

# Convolutional Neural Network with Expert Knowledge for Hyperspectral Remote Sensing Imagery Classification

Chunming Wu<sup>1</sup>, Meng Wang<sup>2</sup>, Lang Gao<sup>2\*</sup>, Weijing Song<sup>2\*</sup>, Tian Tian<sup>2</sup>,  
and Kim-Kwang Raymond Choo<sup>3</sup>

<sup>1</sup> Institute of Geological Survey, China University of Geosciences  
No.388 Lumo Road, Wuhan, 430074, China

<sup>2</sup> Hubei Key Laboratory of Intelligent Geo-Information Processing  
School of Computer Science, China University of Geosciences  
No.388 Lumo Road, Wuhan, 430074, China

[e-mail: langgao@cug.edu.cn, songwj@cug.edu.cn]

<sup>3</sup> Department of Information Systems and Cyber Security  
The University of Texas at San Antonio  
San Antonio, TX, 78249, USA.

[e-mail: raymond.choo@fulbrightmail.org]

\*Corresponding author: Lang Gao and Weijing Song

*Received October 14, 2018; revised January 18, 2019; accepted February 12, 2019;  
published August 31, 2019*

---

## Abstract

The recent interest in artificial intelligence and machine learning has partly contributed to an interest in the use of such approaches for hyperspectral remote sensing (HRS) imagery classification, as evidenced by the increasing number of deep framework with deep convolutional neural networks (CNN) structures proposed in the literature. In these approaches, the assumption of obtaining high quality deep features by using CNN is not always easy and efficient because of the complex data distribution and the limited sample size. In this paper, conventional handcrafted learning-based multi features based on expert knowledge are introduced as the input of a special designed CNN to improve the pixel description and classification performance of HRS imagery. The introduction of these handcrafted features can reduce the complexity of the original HRS data and reduce the sample requirements by eliminating redundant information and improving the starting point of deep feature training. It also provides some concise and effective features that are not readily available from direct training with CNN. Evaluations using three public HRS datasets demonstrate the utility of our proposed method in HRS classification.

---

**Keywords:** Hyperspectral imagery classification, convolutional neural network, principal component analysis, gray-level co-occurrence matrix, differential Mathematical morphology

## 1. Introduction

With constant advances in remote sensing platform and data processing technologies, spectral resolution and spatial resolution of remote sensing images have been significantly enhanced in recent times. Such hyperspectral remote sensing (HRS) images have been used in a broad range of applications ranging from disaster detection [1], urban planning [2, 3] to environmental monitoring [4, 5, 6], and several other fields [30, 31]. However, the availability of such remote sensing information poses new challenges for researchers due to complex data distribution, etc. In addition, effective feature extraction is crucial to the classification of HRS images.

In recent years, the performance of classification for imagery has rapidly increased due to the use of deep convolutional neural networks (CNN). Inspired by the human visual system, deep CNN is able to extract high-level abstract features of imagery by establishing a number of convolutional and pooling layers. Based on the benefits afforded by CNN in extracting deep features and its promising results on natural image interpretation, it has been applied to remote sensing classification and recognition (e.g. see the technical tutorial on deep learning for remote sensing data in [11]). Hu et. al. [21], for example, proposed a deep CNN for HRS classification based on the spectral domain. As the input of CNN, every HRS pixel sample with hundred spectral bands can be regarded as a 2-D image whose height is equal to 1. In [22], a 3-D CNN-based feature extraction model was proposed to extract the high-level features of HRS imagery, and the combined regularization strategy based on sparse constraint was then developed to handle the high dimensionality of the input data.

Due to the distribution of the HRS imagery, it is inefficient to use the CNN method directly. HRS images composed of hundreds of spectral bands and provide rich spectral information. However, such high dimensional data is challenging to work with. On one hand, the data's high dimension can easily lead to the Hughes phenomenon. On the other hand, the redundant structure of HRS images presents challenges to extract high-level features by using the CNN. It not only improves the complexity of the neural network structure, but also reduces the efficiency and quality of the extracted features. Thus, a number of approaches to reduce dimensionality have been introduced, say to preprocess the HRS image before using the CNN to extract the deep features. In [25] for example, the fractional order Darwinian particle swarm optimization was used to select the most informative bands suitable for the designed CNN network. These selected bands are then sent to the classification system to produce the classification map. In [24], the deep spatial features extracted automatically by CNN are integrated with the spectral features based on a balanced local discriminant embedding algorithm to reduce the dimension and classify the HRS data. Atharva et. al [23] proposed a patch-based CNN network for medium-resolution remote sensing image classification. A gated CNN for semantic segmentation for high-resolution remote sensing images was presented in [26]. Other advances include the use of CNN to achieve improvements in image classification [33, 32], segmentation [35], and recognition [34].

However, most existing methods use the prior knowledge of the spectral information [12], and assume that the constructed CNN can extract the efficient deep spectral-spatial features. This assumption is not always satisfied on the classification of HRS images because of the complex distribution characteristics of HRS images [8], especially for those with high spatial resolution. It is easy to know that the more complex distribution of the data, the greater the number of CNN parameters will be used to extract features [37]. However, it is difficult to

obtain sufficient samples in remote sensing area to train a robust classifier based on CNN of complex structures.

Based on the above limitations, we introduce traditional multi handcrafted features, including spectral and spatial expert knowledge information, into the model, as input of the deep CNN in order to further improve the classification accuracy of HRS images. The handcrafted features that can effectively represent raw data while removing unnecessary redundant information [9], reduce the complexity of data distribution and the demand for CNN parameter scale. In other words, we can effectively fit the feature distribution and obtain a promising classification result of the original data by building a simpler network structure. Introducing handcrafted features into our framework also allows us to improve the training starting point of deep features obtained by CNN, and then to achieve the goal of improving the network training efficiency and the quality of the high-level features. Thus, we are able to obtain good classification results using small sample sizes. In order to mitigate the effects of over-fitting problems caused by small samples, L2 regularization and dropout strategies are used during the training stage of the CNN network.

The remainder of this paper is organized as follows. The proposed methodology is discussed in Section 2. The experimental results and analysis are discussed in Section 3. The conclusion is summarized in Section 4.

## 2. Expert Knowledge-based CNN Deep Approach

Deep convolutional neural network has raised the expectation to improve the interpretation performance of HRS image because of its potential for high-level feature extraction. However, it is still a challenge to use CNN directly to effectively improve the classification and recognition performance due to the complex distribution characteristics and the limited samples of HRS images, as discussed earlier. In order to reduce the impact of these two factors, some priori information of HRS images based on expert knowledge (EK) was introduced as the input of a specific CNN in the proposed framework (more details shows in Section 2.1).

EK information used in this pape is obtained by some traditional handcrafted feature extraction methods that have been proposed based on the properties of remote sensing images in the past few decades. These EK information is generally obtained in two different ways: one is to remove the redundant information or noise of the HRS image; another way is to extract some main attribute information of the HRS image. Through the above two ways, the information based on EK can represent ground objects more succinctly and efficiently and contain smaller amounts of data than the original image. The EK information is easier to acquire and more refined than features obtained by training a CNN networks at bottom layers. It means that the training starting point for deep features can be effectively improved by using the EK information as the CNN input. The introduction of the EK information selection will be shown in Section 2.2.

On the premise of obtaining approximate classification results, the effective function parameters needed to characterize the distribution of EK information are far less than those needed to characterize the original HRS distribution. Generally speaking, on the premise of effectively characterizing data distribution, the fewer parameters of the function, the smaller demand for the sample size.

In this article, the demand for the sample size is controlled by designing a appropriate deep CNN that designed to classify HRS images for EK information rather than the original HRS imagery. More details about the CNN shows in Section 2.3.

## 2.1 The Proposed Framework

In the basic common framework with CNN for HRS classification, the band information is usually used as the input of CNN to extract the deep features and obtain the classification result. Such framework as shown in Fig. 1, effectively improves the classification performance of HRS images. To illustrate, the overall accuracy of Indian Pines data set classification is over 97% with 1765 training samples and 6223 testing samples in [22].

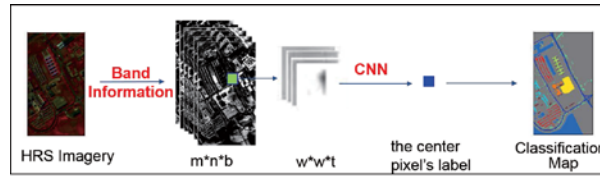


Fig. 1. Common framework with CNN for HRS classification

Based on the above framework, this article introduces the traditional handcrafted HRS features and designs a corresponding CNN structure to handle the HRS classification, as shown in Fig. 2. Firstly we extract the spectral features of the HRS images by using the PCA method, which preserves the main spectral features of the data by removing the redundant information between the bands without changing the shape distribution of the ground objects. After that typical spatial features of the ground objects are extracted as the input of convolution neural network. Finally we use the end-to-end approach to train the CNN and judge the category of pixel in the HRS images with the Softmax classifier. These EK information including typical spatial features can be efficiently and easily obtained through the use of Image Processing Toolbox in MATLAB. Experiments show that the classification accuracy of Indian Pines data set can exceed 95% with 510 training samples and 9856 testing samples at the optimal number of iterations. And we did not deal with the sample size of unbalanced classes for Indian Pines data set by using percentage methods.

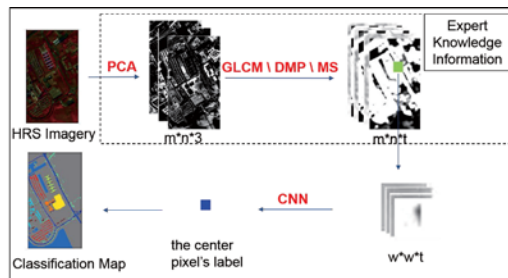


Fig. 2. Pipeline of the proposed deep feature extraction of HRS images

## 2.2 Expert Knowledge Selection

Spectrum and space information is the most concerned and widely used characteristics in pixel-wise HRS classification [16]. In this paper, three main feature extraction techniques are used to obtain EK information: principal component analysis (PCA), gray level co-occurrence matrix (GLCM) and the differential morphological profiles (DMP). The feature parameter values are shown in Table 1.

**Table 1.** The introduction of the feature parameters

Features	Symbol	Instructions	Values
	$B$	Basic images	/
$f_{pca}(B, P)$	$P$	The number of principal components	3
	$T$	The texture features composed of contrast and homogeneity	/
$f_{glcm}(B, T, D, W)$	$D$	The directions of the texture features	$[0^\circ, 45^\circ, 90^\circ, 135^\circ]$
	$W$	The size of the local area window	$[11, 13, 15, 17, 19, 21, 23, 25, 27]$
$f_{dmp}(B, \lambda)$	$\lambda$	The radius of the disk-shaped SEs	$[0, 3, 5, 7, 9, 11]$
$f_{ms}(B, P)$	$T$	The feature of DMP feature and GLCM feature fusion	/

As we all know, HRS images like PaviaU dataset contains rich spectral information with hundreds of bands, which may easily lead to curse of dimensionality for the small ratio between the number of the samples and the number of the features. In our framework, the principal component analysis (PCA) is introduced to extract spectral features firstly for its simple mathematical principles and effective performance. PCA can effectively reduce the data dimension and eliminate the potential threat of dimensionality curse by eliminating the redundant information between HRS bands. The first three principal components for each image are used to present the spectral feature because the first three principal components contain more than 99% of the original image information in our experiments. The feature image acquired based on PCA will be served as a basic image for subsequent processing in the framework.

GLCM CNN-based deep extraction (GLCM-CDE) method, which use GLCM features as the input of the CNN, is proposed for two reasons. For the first one, the GLCM is typical method of texture feature extraction for classification of remote sensing images. From a statistical point of view, the texture information extracted by GLCM can reflect the regularity of distribution of ground objects in the spatial scene, which can not be directly obtained from CNN. HRS image spatial arrangement characteristics obtained by GLCM as the input of the CNN, is not only conducive to obtain the abstractive semantic features, but also enhance the interpretability of the deep features due to the statistical theory. GLCM-based features that mainly preserve texture information greatly reduce the complexity of the original data distribution, which helps to reduce the deep CNN structure complexity and the sample size limitation. The second reason is to consider the effect of patch size of the CNN input data on the classification accuracy and the deep features characterization ability of the deep CNN. The GLCM texture feature is calculated based on local area information and the size of the local area has an important influence on the characterization ability of the feature. We can assume a strong correlation between window size in GLCM that closely related to the local features of the image and patch size of the CNN. Therefore the GLCM-based feature extraction method can be used to explore the contribution of the patch size for the deep feature and the HRS images classification performance. This deep feature extraction method that uses GLCM features as input to the CNN is called GLCM-CDE in this paper.

Morphological profiles is constructed by a series of morphological operations with a family of structuring elements of increasing size [10]. These structure-specific operators that getting such geometric features is not an easy task by training a CNN networks at the bottom

layer. Through the operation of these morphological operator, MP feature not only simplifies the image data, preserves the basic shape information of the image, but also can remove irrelevant structures and remove noise effectively. In order to further enhance the heterogeneity among different objects in the scene, differential morphological profiles based on the morphological gradient theory is used as the input of the CNN in this paper, which has a great impact on the boundary extraction and image segmentation. The introduction of the DMP feature, that mainly retaining the boundary information and fully eliminate unnecessary information, into our deep model as the input of CNN can help to reduce the demand for the model complexity while extracting an effective high-level feature. DMP-based deep geometry features obtained by the proposed CNN further enhance the differences between different ground objects and improve the interpretability of deep features. In this article, the differential morphological profiles-based CNN deep extraction method is referred to as DMP-CDE.

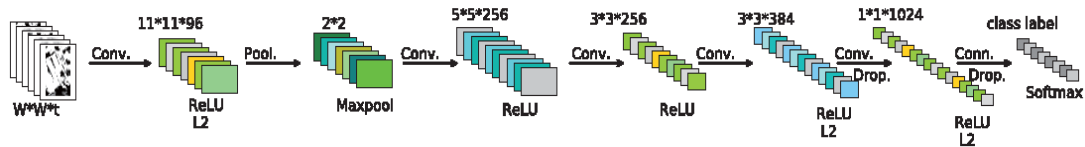
According to the mathematical principle of different features, the features based on GLCM excavated the arrangement rules of ground objects and the DMP features can measure the distance between different ground objects. That means, the features based on the GLCM and the DMP respectively reflect the different attributes of HRS images. It is natural to fuse the DMP feature and the GLCM feature together for the HRS classification because the effective fusion of different features can enhance the difference between different ground objects [20]. The vector stacking method is used to fuse these two spatial features together. Although the amount of the feature data as the input of CNN is slightly increased, the redundancy between data can be effectively controlled because of the difference of attributes between the GLCM-based texture feature and DMP-based shape feature, so that we can further improve the differences between different objects and get an promising classification result on the proposed CNN network. The deep feature extraction method that takes the fused data as CNN input is called multi selected feature-based CNN deep extraction method (MS-CDE), which further enhance the CNN network representation and train a more generalized and robust CNN without increasing the cost of the framework. Based on the direct fusion of GLCM-CDE feature and DMP-CDE feature is not discussed here for three reasons: the first is a challenge for the classifier due to the amount of fusion data, the second is to change the complexity and uniformity of the proposed framework and the third is that the effect of the fusion of deep features at different levels on the classification result is uncertain.

### 2.3 Convolutional Neural Network for EK Information

Convolution neural network is a common network structure in deep learning theory. It consists of multiple convolutional layers and pooling layers, and the input data will be represented by learning the corresponding weights and offsets. Compared with Deep Belief Network (DBN), Deep Auto-Encoder (DAE) and other neural network structures [36], the use of sparse connectivity and weight sharing mechanism within the CNN effectively control the size of network parameters. Deep CNN network has attracted the attention of researchers in the field of remote sensing due to its promising performance in feature extraction and image classification. In this paper, the CNN is used to handle the pixel-wise HRS classification for EK information. The proposed CNN structure is shown in Fig. 3. This neural network containing only five convolutional layers, a pooling layer and a fully\_connected layer for classification has a simpler structure than the classical networks such as AlexNet and VGGNet because of the greatly simplified CNN input data distribution characteristics in our framework. Table 2 shows the difference of the network structures among these three deep neural networks. From the table we can see that the proposed network has a simpler structure and less parameters. In addition, the ReLU function [15, 17] is used in each convolutional layer as the



activation function to extract multilayer abstract features. The max pooling strategy is used to extract the significant characteristics of the data while reducing the computational cost. We improve the abstraction level of features by reducing the size of the convolution kernel layer by layer. Instead of the full connected layer in the traditional CNN structure, convolutional layer with a  $1 \times 1$  convolution kernel [18] is employed to reduce the number of training parameters.  $L_2$  regularization and Dropout strategies [19] are utilized during training process in order to improve the generalization of the CNN network.  $L_2$  regularization is able to minimize the cost function in the training stage by making the sum of the squares of the parameter small. Dropout strategy reduces overfitting in neural networks by preventing complex co-adaptations on training dataset. End-to-end training patterns are used to learn weights and offset values to extract the abstract features of HRS images and improve the classification performance.



**Fig. 3.** The structure of the CNN employed in our experiments

**Table 2.** The comparison of the network structures among the proposed CNN (ProCNN), AlexNet and VGGNet

	ProCNN			AlexNet			VGGNet		
	kernel size	feature number	stride	kernel size	feature number	stride	kernel size	feature number	stride
conv_2d	11*11	96	2	11*11	96	4	3*3	64	1
conv_2d							3*3	64	1
max_pool	2*2		2	3*3		2	2*2		2
conv_2d	5*5	256	1	5*5	256	1	3*3	128	1
conv_2d	3*3	256	1				3*3	128	1
max_pool				3*3		2	2*2		2
conv_2d	3*3	384	1	3*3	384	1	3*3	256	1
conv_2d	1*1	1024	1	3*3	384	1	3*3	256	1
conv_2d				3*3	256	1	3*3	256	1
max_pool				3*3		2	2*2		2
conv_2d							3*3	512	1
conv_2d							3*3	512	1
conv_2d							3*3	512	1
max_pool							2*2		2
conv_2d							3*3	512	1
conv_2d							3*3	512	1
conv_2d							3*3	512	1
max_pool							2*2		2
fully_connected				1*1	4096		1*1	4096	

fully_connected				1*1	4096		1*1	4096	
fully_connected (softmax)	1*1	class number		1*1	class number		1*1	class number	
Layer number	6			8			16		

### 3. Experiments and Results

#### 3.1 Data Description and Evaluation Criteria

In our work, three HRS images are used to evaluate the proposed method: Pavia University (PaviaU) dataset ([Fig. 4\(a\)](#)), Salinas dataset ([Fig. 4\(b\)](#)) and Indian Pine dataset ([Fig. 4\(c\)](#)). The first image is an urban scene and the other two are agricultural scenes, each with different spectral and spatial information. In our experiments 5% of the samples was selected for training and the cross-validation method was used to adjust the parameters of the classifier during the training process. The remaining 95% of the samples were used to test the performance of the algorithm.

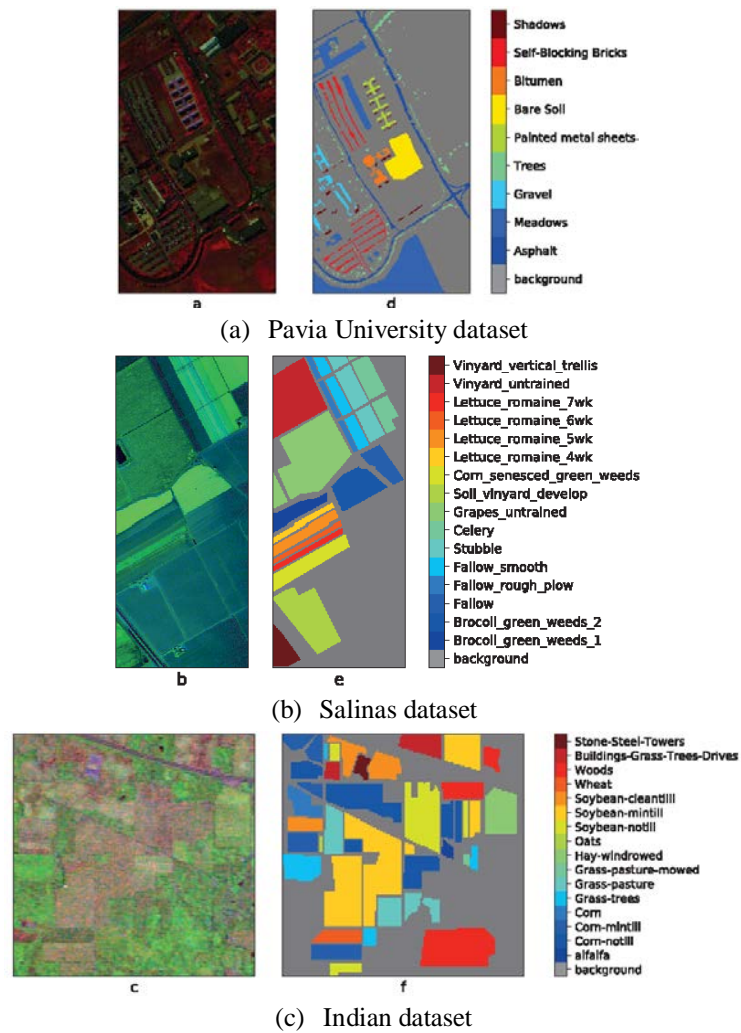
The Pavia University scene was acquired by the Reflective Optics System Imaging Spectrometer (ROSIS) sensor during a flight campaign over Pavia, northern Italy in 2003. The urban scene with 1.3 m/pixel spatial resolution contains  $610 \times 340$  pixels and 115 spectral bands. 103 available spectral bands were used in the experiments. [Fig. 4\(a\)](#) shows the false color image of PaviaU dataset and the corresponding ground truth map with 9 classes respectively. [Table 3](#) provides the numbers of training and testing samples with different classes.

The Salinas scene was collected by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor over Salinas Valley, California, in 1998. This agricultural area covered comprises 512 lines by 217 samples with 3.7-meter pixels spatial resolution. 204 available spectral bands of the Salinas dataset were used in the experiments. There are 16 kinds of objects in the ground truth map ([Fig. 4\(b\)](#)) and the number of training and testing samples with each class is provided in [Table 4](#).

The Indian Pine dataset was gathered by AVIRIS sensor over the Indian Pines test site in North-western Indiana in 1992. This agricultural scene is widely used to measure the performance of HRS image classification algorithm for the unbalanced number of available labeled pixels of each class and the presence of mixed pixels. It consists of  $145 \times 145$  pixels and 224 available spectral bands with 20 m/pixel spatial resolution. The number of bands is reduced to 200 by removing bands for the noise and water absorption. The raw dataset and the corresponding ground truth map with 16 classes are showed in [Fig. 4\(c\)](#). [Table 5](#) provides the numbers of training and testing samples with different classes.

The evaluation criteria for these classification algorithms in this work are overall classification accuracy (OA), per-class average accuracy (AA) and Kappa coefficient (Kappa).





**Fig. 4.** HRS data sets employed in our experiments. a, b and c are images for classification evaluation; d, e and f are the ground truth for these images left to them respectively

**Table 3.** Training and testing pixels for Pavia University dataset

Classes	Training	Testing	Total
Asphalt	331	6300	6631
Meadows	932	17717	18649
Gravel	104	1995	2099
Trees	153	2911	3064
Painted metal sheets	67	1278	1345
Bare Soil	251	4778	5029
Bitumen	66	1264	1330
Self-Blocking Bricks	184	3498	3682
Shadows	47	900	947
Total number	2135	40641	42776

**Table 4.** Training and testing pixels for Salinas dataset

Classes	Training	Testing	Total
Brocoli_green_weeds_1	100	1909	2009
Brocoli_green_weeds_2	186	3540	3762
Fallow	98	1878	1976
Fallow_rough_plow	69	1325	1394
Fallow_smooth	133	2545	2678
Stubble	197	3762	3959
Celery	178	3401	3579
Grapes_untrained	563	10708	11271
Soil_vinyard_develop	310	5893	6203
Corn_senesced_green_weeds	163	3115	3278
Lettuce_romaine_4wk	53	1015	1068
Lettuce_romaine_5wk	96	1831	1927
Lettuce_romaine_6wk	45	871	916
Lettuce_romaine_7wk	53	1017	1070
Vinyard_untrained	363	6905	7268
Vinyard_vertical_trellis	90	1717	1807
Total number	2697	51432	54129

**Table 5.** Training and testing pixels for Indian Pine dataset

Classes	Training	Testing	Total
alfalfa	2	52	54
Corn-notill	71	1363	1434
Corn-mintill	41	793	834
Corn	11	223	234
Grass-trees	24	473	497
Grass-pasture	37	710	747
Grass-pasture-mowed	1	25	26
Hay-windrowed	24	465	489
Oats	1	19	20
Soybean-notill	48	920	968
Soybean-mintill	123	2345	2468
Soybean-cleantill	30	584	614
Wheat	10	202	212
Woods	64	1230	1294
Buildings-Grass-Trees-Drives	19	361	380
Stone-Steel-Towers	4	91	95
Total number	510	9856	10366

### 3.2 The Contribution of the Expert Knowledge to HRS Classification Performance

In order to measure the contribution of handcrafted features, we compared the classification accuracy between the original data and the handcrafted features with the structure of our proposed CNN network. The data sets for all experiments were stored in h5 format.

**Table 6** shows the total size of the training set and testing set as a network input as well as the classification results based on the proposed CNN. As can be seen from the **Table 6**, the

amount of the training set and testing set as the input of CNN based on the GLCM feature or the DMP feature is far less than the original image because handcrafted features are able to represent the original images and eliminate redundant information and undesired noises. Although the amount of the training set and testing set based on the MS feature is slightly higher than the original image on PaviaU dataset, the redundancy and the complexity of data distribution can be effectively controlled because of the difference of attributes between the GLCM-based texture feature and DMP-based shape feature. All these features improve the starting point of the deep features training. The classification results based on handcrafted features with the proposed CNN are much higher than those based on the original images in [Table 6](#). The optimal classification results are obtained on the each dataset based on the MS-CDE feature.

**Table 6.** The performance comparison between the raw images and the handcrafted features based on the CNN

Datasets	CNN Input	Training Size (MB)	Testing Size (GB)	OA	Kappa
PaviaU	RAW	188	3.51	97.7%	0.97
	GLCM	88.1	1.63	98.60%	0.981
	DMP	110	2.04	98.9%	0.985
	MS	198	3.68	99.2%	0.989
Salinas	RAW	472	8.79	97%	0.966
	GLCM	111	2.07	99.3%	0.993
	DMP	139	2.59	99.1%	0.99
	MS	250	4.66	99.8%	0.998
Indian Pine	RAW	112	2.12	84.1%	0.818
	GLCM	27	0.51	95%	0.943
	DMP	33.8	0.64	94.6%	0.938
	MS	60.7	1.14	95.4%	0.948

There are some reasons for the poor results based on the original HRS images. Because of the large amount and complex structure of the original HRS images as well as the simpler structure of the network we proposed, it is difficult to obtain satisfactory classification results on the original data. From the [Table 6](#) we can also see that the classification accuracy on the Indian Pine dataset are worse than on other two datasets because of the more complex data attributes (more details in section 3.1).

### 3.3 Effects of Different Network Structures on Classification Results

In order to evaluate the contribution of the proposed CNN structure, we compare the classification results obtained by different deep convolution neural network: Alexnet, VGGNet and the proposed CNN in Section 2.3. And the difference of network structures between these three deep CNN is shown in [Table 2](#). The raw data and three different handcrafted features including GLCM, DMP and MS are used as the input of each network, respectively.

A summary of findings is presented in [Table 7](#). As can be seen from the Table, the deep features with the proposed CNN achieves higher classification performance than other neural networks for each dataset. The proposed CNN and the VGGNet have similar ability to distinguish between ground objects for PaviaU and Salinas dataset. A large number of small convolution kernels are used in the VGGNet structure, thus making the VGG network have better feature extraction capabilities and achieving good results on these datasets with high spatial resolution. However, for the Indian Pine dataset with more complex distribution

characteristics, VGGNet is difficult to achieve a satisfactory result. The overall classification result for the Indian Pine dataset has made a more significant improvement compared to AlexNet and VGGNet networks. The kappa coefficient values for each dataset from the [Table 7](#) show the effectiveness of our proposed network for HRS image classification. The high-level abstraction features are extracted by gradually reducing the size of the convolution kernel in the proposed CNN and the  $L_2$  regularization and dropout strategy are used in multiple convolutional layers in our network to obtain a better generalization performance.

**Table 7.** Classification accuracy and Kappa coefficient based on different neural network with the optimal number of epochs

Classification performance on PaviaU Dataset (a)

Pavia University	Raw data			GLCM-based deep feature		
	AlexNet	VGGNet	ProCNN	AlexNet	VGGNet	ProCNN
Asphalt	90.70%	95.80%	97.3%	98%	96.30%	97.90%
Meadows	99.80%	99.40%	99.7%	99.60%	99.60%	99.60%
Gravel	59.10%	85.40%	87.9%	90.20%	95.40%	97.30%
Trees	92.20%	99.90%	98.4%	93.60%	93.20%	96.30%
Painted metal sheets	100%	99.80%	99.8%	93.50%	92.90%	98.60%
Bare Soil	77.90%	88.40%	97.9%	96.90%	97.50%	99.20%
Bitumen	67.20%	82.70%	84.8%	90.10%	92.10%	99.40%
Self-Blocking Bricks	76.80%	97.30%	97.1%	95.80%	98.30%	99%
Shadows	91.40%	97.30%	98.1%	73.70%	69%	87.40%
OA	90.10%	96.20%	97.7%	96.70%	96.90%	98.60%
Kappa	0.867	0.949	0.97	0.957	0.959	0.981

Classification performance on PaviaU Dataset (b)

Pavia University	DMP-based deep feature			MS-based deep feature		
	AlexNet	VGGNet	ProCNN	AlexNet	VGGNet	ProCNN
Asphalt	90.8%	97.9%	99%	97.4%	98.7%	100%
Meadows	97.1%	99.9%	99.7%	99.7%	99.7%	99.8%
Gravel	70.1%	98.1%	92.5%	96.1%	96.8%	96.1%
Trees	90.8%	97.6%	98.4%	96.3%	98.2%	98.3%
Painted metal sheets	99.1%	99.8%	99.3%	98.1%	99.8%	99.2%
Bare Soil	73.6%	98.6%	99.8%	100%	99.5%	99.8%
Bitumen	55.5%	97.5%	95.5%	93%	95.7%	98.8%
Self-Blocking Bricks	88.7%	96.7%	98.2%	99.2%	97.7%	97.3%
Shadows	70.4%	90.1%	97.2%	90%	90.9%	95.9%
OA	89%	98.6%	98.9%	98.4%	98.8%	99.2%
Kappa	0.854	0.982	0.985	0.979	0.984	0.989

Classification performance on Salinas Dataset (a)

Salinas	Raw data			GLCM-based deep feature		
	AlexNet	VGGNet	ProCNN	AlexNet	VGGNet	ProCNN
Brocoli_green_weeds_1	96%	99.7%	99.4%	99.3%	99.2%	99.8%
Brocoli_green_weeds_2	97.7%	99.3%	99.7%	98.9%	100%	100%
Fallow	79.3%	99.3%	99.8%	99.3%	99.1%	99.8%
Fallow_rough_plow	93.4%	100%	100%	96.8%	97.4%	99.5%
Fallow_smooth	95.4%	96.7%	99.3%	94.8%	98.5%	99.1%
Stubble	98.7%	100%	100%	99.6%	99.9%	99.9%
Celery	98.7%	99.3%	100%	99.8%	99.9%	100%
Grapes_untrained	74%	93%	87.1%	97.6%	98.6%	99.3%
Soil_vinyard_develop	99.2%	99.3%	99.4%	99.4%	100%	99.9%
Corn_senesced_green_weeds	83.2%	95.4%	99.7%	100%	99.8%	100%
Lettuce_roumaine_4wk	74%	98.3%	98.1%	96.9%	100%	99%
Lettuce_roumaine_5wk	96.6%	95.5%	99.8%	99.5%	99.5%	100%
Lettuce_roumaine_6wk	92.4%	98.2%	100%	97.5%	97.8%	99.5%
Lettuce_roumaine_7wk	74.7%	98.8%	99.7%	98.6%	97.9%	100%
Vinyard_untrained	92.1%	71%	99.6%	99.1%	93.8%	96.9%
Vinyard_vertical_trellis	87.3%	99.5%	97.8%	98.1%	100%	100%
OA	89.1%	93.7%	97%	98.6%	98.6%	99.3%
Kappa	0.879	0.93	0.966	0.984	0.984	0.993

Classification performance on Salinas Dataset (b)

Salinas	Raw data			GLCM-based deep feature		
	AlexNet	VGGNet	ProCNN	AlexNet	VGGNet	ProCNN
Brocoli_green_weeds_1	88.6%	100%	99.5%	97.2%	100%	100%
Brocoli_green_weeds_2	59.9%	99.3%	99.9%	99.7%	100%	100%
Fallow	55.8%	98.4%	99.7%	99.1%	99.6%	100%
Fallow_rough_plow	87.9%	99.2%	100%	97.7%	98.9%	99.9%
Fallow_smooth	66.2%	99.5%	99.2%	96%	98.9%	98.8%
Stubble	58.5%	99.9%	99.4%	99.9%	99.9%	99.9%
Celery	68.7%	99.5%	99.7%	98.5%	99.2%	100%
Grapes_untrained	84.5%	99.9%	98.6%	99.5%	99.4%	99.6%
Soil_vinyard_develop	90.8%	99.7%	98.6%	98.6%	99.9%	100%
Corn_senesced_green_weeds	58.9%	99.6%	99.3%	99.7%	100%	100%
Lettuce_roumaine_4wk	95.3%	99.9%	99%	98.7%	99.9%	99.4%
Lettuce_roumaine_5wk	56.9%	99.6%	100%	99.6%	99.9%	100%
Lettuce_roumaine_6wk	62.6%	98%	100%	94.8%	97.7%	98.7%
Lettuce_roumaine_7wk	91.6%	100%	100%	100%	99.9%	100%
Vinyard_untrained	75.8%	93.7%	98.4%	97.7%	99.1%	99.9%

Vinyard_vertical_trellis	61%	98.7%	98%	98.8%	100%	100%
OA	74.4%	98.8%	99.1%	98.7%	99.6%	99.8%
Kappa	0.715	0.987	0.99	0.986	0.995	0.998

Classification performance on Indian Pine Dataset (a)

Indian Pine	Raw data			GLCM-based deep feature		
	AlexNet	VGGNet	ProCNN	AlexNet	VGGNet	ProCNN
alfalfa	57.7%	40.4%	55.8%	86.5%	78.8%	88.5%
Corn-notill	81.6%	64%	78.1%	96.2%	94.4%	95.2%
Corn-mintill	73.9%	52.6%	76.2%	90.4%	84.7%	92.7%
Corn	62.3%	47.1%	91.9%	95.5%	99.6%	100%
Grass-trees	67.7%	60%	65.3%	82.9%	83.7%	87.1%
Grass-pasture	82.1%	83%	93.7%	89%	88.9%	91.8%
Grass-pasture-mowed	0%	72%	0%	100%	100%	100%
Hay-windrowed	96.8%	91%	99.4%	100%	99.6%	100%
Oats	52.6%	63.2%	42.1%	63.2%	89.5%	78.9%
Soybean-notill	62.2%	64.7%	78.7%	94.6%	84.9%	93.6%
Soybean-mintill	81.7%	77.9%	91.3%	96.2%	97.7%	98.9%
Soybean-cleantill	64.2%	71.9%	67.3%	84.8%	91.1%	88.4%
Wheat	83.2%	88.6%	85.1%	97%	97.5%	97.5%
Woods	97.2%	90.4%	91.7%	97.4%	97%	97.6%
Buildings-Grass-Trees-Drives	75.6%	77.8%	85.3%	93.1%	81.4%	93.1%
Stone-Steel-Towers	53.8%	44%	85.7%	74.7%	68.1%	71.4%
OA	78.9%	73%	84.1%	93.6%	92.4%	95%
Kappa	0.758	0.691	0.818	0.928	0.913	0.943

Classification performance on Indian Pine Dataset (b)

Indian Pine	DMP-based deep feature			MS-based deep feature		
	AlexNet	VGGNet	ProCNN	AlexNet	VGGNet	ProCNN
alfalfa	50%	80.8%	75%	88.5%	98.1%	90.4%
Corn-notill	74.5%	96.6%	93.3%	97%	95.4%	94.6%
Corn-mintill	75.5%	93.6%	94.6%	92.6%	88.9%	91.6%
Corn	90.6%	90.6%	94.6%	96.9%	100%	100%
Grass-trees	80.1%	91.3%	90.1%	80.8%	84.4%	93.2%
Grass-pasture	93.1%	89%	94.8%	96.5%	92.3%	90.7%
Grass-pasture-mowed	24%	100%	76%	100%	80%	16%
Hay-windrowed	90.1%	99.6%	100%	97%	100%	100%
Oats	5.3%	0%	52.6%	78.9%	89.5%	94.7%
Soybean-notill	85%	89.9%	92.7%	91.1%	91.5%	93.7%
Soybean-mintill	76.3%	93.3%	96.2%	93.7%	95.6%	98.7%



Soybean-cleantill	82.7%	83.2%	88.7%	90.4%	87.2%	90.6%
Wheat	99%	99.5%	98%	97.5%	94.1%	98.5%
Woods	90.7%	93.2%	99.3%	98.6%	98.7%	100%
Buildings-Grass-Trees-Drives	89.5%	91.7%	95.8%	90.6%	85.6%	93.1%
Stone-Steel-Towers	95.6%	91.2%	71.4%	59.3%	76.9%	84.6%
OA	82.1%	92.5%	94.6%	93.7%	93.4%	95.4%
Kappa	0.798	0.915	0.938	0.928	0.925	0.948

In order to illustrate the contribution of the proposed CNN in our framework further, we compared the time overhead and the classification performance among the proposed CNN, VGGNet and Alexnet. As we all know that the computational time is influenced by three main factors: experimental platform (software and hardware equipment), experimental data size and classification algorithm. Thus experiments on a specific dataset based on the same handcrafted features should be running on the same platform. There are two computers used during our experimental process. The first computer has two NVIDIA GeForce GTX 1080 Ti and an i7-5930K CPU while the second has a NVIDIA GeForce GTX 960 and an i7-4790K CPU.

Four datasets for each HRS image were used as the input of the deep CNN: raw data, GLAM-based feature, DMP-based feature and MS-based feature. The related experiments of the latter two datasets are running only on the first computer while the experiments on the former two data are running on the two computers. So the valid results for the time consuming is just on the DMP-based feature and MS-based feature. In order to compare the performance of different CNN, the time consuming from the CNN training process to the testing stage are recorded and the number of epochs is 500 for each experiments. The results are shown in [Table 8](#).

**Table 8.** The comparison of computational time and classification performance among AlexNet , vggNet and the proposed CNN (ProCNN)

DMP-based feature were used as the input of CNNs (a)

DMP-based Features	Pavia University dataset			Salinas dataset			Indian Pine dataset		
	AlexNet	vggNet	ProCNN	AlexNet	vggNet	ProCNN	AlexNet	vggNet	ProCNN
Time overhead/s	584	620	548	784	822	720	139	165	130
OA	87%	97.3%	98.3%	63.6%	98.2%	96.9%	77.5%	71.8%	90.7%
kappa	0.828	0.964	0.977	0.594	0.98	0.966	0.746	0.682	0.893

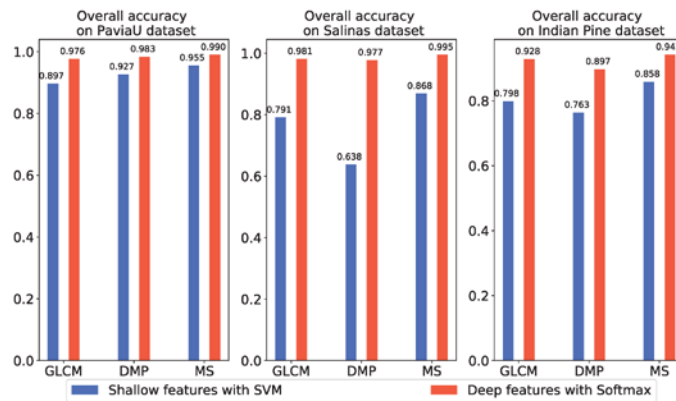
MS-based feature were used as the input of CNNs (b)

MS-based Features	Pavia University dataset			Salinas dataset			Indian Pine dataset		
	AlexNet	vggNet	ProCNN	AlexNet	vggNet	ProCNN	AlexNet	vggNet	ProCNN
Time overhead/s	669	758	651	862	1006	733	149	176	146
OA	95.6%	98.2%	98.9%	96.8%	98.7%	98.5%	86.4%	90.4%	93.3%
kappa	0.941	0.977	0.986	0.964	0.985	0.983	0.845	0.89	0.924

### 3.4 The Relationship between the Deep Features and Corresponding Handcrafted Features on Classification Results

This section explores the relationship between different handcrafted features and corresponding deep features on the image classification results. Three different handcrafted features were used to perform the experiment: texture feature extracted using the GLCM algorithm, shape feature extracted by the DMP algorithm, and the multi handcrafted feature based on the fusion of the GLCM algorithm and the DMP algorithm. **Fig. 8** compares the classification performance of two different classifiers: the deep feature classification results obtained based on the proposed CNN structure and the corresponding handcrafted feature classification results using RBF-SVM classifier. SVM is one of the most widely used classifiers in HRS image classification [27, 29, 28]. It can effectively solve complex nonlinear problems and get promising results by using kernel functions and be easy to generate classification results by using the SVM library directly.

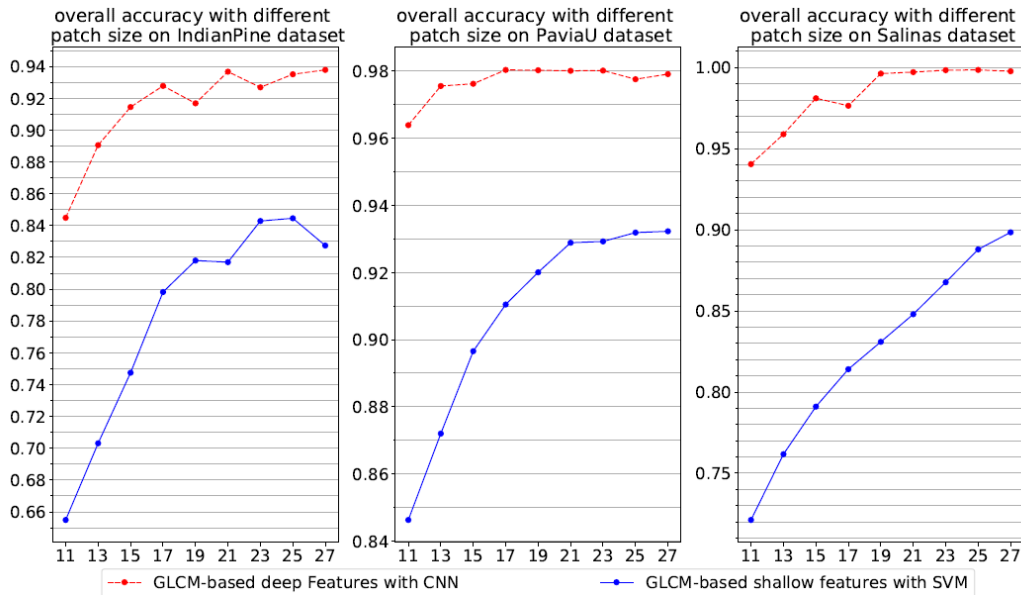
As can be seen that the results of deep features are much higher than the corresponding handcrafted features, because deep learning-based methods can extract the high-level abstract and invariant features of ground objects that help to change the intra- and inner- class differences. By comparing the classification results based on different handcrafted methods, it can be found that the stronger the representation ability of the handcrafted feature is, the higher the classification accuracy will be with the corresponding deep features.



**Fig. 5.** Comparison of classification results using different spatial information between handcrafted features and deep features

### 3.5 The Impact of Patch Size for the Input of CNN

CNN-based deep feature extraction method requires a fixed-size patch as input during training and testing phases. In this experiment, the GLCM-CDE feature is utilized to explore the patch size effect on deep feature learning performance for the correlation between window size in the GLCM feature extraction method and patch size of the CNN input. Then a series of different patch/window size were used and we imply the softmax classifier with GLCM-CDE feature and RBF-SVM classifier with GLCM-based handcrafted feature for HRS datasets.



**Fig. 6.** Comparison of classification results using different window size between handcrafted features and the GLCM-based CNN deep features

The effect of the patch/window size on HRS images classification is provided in [Fig. 6](#). As can be seen that the features extracted by the deep CNN significantly improve the classification accuracy compared with the GLCM-based handcrafted features. By observing, we can find that the classification performance of GLCM-CDE features tends to be stable earlier on the relatively smaller window than the classification performance of handcrafted features. The larger the difference between the window size and the average size of the ground objects is, the more obvious the performance improvement of the classification result based on the GLCM-CDE features is. This means that the features obtained by the CNN network can effectively reduce the distance of the internal homogeneous information of the ground objects and improve the classification results of the HRS images. The square patches of  $15 \times 15$  are used for PaviaU and Salinas datasets and  $17 \times 17$  are used for Indian Pine dataset as the CNN input in all other experiments.

### 3.6 The Effect of Iteration Times

Based on the neural network derived from the data, the training results (network structure, weight and offset) is a representation of the sample data distribution. The application of BP algorithm iterates iteratively to fine-tune the network weights, which enables the network to simulate the distribution characteristics of the overall data through the sample data. Therefore, the degree of learning of the training samples affects the classification result of the data. If the number of training iterations is insufficient, that is, failure to make full use of the sample information, it is difficult to accurately describe the distribution characteristics of the data. On the contrary, if the number of iterations is too large, the network contains information unique to the training samples, which leads to a decrease in the distribution performance of the HRS images by the network. Therefore, we train the neural network with different number of iterations to observe the impact on classification performance: ( $n\_epoch = [500; 1000; 1500; 2000; 3000; 5000; 10000]$ ). The relationship between the epoch number and the optimal classification result is presented in [Table 9](#).

**Table 9.** The relationship between the epoch number and the optimal classification results

Raw data were used as the input of CNNs (a)

Raw data	Pavia University dataset			Salinas dataset			Indian Pine dataset		
	AlexNet	vggNet	ProCNN	AlexNet	vggNet	ProCNN	AlexNet	vggNet	ProCNN
n_epoch	2000	2000	2000	2000	1500	2000	3000	3000	3000
OA	90.1%	96.2%	97.7%	89.1%	93.7%	97%	78.9%	73%	84.1%
kappa	0.867	0.949	0.97	0.879	0.93	0.966	0.758	0.691	0.818

GLCM-based Feature were used as the input of CNNs (b)

GLCM-based Feature	Pavia University dataset			Salinas dataset			Indian Pine dataset		
	AlexNet	vggNet	ProCNN	AlexNet	vggNet	ProCNN	AlexNet	vggNet	ProCNN
n_epoch	1500	1000	2000	2000	1500	1500	1500	1000	2000
OA	96.7%	96.9%	98.6%	98.6%	98.6%	99.3%	93.6%	92.4%	95%
kappa	0.957	0.959	0.981	0.984	0.984	0.993	0.928	0.913	0.943

DMP-based Feature were used as the input of CNNs (c)

DMP-based Feature	Pavia University dataset			Salinas dataset			Indian Pine dataset		
	AlexNet	vggNet	ProCNN	AlexNet	vggNet	ProCNN	AlexNet	vggNet	ProCNN
n_epoch	1500	1500	1500	5000	1500	2000	1500	2000	1500
OA	89%	98.6%	98.9%	74.4%	98.8%	99.1%	82.1%	92.5%	94.6%
kappa	0.854	0.982	0.985	0.715	0.987	0.99	0.798	0.915	0.938

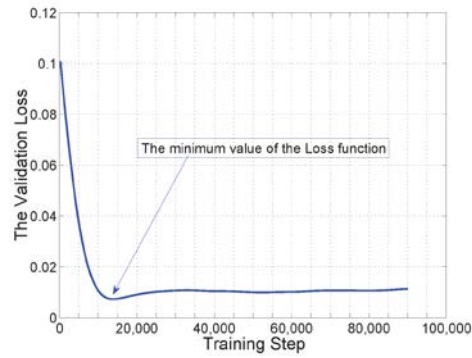
MS-based Feature were used as the input of CNNs (d)

MS-based Feature	Pavia University dataset			Salinas dataset			Indian Pine dataset		
	AlexNet	vggNet	ProCNN	AlexNet	vggNet	ProCNN	AlexNet	vggNet	ProCNN
n_epoch	2000	1000	1500	2000	2000	1500	1500	1500	2000
OA	98.4%	98.8%	99.2%	98.7%	99.6%	99.8%	93.7%	93.4%	95.4%
kappa	0.979	0.984	0.989	0.986	0.995	0.998	0.928	0.925	0.948

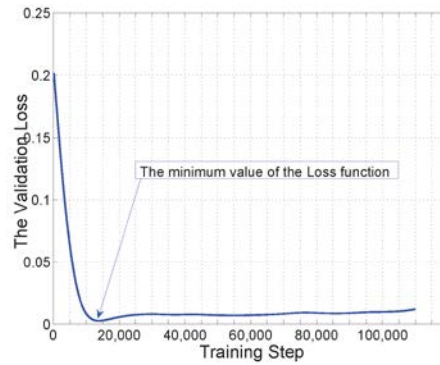
**Fig. 7** shows the curve of the loss function obtained by the method based on MS-CDE feature on the validation dataset during training process with each image. Obviously there will be over-fitting phenomenon when the number of iterations exceeds a certain value. By observing the trend of the curve, we can find the optimal value of the training epoch and get the best classification result. The conversion between training steps and the value of the epoch is calculated as follows. The ratio represents the proportion of the validation set in the training samples during training process and  $n$  represents the number of training samples:

$$step = \left\lceil \frac{(1-ratio)*n}{batch\_size} \right\rceil * n\_epoch \quad (1)$$

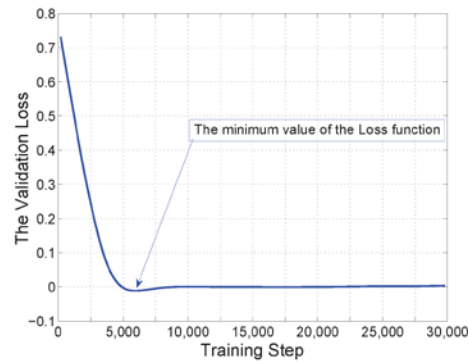
According to **Fig. 7** and the above Eq. (1) a suitable value of the epoch can be obtained to training the network for each dataset and get an optimal classification result for HRS images. **Fig. 8** shows the classification performance of each class of the ground objects under different iterations. It can be seen that in a well-trained network, the difference between the different ground objects can be effectively improved. The classification performance with optimal epoch based on MS-CDE feature is provided in **Fig. 9**.



(a) The curve of the loss value on the PaviaU dataset

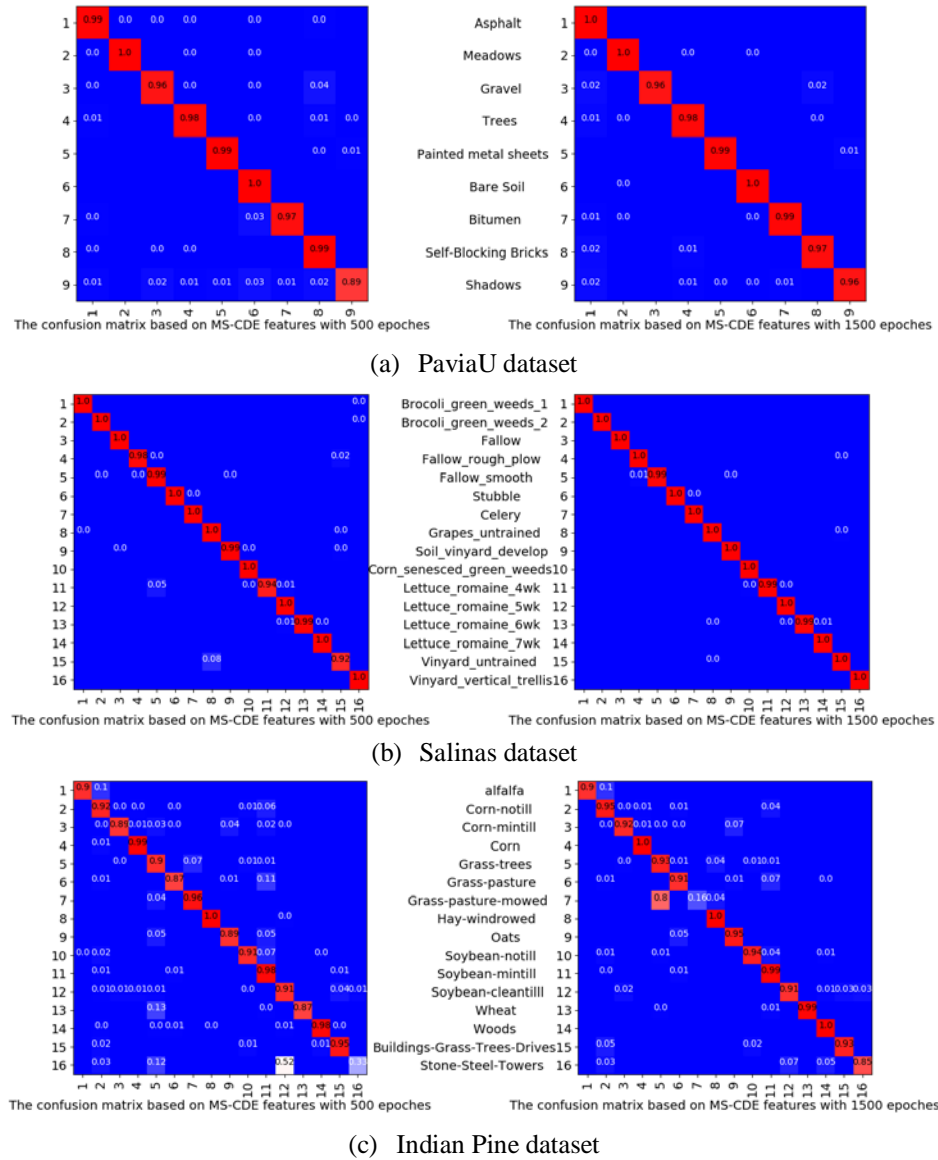


(b) The curve of the loss value on the Salinas dataset



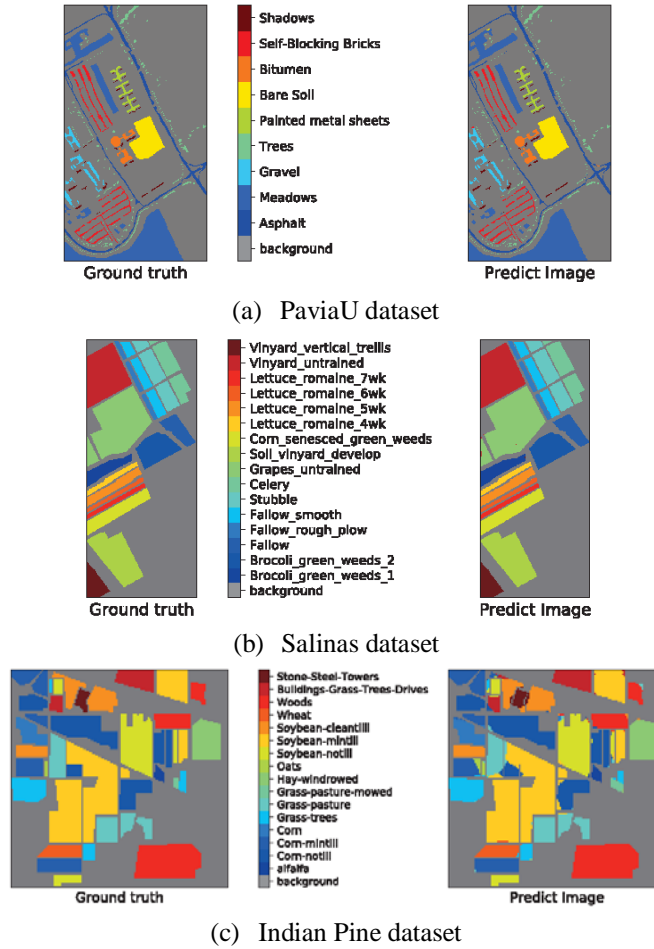
(c) The curve of the loss value on the Indian dataset

**Fig. 7.** The relationship between training epoches and the CNN classification performance



**Fig. 8.** per-class classification result with different epoch





**Fig. 9.** Visualization of the classification performance based on the MS-CDE feature with the optimal epoch

## 4. Conclusion

In this paper, the proposed approaches, fusion of the traditional handcrafted and deep features, are used to handle the inherent challenges in HRS classification. The rationale is that data obtained from the handcrafted features based on expert knowledge only contain the primary information of the original HRS image, with a simplified data distribution. This, in turn, reduces the need for a CNN network structure parameter scale. Spectral features based on the PCA algorithm are used as the basic map with a low dimension. Then, GLCM-based texture features and shape features based on DMP are extracted as the shallow spectral-spatial features for the deep CNN network input. Meanwhile, considering the differences in recognition ability of different handcrafted features on different ground objects, the vectorstacking method is used to further improve the classification performance of the HRS images with the proposed deep CNN.

These handcrafted feature maps, used as the input of CNN, not only improve the training starting point of the deep features, but also offer better guidance and interpretability to the deep features. A simple neural network structure suitable for these HRS images was then

proposed to obtain high-level abstraction features according to the characteristics of these handcrafted data and the theory of convolution neural network, so that satisfactory classification results can be obtained with smaller sample sizes.  $L_2$  regularization and dropout strategies are used during the construction of the CNN network to mitigate the effects of over-fitting problems. In the end, we obtained the optimal network classification results in the situation of a suitable number of iterations and CNN input patch size for each data set.

## Acknowledgments

We thank the Editor and anonymous referees for their constructive feedback. This work was supported by the National Key Research and Development Program of China (No. 2017YFC1404700), the National Natural Science Foundation of China (No. U1711266) and National Natural Science Foundation of China under Grant (No. 41701417). It was also supported by the Youth League Committee of CUG (No. 1610491B24).

## References

- [1] Y. Ren and Y. Liu, "Geological disaster detection from remote sensing image based on experts knowledge and image features," in *Proc. of Geoscience and Remote Sensing Symposium (IGARSS), 2016 IEEE International*, pp. 677–680, 2016. [Article \(CrossRef Link\)](#)
- [2] V. Silvia, C. Jocelyn, Benediktsson, J. Atli, T. Hugues, and W. Björn, "Advanced directional mathematical morphology for the detection of the road network in very high resolution remote sensing images," *Pattern Recognition Letters*, vol. 31, no. 10, pp. 1120–1127, 2010. [Article \(CrossRef Link\)](#)
- [3] V. O. K'are, "Using Deep Convolutional Networks to Detect Roads in Aerial Images," *NTNU*, 2016.
- [4] J. Yan, Y. Ma, and L. Wang, "A cloud-based remote sensing data production system," *Future Generation Computer Systems*, vol. 86, pp. 1–13, 2017. [Article \(CrossRef Link\)](#)
- [5] F. Yuan, K. E. Sawaya, B. C. Loeffelholz, and M. E. Bauer, "Land cover classification and change analysis of the Twin Cities (Minnesota) Metropolitan Area by multitemporal Landsat remote sensing," *Remote sensing of Environment*, vol. 98, no. 2, pp. 317–328, 2005. [Article \(CrossRef Link\)](#)
- [6] H. Masroor, C. Dongmei, C. Angela, W. Hui, and S. David, "Change detection from remotely sensed images: From pixel-based to object-based approaches," *ISPRS Journal of Photogrammetry and Remote Sensing*, no. 80, pp. 91–106, 2013.
- [7] P. Anne, H. Jacky, and W. Christiane, "The utility of texture analysis to improve per-pixel classification for high to very high spatial resolution imagery," *International Journal of Remote Sensing*, vol. 26, no. 4, pp. 733–745, 2005. [Article \(CrossRef Link\)](#)
- [8] W. Han, R. Feng, L. Wang, Y. Cheng, "A semi-supervised generative framework with deep learning features for high-resolution remote sensing image scene classification," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 145, Part A, pp. 23–43, 2018. [Article \(CrossRef Link\)](#)
- [9] L. Wang, K. Lu, P. Liu, R. Ranjan, L. Chen, "IK-SVD: Dictionary Learning for Spatial Big Data via Incremental Atom Update," *Computing in Science and Engineering*, vol. 16, no. 4, pp. 41–52, 2014. [Article \(CrossRef Link\)](#)
- [10] A. H. G'okhan and A. Selim, "Automatic detection of geospatial objects using multiple hierarchical segmentations," *IEEE transactions on Geoscience and Remote Sensing*, vol. 46, no. 7, pp. 2097–2111, 2008. [Article \(CrossRef Link\)](#)
- [11] L. Zhang, L. Zhang, and B. Du, "Deep learning for remote sensing data: A technical tutorial on the state of the art," *IEEE Geoscience and Remote Sensing Magazine*, vol. 4, no. 2, pp. 22–40, 2016. [Article \(CrossRef Link\)](#)

- [12] J. Wei, L. Wang, P. Liu, X. Chen, W. Li, Z. Y. Albert, "Spatiotemporal Fusion of MODIS and Landsat-7 Reflectance Images via Compressed Sensing," *IEEE Trans. Geoscience and Remote Sensing*, vol. 55, no. 12, pp. 7126-7139, 2017. [Article \(CrossRef Link\)](#)
- [13] Y. Chen, X. Zhao, and X. Jia, "Spectral-spatial classification of hyperspectral data based on deep belief network," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 6, pp. 2381-2392, 2015. [Article \(CrossRef Link\)](#)
- [14] T. Tian, L. Gao, W. Song, C. K. Raymond, and J. He, "Feature extraction and classification of VHR images with attribute profiles and convolutional neural networks," *Multimedia Tools and Applications*, vol. 77, no. 14, pp. 18637-18656, 2018. [Article \(CrossRef Link\)](#)
- [15] K. Alex, S. Ilya, and H. Geoffrey, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84-90, 2017. [Article \(CrossRef Link\)](#)
- [16] W. Chen, X. Li, H. He, L. Wang, "A Review of Fine-Scale Land Use and Land Cover Classification in Open-Pit Mining Areas by Remote Sensing Techniques," *Remote Sensing*, vol. 10, no. 1, pp. 1-15, 2018. [Article \(CrossRef Link\)](#)
- [17] Nair. Vinod and H. Geoffrey, "Rectified linear units improve restricted boltzmann machines," in *Proc. of the 27th international conference on machine learning (ICML-10)*, pp. 807-814, 2010.
- [18] M. Lin, Q. Chen, and S. Yan, "Network in network," *arXiv preprint arXiv:1312.4400*, 2013.
- [19] J. Fan, J. Yan. and Y. Ma, "Big Data Integration in Remote Sensing across a Distributed Metadata-Based Spatial Infrastructure," *Remote Sensing*, vol. 10, no. 1, pp. 1-7, 2018. [Article \(CrossRef Link\)](#)
- [20] W. Björn and B. J. Atli, "Fusion of support vector machines for classification of multisensor data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 12, pp. 3858-3866, 2007. [Article \(CrossRef Link\)](#)
- [21] W. Hu, Y. Huang, L. Wei, F. Zhang, and H. Li, "Deep convolutional neural networks for hyperspectral image classification," *Journal of Sensors*, vol. 2015, no. 2, pp. 1-12, 2015. [Article \(CrossRef Link\)](#)
- [22] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 10, pp. 6232-6251, 2016. [Article \(CrossRef Link\)](#)
- [23] A. Sharma, X. Liu, X. Yang, and D. Shi, "A patch-based convolutional neural network for remote sensing image classification," *Neural Networks*, vol. 95, pp. 19-28, 2017. [Article \(CrossRef Link\)](#)
- [24] W. Zhao and S. Du, "Spectral-spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 8, pp. 4544-4554, 2016. [Article \(CrossRef Link\)](#)
- [25] P. Ghamisi, Y. Chen, and X. Zhu, "A self-improving convolution neural network for the classification of hyperspectral data," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 10, pp. 1537-1541, 2016. [Article \(CrossRef Link\)](#)
- [26] H. Wang, Y. Wang, Q. Zhang, S. Xiang, and C. Pan, "Gated Convolutional Neural Network for Semantic Segmentation in High-Resolution Images," *Remote Sensing*, vol. 9, no. 5, pp. 446, 2017. [Article \(CrossRef Link\)](#)
- [27] M. Mattia, C. V. Gustavo, and B. Lorenzo, "A composite semisupervised SVM for classification of hyperspectral images," *IEEE Geoscience and Remote Sensing Letters*, vol. 6, no. 2, pp. 234-238, 2009. [Article \(CrossRef Link\)](#)
- [28] H. Ishida, Y. Oishi, K. Morita, K. Moriwaki, and T. Y. Nakajima, "Development of a support vector machine based cloud detection method for MODIS with the adjustability to various conditions," *Remote Sensing of Environment*, vol. 205, pp. 390-407, 2018. [Article \(CrossRef Link\)](#)
- [29] J. Yan and L. Wang, "Suitability Evaluation for Products Generation from Multisource Remote Sensing Data," *Remote Sensing*, vol. 8, no. 12, pp. 982-995, 2016. [Article \(CrossRef Link\)](#)
- [30] Y. Yuan, H. Lv, and X. Lu, "Semi-supervised change detection method for multi-temporal hyperspectral images," *Neurocomputing*, vol. 148, pp. 363-375, 2015. [Article \(CrossRef Link\)](#)

- [31] A. Ertürk and A. Plaza, “Informative change detection by unmixing for hyperspectral images,” *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 6, pp. 1252–1256, 2015. [Article \(CrossRef Link\)](#)
- [32] J. Cao, Z. Chen, and B. Wang, “Deep convolutional networks with superpixel segmentation for hyperspectral image classification,” in *Proc. of Geoscience and Remote Sensing Symposium (IGARSS), 2016 IEEE International*, pp. 3310–3313, 2016. [Article \(CrossRef Link\)](#)
- [33] Y. Liu, Y. Zhong, F. Fei, and L. Zhang, “Scene semantic classification based on random-scale stretched convolutional neural network for high-spatial resolution remote sensing imagery,” in *Proc. of Geoscience and Remote Sensing Symposium (IGARSS), 2016 IEEE International*, pp. 763–766, 2016. [Article \(CrossRef Link\)](#)
- [34] J. Ding, B. Chen, H. Liu, and M. Huang, “Convolutional neural network with data augmentation for SAR target recognition,” *IEEE Geoscience and remote sensing letters*, vol. 13, no. 3, pp. 364–368, 2016. [Article \(CrossRef Link\)](#)
- [35] Y. Liu, M. Zhang, P. Xu, and Z. Guo, “SAR ship detection using sea-land segmentation-based convolutional neural network,” in *Proc. of Remote Sensing with Intelligent Processing (RSIP), 2017 International Workshop on*, pp. 1–4, 2017. [Article \(CrossRef Link\)](#)
- [36] L. Wang, J. Zhang, P. Liu, C. K. Raymond, and F. Huang, “Spectral-spatial multi-feature-based deep learning for hyperspectral remote sensing image classification,” *Soft Computing*, vol. 21, no. 1, pp. 213–221, 2017. [Article \(CrossRef Link\)](#)
- [37] W. Song, L. Wang, P. Liu, and C. K. Raymond, “Improved t-SNE based Manifold Dimensional Reduction for Remote Sensing Data Processing,” *Multimedia Tools and Applications*, vol. 78, no. 4, pp. 4311–4326, 2019. [Article \(CrossRef Link\)](#)



**Chunming Wu** received his B.S. in Material Science and Engineering from China University of Geosciences (CUG) in 2002, and received Ph.D. in Mineral Resource Assessment from CUG in 2011. He is currently an Associate Professor with Institute of Geological Survey, China University of Geosciences (Wuhan). His major interest include Natural resource survey, evaluation and management, and Geoscience popularization.



**Meng Wang** is an undergraduate student with Spatial Information and Digital Technology, China University of Geosciences (Wuhan). Her major interests include machine learning and remote sensing image processing.



**Lang Gao** received her B.S. in network engineering from China University of Geosciences (Wuhan) in 2016, and she is currently a graduate student with School of Computer Science, China University of Geosciences (Wuhan). Her major interests include computer vision and remote sensing image processing.



**Weijing Song** received her Ph.D. in the Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing, China in 2016. She is now a lecturer at the College of Computer Science, China University of Geosciences (Wuhan). Her research interests include remote sensing image processing, multiple spatiotemporal series analysis including spatio-temporal correlation, data dimensional reduction, and spatio-temporal prediction.



**Tian Tian** received her B.S. in Electronics Information Engineering from Huazhong University of Science and Technology (HUST) in 2009, and received Ph.D. in Control Science and Engineering from HUST in 2015. She is currently an associate professor with College of Computer Science, China University of Geosciences (Wuhan). Her major interests include computer vision and remote sensing image processing.



**Kim-Kwang Raymond Choo** received the Ph.D. in Information Security in 2006 from Queensland University of Technology, Australia. He currently holds the Cloud Technology Endowed Professorship at The University of Texas at San Antonio (UTSA). In 2016, he was named the Cybersecurity Educator of the Year - APAC (Cybersecurity Excellence Awards are produced in cooperation with the Information Security Community on LinkedIn), and in 2015 he and his team won the Digital Forensics Research Challenge organized by Germany's University of Erlangen-Nuremberg. He is the recipient of the 2018 UTSA College of Business Col. Jean Piccione and Lt. Col. Philip Piccione Endowed Research Award for Tenured Faculty, IEEE TrustCom 2018 Best Paper Award, ESORICS 2015 Best Research Paper Award, 2014 Highly Commended Award by the Australia New Zealand Policing Advisory Agency, Fulbright Scholarship in 2009, 2008 Australia Day Achievement Medallion, and British Computer Society's Wilkes Award in 2008. He is also a Fellow of the Australian Computer Society, an IEEE Senior Member, and Co-Chair of IEEE Multimedia Communications Technical Committee (MMTC)'s Digital Rights Management for Multimedia Interest Group.