RESEARCH

Open Access

Spatial-spectral hyperspectral image classification based on information measurement and CNN

Lianlei Lin^{*} , Cailu Chen and Tiejun Xu

Abstract

In order to construct virtual land environment for virtual test, we propose a construction method of virtual land environment using multi-satellite remote sensing data, the key step of which is accurate recognition of ground object. In this paper, a method of ground object recognition based on hyperspectral image (HSI) was proposed, i.e., a HSI classification method based on information measure and convolutional neural networks (CNN) combined with spatial-spectral information. Firstly, the most important three spectra of the hyperspectral image was selected based on information measure. Specifically, the entropy and color-matching functions were applied to determine the candidate spectra sets from all the spectra of the hyperspectral image. Then three spectra with the largest amount of information were selected through the minimum mutual information. Through the above two steps, the dimensionality reduction for hyperspectral images was effectively achieved. Based on the three selected spectra, the CNN network input combined with the spatial-spectral information was designed. Two input strategies were designed: (1) The patch surrounding the pixel to be classified was directly intercepted from the grayscale images of the three selected spectra. (2) In order to highlight the effect of the spectrum of the pixel to be classified, all the spectral components of this pixel were superimposed on the patch obtained by the previous strategy. As a result, a new patch with more prominent spectral components of the pixel to be classified was obtained. Using the two public hyperspectral datasets, Salinas and Pavia Center, the experiments of on both parameter selection and classification performance were performed to verify that the proposed methods had better classification performance.

Keywords: Information measure, Convolutional neural networks, Hyperspectral image classification, Spatial-spectral information

1 Introduction

In recent years, hyperspectral image (HSI) analysis has been widely used in various fields [1], such as land coverage classification and change monitoring [2, 3] and environmental science and mineral development [4]. Hyperspectral sensors can generate three-dimensional hyperspectral images containing spectral information and spatial information by capturing the two-dimensional ground-space images over hundreds of consecutive narrow-band spectra [5]. Hyperspectral sensors

can provide the continuous spectrum [6] covering the spectra from visible to infrared [7], which contains rich spectral information and cannot be provided by many traditional spectral sensors.

The rich spectral information and spatial information of hyperspectral images are useful for the classification of the ground objects. However, the excessively high dimensions and highly redundant information can lead to the dramatic increase in the computational complexity and may affect classification accuracy. Therefore, it is very important to perform the dimensionality reduction of the hyperspectral image before the classification [8]. There are two typical methods for dimensionality reduction, i.e.,

*Correspondence: linlianlei@hit.edu.cn

School of Electronics and Information Engineering, Harbin Institute of Technology, 92 Xidazhi Street, 150001 Harbin, China

feature extraction and band selection. The feature extraction method mainly includes principal component analysis (PCA) [9], linear discriminant analysis (LDA) [10], multidimensional scaling [11], etc. The band selection method mainly constitutes of the examining of correlation [12], the calculation of mutual information [13], etc. In recent years, information-based band selection has become a very popular research topic, which usually uses Shannon entropy or its changes, such as mutual information, as the basis for the measurement of image information. Martnez-Us? et al. proposed a clustering method based on mutual information for the automatic band selection in multispectral images [14]. Wang et al. proposed a supervised classification method based on spatial entropy for the band selection [15]. Moan et al. proposed a new spectral image visualization method to achieve the band selection through the third-order information measurement [16]. Salem et al. proposed a band selection method based on the hierarchical clustering of the spectral layers. A c-means clustering algorithm that combined spatial spectrum information was proposed [17]. A unified framework was proposed to evaluate the four methods from the five perspectives, including feature entropy and classification accuracy [18]. Hossain et al. proposed a dimensionality reduction method (PCA-nMI) that combined PCA with normalized mutual information (nMI) with two constraints [19]. Therefore, the band selection method based on the information measurement can effectively reduce the dimensionality of the hyperspectral image, maximize the selected band information, minimize the redundancy, and achieve the excellent classification performance for the spectral images under the reduced computational complexity.

The same ground objects may have different spectra and the different ground objects may have similar spectral features, thus the classification only by the spectral information may cause errors sometimes [20, 21]. In order to address this problem, the hyperspectral image classification method with spatial-spectral information has appeared in recent years. Using a stack auto-encoder (SAE) to combine the spatial and spectral features, Chen et al. proposed a spatial principal component information classification method [22]. Slavkovikj et al. proposed a CNN framework for hyperspectral image classification, in which spectral features were extracted from a small neighborhood [23]. Makantasis et al. proposed an R-PCA CNN classification method [24], in which the PCA was first used to extract spatial information, and then CNN was used to encode spectral and spatial information. Makantasis et al. used CNN to encode the spectral and spatial information of hyperspectral data. In addition, the multi-layer perceptrons were used for the classification of the hyperspectral image [24]. A regularized feature extraction model based on CNN was introduced to extract the

effective spatial-spectral features for classification [25]. Ghamisi et al. proposed a SICNN model. In the model, CNN was combined with fractional-order Darwin particle swarm optimization (FODPSO) to select the band, which had the largest amount of spatial information and fit the input of CNN model [26]. Mei et al. proposed the SS-CNN classification method and discussed a CNN framework for merging spatial and spectral features [27]. Zhao et al. proposed a classification framework based on the spatial-spectral feature (SSFC). In the frame, the dimensionality reduction and depth learning algorithms were used to extract spectral and spatial features, respectively. In addition, CNN was used to automatically find deep spatial correlation features. The joint features were extracted together by stacking spectral and spatial features, and finally the hyperspectral image was classified based on the trained multi-feature classifier [28]. Therefore, the classification method based on the spatial-spectral information for the hyperspectral image can fully extract the joint spatial-spectral features and achieved higher classification accuracy than the classification method relying on the extraction of only the spectral information or only the spatial information.

As an advanced machine learning technology, deep learning uses deep neural networks to learn hierarchical features of the raw input data from low to advanced level. Deep learning technology has been widely used in image classification [29], agronomy [30], mobile positioning [31], and hyperspectral image classification [32]. Models based on the convolutional neural networks (CNN) could detect local features of the input hyperspectral data, achieving highly accurate and stable classification results [20]. CNN has proven to be superior to SVM in classifying the spectral features extracted from hyperspectral images [33]. Since the feature reliability of each pixel determines the accuracy of classification, it is important to design a feature extraction algorithm dedicated to hyperspectral image classification. Ma et al. proposed a context deep learning algorithm for feature learning, which can better characterize information than the extraction algorithms with predefined features [34]. Pan et al. proposed a novel simplified deep learning model based on regularized greedy forest (RGF) and vertex component analysis network (R-VCANet), which achieved higher accuracy when the number of training samples was insufficient [35]. Mou et al. proposed a new recurrent neural network (RNN) model that can effectively analyze hyperspectral pixels and use them as sequence data to derive information categories through network derivation [36]. A semi-supervised classification method based on multi-decision marking and deep feature learning was proposed to achieve classification tasks using as much information as possible [37]. Zhang et al. proposed a region-based diversified CNN that can encode semantic context-aware

representations to obtain valuable features and improve classification accuracy [38]. Niu et al. proposed a learning to rank model using Siamese convolutional neural networks for 2D and 3D image quality assessment [39]. Therefore, in hyperspectral image classification, the application of deep learning has become more and more extensive and achieved remarkable results. Especially, CNN has been proven to have great advantages in extracting image feature information and realizing image high accuracy classification.

In this paper, a HSI classification method based on information measure and CNN was proposed. Firstly, the uncorrelated spectra were excluded through the calculation of the entropy information, and the preliminary selection of the candidate spectra was completed by combining the color-matching functions. Then, the redundant spectra were identified with the mutual information. In addition, the spectra containing the most useful information were further selected by calculating the minimum normalized mutual information. Moreover, using the selected spectra based on the information measure, the pseudo-color images were generated to achieve the texture information of the ground. Furthermore, a CNN classification method based on the information measure and the enhancement of spectral information was proposed. The integrated spatial-spectral information was put into CNN to achieve high classification accuracy of hyperspectral image.

2 Related work

Virtual test technology is one of the research fields of our team. Virtual test relies on virtual environment, and synthetic natural environment is an important part of virtual environment, which includes atmosphere, land, ocean, and space environment [40]. The virtual land environment for virtual test can not only provide the display function of virtual test scene for the test personnel, but also provide the sensing basis for the virtual sensor, and can interact with other virtual environments [41], which requires the virtual land environment to be able to provide accurate ground information in addition to the traditional texture and elevation information. For this reason, we propose a virtual land environment construction scheme as shown in Fig. 1. It is based on multi-source satellite ground observation data and realized by four steps, including multi-source data fusion, ground object recognition, and so on.

Based on hyperspectral image (HSI), multispectral image (MSI), panchromatic image (PAN), and other optical earth observation data and radar earth observation data such as InSAR, the construction of virtual land environment is completed through four steps: (A) temporal-spatial-spectral fusion, (B) ground truth recognition, (C) digital elevation model (DEM) extraction, and (D) virtual

land environment synthesis. In step A, the HSI is mainly used to obtain the high “temporal-spatial-spectral” resolution HSI which can meet the requirements of building virtual land environment by data-level fusion with PAN, MSI, and other homogeneous remote sensing images with complementary advantages, which is essentially a joint application of multi-sensor data [42, 43]. Step B obtains accurate ground truth information based on the HSI generated in step A. In step C, InSAR image is used to acquire DEM needed for constructing 3D terrain. InSAR is a kind of ground observation technology which has developed rapidly in recent years and has the characteristics of large-scale, high-precision, and all-weather. Finally, step D synthesizes the virtual land environment for the virtual experiment by using the ground truth, elevation, and texture. Among the B step, ground truth recognition, is a key step, which is responsible for providing accurate ground truth information and texture image for the virtual land environment. The proposed HSI classification method in this article belongs to this step, the ground truth information is obtained through the classification of HSI and the texture image is gotten by pseudo-color image synthesis based on the spectra selected by information measure.

3 Methods

3.1 Using information measure to select HIS's spectra

3.1.1 Determination of candidate spectra sets

The amount of entropy was measured by the degree of uncertainty, i.e., the probability of the occurrence of discrete random events [44]. When the amount of information was larger, the redundancy was smaller. Information measurement based on Shannon's communication theory has proven to be very effective in identifying the redundancy of high-dimensional data sets. When these measurements were applied to hyperspectral images, each channel was equivalent to a random variable X . In addition, all the pixels were considered as events x_i of X . The channels with less information were excluded by Shannon entropy to determine the candidate spectra sets as follows:

First, the entropy of each spectrum of the hyperspectral image $H(B_i)$ was calculated. The random variable B_i is the i th spectrum ($i = 1, 2, \dots, n$), x_i is the pixel of the i th spectrum, $p_{B_i}(x_i)$ is the probability density function of the spectrum B_i , b is the logarithmic order.

$$H(B_i) = - \sum_{i=1}^n p_{B_i}(x_i) \log_b[p_{B_i}(x_i)] \quad (1)$$

Second, the local average of the entropy of each spectrum was defined. In the equation, m is the window size, indicating the size of the neighborhood

$$\overline{H_m(B_i)} = \frac{1}{m} \sum_{p=-m/2}^{m/2} H(B_{i+p}) \quad (2)$$

Fig. 1 The process of constructing a virtual land environment using multi-source satellite Earth observation data

Finally, the spectrum B_i that met the following conditions was retained. In the equation, σ is the threshold factor. The spectrum was considered redundant if its entropy was higher or lower than the upper or the lower floating threshold factor σ of the local average $\overline{H_m(B_i)}$.

$$H(B_i) \in [\overline{H_m(B_i)} \times (1 - \sigma), \overline{H_m(B_i)} \times (1 + \sigma)] \quad (3)$$

As shown in Fig. 2, the horizontal axis represents the number of spectra, i.e., the spectral dimension, the vertical axis represents the entropy of each spectrum, and the blue curve is the entropy curve. The smoothness of the entropy curve determined the values of the window size m the threshold factor σ . If the curve was smooth,

the change of the adjacent spectra information of the hyperspectral image was small, the uncertainty of the spectrum information was small, the number of spectra falling outside the relevant range was small, the probability of having an uncorrelated spectrum was small, and the number of redundant spectra was small. In this case, the smaller values of σ and m were chosen to improve the ability to exclude redundant spectra. On the contrary, if the curve fluctuated greatly, the information of the adjacent spectra of the hyperspectral image changed dramatically, the uncertainty of the spectrum information was large, the number of spectra falling outside the correlation range was large, the probability of having an

Fig. 2 Elimination of uncorrelated spectra on the Indian pines dataset

uncorrelated spectrum was large, and there were many redundant spectra. In this case, the larger values of σ and m were chosen to prevent the elimination of the spectra with valid information.

Next, based on the calculated entropy, the CIE 1931 Standard Chroma Observer Color Matching Function (CMF) [45], which described the visual color characteristics of human eyes, was used to complete the determination of candidate spectra. The CMF determined the amount of light (red, blue, and green) to achieve the same visual effect as the monochromatic light corresponding to the wavelength. By applying CIE color matching to the hyperspectral images in the visible range, the hyperspectral images can be visualized as images with the correct color matching [46]. By setting the threshold value t for the CMF coefficients of the three primary colors, the candidate spectra based on the three primary color channels, Set_R^t , Set_B^t , and Set_G^t , were obtained.

In Fig. 3, two spectral thresholds ($t = 0.1$, $t = 0.5$) were set for the trend curve of the red light CMF coefficient. When the CMF coefficient was above the threshold, the corresponding spectra were preserved. However, it was very challenging to set the parameter t without a specific application. In this paper, an automatic threshold method was used, in which the optimal threshold of t was defined as the value to maximize the amount of discarded information. S_{discard}^t was defined as the set of channels that were discarded by thresholding the CMF. S_{selected}^t was the complementary set of S_{discard}^t . The optimal threshold was defined as:

$$t_{\text{opt}} = \arg\max(t) \quad \text{subject to} \quad H(S_{\text{discard}}^t) < H(S_{\text{selected}}^t) \quad (4)$$

where $H(S_{\text{discard}}^t)$ is the total entropy of the discarded spectra obtained by the above derivation, $H(S_{\text{selected}}^t)$ is the total entropy of the selected spectra. Based on these results, the spectra were initially selected.

3.1.2 Spectra selection

Mutual information is a measure of useful information. It is defined as the contained amount of information in a random variable another random variable. The mutual information between two random variables X and Y is defined as follows:

$$I(X, Y) = \sum_{i=1 \dots n} \sum_{j=1 \dots m} P_{X,Y}(x_i, y_j) \log \frac{P_{X,Y}(x_i, y_j)}{P_X(x_i) \cdot P_Y(y_j)} \\ = H(X) + H(Y) - H(X, Y) \quad (5)$$

where $P_X(x_i)$ is the probability density function of X , $P_Y(y_i)$ is the probability density function of Y , and $P_{X,Y}(x_i, y_j)$ is the joint probability density of X and Y . $H(X)$ is the entropy of the random variable X , $H(Y)$ is the entropy of the random variable Y . $H(X, Y)$ is the joint entropy of two random variables X and Y .

Further, Bell proposed that the mutual information of three random variables X , Y , and Z can be defined as [47]

$$I(X, Y, Z) = H(X, Z) + H(Y, Z) - H(X, Y, Z) \\ = -H(Z) - I(X, Y) \quad (6)$$

Fig. 3 Changes of CMF coefficient of red light

where $H(X, Y, Z)$ is the third-order joint entropy of the three random variables X , Y , and Z .

The above principle is also applicable to hyperspectral images. The information of one channel can increase the mutual information between the other two channels. In this case, when the overlapped information between the two channels is less, the interdependence degree between the two random variables is lower, and more information is contained. Two criteria need to be considered to reduce the dimensionality of hyperspectral images, i.e., maximum information and minimum redundancy.

Pla proposed to standardize the mutual information [48]. In this paper, the k th-order normalized information (NI) of the spectrum $S = \{B_1, \dots, B_k\}$ was used as the standardized mutual information. NI was defined as follows:

$$NI_k(S) = \frac{K \times I(S)}{\sum_{i=1}^k H(B_i)} \quad (7)$$

where $I(S)$ is the mutual information among the spectra of B_1 to B_k and $H(B_i)$ is the entropy of the spectrum B_i .

In the previous section, the threshold t was set for the CMF coefficients of the three primary colors to obtain three candidate spectra, i.e., Set_R^t , Set_B^t , and Set_G^t . When the value of the mutual information was smaller, the amount of information contained in the selected spectra was larger, and the dimensionality reduction effect on the hyperspectral image was better. Based on the mutual information, the three spectra, i.e., x_R, y_B, z_G ($x_R \in \text{Set}_R^t, y_B \in \text{Set}_B^t, z_G \in \text{Set}_G^t$) were obtained to minimize $NI_3(x_R, y_B, z_G)$. These three spectra contained the most important information of the hyperspectral images.

3.2 Pseudo-color images synthesis

To get the texture information of in the construction of virtual land environment, we adopt the pseudo-color images synthesis technology, which is based on the three spectra selected from the previous section. Besides, the pseudo-color images can also be convenient for human eyes to observe. As an image enhancement technology, the pseudo-color images synthesis technology can synthesize multispectral monochrome images into pseudo-color images by adding or subtracting colors.

Using the three grayscale images, x_R, y_B, z_G , the pseudo-color images synthesis can be performed based on information measurement according to Section 3.1.1.

As shown in Fig. 4, three grayscale images of the hyperspectral three-dimensional color image were saved by the three-color synthesis method [49], and the three color-changing functions corresponding to the three colors of red, green, and blue were set to $R(x, y)$, $G(x, y)$, and $B(x, y)$.

$$\begin{aligned} R(x, y) &= \text{Red}(\text{Gray}_1(x, y)), \\ G(x, y) &= \text{Green}(\text{Gray}_2(x, y)), \\ B(x, y) &= \text{Blue}(\text{Gray}_3(x, y)) \end{aligned} \quad (8)$$

where $\text{Gray}_i(x, y)$, ($i = 1, 2, 3$) represents the grayscale data of three grayscale images, $\text{Red}(\text{Gray}_1(x, y))$, $\text{Green}(\text{Gray}_2(x, y))$, and $\text{Blue}(\text{Gray}_3(x, y))$ indicate that the color conversion of red, green, and blue is corresponding to the three grayscale images, respectively. Finally, the three color-converted images were combined into a pseudo-color image. The generated pseudo-color image can also improve the fidelity through color correction technology [50].

$$(R, G, B) = (R(x, y), G(x, y), B(x, y)) \quad (9)$$

3.3 HSI classification based on CNN

3.3.1 HSI classification based on information measure

As shown in Fig. 5, the classification process of hyperspectral image based on information measure (IM for short) can be divided into the following two main steps: Firstly, three candidate spectra were selected based on entropy and color matching function and three most important spectra were selection based on the minimum mutual information. Then the grayscale images of these three spectra were synthesized into the pseudo-color images. These three spectra contained the most important spectral information. Then neighborhood of the pixel to be classified was finally extracted to generate a patch with spatial-spectral information. Secondly, the patches were input into the CNN for training and testing.

As shown in Fig. 6, assume that the size of hyperspectral data is $I_1 \times I_2 \times I_3$. From the three spectra x_R, y_B, z_G , a $m \times m \times 3$ patch with spatial-spectral information was extracted and put into CNN. From the perspective of the two spatial dimension $I_1 \times I_2$, the patch to CNN contained three layers of spatial information with the size of $m \times m$. From the perspective of spectral dimension I_3 , the patch contained all the spectral information of the three spectra x_R, y_B, z_G . Therefore, the method uses all the spatial-spectral information of the three spectra x_R, y_B, z_G , and the number of input channels to CNN was $\text{in_channels}=3$. CNN network consists of two convolution layers (Conv1 and Conv2), two pooling layers (Max-pooling1 and Max-pooling2), and two full connection layers (Full_connected1 and Full_connected2). Finally, the classification of hyperspectral images is realized by using Softmax classifier. The training stage is as follows:

Step 1. Random initialization of network parameters, including Conv1, Conv2, Max-pooling1, Max-pooling2, and so on.

Step 2. The input patch is propagated forward through the convolution layers, the pooling layers, and the full connection layers to get the output value.

Fig. 4 Pseudo-color images synthesis by adding red, green, and blue colors

Step 3. Calculate the error between the output value and the target value of the network.

Step 4. When the error is greater than the expected value, the error is propagated back to the network, and the errors of the full connection layers, the pooling layers, and the convolution layers are obtained in turn; when the error is equal to or less than the expected value, the training is ended.

Step 5. Update the network parameters according to the obtained error and go to step 2).

3.3.2 HSI classification based on information measure and enhanced spectral information

On the basis of the previous methods, we propose HSI classification based on information measure and enhanced spectral information (IM_ESI for short) as shown in Fig. 7. Compared with the IM method, the

main difference of IM_ESI is extracting all the spectral information of the pixel to be classified (the patch central pixel) and then attach them to previous $m \times m \times 3$ patch by combing, intercepting, and deforming the spectral information, generating a new $m \times m \times 6$ patch for CNN' input finally.

As shown in Fig. 8, Firstly, the one-dimensional spectral information of the patch center with the size of $1 \times 1 \times I_3$ was repeatedly superimposed by n times to obtain the one-dimensional spectral information with the size of $1 \times (n \times I_3)$. A one-dimensional spectral vector with the same size as the two-dimensional spatial information $m \times m$ was intercepted. The one-dimensional spectrum with the size of $1 \times (m \times m)$ was deformed into a two-dimensional spectral matrix of $m \times m$ and superimposed in three layers. The obtained results were combined with the two-dimensional spatial information with the size of

Fig. 5 Flow chart of IM method

Fig. 6 Schematic diagram of IM method

$m \times m$ in three layers of the three spectra x_R, y_B, z_G to obtain the spatial-spectral patch with the size of $m \times m \times 6$. The obtained spatial-spectral patch was put into the CNN. The number of input channel in CNN model was $\text{in_channels}=6$. The structure of CNN network is the same as that of the previous network.

4 Results and discussion

4.1 Datasets

The Salinas dataset was acquired by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor in the Salinas Valley, California, USA. The uncorrected

hyperspectral image data contained 224 spectra, and the corresponding two-dimensional floor space of each spectrum contained 512×217 pixels with a spatial resolution of 3.7 m per pixel. The corrected spectral dimension was reduced to 204 by removing the 20 spectra (108–112, 154–167, and 224) which covered the water absorption region. The ground objects of the hyperspectral image were classified into 16 categories. The Pavia Center dataset was obtained by the Reflex Optical System Imaging Spectrometer (ROSIS) hyperspectral sensor over the city center of Pavia in southern Italy. The spatial resolution was 1.3 m per pixel, and

Fig. 7 Flow chart of IM_ESI method

the hyperspectral image data had 102 spectra after the noise spectrum was deleted. In addition, the two-dimensional ground space corresponding to each spectrum contained 1096×1096 pixels. The area contained 9 types of ground objects. The datasets are gotten from the website (http://www.ehu.es/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes).

There were both similarities and differences between the two datasets. Both the Salinas and Pavia Center datasets had large sizes, but the size of Pavia Center dataset was extremely large. The 2D image size of Pavia Center dataset was about 11 times as large as the size of the Salinas dataset. Salinas dataset mainly reflected the information of vegetation (such as Fallow and Celery) and included a rich variety of features with the regular distribution. Pavia Center dataset mainly reflected the landscape information of the urban landscape and included small types of features with strongly irregular distribution. The Salinas dataset contained rich spectral information while the Pavia Center dataset contained less spectral information. The spatial resolution of both datasets was high, and the spatial resolution of the Pavia Center dataset was higher than that of the Salinas dataset. In addition, through principal component analysis, it was found that the first principal component of both two

datasets contained far more information than the other bands, indicating that the image information had a concentrated distribution. According to the characteristics of the two datasets, 5% of all labeled pixels in the Salinas data set were randomly selected as training samples, and the remaining 95% of the pixels were used as test samples. The distribution of sample size in Salinas dataset is shown in Table 1. Nine percent of the pixels of the Pavia Center dataset were randomly selected and used as training samples, and the remaining 91% of the pixels were used as test samples. The distribution of sample size in Pavia Center is shown in Table 2.

4.2 Experiments and result analysis

4.2.1 Parameter selection experiments

In the spectra selection algorithm based on information measure, two key parameters needed to be manually adjusted, i.e., the window size m and the threshold σ . Both parameters determined the excluded spectra and had a significant impact on the CNN classification results of hyperspectral images.

Experiment on parameter σ As shown in Table 3, for the Salinas dataset, the window size was set to $m = 11$. Under different σ values, the classification

Table 1 Samples distribution of Salinas dataset

Class	Name	Total	Train	Test
1	Brocoli_green_weeds_1	2009	101	1908
2	Brocoli_green_weeds_2	3726	187	3539
3	Fallow	1976	99	1877
4	Fallow_rough_plow	1394	70	1324
5	Fallow_smooth	2678	134	2544
6	Stubble	3959	198	3761
7	Celery	3579	179	3400
8	Grapes_untrained	11,271	564	10,707
9	Soil_vinyard_develop	6203	311	5892
10	Corn_senesced_green_weeds	3278	164	3114
11	Lettuce_romaine_4wk	1068	54	1014
12	Lettuce_romaine_5wk	1927	97	1830
13	Lettuce_romaine_6wk	916	46	870
14	Lettuce_romaine_7wk	1070	54	1016
15	Vinyard_untrained	7268	364	6904
16	Vinyard_vertical_trellis	1807	91	1716
Total	Total	54,129	2713	51,416

accuracy of the IM method (using OA as an example) was different.

From the results in Table 3, the classification performance was the best when $\sigma = 0.05$ and $\sigma = 0.15$. However, σ had large fluctuations after these two peaks, resulting in the decrease of OA. In addition, when $\sigma > 0.2$, the OA results were the independent on the change of σ . Thus, when σ was large enough, different values of σ had almost the same effect on the classification results. Through repeated experiments, the threshold σ was set to a medium value, i.e., $\sigma = 0.1$.

Experiment on parameter m The window size m also had an important influence on the proposed method. The threshold was set to $\sigma = 0.1$. The effect of different window sizes m on OA is shown in Table 4.

Table 2 Samples distribution of Pavia Center dataset

Class	Name	Total	Train	Test
1	Water	824	74	750
2	Trees	820	74	746
3	Asphalt	816	73	743
4	Self-blocking bricks	808	73	735
5	Bitumen	808	73	735
6	Tiles	1260	113	1147
7	Shadows	476	43	433
8	Meadows	824	74	750
9	Bare soil	820	74	746
Total	Total	7456	671	6785

Table 3 Overall classification accuracy (OA) (%) at different σ ($m = 11$) on the Salinas dataset

σ	0.05	0.08	0.1	0.15	0.2	0.3	0.4
OA	98.87	97.82	97.82	98.16	97.61	97.61	97.61

By comparing the OA in Table 4, when $m > 14$, the overall classification accuracy rate was maintained at a high level. Since the entropy curvature of the Salinas dataset changed abruptly, a larger window size ($m = 20$) and a medium σ ($\sigma = 0.1$) were more suitable. Because the distributions of the image information in the Salinas dataset and the Pavia Center dataset were relatively concentrated, the window size was set to $m = 20$ and the threshold was set to $\sigma = 0.1$.

4.2.2 Pseudo-color images synthesis experiments

For the Salinas dataset, the three spectra x_R, y_B, z_G selected by the above parameters based on the information measure corresponded to the 32nd, 61st, and 66th spectra, and the mutual information $NI_3(x_R, y_B, z_G)$ was minimized to $NI_3(32, 61, 66) = 0.545$. For the Pavia Center dataset, the three selected spectra were the 6th, 68th, and 28th spectra, and the minimum mutual information was $NI_3(6, 68, 28) = 0.5358$. A pseudo-color image was generated from the grayscale images of the three selected spectra on the two datasets as shown in Fig. 9.

4.2.3 Classification performance comparison experiments

In the IM method, the patch sizes for the Salinas dataset and the Pavia Center dataset were $27 \times 27 \times 3$ and $21 \times 21 \times 3$. In the IM_ESI method, the patch size of the spatial information for the Salinas dataset was a 27×27 , and the spectral information of each pixel point was represented by a one-dimensional vector with the size of 1×204 . Thus, the spectral information was repeated four times and accumulated to obtain 1×816 . Then a one-dimensional spectral vector with the size of 1×729 ($27 \times 27 = 729$) was intercepted and transformed into a 27×27 two-dimensional spectral matrix. Copying them three times and get a $27 \times 27 \times 3$ three-dimensional spectral patch. Combining them with the $27 \times 27 \times 3$ patch obtained by IM method, then getting a new $27 \times 27 \times 6$ patch for CNN's input. Similarly, the patch size for the Pavia Center dataset in the IM_ESI method was $21 \times 21 \times 6$.

As shown in Table 5, two CNN networks were created for the two hyperspectral image datasets. In both networks, the ReLU activation function and the maximum pooling mode were selected. The random inactivation (dropout) prevented or mitigated the overfitting. The

Table 4 Overall classification accuracy (OA) (%) at different m ($\sigma = 0.1$) on the Salinas dataset

m	5	8	11	14	17	20
OA	98.21	98.50	97.82	98.80	98.86	98.95

(a) (b)

Fig. 9 Pseudo-color image generated by grayscale image of the three selected spectra based on information measure. **a** Pseudo-color image generated by the grayscale image of the 32nd, 61st, and 66th spectra of the Salinas dataset. **b** Pseudo-color image generated by the grayscale image of the 6th, 68th, and 28th spectra of the Pavia Center dataset

value of keep_prob (keep_prob = 0.5) indicated the probability that the neurons were selected. In the situation of keep_prob = 0.5, 50% of the data was discarded. Both networks were randomly initialized by the normal distribution with the known mean and standard deviation. After the initialization was completed, the training samples were input into the networks and the weights of the networks were updated. The convolutional layer (Conv1, Conv2), the pooling layer (Maxpool1, Maxpool2), and the full connection layer (Fc1_units, Fc2_units) of the CNN model were set according to Table 5. In the above table, 128@5 × 5 indicated that there were 128 convolution kernels with the size of 5 × 5 in this layer, and strides = 2 meant that the step size was 2. The learning rate in the training of CNN networks for both datasets was set to 0.005, the training times were set to 260 for Salinas dataset and 100 for Pavia Center dataset. After the network was trained, the objects to be classified can be input into the corresponding CNN classification model to predict the category of the object.

Table 5 CNN network parameter setting

Parameter	Conv1	Maxpool1	Conv2	Maxpool2	Fc1_units	Fc2_units
Value	128@ 5 × 5	Strides=2 2 × 2	256@ 3 × 3	Strides=2 2 × 2	200	84

In order to fully verify the validity of the method, the proposed IM and IM_ESI methods were compared with the following similar methods: (1) The CNN classification method based on spectral information of hyperspectral image (referred to as SPE), which was developed according to the concept of “classification method based on spectral information gray image” in [51]; (2) CNN classification method based on hyperspectral image spatial information (referred to as PCA1), which was developed according to [25] using 2D-CNN to extract spatial features; (3) two types of CNN classification methods for hyperspectral image with fused spatial-spectral information, i.e., CNN classification method with the fusion of first principal component spatial information and spectrum information (referred to as PCA1_SPE) and CNN classification method based on PCA first three principal components spatial-spectral information (referred to as PCA3), which were obtained according to the combination idea of spatial-spectral information proposed by [52]. In the experiment, three evaluation indicators (total classification accuracy (OA), average classification accuracy (AA), and Kappa coefficient) were used to evaluate the classification performance [53]. OA is the ratio of the number of correctly classified pixels to the total number of pixels. AA is the ratio of the number of correctly classified pixels in each class to the total number of pixels in such class. Kappa coefficient is an index based on the confusion matrix to measure the classification accuracy. The kappa

Table 6 The classification results on Salinas dataset (%)

Class	SPE	PCA1	PCA1_SPE	PCA3	IM	IM_ESI
1	99.20	99.32	99.75	99.69	99.85	99.45
2	97.91	99.53	99.43	99.83	99.99	99.95
3	98.84	99.69	99.65	99.94	99.99	100.00
4	99.39	99.73	99.89	99.76	99.91	99.88
5	97.59	99.66	99.44	99.74	99.62	99.79
6	99.92	99.68	99.91	99.90	99.99	99.95
7	99.58	99.50	98.91	99.74	99.99	99.72
8	85.50	92.26	93.39	97.00	98.18	98.87
9	99.67	98.67	99.30	99.96	99.90	99.56
10	94.98	99.02	98.86	99.69	99.79	99.38
11	94.79	99.61	99.20	99.34	99.29	99.80
12	99.35	99.65	99.83	99.83	99.92	99.97
13	98.65	99.78	99.44	99.77	99.59	99.98
14	97.27	98.78	98.68	98.83	99.84	99.58
15	73.41	91.93	91.58	97.68	98.97	99.24
16	95.93	95.56	98.55	99.96	99.92	99.86
OA	87.06	94.44	95.02	98.12	98.95	99.13
AA	95.75	98.27	98.49	99.42	99.67	99.69
Kappa	85.58	93.81	94.45	97.93	98.84	99.02

coefficient can be used for consistency tests. Supposing N is the total number of samples, c is the total number of classes, X_{ii} is the diagonal element of the matrix, X_{i+} is the sum of all rows, X_{+i} is the sum of all columns, then $\sum_{i=1}^c X_{ii}$ is the total number of correctly classified samples, and $\sum_{i=1}^c (X_{i+} \cdot X_{+i})$ is the sum of the products of X_{i+} and X_{+i} .

$$\text{Kappa} = \frac{N \cdot \sum_{i=1}^c X_{ii} - \sum_{i=1}^c (X_{i+} \cdot X_{+i})}{N^2 - \sum_{i=1}^c (X_{i+} \cdot X_{+i})} \quad (10)$$

Experiment on Salinas dataset The comparison of the classification performance on the Salinas dataset is shown in Table 6. From Table 6, the IM and IM_ESI methods had the best classification accuracy in the Salinas dataset. Especially, compared to the SPE, PCA1, and PCA3 methods, the value of OA using the IM_ESI method was improved by 12.07%, 4.69%, and 1.01%, respectively. The excellent performance on AA and kappa coefficients fully demonstrated the stability and accuracy of IM and IM_ESI methods

Figure 10 a–f are the classification results of the SPE, PCA1, PCA1_SPE, PCA3, IM, and IM_ESI methods on the Salinas dataset, and g is the ground truth of Salinas dataset.

Fig. 10 The classification results of Salinas dataset. **a** SPE. **b** PCA1. **c** PCA1_SPE. **d** PCA3. **e** IM. **f** IM_ESI. **g** The ground truth

Table 7 The classification results on Pavia Center dataset (%)

Class	SPE	PCA1	PCA1_SPE	PCA3	IM	IM_ESI
1	99.21	99.67	99.93	99.84	99.90	99.97
2	77.03	89.76	96.33	97.07	98.99	99.31
3	94.39	92.84	98.98	97.99	98.63	98.46
4	85.16	95.79	98.46	99.92	99.93	99.89
5	84.27	96.49	98.71	99.12	99.65	99.62
6	95.13	96.89	99.03	99.01	99.61	99.43
7	89.49	97.45	96.87	98.32	99.30	99.41
8	96.25	90.09	99.28	99.29	99.72	99.66
9	97.64	99.70	99.97	99.19	99.12	99.93
OA	91.50	92.14	98.59	98.73	99.46	99.50
AA	90.95	95.41	98.62	98.86	99.43	99.52
Kappa	88.07	89.06	98.01	98.21	99.23	99.29

From Fig. 10, it can be seen more intuitively that the overall classification effect by the IM and IM_ESI methods on the Salinas dataset is significantly better than that by the SPE, PCA1, PCA1_SPE, and PCA3 methods. Especially, for the highly similar feature classes, such as classes 1 and 2 (Broccoli_green_weeds_1 and Broccoli_green_weeds_2) and classes 8 and 15 (Grapes_untrained and Vinyard_untrained), the classification effect was significantly improved, and misclassification for the scenario of highly similar features was significantly reduced. The results indicated that the IM and IM_ESI methods had outstanding advantages in classifying highly similar feature information.

Experiment on Pavia Center dataset The comparison of the classification performance of the SPE, PCA1, PCA1_SPE, PCA3, IM, and IM_ESI methods on the Pavia Center dataset is shown in Table 7. From Table 7, the IM and IM_ESI methods had the best classification accuracy, both of which were above 99%. Especially, the classification accuracy of IM_ESI method was 8%, 7.36%, and 0.91% higher than that of the SPE, PCA1, and PCA1_SPE methods, respectively. The average classification accuracy and kappa coefficient of IM and IM_ESI were also the highest, indicating that the predicted classification results of various types of features were more consistent with the real object information. Therefore, the results on the Pavia Center dataset further demonstrated the stability and accuracy of the classification methods IM and IM_ESI on large data sets.

Figure 11 a–f are the classification results of the SPE, PCA1, PCA1_SPE, PCA3, IM, and IM_ESI methods on the Pavia Center dataset, and g is the ground truth of Pavia Center dataset.

From Fig. 11, the IM and IM_ESI methods had outstanding classification performance with the sufficient

training samples in the Pavia Center dataset. In both methods, almost all the samples were correctly classified.

Based on the performance comparison between Salinas and Pavia Center datasets, the hyperspectral image classification method based on information measurement (IM and IM_ESI) had the highest and most stable classification results with both normal and very sufficient sample size. With the normal sample size, the classification performances was good, the classification accuracy of highly similar feature information was high, and the misclassification of information was very rare or almost none. As the sample size got richer, the classification performance was even more superior. The features extracted by the traditional PCA are the linear addition of the original features. It requires that the data to be dimensionally reduced be linearly correlated. It cannot solve the problem that the data has no linear correlation. Therefore, the nonlinear features embedded in the HSI data cannot be preserved by the linear PCA model. In this paper, from the perspective of information measurement, three spectra with the largest amount of information and the most complementary information are selected. Therefore, experiments show that the information measurement methods used in this paper outperformed the PCA ones. Research shows that feature extraction can help machine learning to improve the generalization performance. As we all know, besides PCA, there are some common dimensionality reduction methods, such as Kernel-PCA (KPCA) [54], locally linear embedding (LLE) [55], and so on. KPCA generally performs better than PCA, the reason lies in the fact that KPCA can explore higher order information of the original inputs than PCA. KPCA implicitly considers the high order information of the original inputs and extracts more principal components, more principal components could also be extracted in KPCA, eventually resulting in the best generalization performance. LLE is much better than PCA in dealing with so-called manifold dimensionality reduction. LLE maps its inputs into a single global coordinate system of lower dimensionality, and its optimizations do not involve local minima. By exploiting the local symmetries of linear reconstructions, LLE can learn the global structure of nonlinear manifolds, such as those generated by images of faces or documents of text. So, in the future research, we can try to use KPCA, ICA, and LLE to replace the methods based on information measurement or compare our in this paper with the classification methods based on these dimensionality reduction methods for further experimental testing.

4.3 Discussion

The proposed HSI classification method that performs better than the SPE, PCA1, PCA1_SPE, and PCA3 methods is to provide accurate ground truth information and texture image for the virtual land environment. Since our

Fig. 11 The classification results of Pavia Center dataset. **a** SPE. **b** PCA1. **c** PCA1_SPE. **d** PCA3. **e** IM. **f** IM_ESI. **g** The ground truth

method is based on spatial-spectral information of three spectra that contain the most important spectral information. If the spectral information is scattered that three spectra cannot contain the most important spectral information, our method may not work well. The method has certain limitations on the application of the actual scene and may not be as good as the effect obtained on the experimental image. In view of the limitation, we will further improve our method to enhance its applicability and generality. Although this method has certain limitations, it can provide a reference direction for us to study

the classification method of hyperspectral images in the future.

5 Conclusions

In this paper, a CNN classification method based on information measure (IM) was proposed. In addition, the principle and complete implementation process of the classification method was introduced. Firstly, the candidate spectral sets was determined based on entropy and color-matching function. The minimum mutual information was calculated to achieve the accurate selection of three

spectra, and the grayscale image of the selected spectrum was synthesized to pseudo-color image for the texture information of virtual land environment. Secondly, a hyperspectral image classification method (IM_ESI) was proposed by combining spatial-spectral information of the IM method and enhanced spectral information. The spatial-spectral information of the selected spectra based on information measure was combined with all spectral information of the pixel to be classified after deformation and superposition to obtain new spatial-spectral patch, which was input into the CNN network. Through the parameter selection experiment, the influence of two key parameters (window size m , threshold σ) on the performance of the proposed method was analyzed, and the CNN network parameters of IM and IM_ESI were reasonably selected. Finally, a comparison experiment was performed on Salinas and Pavia Center datasets. The results showed that the proposed method (IM and IM_ESI) had good classification performance regardless of the sample size and distribution. The classification accuracy was high for the information with similar features, and the classification performance on the data with sufficient samples was more superior.

Abbreviations

AA: Average classification accuracy; AVIRIS: Airborne Visible/Infrared Imaging Spectrometer; CMF: Color-matching function; CNN: Convolutional neural networks; FODPSO: Fractional-order Darwin particle swarm optimization; HSI: Hyperspectral image; IM: Information measure; IM_ESI: Information measure and enhanced spectral information; LDA: Linear discriminant analysis; MSI: Multispectral image; NI: Normalized information; nMI: Normalized mutual information; OA: Total classification accuracy; PAN: Panchromatic image; PCA-nMI: PCA with normalized mutual information; PCA: Principal component analysis; PCA1: CNN classification method based on hyperspectral image spatial information; PCA1_SPE: CNN classification method with the fusion of first principal component spatial information and spectrum information; PCA3: CNN classification method based on PCA first three principal components spatial-spectral information; R-VCANet: Regularized greedy forest and vertex component analysis network; RGF: Regularized greedy forest; RNN: Recurrent neural network; ROSIS: Reflex Optical System Imaging Spectrometer; SAE: Stack auto-encoder; SPE: CNN classification method based on spectral information of hyperspectral image; SSFC: Spatial-spectral feature

Acknowledgements

Not applicable.

Authors' contributions

LL and CC contributed to the conceptualization and methodology, obtained the resources, and wrote and prepared the original draft. CC contributed to the validation and formal analysis. CC and TX wrote, reviewed, and edited the manuscript. LL supervised the study. The authors read and approved the final manuscript.

Funding

This work was supported by the National Natural Science Foundation of China (Grant No. 61671170) and National Key R & D Plan (Grant No. 2017YFB1302701).

Availability of data and materials

We declare that the MATLAB code used for the simulation will not be shared, and we assure that we will send it on demand. The datasets used or analyzed during the current study could be downloaded through the following hyperlink: http://www.ehu.es/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes.

Competing interests

The authors declare that they have no competing interests.

Received: 18 October 2019 Accepted: 13 February 2020

Published online: 06 March 2020

References

1. F. Bei, L. Ying, H. Zhang, C. W. Chan, Semi-supervised deep learning classification for hyperspectral image based on dual-strategy sample selection. *Remote. Sens.* **10**(4), 574 (2018)
2. F. M. Lacar, M. M. Lewis, I. T. Grierson, in *IEEE International Geoscience & Remote Sensing Symposium*. Use of hyperspectral imagery for mapping grape varieties in the barossa valley, south australia, (2001). <https://doi.org/10.1109/igarss.2001.978191>
3. L. Mou, P. Ghamisi, X. Zhu, L. Mou, P. Ghamisi, X. Zhu, L. Mou, P. Ghamisi, X. Zhu, Unsupervised spectral-spatial feature learning via deep residual conv-deconv network for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **PP**(99), 1–16 (2018)
4. W. Chen, D. Bo, L. Zhang, Slow feature analysis for change detection in multispectral imagery. *IEEE Trans. Geosci. Remote Sens.* **52**(5), 2858–2874 (2014)
5. H. Lyu, L. Hui, in *Geoscience & Remote Sensing Symposium*. Learning a transferable change detection method by recurrent neural network, (2016). <https://doi.org/10.1109/igarss.2016.7730344>
6. D. Chutia, D. K. Bhattacharyya, K. K. Sarma, R. Kalita, S. Sudhakar, Hyperspectral remote sensing classifications: a perspective survey. *Trans. GIS* **20**(4), 463–490 (2015)
7. L. Ying, H. Zhang, S. Qiang, Spectral-spatial classification of hyperspectral imagery with 3d convolutional neural network. *Remote. Sens.* **9**(1), 67 (2017)
8. G. E. Hinton, R. R. Salakhutdinov, Reducing the dimensionality of data with neural networks. *Science* **313**(5786), 504–507 (2006). <https://doi.org/10.1126/science.1127647>. <https://science.sciencemag.org/content/313/5786/504.full.pdf>
9. J. Jiang, J. Ma, C. Chen, Z. Wang, Z. Cai, L. Wang, Superpca: A superpixelwise pca approach for unsupervised feature extraction of hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.* **56**(8), 4581–4593 (2018)
10. H. S. Chu, B. C. Kuo, C. H. Li, C. T. Lin, in *IEEE International Conference on Fuzzy Systems*. A semisupervised feature extraction method based on fuzzy-type linear discriminant analysis, (2011). <https://doi.org/10.1109/fuzzy.2011.6007733>
11. B. Jiao, C. Fan, Z. Wang, Multidimensional scaling used for image classification based on binary partition trees. *Comput. Eng. Appl.* (2015)
12. X. Zhou, B. Xiang, M. Zhang, Novel spectral interval selection method based on synchronous two-dimensional correlation spectroscopy. *Anal. Lett.* **46**(2), 340–348 (2013)
13. B. Guo, S. R. Gunn, R. I. Damper, J. D. B. Nelson, Band selection for hyperspectral image classification using mutual information. *IEEE Geosci. Remote. Sens. Lett.* **3**(4), 522–526 (2006)
14. A. Martínez-Usó, F. Pla, P. García-Sevilla, J. M. Sotoca, *Automatic Band Selection in Multispectral Images Using Mutual Information-Based Clustering*, (2006). https://doi.org/10.1007/11892755_67
15. B. Wang, W. Xin, Z. Chen, A hybrid framework for reservoir characterization using fuzzy ranking and an artificial neural network. *Comput. Geosci.* **57**(57), 1–10 (2013)
16. S. L. Moan, A. Mansouri, Y. Voisin, J. Y. Hardeberg, A constrained band selection method based on information measures for spectral image color visualization. *IEEE Trans. Geosci. Remote. Sens.* **49**(12), 5104–5115 (2011)
17. M. B. Salem, K. S. Ettabaa, M. S. Bouhlel, in *Image Process. Appl. Syst.* Hyperspectral image feature selection for the fuzzy c-means spatial and spectral clustering, (2016). <https://doi.org/10.1109/ipas.2016.7880114>
18. W. Bo, C. Chen, T. M. Kechadi, L. Sun, A comparative evaluation of filter-based feature selection methods for hyper-spectral band selection. *Int. J. Remote Sens.* **34**(22), 7974–7990 (2013)
19. M. A. Hossain, X. Jia, M. Pickering, Subspace detection using a mutual information measure for hyperspectral image classification. *IEEE Geosci. Remote. Sens. Lett.* **11**(2), 424–428 (2014)
20. M. Zhang, W. Li, Q. Du, Diverse region-based CNN for hyperspectral image classification. *IEEE Trans. Image Process. Publ. IEEE Sig. Process. Soc.* **27**(6), 2623 (2018)
21. L. Bing, X. Yu, P. Zhang, A. Yu, X. Wei, Supervised deep feature extraction for hyperspectral image classification. *IEEE Trans. Geosci. Remote. Sens.* **56**(4), 1909–1921 (2018)

22. Y. Chen, Z. Lin, Z. Xing, W. Gang, Y. Gu, *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens.* **7**(6), 2094–2107 (2014)
23. V. Slavkovikj, S. Verstockt, W. D. Neve, S. V. Hoecke, R. V. D. Walle, Hyperspectral image classification with convolutional neural networks (2015)
24. K. Makantasis, K. Karantzas, A. Doulamis, N. Doulamis, in *Geosci. Remote. Sens. Symp.* Deep supervised learning for hyperspectral data classification through convolutional neural networks, (2015). <https://doi.org/10.1109/igarss.2015.7326945>
25. Y. Chen, H. Jiang, C. Li, X. Jia, P. Ghamisi, Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. *IEEE Trans. Geosci. Remote. Sens.* **54**(10), 6232–6251 (2016)
26. P. Ghamisi, Y. Chen, X. Z. Xiao, A self-improving convolution neural network for the classification of hyperspectral data. *IEEE Geosci. Remote. Sens. Lett.* **13**(10), 1537–1541 (2017)
27. S. Mei, J. Ji, J. Hou, L. Xu, D. Qian, S. Mei, J. Ji, J. Hou, L. Xu, D. Qian, Learning sensor-specific spatial-spectral features of hyperspectral images via convolutional neural networks. *IEEE Trans. Geosci. Remote. Sens.* **55**(8), 4520–4533 (2017)
28. W. Zhao, S. Du, Spectral–spatial feature extraction for hyperspectral image classification: a dimension reduction and deep learning approach. *IEEE Trans. Geosci. Remote. Sens.* **54**(8), 4544–4554 (2016)
29. G. Ososkov, P. Goncharov, Shallow and deep learning for image classification. *Opt. Mem. Neural Netw.* **26**(4), 221–248 (2017)
30. C.-H. Chen, H.-Y. Kung, F.-J. Hwang, Deep learning techniques for agronomy applications. *Agronomy*, **9**, 142 (2019). <https://doi.org/10.3390/agronomy9030142>
31. L. Wu, C.-H. Chen, Q. Zhang, A mobile positioning method based on deep learning techniques. *Electronics*, **8**, 59 (2019). <https://doi.org/10.3390/electronics8010059>
32. W. Hao, S. Prasad, Convolutional recurrent neural networks for hyperspectral data classification. *Remote. Sens.* **9**(3), 298 (2017)
33. H. Wei, Y. Huang, W. Li, Z. Fan, H. Li, Deep convolutional neural networks for hyperspectral image classification. *J. Sens.* **2015**(2), 1–12 (2015)
34. X. Ma, G. Jie, H. Wang, Hyperspectral image classification via contextual deep learning. *Eurasip J. Image Video Process.* **2015**(1), 20 (2015)
35. B. Pan, Z. Shi, X. Xia, R-vanet: A new deep-learning-based hyperspectral image classification method. *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens.* **10**(5), 1975–1986 (2017)
36. L. Mou, P. Ghamisi, X. Z. Xiao, Deep recurrent neural networks for hyperspectral image classification. *IEEE Trans. Geosci. Remote. Sens.* **55**(7), 1–17 (2017)
37. X. Ma, H. Wang, J. Wang, Semisupervised classification for hyperspectral image based on multi-decision labeling and deep feature learning. *Isprs J. Photogramm. Remote. Sens.* **120**, 99–107 (2016)
38. Y. Zhong, L. Zhang, An adaptive artificial immune network for supervised classification of multi-/hyperspectral remote sensing imagery. *IEEE Trans. Geosci. Remote Sens.* **50**(3), 894–909 (2012)
39. Y. Niu, D. Huang, Y. Shi, X. Ke, Siamese-network-based learning to rank for no-reference 2D and 3D image quality assessment. *IEEE Access*, **7**, 101583–101595 (2019). <https://doi.org/10.1109/ACCESS.2019.2930707>
40. W. H. Liu, X. R. Wang, L. I. Ning, Modeling and simulation of synthetic natural environment. *Acta Simul. Syst. Sin.* **16**(12), 2631–2635 (2004)
41. L. Lin, C. Chen, J. Yang, S. Zhang, Deep transfer HSI classification method based on information measure and optimal neighborhood noise reduction. *Electronics*, **8**(10) (2019). <https://doi.org/10.3390/electronics8101112>
42. L. Kong, J. Pan, V. Snäsel, P. Tsai, T. Sung, An energy-aware routing protocol for wireless sensor network based on genetic algorithm. *Telecommun. Syst.* **67**(3), 451–463 (2018). <https://doi.org/10.1007/s11235-017-0348-6>
43. T. Nguyen, J. Pan, T. Dao, An improved flower pollination algorithm for optimizing layouts of nodes in wireless sensor network. *IEEE Access*, **7**, 75985–75998 (2019). <https://doi.org/10.1109/ACCESS.2019.2921721>
44. C. E. Shannon, A mathematical theory of communication. *Bell Syst. Tech. J.* **27**(4), 623–656 (2014)
45. M. Q. Shaw, M. D. Fairchild, in *Aic*. Evaluating the 1931 CEI color matching functions, (2002). <https://doi.org/10.1002/col.10077>
46. N. P. Jacobson, M. R. Gupta, Design goals and solutions for display of hyperspectral images. *IEEE Trans. Geosci. Remote. Sens.* **43**(11), 2684–2692 (2005)
47. A. J. Bell, The co-information lattice. *Proc. Int. Symp. Indep. Component Anal. Blind Source Sep.*, 921–926 (2003)
48. A. Martinez-Uso, F. Pla, J. M. Sotoca, P. García-Sevilla, Clustering-based hyperspectral band selection using information measures. *IEEE Trans. Geosci. Remote. Sens.* **45**(12), 4158–4171 (2007)
49. J. F. Gilmore, *Automated Fake Color Separation: Combining Computer Vision and Computer Graphics*, (1987). <https://doi.org/10.1117/12.940667>
50. Y. Niu, P. Liu, T. Zhao, Y. Fan, Matting-based residual optimization for structurally consistent image color correction. *IEEE Trans. Circ. Syst. Video Technol.*, 1–1 (2019). <https://doi.org/10.1109/TCSVT.2019.2949587>
51. S. Xinyi, *Hyperspectral Image Classification Based on Convolutional Neural Networks*, PhD thesis. (Harbin Institute of Technology, Heilongjiang, 2016)
52. H. Lishuan, *Study of Dimensionality Reduction and Spatial-Spectral Method for Classification of Hyperspectral Remote Sensing Image*, PhD thesis. (China University of Geosciences, Beijing, 2018)
53. W. D. Thompson, S. D. Walter, Kappa and the concept of independent errors. *J. Clin. Epidemiol.* **41**(10), 969–970 (1988)
54. L. J. Cao, K. S. Chua, W. K. Chong, H. P. Lee, Q. M. Gu, A comparison of PCA, KPCA and ICA for dimensionality reduction in support vector machine. *Neurocomputing*, **55**(1), 321–336 (2003)
55. S. T. Roweis, L. K. Saul, Nonlinear dimensionality reduction by locally linear embedding. *Science*, **290**(5500), 2323–2326 (2000)

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)