

# Big Data Classification of Remote Sensing Image Based on Cloud Computing and Convolutional Neural Network

Xiaobo Wu (✉ [wuxb@lingnan.edu.cn](mailto:wuxb@lingnan.edu.cn))  
Lingnan Normal University

---

## Research Article

**Keywords:** Remote Sensing Image, Cloud Computing, Convolutional Neural Network, Big Data Classification

**Posted Date:** August 9th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-770665/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.  
[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at Soft Computing on January 11th, 2022. See the published version at <https://doi.org/10.1007/s00500-021-06562-y>.

# Big Data Classification of Remote Sensing Image Based on Cloud Computing and Convolutional Neural Network

Xiaobo Wu<sup>1, a\*</sup>

<sup>1</sup>*School of Geographical Sciences, Lingnan Normal University, Zhanjiang 524048, Guangdong, China*

<sup>a</sup> *Email: wuxb@lingnan.edu.cn*

\*Corresponding author

**Abstract:** In recent years, with the continuous development of cloud computing and big data technology, information technology is penetrating all corners of enterprise development at a speed that ordinary people cannot imagine. Today's remote sensing image classification plays an important role in many applications. This article is based on cloud computing and convolutional neural network to study the remote sensing image big data classification. This article takes cloud computing technology and convolutional neural network technology as the technical points. First, it introduces cloud computing technology in more detail from the basic concepts, characteristics and classification of cloud computing, and then compares traditional methods from the selection of feature extraction methods and classifiers. The remote sensing image classification method is described, and finally the convolutional neural network model that needs to be applied in this research is explained. In the experiment, the features extracted by multiple pre-training networks are fused. Through the study of the three feature fusion methods, it is found that the classification accuracy rate has been further improved on the three data sets. The experimental results in this paper show that the remote sensing image big data classification method based on cloud computing and convolutional neural network has more than a little improvement in accuracy than the traditional remote sensing image classification method, and the accuracy rate has reached 86.3%.

**Keywords:** Remote Sensing Image, Cloud Computing, Convolutional Neural Network, Big Data Classification

## 1. Introduction

In recent years, the instruments and technologies that can be used for earth observation (including multi-spectrometer, hyper-spectrometer, synthetic aperture radar, etc.) have been developed rapidly. At the same time, these devices also receive a large number of aerial and satellite remote images with image resolution. Spectral analysis and different time analysis of these high-quality remote sensing images provide important information support for the introduction of intelligent ground observation. At the same time, high-quality, high-performance remote sensing images also indicate a higher demand for related research. Remote sensing image classification has always been an active research topic in the field of remote sensing image analysis. The classification of remote sensing images requires that images be classified into a set of meaningful categories based on the content of remote sensing images. Scene classification of telemetry images plays an important role in many applications. In the research of natural disaster detection, the determination of coverage area, spatial object detection, geographic image extraction, vegetation map, environmental monitoring and urban planning, is an important step in the sorting of remote sensing scene images.

In the past few years this was because it was easier to download high-resolution remote sensing

images. Therefore, more and more remote sensing image data sets are highlighted and shared. These data sets tend to have high class variance and small class variance. The image resolution and zoom ratio are different. The amounts of data in each category and the number of scene categories have also increased. The remote sensing image classification algorithm puts forward higher requirements. For these high-demand, high-definition remote scene data sets. Traditional classification methods may not achieve the desired classification accuracy. Therefore, how to quickly and accurately classify these high-resolution remote sensing images is a question of many applications and technical requirements.

Remote sensing image classification plays an important role in a wide range of applications, so it has received great attention. In the past few years, researchers have made major efforts to develop various data sets or propose various methods for scene classification from remote sensing images. Hao S solves the problem of unsupervised visual domain that is used to transfer images. A discriminative cross-view subspace alignment algorithm is introduced. Then, it's by minimizing the Bregman matrix divergence function to learn to align the mapping of the source subspace and the target subspace. Finally, the labeled source data is projected into the target-aligned source subspace, and the unlabeled target data is projected into the target subspace and classified. However, this research can only be used for a set of scene category models or scene attribute models to classify VHR remote sensing images [1]. Gong C first comprehensively reviewed recent developments. Then, a large-scale data set named "NWPU-RESISC45" was proposed, which contains 31,500 images, covering 45 scene classes, and each class has 700 images. The creation of this data set will enable the community to develop and evaluate various data-driven algorithms. Finally, several representative methods were evaluated using the suggested data set, and the results were reported as a useful baseline for future research. However, the data set proposed by this research still has certain limitations [2]. Due to the collection principle of remote sensing images and big data, Huang X proposed an encryption scheme using a chaotic system to build a compressed sensing framework. Through further arrangement operation function, the compressed remote sensing image can also be encrypted. The key space consists of the initial conditions in the chaotic system. The experimental results show that when a chaotic system is used to generate a random measurement matrix, the proposed method can be applied to compress sensing. Although this research uses permutation operations, it is mainly used to enhance the security of cipher text [3].

The innovations of this paper are: (1) A M-ConvNet convergence neural network model is provided, which performs well in classifying high-resolution SAR images. The algorithm and training settings of the model can be introduced in detail. Finally, experiments and conclusive inspections on the MSTAR high-resolution image data set are carried out; (2) the classification of remote sensing scene data applies the deep residual pre-training network classification method to the image classification problem. The pre-trained network is the feature, and the output of the extractor is more representative of the vector attributes of the image.

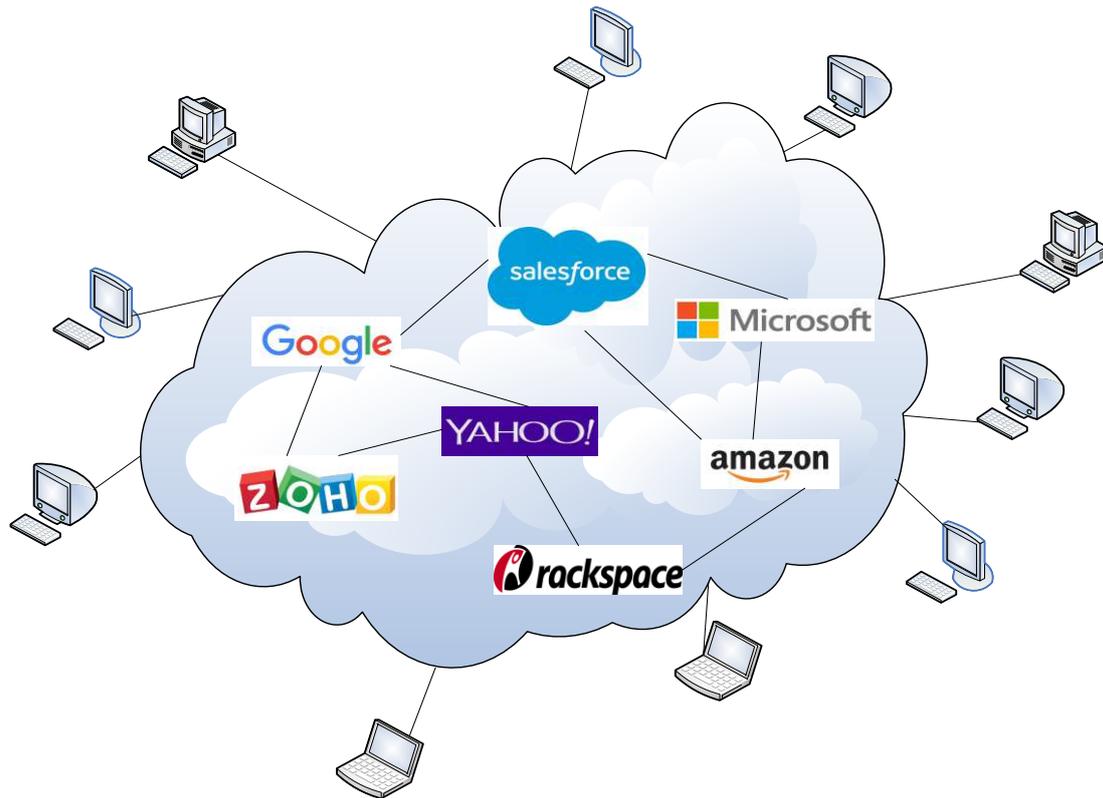
## **2. Research Methods of Remote Sensing Image Big Data Classification**

### *2.1 Cloud Computing*

#### (1) Basic concepts of cloud computing

Cloud computing was first proposed by Google's CEO Eric Schmidt in 2006, but cloud computing is not a brand-new concept. It conveys an important trend that has existed for a long time: computing across the network [4-5]. Although cloud computing was proposed and rising in recent years, the origin of cloud computing technology can be traced back to grid computing [6]. In order to facilitate the

understanding of the "cloud" thought, it can be explained in Figure 1.



**Figure1.** Overview of cloud computing

It is easy to see from Figure 1, cloud is very simple, it can provide cloud computing services, cloud storage and various application services [7]. Users in cloud computing do not need to care about what is in the cloud, what model of cloud processor is, the storage space of the hard disk, the specific geographic location of the server, how the computer is connected, or who developed the application software. Users only need to care about the reliability and credibility of the service, and they can enjoy cloud computing services at any time and any place. Based on the concept of cloud computing first proposed by Google, it refers to the computing power of computers [8-9]. By assigning computers to resource pools composed of multiple computers. Every application can request computing resources. Storage and various customized services cloud computing can be divided into two categories: cloud computing in a narrow sense and cloud computing in a broad sense [10]. Cloud computing, in a narrow sense, can be understood as the use and delivery method of IT infrastructure, and the required resources are obtained on demand through the network. In a broad sense, cloud computing can be understood as a method of use and service [11]. The required services can be obtained on demand through the network. There are no restrictions on this type of service. Perhaps it is software network resources Internet or other services cloud computing describes a new model of increasing, utilizing and providing Internet-based IT services through the Internet. There must be dynamic expansion, easy to use, and frequent use of virtual resources.

#### (2) Features and classification of cloud computing

Cloud computing has several outstanding characteristics [12-13]: 1) Data is safe and reliable. Cloud computing is the most reliable and secure data storage center. Users do not have to worry about data loss, virus intrusion and other issues. 2) Virtualization technology. The biggest feature of any existing cloud platform is its virtualization technology. It can manage the virtualization, scheduling and

application of server hardware resources. 3) Flexible customized services. Users can customize the services, applications, and resources on the cloud platform according to their own needs. Cloud platform can deployment resources, services, and computing capabilities. In addition to related applications on-demand by users, users can also manage custom services. 4) High cost performance. Compared with the traditional way of providing services, building a cloud computing platform does not require the purchase of large and expensive servers. The cloud computing platform requires very low user-side hardware equipment and is very convenient to use. A server cluster composed of ordinary PCs It can provide users with a variety of personalized services, and the computing power exceeds that of large mainframes, so users only need a small amount of investment in this area.

Cloud computing can be divided into three categories: public cloud, private cloud and hybrid cloud [14].

#### 1) Public cloud

Public cloud or external cloud is described from the perspective of mainstream awareness, which dynamically places resources on a fine-grained, self-service infrastructure on the Internet, and obtains third-party providers through network applications or network services to share resources and services [15-16]. Users access these resources and services through the Internet, but users do not own the resources of cloud computing.

#### 2) Private cloud

Private cloud and internal cloud are new words that some vendors have recently used to describe cloud computing providers that simulate private networks. Private cloud is usually the only internal cloud computing platform [17]. Private cloud provides effective control over data reliability. The company can control the applications running on that infrastructure through the company-owned infrastructure. Make full use of and manage the use of cloud resources to create the necessary know-how for the company's development and construction.

#### 3) Hybrid cloud

As the name suggests, a hybrid cloud environment is composed of multiple internal and/or external clouds [18].

### *2.2 Traditional Remote Sensing Image Classification Method*

#### (1) Feature extraction method

The spectral dimension of a remotely detected image is the spectral response value obtained from the reflection of hundreds of contiguous narrow bands. Therefore, the adjacent spectral feature points have a clear relationship and redundancy of data. Therefore, consider extracting the data from the high-dimensional space to robust characteristics of dimensionality [19-20].

#### 1) Band selection

Regardless of linear or non-line, the dimensionality reduction operation of spectral features destroys the physical characteristics of the original data to a certain extent, so the band selection has gradually been widely used by many researchers in the field of remote sensing [21]. Band selection, as the name suggests, is to select the appropriate number of best bands from all bands as features, because hundreds of bands have partial overlap, and there is a certain correlation between adjacent bands. Not all bands are equivalent for classification. It is important, so it aims to filter out suitable bands for subsequent classification through certain band selection criteria. From the perspective of classification, the greater the degree of difference between each category, the easier it is to distinguish, for example, the mixing distance shown in formula (1):

$$Z_{ij} = \sum_k^P |A_{ik} - A_{jk}| \quad (1)$$

Among them,  $A_{ik}$  and  $A_{jk}$  respectively represent the sum of the spectral values belonging to  $i$  and  $j$  categories in the  $k$ -th band, and  $P$  represents the total number of bands in the hyperspectral image. The greater the mixing distance between the two, the easier it is to distinguish the two categories. The results of using absolute distance and Euclidean distance as the measurement criteria are the same, but this measurement method only considers the distance of the spectrum, and does not consider the correlation of the spectrum, because if the angle between the two spectral characteristic directions of different categories is very large It is small, but the distance is relatively close in space, and it is easy to misjudge, and the Mahalanobis distance is a kind of variance that takes into account the overall data, as shown in equation (2):

$$Z_{ij} = \sqrt{(a_i - a_j)^T S^{-1} (a_i - a_j)} \quad (2)$$

Among them,  $a_i$  and  $a_j$  represent pure pixels belonging to categories  $i$  and  $j$ , respectively. Here, the calculation of pure pixels can be the mean vector of all categories, or it can be manually selected.

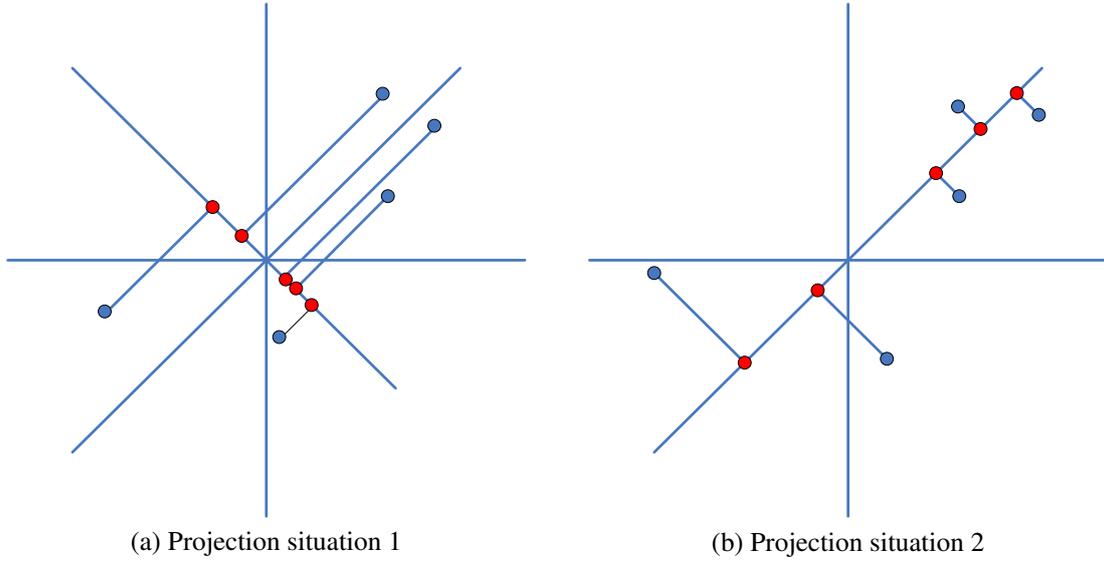
## 2) Dimensionality reduction

Dimensionality reduction aims to project high-dimensional features into a low-dimensional space so that most of the information can be retained. We assume that there are a number of discrete points in the three-dimensional space, and these points all fall on a cross-section of the origin, starting from the three-dimensional space. From the point of view, these discrete points can be represented by three-dimensional coordinates, but if we look at it from the point of view of the cross section, since they are all on the cross section, in the two-dimensional space of the cross section, these points can be It is expressed in two-dimensional coordinates, so if a transformation formula is found to transform the original three-dimensional space coordinates  $(x, y, z)$  into the cross-sectional coordinate space  $(x', y')$ , then we have achieved dimensionality reduction Operation, and the core technology of the dimensionality reduction method is to find this transformation method. For example, linear transformation is to realize the projection operation through the transformation matrix.

Principal component analysis (PCA) [22], and it is called KL transformation. It is a decorrelation technique widely used in data compression and data interpretation, and it retains the direction of the maximum variance of the data as its first principal. Component vector, the idea of PCA is to map  $m$ -dimensional features to  $n$ -dimensional, where  $n < m$ , this  $n$ -dimensional feature is its  $n$  principal components, which are derived through linear transformation, rather than simply from  $m$ -dimensional features Select  $n$ -dimensional features. Because in the field of signal analysis, it is believed that the signal has a large variance, and the noise has a small variance. The signal-to-noise ratio (SNR) is the ratio of the signal variance and the noise variance. The larger the value of SNR, the less the information contained in the signal. Noise, as shown in Figure 2, according to the theory of maximizing variance, the projection direction of Figure 2(b) is better than that of Figure 2(a). Here, suppose the covariance  $Cov$  of the data sample is as shown in equation (3):

$$Cov = \frac{1}{N} \sum_k^N (x_k - m)(x_k - m)^T \quad (3)$$

Where  $x_k$  represents the spectral vector of the  $k$ -th pixel,  $m$  represents the mean vector of the sample,  $N$  represents the total number of pixels,  $l$  represents the number of bands of the pixel, and the best view of the data is the feature value with the largest covariance.



**Figure2.** Data projection

## (2) The choice of classifier

After extracting the features of the pixels through the appropriate feature extraction method, select the appropriate classifier to classify [23]. Different classifiers are suitable for different scenes. According to the feature selection, if the classifier is selected properly, good classification results can be obtained. Here will introduce two common classifiers for hyper spectral remote sensing images.

### 1) Support vector machine

For a two-dimensional space, the support vector machine is just a straight line. It extends to a multi-dimensional space to be a plane or a hyper plane. Its purpose is to separate two linearly separable classes, but not all of them can separate the two. The straight lines or planes distinguished by the classes are correct. We need a dividing line/surface that can separate the two classes. The distance from the closest point to it in the two classes is equal to the dividing line/surface, and we want this interval to be as long as possible [24].

Assuming that a given batch of lower data set  $B = \{(x_1, y_1), \dots, (x_n, y_n)\}$  with  $n$  sample points,  $x_i$  is a multi-dimensional data point, and its corresponding  $y_i \in \{+1, -1\}$  is its label. In the space formed by the data, the interface can be expressed as:

$$A^T x + b = 0 \quad (4)$$

Then there are:

$$A^T x_i + b \geq 0 \quad \text{if } y_i = +1 \quad (5)$$

$$A^T x_i + b \leq 0 \quad \text{if } y_i = -1 \quad (6)$$

The distance from the data point to the interface is:

$$\frac{1}{\|A\|} (A^T x_i + b) \quad (7)$$

Then the solution to the interface can be expressed as:

$$\arg \max_{a,b} \left\{ \frac{1}{\|A\|} \min_n [y_i (A^T x_i + b)] \right\} \quad (8)$$

That is to say, the interface we hope to find can accurately divide the different types of data, while

being as far away as possible from the closest points of the two types. In order to facilitate the derivation, the denominator of the objective function expression is flipped up and rewritten as:

$$\arg \min_{a,b} \frac{1}{2} \|A\|^2 \quad (9)$$

$$s. t. y_i (A^T x_i + b) \geq 1 \quad (10)$$

## 2) Polynomial logistic regression

For linear regression prediction, as shown in the formula:

$$z = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n = \alpha^T x \quad (11)$$

Linear regression is to fit the original data to make possible prediction results for subsequent data, while logistic regression is mainly used for classification and is also based on linear regression, as in the formula:

$$g_\alpha(x) = \frac{1}{1+e^{-\alpha^T x}} \quad (12)$$

$$y = \frac{1}{1+e^{-x}} \quad (13)$$

The activation function of formula (13) is also called the sigmoid function. It maps the value of linear prediction so that it will eventually fall between [0, 1], which is sometimes regarded as a probability. When the output value is greater than 0.5 It belongs to one category, and it belongs to another category if it is less than or equal to 0.5, so it is a two-category, and polynomial logistic regression is the promotion of logistic regression on multi-classification problems. Its predictive category probability formula is:

$$w(y^i = j | x^i; \alpha) = \frac{e^{\alpha_j^T x^i}}{\sum_{l=1}^k e^{\alpha_l^T x^i}} \quad (14)$$

It means to predict the probability of belonging to category j for the i-th sample, where  $\alpha_l^T$  represents the parameters belonging to the l-th category, and then we build the loss function based on the maximum likelihood estimation:

$$R(\alpha) = -\frac{1}{m} \left[ \sum_{i=1}^m \sum_{j=1}^k l\{y^i = j\} \log \frac{e^{\alpha_j^T x^i}}{\sum_{l=1}^k e^{\alpha_l^T x^i}} \right] \quad (15)$$

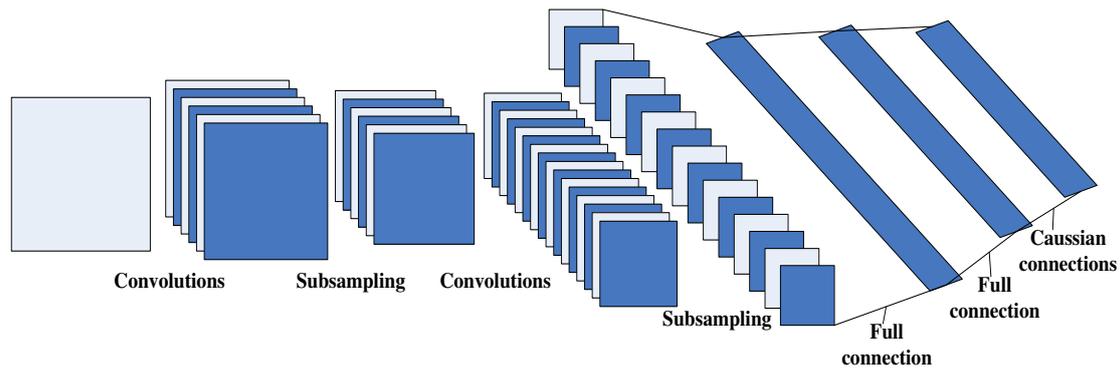
Because  $R(\alpha)$  is a convex function, algorithms such as gradient descent and L-BFGS can be used to iteratively find the global optimal solution. In the early years, polynomial logistic regression has been widely used in the field of remote sensing image classification. At the same time, it is also called soft max and has been successfully applied in many fields.

### 2.3 Convolutional Neural Network

Convolutional Neural Network (CNN) is a well-known deep learning model [25]. The name comes from the introduction of assembly functions in this model, and CNN can be classified as a multilayer neural network model. But different from the traditional multi-level power supply neural network, the input of CNN is in two-dimensional (image-like) form, the connection weight is also two-dimensional, as the core of the assembly) two-dimensional convergence and concentration because CNN can directly handle two-dimensional patterns. Therefore, it has been widely used in computer vision. For example, CNN has achieved success in many fields such as image sorting. In early 1998,

LeCun et al. proposed a neural network model called LeNet for handwritten digit recognition. Just because of the lack of a large amount of training data at the time and the limited computer processing power, LeNet did not perform well in solving complex problems (such as the sorting of large pictures and videos). Finally, in the context of a high-performance computing platform, the researchers proposed a convolutional neural network model.

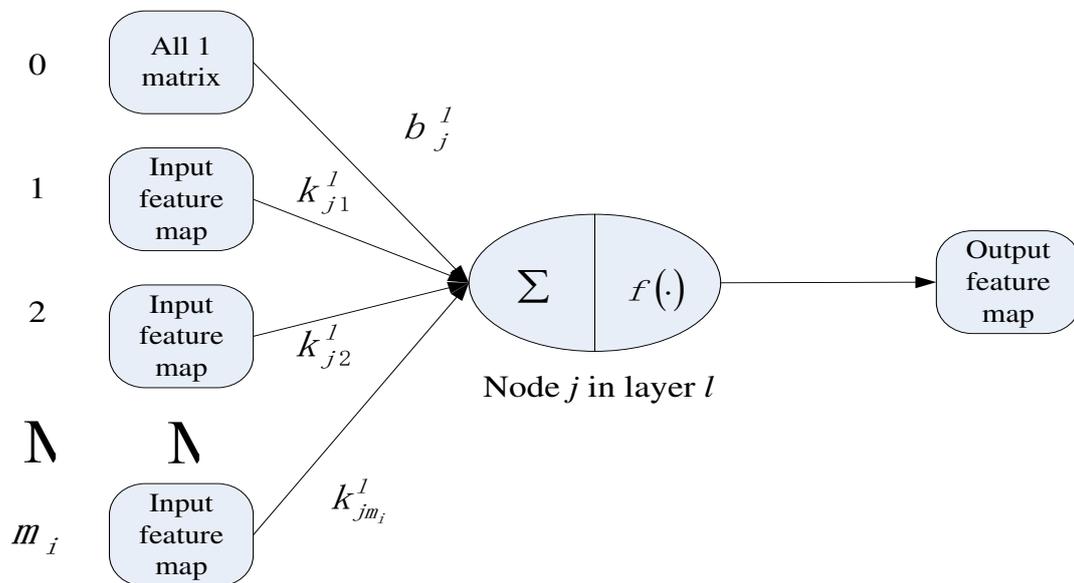
Convolutional neural network is a kind of multilayer neural network. Its basic components are convolutional layer, pooling layer and fully connected layer. Figure 3 is the topology diagram of the convolutional neural network.



**Figure3.** The structure diagram of the convolutional neural network LeNet

As shown in Figure 3, the input plane of the LeNet neural network is a two-dimensional model. The curve layer is used to extract different attributes of the image, and the pooling layer is used for degradation. The conventional class results and sampling levels are called feature maps.

Generally speaking, each convolutional layer is composed of many nodes, and the structure of the convolutional layer nodes is shown in Figure 4.



**Figure4.** The structure of the convolutional layer node

In Figure 4, the input of the convolutional layer node is a feature map, because formally the feature map is a two-dimensional matrix, so the all-one matrix can also be regarded as a feature map.  $b_j^l$  is the bias of  $l$  node  $j$ , which is a scalar.  $k_{js}^l (s = 1, 2, \dots)$  is a learnable convolution kernel matrix, and the elements in the matrix are the weights. Each convolution kernel matrix  $k_{js}^l$  and its

corresponding input feature map are convolved to obtain a two-dimensional pattern (matrix), so that a total of  $m_i$  two-dimensional patterns (matrices) are obtained. Then, the sum of these matrices, plus the bias, is used as the input of the activation function, and finally the output characteristic map of the node is obtained. The output of the convolutional layer node can be expressed by formula (16):

$$x_j^l = f(\sum_{i=1}^{m_i} x_j^{l-1} * k_{ji}^l + b_j^l T) \quad (16)$$

The pooling layer reduces the dimensionality of the image through various grouping functions and improves the continuous change of image characteristics. Typical grouping operations include maximum grouping and average grouping. The connection energy is similar to the transformation energy, that is, the connection window moves from top to bottom and from left to right on the input attribute graph according to certain rules.

### 3. Remote Sensing Image Big Data Classification Research Experiment

#### 3.1 Experimental Setup

In order to test the effect, made a lot of experimental comparisons, and tested three representative remote sensing scene data sets. These three data sets include the UCMERGED data set, the RSSCN7 data set and the NWPU-RESISC45 data set. At the same time, compared with four low-level feature expression methods and two intermediate feature expression methods: SIFT, CH, LBP, and BoVW and LLC. In order to train faster, we performed in the experimental environment shown in Table 1:

**Table1.** Experimental environment parameters

Processor	Intel® Xeon(R) CPU E5-2630 v4 @ 2.20GHz × 16
RAM	62.8G
Graphics card	GeForce GTX 1080/PCIe/SSE2
Operating system	Ubuntu 14.04 LTS 64 bit
Run the software	ANACONDA (python2.7)
Deep learning framework	Caffe

#### 3.2 Experimental Data

##### (1) UC MERCED data set

The UC Merced Land Use Data Set (UCMD) data set contains 21 scene categories, each of which contains 100 optical remote sensing scenes with a size of 256×256 pixels, including fields, airplanes, baseball fields, beaches, buildings, bushes, dense houses, forests, highways, golf courses, ports and intersections, blocks, crosswalks, parking lots, Rivers, corridors, scattered residential areas, oil storage tanks and tennis courts and other 21 types of scenes such as. An example image of the UCMD data set is shown in Figure 5.



**Figure5.** Part of the image of the UC MERCE remote sensing scene dataset

(2) RSSCN7 data set

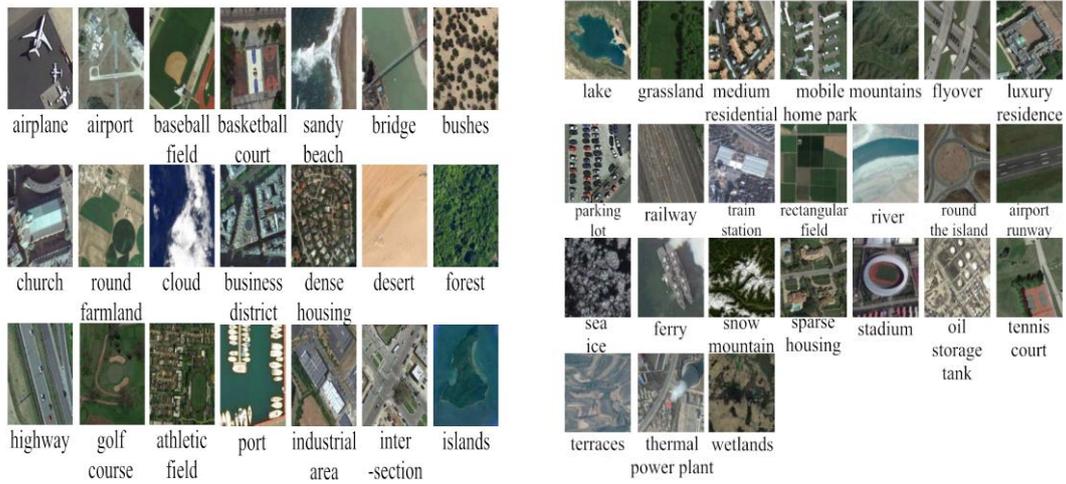
The RSSCN7 dataset is constructed by extracting satellite images from Google Earth, and consists of seven typical scene categories including grassland, forest, farmland, parking lot, residential area, industrial area, river and lake. Each category has 400 images with  $400 \times 400$  pixels. These 400 images are sampled from 4 images with different zoom ratios. Each image has a different imaging angle. This makes the RSSCN7S data set better than the UCMD data set. Some example images of the RSSCN7 data set are shown in Figure 6.



**Figure6.** Part of the image of the RSSCN7 remote sensing scene dataset

(3) NWPU-RESISC45 data set

The NWPU-RESISC45 dataset is constructed by extracting satellite images from Google Earth. It consists of a total of 31,500 remote sensing images, divided into 45 scene categories, each category contains 700  $256 \times 256$  pixel images, 45 RGB color spaces. Some example images of the NWPU-RESISC45 data set are shown in Figure 7.



**Figure7.** Part of the image of the remote sensing scene dataset of the NWPU-RESISC45 dataset

#### 4. Experimental Data Analysis of Remote Sensing Image Big Data Classification Research

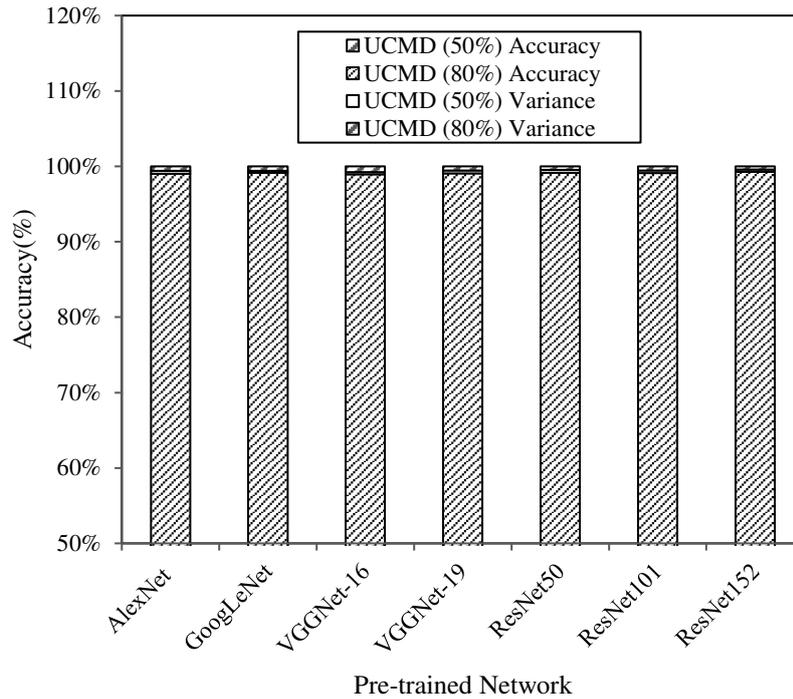
##### 4.1 Experimental Results and Analysis of the UCMERCE Data Set

Table 2 is the classification results of UCMERCE data set after feature extraction using four low-level feature expression methods and two intermediate-level feature expression methods. Algorithms include SIFT, CH, LBP, GIST, and BoVW and LLC.

**Table2.** Classification accuracy and variance of the UCMERCE data set based on the expression of middle and low-level features

Model		SIFT	LBP	CH	BoVW	LLC
UCMD (50%)	Accuracy	29.98	35.57	43.09	71.47	71.12
	Variance	0.96	1.37	1.13	1.31	1.01
UCMD (80%)	Accuracy	33.10	37.29	47.21	76.52	73.55
	Variance	1.94	1.89	0.94	2.01	1.50

The classification accuracy and variance of seven pre-trained networks of UCMERCE data set using AlexNet, GoogLeNet, VGGNet-16, VGGNet-19, ResNet-50, ResNet-101, and ResNet-152 for feature extraction are shown in Figure 8.



**Figure8.** The classification accuracy and variance of the UCMERCE dataset based on the feature extraction of the pre-trained network

The accuracy and variance of classification after splicing the features extracted by pre-training on the UCMERCE data set are shown in Table 3.

**Table3.** Classification accuracy and variance of the UCMERCE dataset based on the splicing and fusion of multiple pre-training features

Model	UCMD (50%)		UCMD (80%)	
	Accuracy	Variance	Accuracy	Variance
ResNet-50& ResNet-101& ResNet-152	96.85	0.42	97.96	0.67
AlexNet&VGGNet-16& VGGNet-19	96.68	0.87	97.73	0.64

The classification accuracy and variance of the addition of the pre-training feature values of the UCMERCE data set are shown in Table 4.

**Table4.** Classification accuracy and variance of the UCMERCE data set based on the addition of pre-trained feature values

Model	UCMD (50%)		UCMD (80%)	
	Accuracy	Variance	Accuracy	Variance
ResNet-50& ResNet-101& ResNet-152	96.89	0.37	97.86	0.72
AlexNet&VGGNet-16& VGGNet-19	96.51	0.59	97.42	0.73

Table 5 shows the accuracy and variance of the UCMERCE data set using different pre-training networks to extract features and then the classification results for voting.

**Table5.** Classification accuracy and variance of the UCMERCE dataset based on different pre-trained network voting

Model	UCMD (50%)		UCMD (80%)	
	Accuracy	Variance	Accuracy	Variance
ResNet-50& ResNet-101& ResNet-152 & AlexNet&VGGNet-16&VGGNet-19	97.20	0.32	98.12	0.57

ResNet-50& ResNet-101& ResNet-152	96.37	0.54	97.67	0.69
AlexNet&VGGNet-16& VGGNet-19	95.72	0.27	97.36	0.71

It can be seen from the above experiments that the pre-trained network trained with ImageNet data is used as a feature extractor, and the accuracy of the extracted feature vector using a linear classifier for classification is much higher than that obtained by using low-level manual feature descriptors. The accuracy of traditional methods are all less than 80%, while the feature extraction using pre-training network has an accuracy of more than 90% on UCMERCE data. Compared with the features extracted by the AlexNet, GoogLeNet, VGGNet-16, and VGGNet-19 pre-trained networks, ResNet-50, ResNet-101, and ResNet-152 are more efficient, especially the ResNet-50 network model. The features extracted by the six pre-training networks of AlexNet, VGGNet-16, VGGNet-19, ResNet-50, ResNet-101, and ResNet-152 are fused, and the two ways of connecting and adding the values in the dimensions of the features Compared with the classification based on single network feature extraction, the accuracy has been improved. The best effect is the fusion of the features provided by the three pre-training networks of ResNet-50, ResNet-101, and ResNet-152, and the classification accuracy has increased by more than 1%. In the way of voting on the classification results of different pre-trained network features, the accuracy of voting on all the results of the network is higher, but the two results are basically the same, and the difference is negligible.

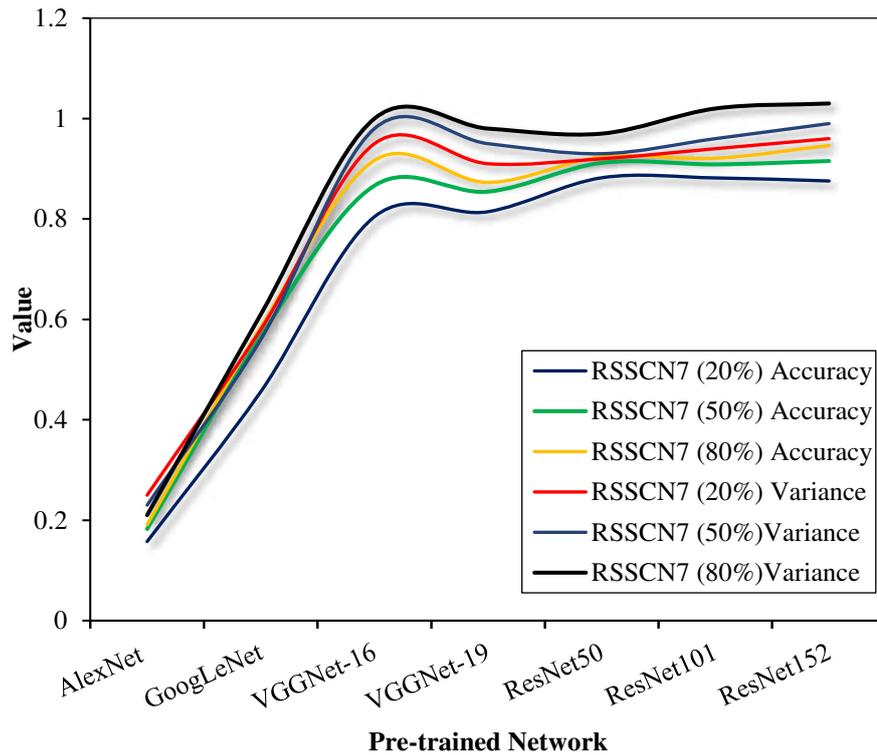
#### 4.2 Experimental Results and Analysis of the RSSCN7 Data Set

Through the application of four low-level feature expression methods and two intermediate-level feature expression methods, the classification results of the RSSCN7 data set after feature extraction are performed. The algorithm includes SIFT, CH, LBP, and BoVW and LLC. The results are shown in Table 6:

**Table6.** The classification accuracy and variance of the RSSCN7 dataset based on the expression of middle and low-level features

Model		SIFT	LBP	CH	BoVW	LLC
RSSCN7 (50%)	Accuracy	32.79	60.41	60.54	81.25	77.08
	Variance	1.23	1.01	0.98	1.16	1.28

The RSSCN7 data set is applied to seven pre-trained networks of AlexNet, GoogLeNet, VGGNet-16, VGGNet-19, ResNet-50, ResNet-101, and ResNet-152 for feature extraction. The classification accuracy and variance are shown in Figure 9.



**Figure9.** The classification accuracy and variance of the RSSCN7 data set based on the feature extraction of the pre-trained network

It can be seen from the above experiments that the pre-trained network trained with ImageNet data is used as a feature extractor, and the accuracy of the extracted feature vector using a linear classifier for classification is much higher than that obtained by using low-level manual feature descriptors. The accuracy of traditional methods are all less than 80%, and the feature extraction using pre-training network is more than 90% on the RSSCN7 data. Compared with the features extracted by the AlexNet, GoogLeNet, VGGNet-16, and VGGNet-19 pre-trained networks, ResNet-50, ResNet-101, and ResNet-152 are more efficient, especially the ResNet-50 network model. The features extracted by the six pre-training networks of AlexNet, VGGNet-16, VGGNet-19, ResNet-50, ResNet-101, and ResNet-152 are fused, and the two ways of connecting and adding the values in the dimensions of the features Compared with the classification based on single network feature extraction, the accuracy has been improved. The best effect is the fusion of the features of the three pre-training networks ResNet-50, ResNet-101, and ResNet-152, and the accuracy has increased by more than 2%. In the way of voting on the classification results after different pre-training network features, the results of all networks are voted uniformly, the accuracy rate is higher, and the accuracy rate is improved by nearly 0.5%.

#### 4.3 Experimental Results and Analysis of the NWPU-RESISC45 Data Set

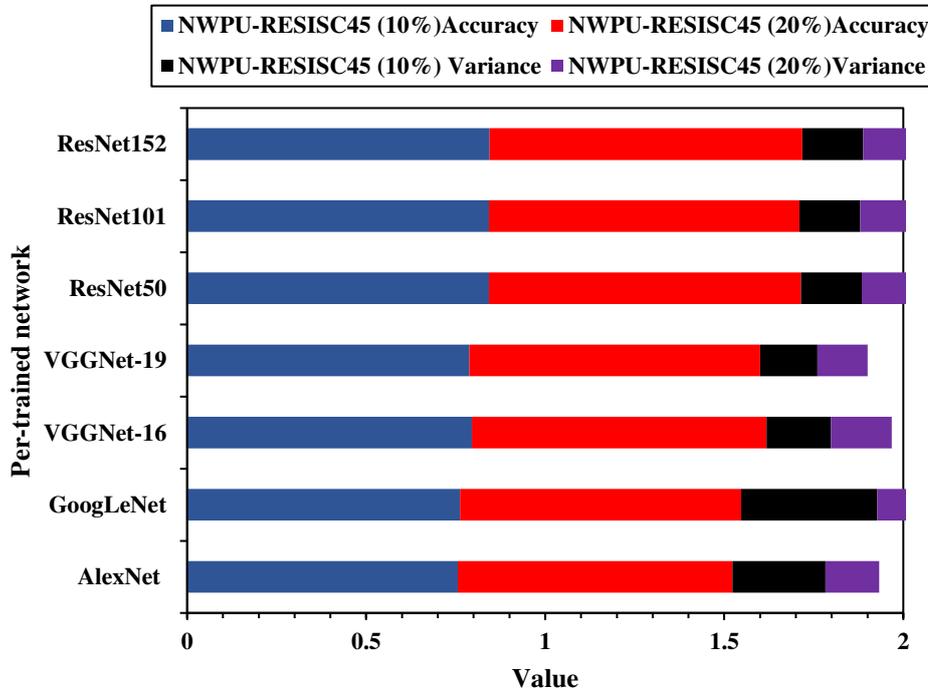
Through the application of four low-level feature expression methods and two intermediate-level feature expression methods, the NWPU-RESISC45 data set is classified as a result of feature extraction. The algorithm includes CH, LBP, GIST, and BoVW and LLC. The results are shown in Table 7:

**Table7.** Classification accuracy and variance of the NWPU-RESISC45 dataset based on the expression of middle and low-level features

Model		GIST	LBP	CH	BoVW	LLC
NWPU-RESISC45	Accuracy	15.87	19.22	24.81	41.72	38.78

(10%)	Variance	0.25	0.43	0.21	0.19	0.20
NWPU-RESISC45	Accuracy	17.85	27.11	27.47	44.93	38.57
(20%)	Variance	0.24	0.16	0.21	0.26	0.19

The UCMERCE data set uses seven pre-trained networks, AlexNet, GoogLeNet, VGGNet-16, VGGNet-19 classification accuracy. The experimental results are shown in Figure 10.



**Figure10.** The classification accuracy and variance of the NWPU-RESISC45 data set based on the feature extraction of the pre-trained network

It can be seen from the above experiments that the pre-trained network trained with ImageNet data is used as a feature extractor, and the accuracy of the extracted feature vector using a linear classifier for classification is much higher than that obtained by using low-level manual feature descriptors. The accuracy of traditional methods are all less than 50%, and the feature extraction using pre-training network is more than 75% on the NWPU-RESISC45 data. Compared with the features extracted by AlexNet, GoogLeNet, VGGNet-16, and VGGNet-19 networks, are more efficient, especially the ResNet-50 network model. The features extracted by the six pre-training networks of AlexNet, VGGNet-16, VGGNet-19 are fused, and the two ways of connecting and adding the values in the dimensions of the features Compared with the classification based on single network feature extraction, the accuracy has been improved. The best effect is the fusion of the three pre-trained networks of ResNet-50, ResNet-101, and ResNet-152, which is improved by more than 2.5%. In the way of voting on the classification results of different pre-trained network features, the accuracy of voting on all the results of the network is higher, which is about 0.5% higher.

## 5. Conclusions

In recent years, remote sensing technology has made great progress. It is more difficult to distinguish within a few years, and remote scene data sets are tagged and shared. These data sets tend to differ in higher categories. And the difference between categories will be smaller, which also increases the need for remote sensing image algorithm research. For these high-demand remote scene data sets, traditional classification methods may not achieve the desired classification accuracy.

Therefore, how to quickly and accurately identify these remote sensing images is a problem of many applications and technical requirements. This article focuses on the classification of remote sensing images based on neural networks. First, the existing remote sensing image classification algorithm and search mode are introduced. Then an effective neural network model is proposed for the classification of high-resolution SAR images. The method part introduces and discusses some theoretical foundations of cloud computing technology, and proposes a convolutional neural network model, which performs well in classifying high-score SAR images, and the model is proved through experiments. The experimental part uses three models to extract features from remote sensing scene images. Before training multiple networks, three methods are performed on feature networks separated by multiple networks to prove that the method proposed in this paper achieves more efficient results under the constraints of experimental conditions.

## **6. Declarations**

### *6.1 Ethics Approval and Consent to Participate*

This article is ethical, and this research has been agreed.

### *6.2 Consent for Publication*

The picture materials quoted in this article have no copyright requirements, and the source has been indicated.

### *6.3 Availability of Data and Material*

Data sharing does not apply to this article because no data set was generated or analyzed during the current research period

### *6.4 Competing Interests*

The authors declare that they have no competing interests

### *6.5 Funding*

The author(s) received no financial support for the research, authorship, and/or publication of this article.

## **References**

- [1]Hao S , Shuai L , Zhou S , et al. Unsupervised Cross-View Semantic Transfer for Remote Sensing Image Classification[J]. IEEE Geoscience and Remote Sensing Letters, 2016, 13(1):13-17.
- [2]Gong C , Han J , Lu X . Remote Sensing Image Scene Classification: Benchmark and State of the Art[J]. Proceedings of the IEEE, 2017, 105(10):1865-1883.
- [3]Huang X , Ye G , Chai H , et al. Compression and encryption for remote sensing image using chaotic system[J]. Security & Communication Networks, 2016, 8(18):3659-3666.
- [4]Silvia V , David M , Jordi I , et al. Production of a Dynamic Cropland Mask by Processing Remote Sensing Image Series at High Temporal and Spatial Resolutions[J]. Remote Sensing, 2016, 8(1):1-21.
- [5]Pandit V , Bhiwani R J . Remote Sensing Image Fusion Using High-Resolution Scale Regression And Half-Gradient[J]. International Journal of Scientific & Technology Research, 2020, 9(2):141-149.
- [6]Chen S , Zhang L , Feng R , et al. High-Resolution Remote Sensing Image Classification with RmRMR-Enhanced Bag of Visual Words[J]. Computational Intelligence and Neuroscience, 2021, 2021(1):1-10.
- [7]Yang, Pinglv, Zhou, et al. Segmentation for remote sensing image with shape and spectrum prior.[J]. Neuruting, 2016, 216(C):746-755.
- [8]Wang Y , Wang C . High resolution remote sensing image segmentation based on multi-features fusion[J]. Engineering Review, 2017, 37(3):289-297.

- [9]Maulik U , D Chakraborty. Remote Sensing Image Classification: A survey of support-vector-machine-based advanced techniques[J]. IEEE Geoscience & Remote Sensing Magazine, 2017, 5(1):33-52.
- [10]Fan C , Wang L , Liu P , et al. Compressed sensing based remote sensing image reconstruction via employing similarities of reference images[J]. Multimedia Tools & Applications, 2016, 75(19):1-25.
- [11]Lu X , Wang B , Zheng X , et al. Exploring Models and Data for Remote Sensing Image Caption Generation[J]. IEEE Transactions on Geoscience & Remote Sensing, 2017, PP(99):1-13.
- [12]Wang Y , Li Y , Zhao Q H . Region-based multiscale segmentation of panchromatic remote sensing image[J]. Kongzhi yu Juece/Control and Decision, 2018, 33(3):535-541.
- [13]Jean-Claude B , Rigobert T , Joachim E , et al. Geological Context Mapping of Batouri Gold District (East Cameroon) from Remote Sensing Imaging, GIS Processing and Field Works[J]. Journal of Geographic Information System, 2019, 11(6):766-783.
- [14]Tang S , Shen C , Zhang G . Adaptive regularized scheme for remote sensing image fusion[J]. Frontiers of Earth Science, 2016, 10(2):236-244.
- [15]Yang Y , Lu H , Huang S , et al. Remote Sensing Image Fusion Based on Fuzzy Logic and Saliency Measure[J]. IEEE Geoscience and Remote Sensing Letters, 2020, 17(11):1943-1947.
- [16]H Ismanto, Doloksaribu A , Susanti D S , et al. The Accuracy of Remote Sensing Image Interpretation On Changes In Land Use Suitability In Merauke Regency Papua[J]. International Journal of Engineering Trends and Technology, 2020, 68(10):42-47.
- [17]Yi C , Yan L , Tao W , et al. Remote Sensing Image Stripe Noise Removal: From Image Decomposition Perspective[J]. IEEE Transactions on Geoscience & Remote Sensing, 2016, 54(12):7018-7031.
- [18]H Song, Yang W , Dai S , et al. Multi-source remote sensing image classification based on two-channel densely connected convolutional networks[J]. Mathematical Biosciences and Engineering, 2020, 17(6):7353-7377.
- [19]Lu H , Wei J , Wang L , et al. Reference Information Based Remote Sensing Image Reconstruction with Generalized Nonconvex Low-Rank Approximation[J]. Remote Sensing, 2016, 8(6):1-20.
- [20]Sun F , Li H , Liu Z , et al. Arbitrary-angle bounding box based location for object detection in remote sensing image[J]. European Journal of Remote Sensing, 2021, 54(1):102-116.
- [21]Binisha R J . A Content based Remote Sensing Image Retrieval Using A Modified Local Color Descriptors[J]. Journal of Advanced Research in Dynamical and Control Systems, 2020, 12(3):244-248.
- [22]Liu X , Yu J , Song W , et al. Remote Sensing Image Classification Algorithm Based on Texture Feature and Extreme Learning Machine[J]. Computers, Materials and Continua, 2020, 65(2):1385-1395.
- [23]Wang H , Wang J , Yao F , et al. Multi-Band Contourlet Transform For Adaptive Remote Sensing Image Denoising[J]. The Computer Journal, 2020, 63(7):1084-1098.
- [24]Duan Y , Tao X , Liu X , et al. Robustness evaluation of remote-sensing image feature detectors with TH priori-information data set[J]. China Communications, 2020, 17(10):218-228.
- [25]Lei P , Liu C . Inception residual attention network for remote sensing image super-resolution[J]. International Journal of Remote Sensing, 2020, 41(24):9565-9587.



Xiaobo Wu was born in Lanzhou, Gansu, P.R. China, in 1981. She received the Master

degree from Chongqing University, P.R. China. Now, she works in School of Geographical Sciences, Lingnan Normal University. Her research interests include neural network, cloud security, and big data analysis.

E-mail: [wuxb@lingnan.edu.cn](mailto:wuxb@lingnan.edu.cn)