**
	Bare soil	57.95	76.75	68.17	65.91	**78.59**
	Road	65.04	69.37	72.72	86.81	**95.01**
	Bright object	72.49	69.59	68.05	**93.44**	92.66
	Water	95.56	97.03	97.71	99.35	**99.94**
OA (%)		77.61	85.63	86.53	87.13	**93.95**
KAPPA		0.7414	0.8330	0.8435	0.8497	**0.9290**
AA (%)		73.46	83.21	84.37	84.93	**92.69**

superior performance in precise crop classification with the WHU-Hi-HongHu dataset.

5. Discussion

5.1. Sensitivity in relation to training set size

The influence of different set sizes for training and testing was analyzed with the WHU-Hi-LongKou, WHU-Hi-HanChuan, and WHU-Hi-HongHu datasets. We selected different numbers (25, 50, 100, 150, 200, 250, 300) of labeled pixels from each category at random as training samples, and the rest were used as test samples. The overall classification accuracies of the different classification methods using different samples for training are shown in Fig. 17. The curves demonstrate that, as the number of training samples increases, the performance of all the different methods improves. Furthermore, a superior performance is obtained with the proposed CNNCRF method under most of the training sample sizes. In particular, for the complex scenes in the WHU-Hi-HanChuan and WHU-Hi-HongHu datasets, the proposed method shows a more significant advantage over the other classifiers.

5.2. Sensitivity in relation to spatial patch size

The influence of different spatial patch sizes for the benchmark CNN and CNNCRF methods was analyzed with the WHU-Hi-LongKou, WHU-Hi-HanChuan, and WHU-Hi-HongHu datasets. The different spatial patch sizes ($3 \times 3 \times d$, $5 \times 5 \times d$, $7 \times 7 \times d$, $9 \times 9 \times d$, $11 \times 11 \times d$, and $13 \times 13 \times d$, where d is the band number) were selected as the input data of the benchmark CNN and CNNCRF methods. The classification accuracies and training times of the benchmark CNN and CNNCRF methods using the different spatial patch sizes are shown in Fig. 18. The curves demonstrate that the OAs of CNNCRF are better than those of the benchmark CNN under all the spatial patch sizes. In addition, the benchmark CNN is more affected by the spatial patch size of the input data, which is due to the benchmark CNN only using the local spatial information from the spatial patches, whereas CNNCRF can incorporate the spatial-contextual information. The optimal spatial patch sizes for the WHU-Hi-LongKou, WHU-Hi-HanChuan, and WHU-Hi-HongHu datasets are 9, 13, and 11, respectively. The reason for this

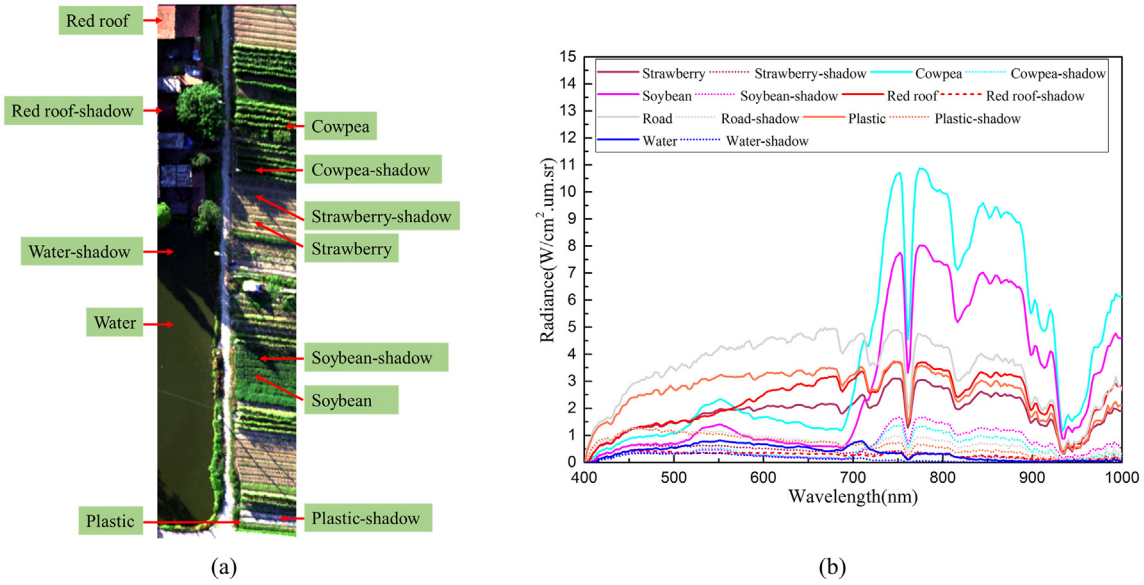

Fig. 14. (a) The shadow-covered ground objects in the WHU-Hi-HanChuan dataset. (b) The spectral curves of the ground objects and shadow-covered ground objects.

is that the optimal spatial patch size is affected by numerous factors, including the spatial resolution of the imagery and the spatial distribution of the ground objects. In general, with the same ground object distribution, the higher the spatial resolution of the image, the larger the optimal patch size. For the CNNCRF method in Fig. 18, a spatial patch size from 9 to 13 is a stable and reasonable value. However, with the increase of the spatial patch size, the internal memory requirement, graphics memory requirement, and training times for the CNNCRF method all increase significantly. To make a fair comparison and balance the calculation cost and accuracy, we set the same spatial patch size of $9 \times 9 \times d$ in all the WHU-Hi datasets.

5.3. Computational efficiency

The computational efficiency of the training and testing for the different classification methods was analyzed using the WHU-Hi-LongKou, WHU-Hi-HanChuan, and WHU-Hi-HongHu datasets. The SVM, FNEA-OO, and SVRFMC methods were implemented in MATLAB 2017 and eCognition 8.0 on a desktop PC with an Intel Xeon E3–1220 v5 CPU. The benchmark CNN and CNNCRF methods were implemented in the TensorFlow toolbox and MATLAB 2017 under CUDA 9 with an NVIDIA GeForce RTX 2080 Ti GPU and an Intel Xeon E5–2690 v4 CPU. Table 9 lists the average run times for the different classification methods with the WHU-Hi dataset. In the experiment, the probabilistic images obtained by the SVM and benchmark CNN methods were used as the unary potential term of the SVRFMC and CNNCRF methods, respectively. Therefore, the training times of the SVRFMC and CNNCRF methods were the same as those of the SVM and benchmark CNN methods. With the increase of the image size and the total number of training samples, the training and test times of the benchmark CNN (GPU version) do not increase by much, whereas the training times of the SVM, FNEA-OO, and SVRFMC methods increases significantly. Because the CRF model has to infer the relationship between each pixel and the neighborhood pixels, the test speed of SVRFMC and CNNCRF is relatively slow.

5.4. The generality and applicability of the CNNCRF method

To investigate the generality and applicability of the proposed CNNCRF method for precise crop mapping, the Salinas AVIRIS dataset, which is a public dataset for precision agriculture classification, was used in this experiment. The Salinas AVIRIS dataset was acquired on October 9, 1998, in the south of the city of Greenfield in the Salinas Valley of California, America, by the AVIRIS sensor equipped on a manned air-borne platform (Gualtieri et al., 1999). The size of the Salinas dataset is 550 × 400 pixels, there are 224 bands from 0.4 to 2.5 um, and the spatial resolution is 3.7 m. The image contains soil, bare ground, vegetables, and vineyards, with 16 subclasses. Because the Salinas AVIRIS dataset includes a lot of subclasses of the same family, it is a challenging task to identify the different crops. An overview of this dataset is provided in Fig. 19. The spectra of the 16 agricultural types are respectively presented in Fig. 20. In this experiment, 15 labeled pixels for each class were randomly selected as training samples. The detailed class information for the Salinas AVIRIS dataset is provided in Table 10.

Fig. 21c–f are the classification maps of the different classification methods (SVM, FNEA-OO, SVRFMC, benchmark CNN, and CNNCRF) for the Salinas AVIRIS dataset, and Table 11 lists the corresponding

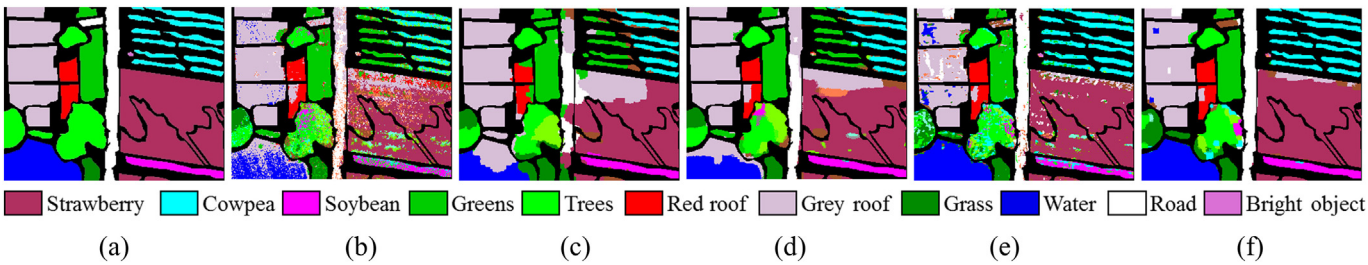

Fig. 15. The classification results for the sub-image of the WHU-Hi-HanChuan dataset: (a) Ground-truth image. (b) SVM. (c) FNEA-OO (d) SVRFMC. (e) Benchmark CNN. (f) CNNCRF.

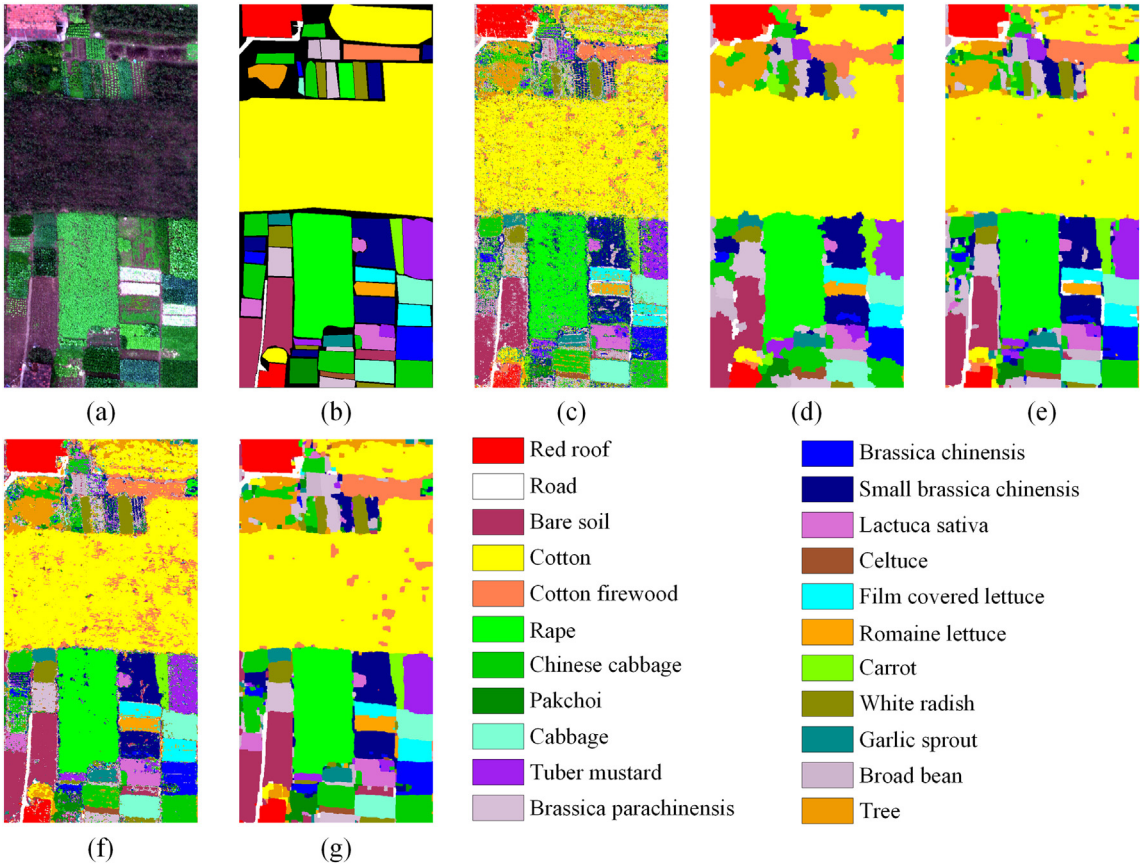

Fig. 16. The classification results for the WHU-Hi-HongHu dataset. (a) True-color image. (b) Ground-truth image. (c) SVM. (d) FNEA-OO (e) SVRFMC. (f) Benchmark CNN. (g) CNNCRF.

quantitative evaluation results. The OA value is 78.41% in the classification result of SVM, and there is severe SP noise in the classification map of SVM due to it not considering the spatial information or the spatial-contextual information. In addition, because of the very similar subclasses, there is serious misclassification between the grapes-untrained and vineyard-untrained classes, and the brocoli_green_weeds_1 and brocoli_green_weeds_2 classes. The OA of the FNEA-OO method shows a smoother classification map and an 8.71% improvement over

Table 8
Classification accuracies for the WHU-Hi-HongHu dataset, the best results are highlighted in bold font.

Class		SVM	FNEA-OO	SVRFMC	Benchmark CNN	CNNCRF
Per-class accuracy (%)	Red roof	83.93	96.24	95.26	93.85	**98.32**
	Road	87.34	46.51	95.02	91.68	**95.93**
	Bare soil	72.85	82.29	82.23	89.94	**95.66**
	Cotton	78.96	**97.55**	96.13	86.58	94.76
	Cotton firewood	77.25	92.24	97.66	84.37	**99.75**
	Rape	81.95	91.20	**92.84**	90.38	92.54
	Chinese cabbage	59.25	70.17	77.66	76.99	**90.46**
	Pakchoi	41.63	60.29	64.19	62.54	**81.36**
	Cabbage	90.86	93.20	95.44	93.11	**95.90**
	Tuber mustard	54.08	86.81	84.93	75.26	**95.31**
	Brassica parachinensis	48.31	69.35	71.92	70.52	**93.08**
	Brassica chinensis	61.31	83.67	83.13	72.13	**84.20**
	Small Brassica chinensis	49.86	72.93	70.56	72.96	**83.80**
	Lactuca sativa	63.78	61.49	73.65	88.66	**96.51**
	Celtuce	85.92	89.14	99.89	98.12	**100.00**
	Film covered lettuce	78.01	**97.89**	93.91	94.11	96.34
	Romaine lettuce	70.65	91.65	80.65	90.72	**98.87**
	Carrot	79.24	90.86	**99.13**	90.50	98.17
	White radish	68.22	72.88	84.61	91.18	**95.65**
	Garlic sprout	77.85	92.29	96.84	93.18	**98.64**
	Broad bean	74.67	96.01	100.00	91.45	**100.00**
	Tree	81.14	94.24	99.92	97.39	**100.00**
OA (%)		73.55	88.83	89.86	85.43	**93.74**
KAPPA		68.05	85.90	87.28	82.10	**92.17**
AA (%)		71.23	83.13	87.98	86.16	**94.78**

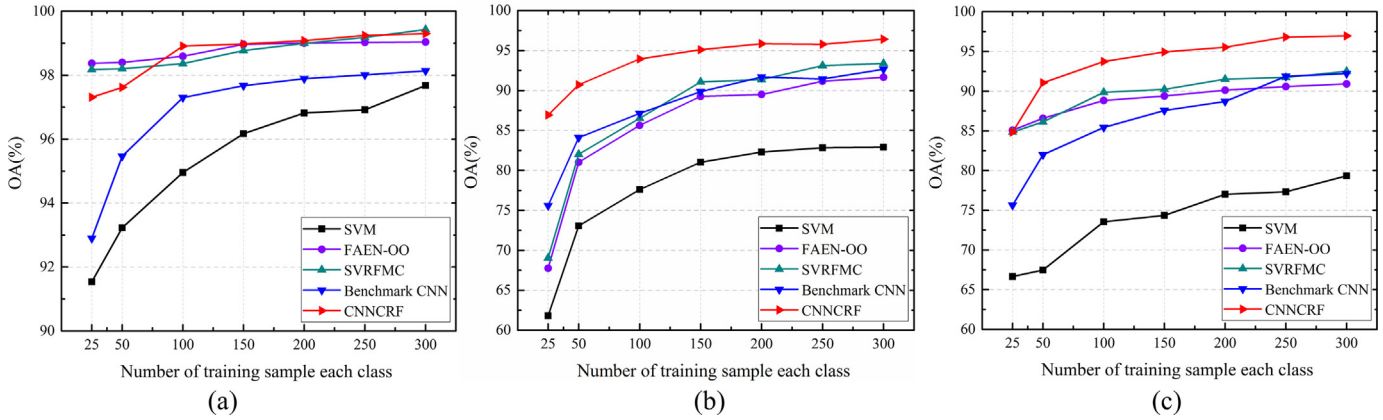

Fig. 17. Effect of different numbers of training samples for SVM, FNEA-OO, SVRFMC, CNN, and CNNCRF on (a) the WHU-Hi-LongKou dataset, (b) the WHU-Hi-HanChuan dataset, and (c) the WHU-Hi-HongHu dataset.

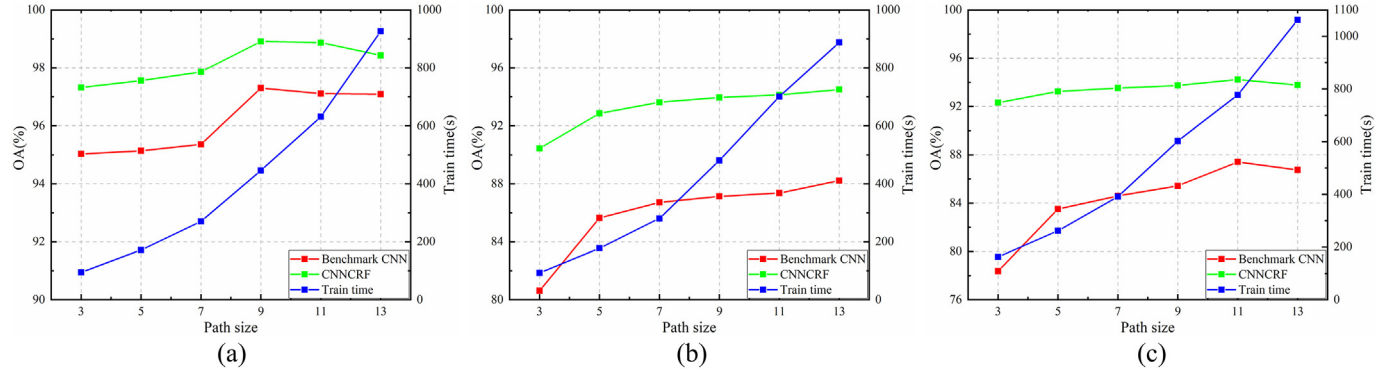

Fig. 18. Effect of different spatial patch sizes for the benchmark CNN and CNNCRF methods on (a) the WHU-Hi-LongKou dataset, (b) the WHU-Hi-HanChuan dataset, and (c) the WHU-Hi-HongHu dataset.

Table 9
Average run times for the different classification methods with the WHU-Hi dataset.

Dataset		SVM	FNEA-OO	SVRFMC	Benchmark CNN	CNNCRF
WHU-Hi-LongKou	***Train time(s)***	46.154	81.393	46.154	446.282	446.282
	Test time(s)	34.928	43.673	87.078	69.367	121.517
WHU-Hi-HanChuan	***Train time(s)***	178.096	370.444	178.096	480.761	480.761
	Test time(s)	216.025	226.149	513.345	114.924	412.244
WHU-Hi-HongHu	***Train time(s)***	303.298	625.014	303.298	502.210	502.210
	Test time(s)	452.479	467.349	1131.454	132.127	811.102

SVM, but there are still serious misclassifications between the grapes_untrained and vineyard_untrained, lettuce_romaine_4wk and lettuce_romaine_5wk classes due to the difficulty to select the segmentation scale for various categories. The SVRFMC method achieves higher classification accuracy and less SP noise than SVM, but there are still many grapes_untrained are classified as vineyard_untrained. The benchmark CNN method achieves a 6.80% accuracy improvement over the SVM classifier, but many isolated regions are still found in the grapes_untrained and vineyard_untrained classes. The CNNCRF method achieves the highest classification accuracies (OA, AA, and kappa), and the OA shows an 11.64% improvement over SVM. However, the grapes_untrained and vineyard_untrained classes still show serious misclassification, because the spectral and texture features are very similar between grapes_untrained and vineyard_untrained. In summary, this experiment shows that CNNCRF shows superior generality and applicability in precise crop classification for the aerial hyperspectral imagery from the AVIRIS sensor with a meter-level spatial resolution.

5.5. The application prospects of the WHU-Hi dataset

At present, an increasing number of studies are focusing on agricultural remote sensing via UAV-borne hyperspectral systems. However, there is still a lack of public benchmark datasets for crop classification based on UAV-borne H^2 imagery. The WHU-Hi dataset is a new public UAV-borne hyperspectral benchmark dataset for precise crop classification, which can be used to test the effectiveness of crop classification methods, and even some of the state-of-the-art hyperspectral classification methods, such as the deep learning based methods of Spectral–Spatial Residual Network (SSRN) (Zhong et al., 2018b), Spectral-Spatial Attention Networks (SSAN) (Mei et al., 2019), and Fast Patch-Free Global Learning Framework (FPGA) (Zheng et al., 2020). The CNNCRF method considers the in-depth spectral and local spatial features and the spatial-contextual information to achieve competitive classification results. The classification accuracy of the CNNCRF method can thus be considered as a baseline. In summary, the purpose of this article was to encourage more researchers to focus on

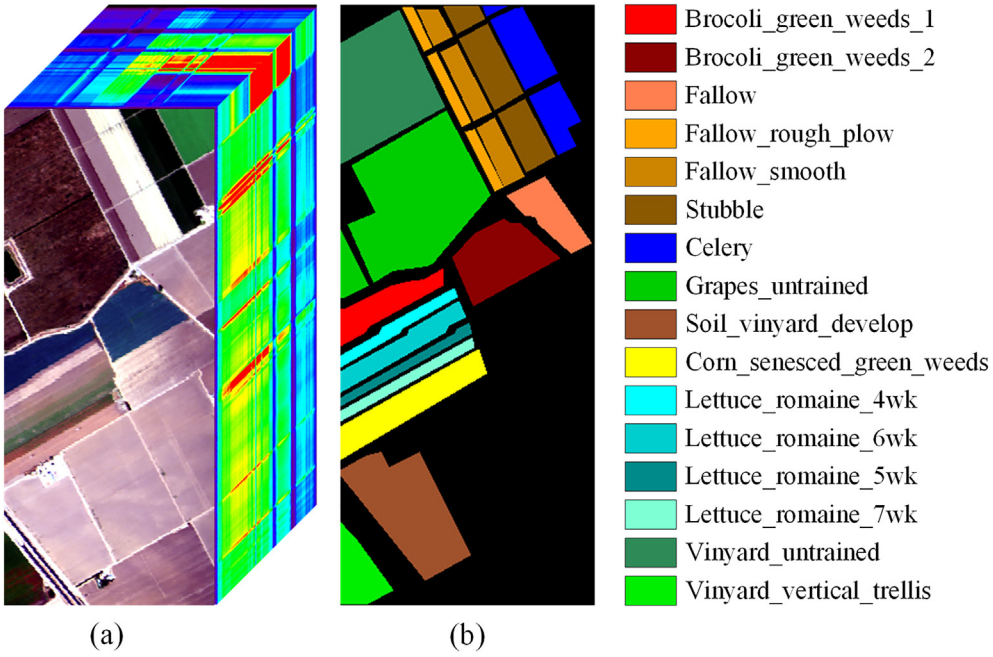

Fig. 19. Salinas AVIRIS dataset. (a) RGB false-color image. (b) Ground-truth image.

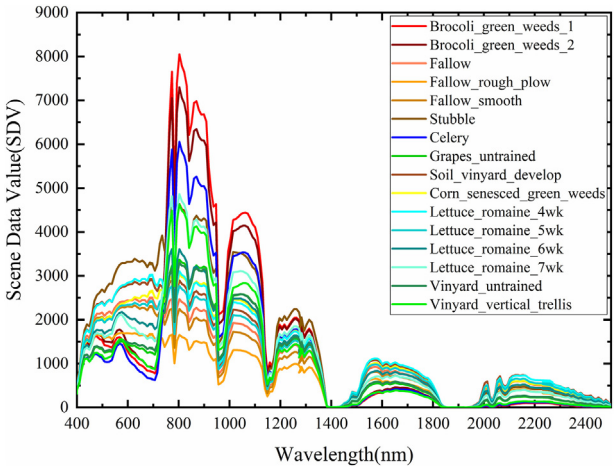

Fig. 20. Spectra of the Salinas AVIRIS dataset.

crop classification based on UAV-borne H^2 imagery. In this regard, this article provides both a benchmark dataset and a baseline classification result.

6. Conclusion

In this article, aiming at the planting characteristics of crop heterogeneity under land fragmentation, we have proposed a crop monitoring strategy based on UAV-borne H^2 imagery, and a benchmark dataset, the WHU-Hi dataset, for precise crop classification, which consists of three individual UAV-borne hyperspectral datasets. The deep convolutional neural network with conditional random field classifier (CNNCRF) framework was also proposed for precise crop classification with UAV-borne H^2 imagery. The proposed approach first uses a virtual sample augmentation strategy based on the hyperspectral imaging mechanism, to lessen the limitation of the limited labeled samples. Following this, the high-level robust spectral-spatial features are fused and extracted by the deep CNN. The CRF model then further integrates the spatial-contextual information, to improve the problem of holes and isolated regions in the classification map. The experimental results obtained with the WHU-Hi dataset demonstrated that the proposed CNNCRF classification framework is effective and efficient in precise crop classification based on UAV-borne H^2 imagery. Our future research work will pay attention to the practical application of precise crop classification based on UAV-borne H^2 imagery.

Declaration of Competing Interest

None.

Acknowledgements

The authors would like to thank the editor, associate editor, and anonymous reviewers for their helpful comments and advice. This work

Table 10
Class information for the Salinas AVIRIS dataset.

No.	Class name	Training samples	Verification samples	No.	Class name	Training samples	Verification samples
C1	Brocoli_green_weeds_1	15	1994	C9	Soil_vineyard_develop	15	6188
C2	Brocoli_green_weeds_2	15	3711	C10	Corn_senesced_green_weeds	15	3263
C3	Fallow	15	1961	C11	Lettuce_romaine_4wk	15	1053
C4	Fallow_rough_plow	15	1379	C12	Lettuce_romaine_5wk	15	1912
C5	Fallow_smooth	15	2663	C13	Lettuce_romaine_6wk	15	901
C6	Stubble	15	3944	C14	Lettuce_romaine_7wk	15	1055
C7	Celery	15	3564	C15	Vineyard_untrained	15	7253
C8	Grapes_untrained	15	11,256	C16	Vineyard_vertical_trellis	15	1792

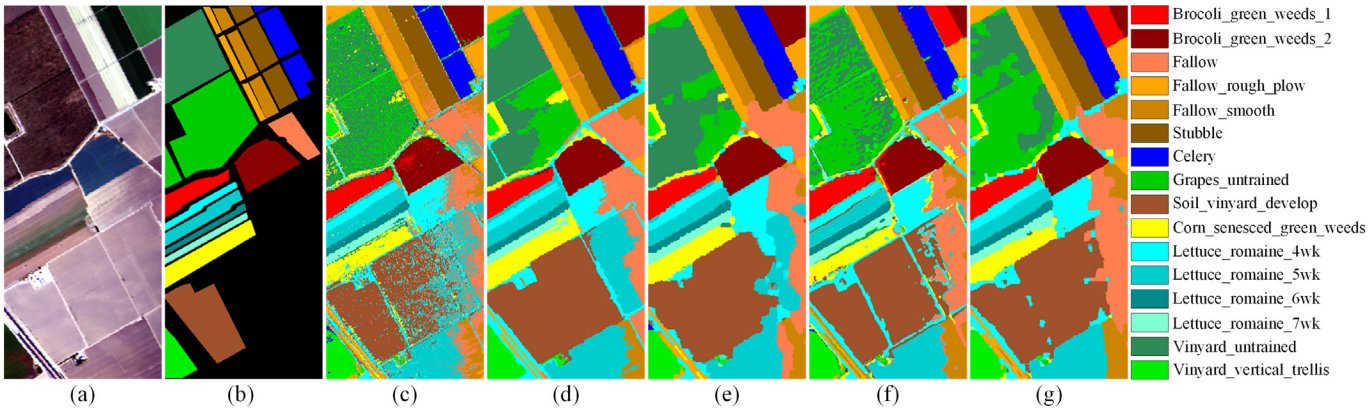

Fig. 21. The classification results for the Salinas AVIRIS dataset. (a) True-color image. (b) Ground-truth image. (c) SVM. (d) FNEA-OO (e) SVRFMC. (f) Benchmark CNN. (g) CNNCRF.

Table 11
Classification accuracies for the Salinas AVIRIS dataset, the best results are highlighted in bold font.

Class		SVM	FNEA-OO	SVRFMC	Benchmark CNN	CNNCRF
Per-class accuracy (%)	Brocoli_green_weeds_1	97.69	99.85	100	98.29	**100**
	Brocoli_green_weeds_2	92.59	**100.00**	100	98.44	**100**
	Fallow	75.52	**100.00**	95.72	88.22	89.6
	Fallow_rough_plow	98.19	**100.00**	99.71	97.32	98.4
	Fallow_smooth	83.55	**96.92**	85.77	91.66	94.93
	Stubble	99.67	**99.47**	99.97	99.97	99.97
	Celery	99.35	99.13	**100**	99.35	**100**
	Grapes_untrained	44.00	51.00	45.85	69.85	**80.72**
	Soil_vineyard_develop	96.38	**100.00**	**100**	97.95	99.03
	Corn_senesced_green_weeds	81.67	**96.94**	84.4	79.62	91.36
	Lettuce_romaine_4wk	96.11	90.12	96.11	99.43	**99.81**
	Lettuce_romaine_5wk	98.59	**100.00**	99.84	99.63	99.95
	Lettuce_romaine_6wk	98.11	97.11	96.78	99.45	**99.22**
	Lettuce_romaine_7wk	90.43	89.95	98.01	97.25	**99.62**
	Vineyard_untrained	59.62	86.87	**93.58**	57.4	66.15
	Vineyard_vertical_trellis	94.87	**100.00**	99.16	97.71	99.05
OA (%)		78.41	87.12	85.81	85.21	**90.05**
KAPPA		0.7613	0.8577	0.8433	0.8356	**0.8892**
AA (%)		87.90	94.21	93.43	91.97	**94.86**
Train time(s)		21.354	22.050	21.354	393.317	393.317
Test time(s)		34.097	61.147	147.451	18.505	131.859

was supported by National Key Research and Development Program of China under Grant No. 2017YFB0504202, National Natural Science Foundation of China under Grant Nos. 41771385, 41820104006 and 61871299, and by the China Postdoctoral Science Foundation.

References

Adão, T., Hruška, J., Pádua, L., Bessa, J., Peres, E., Morais, R., Sousa, J., 2017. Hyperspectral imaging: a review on UAV-based sensors, data processing and applications for agriculture and forestry. Remote Sens. 9, 1110.

Baatz, M., Schäpe, A., 2000. Multiresolution segmentation:an optimization approach for high quality multi-scale image segmentation. Angewandte Geographische Informationsverarbeitung XII, 12–23.

Cai, Y., Guan, K., Peng, J., Wang, S., Seifert, C., Wardlow, B., Li, Z., 2018. A high-performance and in-season classification system of field-level crop types using time-series Landsat data and a machine learning approach. Remote Sens. Environ. 210, 35–47.

Chang, C.-C., Lin, C.-J., 2011. LIBSVM: a library for support vector machines. ACM Trans. Intell. Syst. Technol. 2, 1–27.

Chen, Y., Jiang, H., Li, C., Jia, X., Ghamisi, P., 2016. Deep feature extraction and classification of Hyperspectral images based on convolutional neural networks. IEEE Trans. Geosci. Remote Sens. 54, 6232–6251.

Ciregan, D., Meier, U., Schmidhuber, J., 2012. Multi-column deep neural networks for image classification. Comput. Vis. Pattern Recogn. 3642–3649.

Galvao, L.S., Formaggio, A.R., Tisot, D.A., 2005. Discrimination of sugarcane varieties in Southeastern Brazil with EO-1 Hyperion data. Remote Sens. Environ. 94, 523–534.

Goetz, A.F., Vane, G., Solomon, J.E., Rock, B.N., 1985. Imaging spectrometry for earth remote sensing. Science 228, 1147–1153.

Gualtieri, J., Chettri, S.R., Cromp, R., Johnson, L., 1999. Support vector machine classifiers as applied to aviris data. In: Proc. Eighth JPL Airborne Geoscience Workshop, pp. 217–227.

Honkavaara, E., Saari, H., Kaivosoja, J., Pölönen, I., Hakala, T., Litkey, P., Mäkynen, J., Pesonen, L., 2013. Processing and assessment of spectrometric, stereoscopic imagery collected using a lightweight UAV spectral camera for precision agriculture. Remote Sens. 5, 5006–5039.

Jiang, Z., Huete, A.R., Chen, J., Chen, Y., Li, J., Yan, G., Zhang, X., 2006. Analysis of NDVI and scaled difference vegetation index retrievals of vegetation fraction. Remote Sens. Environ. 101, 366–378.

Kato, N., Suzuki, M., Omachi, S.i., Aso, H., Nemoto, Y., 1999. A handwritten character recognition system using directional element feature and asymmetric Mahalanobis distance. IEEE Trans. Pattern Anal. Mach. Intell. 21, 258–262.

Kumar, S., 2003. Discriminative random fields: A discriminative framework for contextual interaction in classification. In: Proc. Int. Conf. Computer Vision. IEEE, pp. 1150–1157.

Lafferty, J., McCallum, A., Pereira, F.C., 2001. Conditional random fields: probabilistic models for segmenting and labeling sequence data. Proc. ICML 1, 282–289.

LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. Proc. IEEE 86, 2278–2324.

Li, S., Song, W., Fang, L., Chen, Y., Ghamisi, P., Benediktsson, J.A., 2019. Deep learning for Hyperspectral image classification: an overview. IEEE Trans. Geosci. Remote Sens. 57, 6690–6709.

Löw, F., Michel, U., Dech, S., Conrad, C., 2013. Impact of feature selection on the accuracy and spatial uncertainty of per-field crop classification using support vector machines. ISPRS J. Photogramm. Remote Sens. 85, 102–119.

Lv, P., Zhong, Y., Zhao, J., Jiao, H., Zhang, L., 2016. Change detection based on a multifeature probabilistic ensemble conditional random field model for high spatial resolution remote sensing imagery. IEEE Geosci. Remote Sens. Lett. 13, 1965–1969.

Mei, X., Pan, E., Ma, Y., Dai, X., Huang, J., Fan, F., Du, Q., Zheng, H., Ma, J., 2019. Spectral-spatial attention networks for Hyperspectral image classification. Remote

Sens. 11, 963.
Meng, S., Zhong, Y., Luo, C., Hu, X., Wang, X., Huang, S., 2020. Optimal temporal window selection for winter wheat and rapeseed mapping with Sentinel-2 images: a case study of Zhongxiang in China. Remote Sens. 12, 226.
Nidamanuri, R.R., Zbell, B., 2011. Use of field reflectance data for crop mapping using airborne hyperspectral image. ISPRS J. Photogramm. Remote Sens. 66, 683–691.
Piiroinen, R., Heiskanen, J., Mõttus, M., Pellikka, P., 2015. Classification of crops across heterogeneous agricultural landscape in Kenya using AisaEAGLE imaging spectroscopy data. Int. J. Appl. Earth Obs. Geoinf. 39, 1–8.
Pinter Jr., P.J., Hatfield, J.L., Schepers, J.S., Barnes, E.M., Moran, M.S., Daughtry, C.S., Upchurch, D.R., 2003. Remote sensing for crop management. Photogramm. Eng. Remote Sens. 69, 647–664.
Rao, N., Garg, P.K., Ghosh, S.K., 2007. Development of an agricultural crops spectral library and classification of crops at cultivar level using hyperspectral data. Precis. Agric. 8, 173–185.
Shu-hao, T., Fu-tian, Q., Heerink, N., 2003. Causes and determinants of land fragmentation. China Rural Surv. 6, 24–30.
Son, N.-T., Chen, C.-F., Chen, C.-R., Duc, H.-N., Chang, L.-Y., 2014. A phenology-based classification of time-series MODIS data for rice crop monitoring in Mekong Delta, Vietnam. Remote Sens. 6, 135–156.
Suomalainen, J., Anders, N., Iqbal, S., Roerink, G., Franke, J., Wenting, P., Hünniger, D., Bartholomeus, H., Becker, R., Kooistra, L., 2014. A lightweight hyperspectral mapping system and photogrammetric processing chain for unmanned aerial vehicles. Remote Sens. 6, 11013–11030.
Tatsumi, K., Yamashiki, Y., Torres, M.A.C., Taipe, C.L.R., 2015. Crop classification of upland fields using random forest of time-series Landsat 7 ETM+ data. Comput. Electron. Agric. 115, 171–179.
Tong, Q., Xue, Y., Zhang, L., 2013. Progress in hyperspectral remote sensing science and technology in China over the past three decades. IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens. 7, 70–91.
Tu, Y., Bian, M., Wan, Y., Fei, T., 2018. Tea cultivar classification and biochemical parameter estimation from hyperspectral imagery obtained by UAV. PeerJ 6, e4858.
Wang, S., Azzari, G., Lobell, D.B., 2019. Crop type mapping without field-level labels: random forest transfer and unsupervised clustering techniques. Remote Sens. Environ. 222, 303–317.
Wardlow, B.D., Egbert, S.L., 2008. Large-area crop mapping using time-series MODIS 250 m NDVI data: an assessment for the US central Great Plains. Remote Sens. Environ. 112, 1096–1116.
Xiao, L., Xianjin, H., Taiyang, Z., Yuntai, Z., Yi, L., 2013. A review of farmland fragmentation in China. J. Resour. Ecol. 4, 344–353.
Yu, S., Jia, S., Xu, C., 2017. Convolutional neural networks for hyperspectral image classification. Neurocomputing 219, 88–98.
Zhang, C., Kovacs, J.M., 2012a. The application of small unmanned aerial systems for precision agriculture: a review. Precis. Agric. 13, 693–712.
Zhang, C., Kovacs, J.M., 2012b. The application of small unmanned aerial systems for precision agriculture: a review. Precis. Agric. 13, 693–712.
Zhang, X., Sun, Y., Shang, K., Zhang, L., Wang, S., 2016. Crop classification based on feature band set construction and object-oriented approach using Hyperspectral images. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 9, 4117–4128.
Zhao, W., Du, S., 2016. Spectral–spatial feature extraction for hyperspectral image classification: a dimension reduction and deep learning approach. IEEE Trans. Geosci. Remote Sens. 54, 4544–4554.
Zhao, J., Zhong, Y., Zhang, L., 2015. Detail-preserving smoothing classifier based on conditional random fields for high spatial resolution remote sensing imagery. IEEE Trans. Geosci. Remote Sens. 53, 2440–2452.
Zhao, J., Zhong, Y., Jia, T., Wang, X., Xu, Y., Shu, H., Zhang, L., 2018. Spectral-spatial classification of hyperspectral imagery with cooperative game. ISPRS J. Photogramm. Remote Sens. 135, 31–42.
Zhao, J., Zhong, Y., Hu, X., Wei, L., Zhang, L., 2020. A robust spectral-spatial approach to identifying heterogeneous crops using remote sensing imagery with high spectral and spatial resolutions. Remote Sens. Environ. 239, 111605.
Zheng, Z., Zhong, Y., Ma, A., Zhang, L., 2020. FPGA: fast patch-free global learning framework for fully end-to-end hyperspectral image classification. IEEE Trans. Geosci. Remote Sens. 1–15.
Zhong, P., Wang, R., 2011. Modeling and classifying hyperspectral imagery by CRFs with sparse higher order potentials. IEEE Trans. Geosci. Remote Sens. 49, 688–705.
Zhong, Y., Lin, X., Zhang, L., 2014. A support vector conditional random fields classifier with a Mahalanobis distance boundary constraint for high spatial resolution remote sensing imagery. IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens. 7, 1314–1330.
Zhong, Y., Wang, X., Xu, Y., Wang, S., Jia, T., Hu, X., Zhao, J., Wei, L., Zhang, L., 2018a. Mini-UAV-borne Hyperspectral remote sensing: from observation and processing to applications. IEEE Geosci. Remote Sens. Mag. 6, 46–62.
Zhong, Z., Li, J., Luo, Z., Chapman, M.A., 2018b. Spectral–spatial residual network for hyperspectral image classification: a 3-D deep learning framework. IEEE Trans. Geosci. Remote Sens. 56, 847–858.
Zhou, L., Cao, G., Li, Y., Shang, Y., 2016. Change detection based on conditional random field with region connection constraints in high-resolution remote sensing images. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 9, 3478–3488.