Neurocomputing 410 (2020) 211–219

Contents lists available at ScienceDirect

Neurocomputing

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

Brief papers

Multidimensional relation learning for hyperspectral image classification

Jie Fang [a,*], Xiaoqian Cao [b,*]

[a] University of Chinese Academy of Sciences, Beijing 100049, PR China
[b] College of Electrical and Information Engineering, Shaanxi University of Science and Technology, Xi'an 710021, Shaanxi, PR China

ARTICLE INFO

Article history:
Received 8 August 2019
Revised 3 April 2020
Accepted 11 May 2020
Available online 26 May 2020
Communicated by B. Hu

Keywords:
Hyperspectral image classification
Spatial relation learning
Spectral relation learning
Category relation learning
Multidimensional relation learning

ABSTRACT

Classification is a fundamental technique in hyperspectral image understanding field, which is to give each pixel a specific semantic label based on its characteristics automatically. Recently, even though convolutional neural network (CNN) based methods have been applied to hyperspectral image classification, their performances are not in accordance with our expectation. The reason is that, most of the existing methods can not exploit the intrinsic characteristics of different pixels in hyperspectral image effectively, they often ignore the inherent relationships among different spatial pixels, different spectral bands and different category dependencies. Specifically, the relationships among different spatial pixels mean that the similarities of attribute and location information among different pixels. The relationships among different spectral bands mean that the different contributions of them to the final classification. The relationships among different categories mainly mean that the severe sample imbalance problem in hyperspectral image classification task. To address the aforementioned problems, a deep convolutional network based multidimensional relation learning method is proposed in this paper. Firstly, we propose a weighted gaussian mask based spatial relation learning module to alleviate the spatial diffusion issue. Besides, we propose a channel attention alike spectral relation learning module to alleviate the interferences from noisy and redundant bands. Additionally, we propose an integrated decision based category relation learning module to alleviate the ill-conditioned classifying issue from imbalanced samples. Finally, we test the proposed method on four public and challenging datasets, and the experimental results validate the effectiveness and robustness of our method.

1. Introduction

Hyperspectral imaging (HSI) provides useful information for monitoring the surface of Earth [1], which attracts many researchers' interests because of its applications on various aspects such as environmental management [2], surveillance [3] and agriculture [4]. As a fundamental technique in hyperspectral image understanding field, hyperspectral image classification aims to give each pixel a specific semantic label according to its own characteristics automatically. During the last decade, several different approaches are proposed for hyperspectral image classification [5–8], these methods [9–12] especially the convolutional neural network (CNN) based ones [13–16] have achieved relatively satisfactory performances. Nevertheless, there are still three pendent issues which hinder the accuracy improvements of hyperspectral image classification, including spatial diffusion, band interference and sample imbalance. Firstly, in order to utilize the spatial information effectively, most existing 2D-CNN and 3D-CNN based methods directly divide the original image into several patches in overlapping way, then predict the label of each patch and treat it as that of the center pixel in the corresponding patch. These methods bring severe spatial diffusion problem due to the overmuch consideration of texture information, especially for pixels in boundary or near-boundary regions. Secondly, hyperspectral images are often accompanied by lots of noisy and redundant bands. Most existing methods utilizes two-stage strategy, band selection and pixel classification, to reduce the interferences and further improve the classification accuracy. Even though these two stage methods can alleviate the influences of noisy and redundant to a certain extent, they ignore the relevance between band selection and classification then may lose some potential discriminant information. Thirdly, sample imbalance issue is a common problem for hyperspectral image classification task, which usually leads to the

* Corresponding authors.
E-mail addresses: jackfang713508@gmail.com (J. Fang), caoxiaoqian@sust.edu.cn (X. Cao).

https://doi.org/10.1016/j.neucom.2020.05.034
0925-2312/

ill-conditioned classifier and further results in unexpected classification performances.

To address the aforemetioned problems, we propose a multidimensional relation learning method for hyperspectral image classification. Specifically, we propose a weighted gaussian mask based spatial relation learning module to alleviate the spatial diffusion issue, which gives different importance weights to each pixel in the patch according to the physical distance and spectral similarity between the center pixel and the current one. Then, we propose a channel attention-alike spectral relation learning module to alleviate the interferences from noisy and redundant bands. Specifically, we utilize a convolutional neural network to learn the importance weight of each band, and then multiply it with the original corresponding spectral band as the selected band to predict the category label. Additionally, we propose an integrated decision based category relation learning module to alleviate the ill-conditioned classifying issue from imbalanced samples, and further improve the classification accuracy especially for the categories with small scale samples. Specifically, we replace the C–classifier with C binary classifier and train each binary classifier with the focal loss in a joint learning way. In summary, the contributions of this paper can be concluded as follows:

1. We propose a weighted gaussian mask based spatial relation learning module to alleviate the spatial diffusion issue.
2. We propose a channel attention-alike spectral relation learning module to alleviate the interferences from noisy and redundant bands.
3. We propose an integrated decision based category relation learning module to alleviate the ill-conditioned classifying issue from imbalanced samples.

The remainder of this paper is organized as follows. In Section 2, we introduce some existing methods for hyperspectral image classification, including unsupervised ones and supervised ones. Section 3 describes the proposed method. We report the experimental results in Section 4 and conclude the paper in Section 5.

2. Related works

This section introduces some existing methods for hyperspectral image classification. According to different learning forms, these methods can be divided into two main subbranches, unsupervised learning methods and supervised learning ones, which are described in Sections 2.1 and 2.2 respectively.

2.1. Unsupervised learning methods for hyperspectral image classification

Unsupervised hyperspectral image classification methods do not need labeled data and they only use the inherent characteristics of raw image to divide pixels into different collections. At the very beginning, unsupervised methods like K-means [17] and Spectral Cluster [18,19] were widely used for hyperspectral image classification. Recently, many sophisticated unsupervised approaches have been proposed to extract more representative features of remote sensing images. These methods can be used to tackle hyperspectral images (HSIs) because they can explore hidden interactions and correlations of complex data. For example, Marinoni et al. proposed an unsupervised feature extraction method for hyperspectral image classification [20], which utilizes mutual information maximization to search the most relevant representation.

2.2. Supervised learning methods for hyperspectral image classification

Even though supervised methods achieved relatively satisfactory performances for hyperspectral image classification, supervised methods have shown to obtain more accurate results through learning the projection relationships between original image and category label map [21]. During the past years, many supervised machine learning methods have been proposed for hyperspectral image classification task [22]. Kernel based methods [23] such as support vector machine (SVM), statistical methods such as principal component analysis (PCA) [24] are widely used for hyperspectral image classification. Nonetheless, because of the inherent complex spectral and spatial structure, these traditional machine learning methods can not effectively exploit the information of hyperspectral image, and the classification results can not meet our expectations further. Recently, with the rapid development of artificial intelligence and machine learning, because of their strong feature extraction and representation capabilities, deep neural networks especially deep convolutional neural networks (CNNs) have achieved significant performances on many computer vision tasks, such as scene recognition [25–27], object detection [28], semantic image segmentation [29], *etc*. As for hyperspectral image classification task, Chen et al. proposed to use convolutional neural networks for hyperspectal image classification, which improved the classification performances by effectively exploiting the spatial information. Mei et al. [30] proposed a CNN architecture which integrates spectral signatures and spatial context through preprocessing each pixel in the image. Li et al. [31] proposed to learn discriminative features through combining CNN model with pixel pairs, and they obtained the classification label through the conventional majority voting mechanism. Zhang et al. [32] proposed to use two different CNNs to exploit spatial and spectral information of image separately, then combined the features together and obtained the final classification label with a softmax classifier. Zhao et al. [33] proposed to combine a CNN based spatial feature representation module and a balanced local discriminant embedding based spectral feature representation module to obtain the final representation for the classification. Even though these methods utilize spatial and spectral information for classification, the feature representation process is not in a joint learning way and can not exploit the spatial-spectral information effectively further. In this case, Zhong et al. [34] proposed to use the 3D CNN to extract the spatial and spectral information of the image simultaneously, which calculates the spatial neighbourhood and the corresponding spectral information for each pixel with 3D convolution kernels.

However, because of the sample scale and vanishing gradient problem [35], it is difficult to train very deep CNNs with hyperspectral images. Specifically, gradients obtained from deep layers of the network tend to be smaller, which makes the poor propagation of activations and elongates the loss function. In this case, the performances of deep CNNs are saturated and decrease rapidly. Recently, advanced deep CNN architectures have been proposed to exploit highly discriminative spectral-spatial features of hyperspectral images. A typical example is the residual network (ResNet) [36], which designs a novel CNN architecture based on the residual blocks [37] to address the vanishing gradient problem. The residual blocks in the ResNet have the same topology that performs a set of transformations whose outputs are aggregated by summation. Actually, ResNet can be considered as a ensemble of much shallower networks, which creates a much deeper architecture than its plain counterparts. Specifically, ResNet ensures a minimum loss of information by modeling each block closer to an identity mapping than to a zero mapping, and adds shortcut connections between each residual block to obtain more detailed information rather than just abstract ones. Due to the aforementioned

advantages of ResNet, this architecture based methods [38] have achieved more competitive performances for hyperspectral image classification.

3. Proposed method

This section details the proposed multidimensional relation learning method for hyperspectral image classification, which mainly contains three components, spatial relation learning, spectral relation learning and category relation learning. The overview architecture of the proposed method is introduced in Section 3.1. Spatial relation learning, spectral relation learning and category relation learning are described in Sections 3.3, 3.4 and 3.5 respectively.

3.1. Overview

Following the CNN based hyperspectral image classification methods, we also divide the original image into different patches in an overlapping way, then predict the semantic label of each patch and consider it as that of center pixel in the very patch. Differently, we propose a weighted gaussian mask based spatial relation learning module to alleviate the spatial diffusion issue, we propose a channel attention-alike spectral relation learning module to alleviate the interferences from noisy and redundant spectral bands, we propose integrated decision based category relation learning module to alleviate the ill-conditioned classifying issue from imbalanced samples. The overview flowchart of the proposed multidimensional relation learning method is shown in Fig. 1.

3.2. Data preparation

In order to consider the spectral and spatial information of hyperspectral image, we divide the image into several blocks in an overlapping way then predict the category of each block, and use it to represent the label of the center pixel in the block. Specifically, we denote a hyperspectral image with $\mathbf{X} \in \mathbf{R}^{W\times H\times C}$, where W, H, and C denotes the width, height, and band number of the image respectively. The $(i,j)_{th}$ pixel in image $\mathbf{X}$ can be defined as a spectral vector $\mathbf{x}_{i,j} \in \mathbf{R}^{C} = [x_{i,j,1}, x_{i,j,2}, \cdots, x_{i,j,C}]$, where $x_{i,j,k}$ represents the element of location with coordinate (x,y) in k_{th} band of image $\mathbf{X}$. Additionally, we define a neighbouring region $\mathbf{w}_{i,j} \in \mathbf{R}^{d\times d\times C}$ around $\mathbf{x}_{i,j}$, and use it to consider the texture spectral and spatial information of $\mathbf{x}_{i,j}$ simultaneously. It is worth to note that the neighbouring region of the boundary and near-boundary pixels can not constructed with the aforementioned strategy. In this case, we propose a new neighbouring region construction strategy, which is described as follows.

Firstly, expand the original hyperspectral image $\mathbf{X} \in \mathbf{R}^{W\times H\times C}$ to. $\mathbf{X}_e \in \mathbf{R}^{(W+K-1)\times(H+K-1)\times C}$ in a conventional zero padding way, where K is the size of neighbouring region. Then, for the padding set.

$S^e = \left\{\mathbf{x}^e_{i,j} \middle| i,j \leqslant p \quad or \quad i \geqslant W+p+1 \quad or \quad j \geqslant H+p+1\right\}$, we utilize the symmetric fulling method to replace its 0 values. Where $p = \frac{K-1}{2}$, and $\mathbf{x}^e_{i,j}$ represents the spectral vector in $(i,j)_{th}$ location of the expanded hyperspectral image $\mathbf{X}_e$.

3.3. Spatial relation learning

This subsection details the proposed spatial relation learning module. As is known, spatial information can contribute to the performance of hyperspectral image classification, especially for the methods based on convolutional neural networks (CNNs). However, for the spectral vector of each pixel, the overconsideration of texture information may suppress the inherent characteristics itself, and further result in the misclassification, especially for the boundary and near-boundary pixels (spatial diffusion problem). In order to visualize this problem, we give an example in Fig. 2.

To address the spatial diffusion problem, we propose a weighted gaussian mask based spatial relation learning module to emphasize the pixel to be classified, while consider its texture information simultaneously according to their similarities. Specifically, if the spectral vector is similar to ones in its neighbouring region, we can use more texture information to predict its semantic label and further to improve the regional connectivity of the whole predicted map. Otherwise, we need to highlighted the center pixel and suppress its surroundings to alleviate the spatial diffusion problem, especially for the boundary and near-boundary regions. The aforementioned two mechanisms are realized through a set of dynamic weighted gaussian masks $\underset{1\leqslant l\leqslant WH}{G^l}$. Additionally, the element of G^l is defined as Eq. (1),

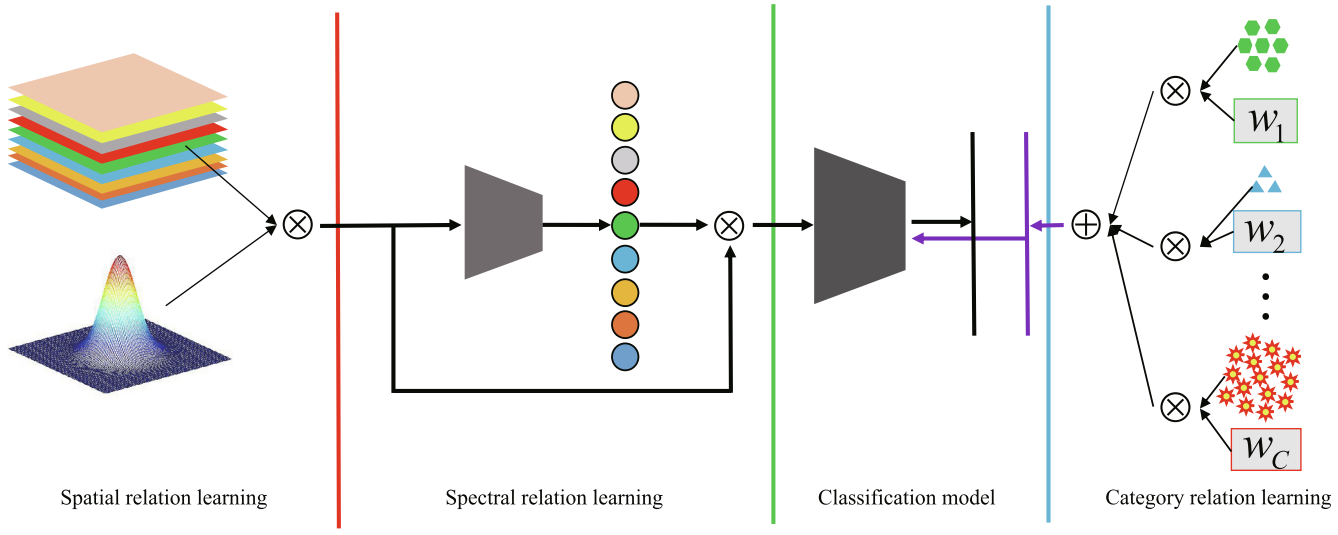

Fig. 1. The proposed multidimensional relation learning method for hyperspectral image classification, which mainly contains three relation learning modules and a conventional convolutional neural network (CNN).

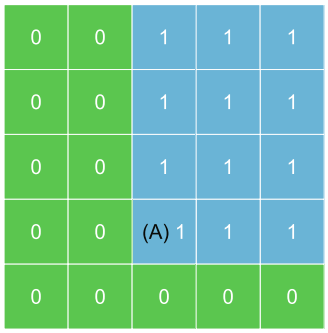

Fig. 2. The schematic diagram of spatial diffusion problem. Specifically, if a 3×3 block of neighbour A is used to predict its category, and a 3×3 average filter is directly applied to his block to extract its feature, point A will be misclassified to category 0 since the filter value in this point is 0.44 (which is much closer to 0 but not 1). Similarly, pixels distributed in boundary and near-boundary may suffer from this problem, which will lead to the coarse-boundary predicted mask, and we call this phenomenon spatial diffusion.

$$G^l_{w,h} = \exp\left(-\frac{(w-w_c)^2 + (h-h_c)^2}{2\sigma_l^2}\right), \tag{1}$$

where $G^l_{w,h}$ represents the $(w,h)_{th}$ element of l_{th} gaussian mask, and (w_c, h_c) represents the center coordinate of G^l. σ_l is a hyperparameter to control the window shape of G^l, which can be calculated with Eq. (2),

$$\sigma^l = \lambda E^l_s + (1-\lambda) C^l_s. \tag{2}$$

In Eq. (2), E^l_s and C^l_s represent the normalized Euclidean and Cosine similarities among center pixel and ones in its neighbouring region for l_{th} patch of hyperspectral image $\mathbf{X}$, and they are defined as Eqs. (3) and (4) respectively. Additionally, λ is a hyperparameter to balance these two similarity terms.

$$E^l_s = 1 - \frac{1}{K_w K_h}\sum_{w=1}^{K_w}\sum_{h=1}^{K_h}\frac{||I^l_{w,h} - I^l_{w_c,h_c}||_1}{||I^l_{w,h}||_1 + ||I^l_{w_c,h_c}||_1 + \varepsilon}, \tag{3}$$

$$C^l_s = \frac{1}{K_w K_h}\sum_{w=1}^{K_w}\sum_{h=1}^{K_h}\frac{||I^l_{w,h} \cdot I^l_{w_c,h_c}||_1}{||I^l_{w,h}||_2 \cdot ||I^l_{w_c,h_c}||_2 + \varepsilon}, \tag{4}$$

In Eqs. (3) and (4), K_w and K_h represent the width and height of the neighbouring region respectively. $I^l_{w,h}$ represents the $(w,h)_{th}$ spectral vector of l_{th} patch in hyperspectral image $\mathbf{X}$. $||\cdot||_1$ and $||\cdot||_2$ represent the 1-norm and 2-norm for vector respectively. Additionally, ε is a hyperparameter to avoid the division by zero.

After the aforementioned procedures, we multiply the weighted gaussian mask with the corresponding original patch to obtain the self-emphasized patch, which is defined in Eq. (5). And then input it to the classification network to predict the semantic label of the patch.

$$I^l_{se} = I^l \cdot G^l \tag{5}$$

In Eq. (5), I^l_{se} represents the l_{th} self-emphasized patch, I^l represents the l_{th} patch of the original hyperspectral image $\mathbf{X}$ and G^l represents the corresponding weighted gaussian mask. Additionally, $\cdot$ represents the element-wise multiplication operation.

After obtain I^l_{se}, let us revisit the case referred in Fig. 2. Simply, when we set σ^l to 1, the filter value of block of neighbour A can be calculated with Eq. (6),

$$v^l = \frac{1 + \frac{2}{e} + \frac{1}{e^2}}{1 + \frac{4}{e} + \frac{4}{e^2}} = 0.62 > 0.5, \tag{6}$$

and point A is correctly classified to category 1. The reason is that, the Gaussian mask highlights the intrinsic attribute information of center pixel itself while suppresses the influences of ones in its neighbour to a certain extent.

3.4. Spectral relation learning

This subsection details the proposed spectral relation learning module. As is aforementioned in Section 1, noisy and redundant bands are often exist in hyperspectral images, and they influence the classification results to a large extent. Most of the traditional band selection methods are independent to the classification, and since they can not exploit the discriminative information from the original bands effectively. In this case, we propose a self-attention alike spectral relation learning module to jointly optimize the band selection and spectral classification processes simultaneously. In order to facilitate the understanding of our spectral relation learning module, we first review the channel attention strategy.

As is shown in Fig. 3, channel attention module actually implements a re-characterization for the input feature cube, which improves the representation capability of the feature cube through emphasizing the high-energy channels and suppressing the low-energy ones in the feature cube simultaneously. Specifically, a conventional global average pooling or global max pooling operation is applied to each channel of the input feature cube, and value after pooling operation of each channel is considered as its own weight coefficient. Then the weight coefficients are multiplied to the corresponding channels of the feature cube, and the attentioned feature cube can be obtained.

Differ from the traditional channel attention architecture, the input of our expected spectral relation learning module is the fixed bands of hyperspectral image but not the dynamic feature maps. In this case, we can not directly use the popular channel attention module to select the discriminative bands to improve the final classification performance. If so, both the band information and the corresponding coefficient can not be updated in the training procedure, and the selection mechanism will lose efficacy.

To address the aforementioned problem, we propose channel attention alike spectral relation learning module to learn the weight coefficient of each band dynamically, as is shown in Fig. 4. More importantly, based on the traditional channel attention

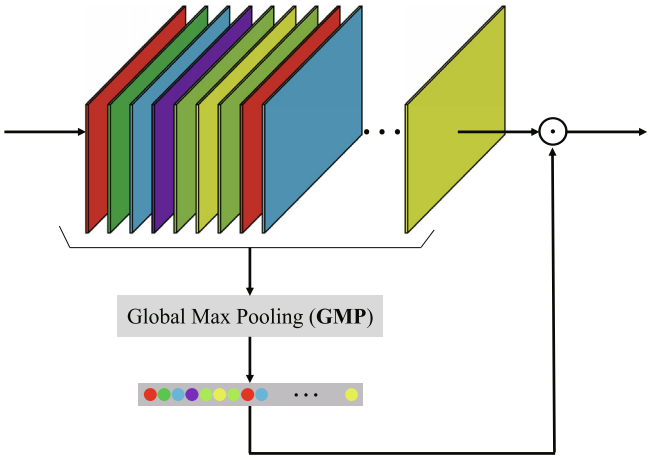

Fig. 3. The traditional channel attention architecture, GMP value of each feature channel is multiplied to the corresponding channel to enhance the representation, which can effectively alleviate the interferences of noisy and redundant channels. In addition, $\odot$ represents the multiplication of feature maps and their corresponding weights.

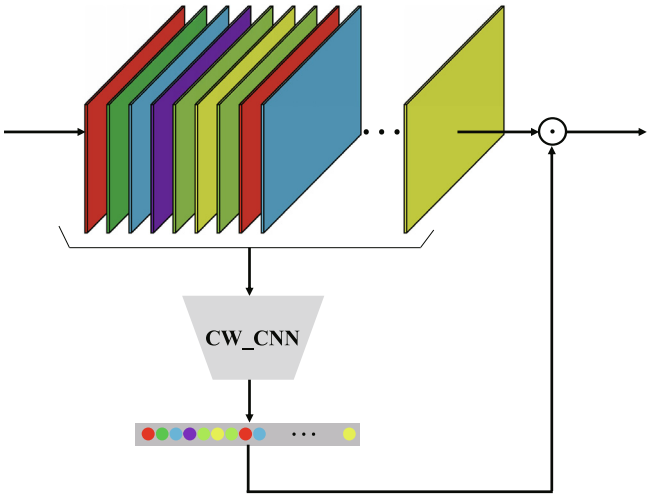

Fig. 4. The architecture of our spectral relation learning module, where *CW_CNN* is used to obtain weights of different bands according to their contributions to the final classification. This module is actually similar to a dimension reduction mechanism, which can suppress the influences of redundant bands to the subsequent encoding process.

architecture, our strategy of adjustment is simple and easy to comprehend. Specifically, we replace the global average pooling (GAP) or global max pooling (GMP) with a set of channel-weight convolutional neural networks (*CW_CNN*), which is a regression based CNN. The output of *CW_CNN* is a weight coefficient in the closed interval [0, 1], which reflects the importance of the current band for the final classification. Similar to the traditional channel attention module, each weight coefficient only related to the corresponding band information directly. Additionally, all *CW_CNNs* are weighted shared, which is to avoid the overfitting problem.

3.5. Category relation learning

This subsection details the proposed category relation learning module. As is aforementioned in Section 1, sample imbalance problem usually exists in hyperspectral images, and which may result in the ill-conditioned classifier and further influence the classification results, especially for the categories with small scale samples. In this case, we propose a category relation learning module to address the ill-conditioned classifier problem. Firstly, inspired by the conventional Focal loss, we transform the original N-classification problem into N-1 binary classification ones. Specifically, we set the category with the largest scale of samples as the background (c_0), and classify each category of the rest with it by a specific binary classifier. Additionally, each binary classifier is equipped with a focal loss alike loss function to avoid the sample imbalance problem. The difference is that, we utilize a automatic setting mechanism for hyperparameters but not choose it manually, compared to the traditional focal loss. Finally, we use the conventional nonmax suppress strategy to obtain the final semantic label of each pixel in the hyperspectral image.

Specifically, according to the aforementioned description of our category relation learning module, the loss function of the proposed method is defined as Eq. (7),

$$\ell = -\sum_{i=1}^{N-1}(\mu_i \cdot y^i \log \hat{y}^i + (1-\mu_i)\cdot(1-y^i)\log(1-\hat{y}^i)). \tag{7}$$

In Eq. (7), N represents the number of categories. y^i and $\hat{y}^i$ represents the groundtruth and predicted probability of the sample belong to category i respectively. Additionally, μ_i represents the hyperparameter which is to balance the loss of the groundtruth and the category i, which can be calculated with Eq. (8),

$$\mu_i = \frac{\sum_{w=1}^{W}\sum_{h=1}^{H} \mathrm{I}\{C_{w,h} = c_0\}}{\sum_{w=1}^{W}\sum_{h=1}^{H} \mathrm{I}\{C_{w,h} = c_i\}}. \tag{8}$$

In Eq. (8), W and H represents the width and height of the whole hyperspectral image respectively. $\mathrm{I}\{\cdot\}$ denotes the indicator function, it equals to 1 when the condition satisfies and 0 otherwise. Additionally, $C_{w,h}$ represents the category identifier of $(w,h)_{th}$ pixel in the original hyperspectral image, c_i represents the label index of i_{th} category. In the testing stage, we use the nonmax suppress strategy to obtain the final semantic label of the current patch (actually the semantic label of the center pixel in the patch), and the formulation is defined as Eq. (9),

$$i = \underset{0\leqslant i\leqslant N-1}{\arg\max} \hat{y}^i. \tag{9}$$

4. Experiments

This section details the experiments, including datasets, experimental settings, evaluation metrics, contrasting methods and experimental results.

4.1. Datasets

This subsection introduces four public datasets used in this paper, including Indian Pines (IP) [39], University of Pavia (UP) [40], Salinas Valley (SV) and Kennedy Space Center (KSC).

Indian Pines (IP) dataset was collected in 1992 by Airborne/Infrared Imaging Spectrometer (AVIRIS) sensor over an agricultural area in Northwestern Indiana. Specifically, IP dataset covers a set of agricultural fields with regular geometry and irregular forest areas. The selected scene contains 145×145 pixels, with a total of 224 spectral bands in the wavelength range from 400 to 2500 nm, and spatial resolution of 20 meters per pixel. Additionally, about half of the data, 10249 from total of 21025, contains the label information from 16 different categories.

University of Pavia (UP) dataset was collected by the Reflective Optics System Imaging Spectrometer sensor over the campus as the university of Pavia, Northern Italy. This dataset mainly contains an urban environment with multiple solid structures, natural objects and shadows. The considered scene contains 610×340 pixels, with a total 103 spectral bands in the wavelength range from 0.43 to 0.86 μm, and spatial resolution of 1.3 m per pixel. Additionally, about 20% of the data, 42776 from total of 207400, contains the label information from nine categories.

Salinas Valley (SV) dataset was collected by the AVIRIS sensor over the Salinas Valley, CA, USA. The area covered by SV comprises 512 lines by 217 samples. Additionally, this dataset was only available as at-sensor radiance data, which contains the label information from 16 categories.

Kennedy Space Center (KSC) dataset was collected in 1996 by the AVIRIS sensor over the Kennedy Space Center in FL, USA. The KSC dataset contains 512×614 pixels, with a total 176 bands in the wavelength range from 400 to 2500 nm, and spatial resolution of 20 m per pixel. Additionally, about 1.6% of the data, contains label information from 13 categories.

4.2. Experimental settings

In this subsection, we introduce the experimental settings, including data partition, hyperparameter settings and testing platform, which are described as followings.

4.2.1. Data partition

For each dataset tested in this paper, we randomly choose 10% samples as the training set and the left as the testing set.

Additionally, in order to validate the robustness of the proposed method, we repeat the experiment ten times on each dataset, including the (data participation, training and testing), and calculate the expectation and variance to reflect the performance of our method.

4.2.2. Hyperparameter settings

In this subsection, we details the hyperparameter settings in this paper. Actually, there is only one hyperparameter to be confirmed in this paper, λ in Eq. (2), which is to balance the importance of two loss terms. For convenience, λ is set to 0.5 in this paper. In other words, we think that the Euclidean similarity and the cosine similarity have the same contribution to the σ^l. In addition, the patch size is set to 9×9. For each of the contrasting algorithms, we train the model through min-batch SGD with batch size 128 for 1000 epoches. The initial learning rate is set to 2×10^{-4}, and the momentum is set to 0.9.

4.2.3. Testing platform

The algorithms are implemented with Pytorch, and the testing platform is X99UD4 of GIGABTYE, GPU (8G×8) of Titan X.

4.3. Evaluation metrics

In order to evaluate the experimental results, two commonly used quantitative metrics are used to evaluate the classification performance, overall accuracy (OA) and average accuracy (AA). Additionally, we calculate the standard deviation to demonstrate the stability and robustness of the methods.

4.4. Contrasting methods

To validate the effectiveness of the proposed method, we compare it with six state-of-the-art ones. Additionally, all of the six contrasting methods are based on deep neural networks, including multilayer perception (MLP), 1D-CNN, 2D-CNN, deep learning ensemble (DLE) [41], cascaded recurrent neural network (CasRNN) [42], and 3D-CNN. Specifically, MLP and 1D-CNN predict the semantic label only using the spectral information. 2D-CNN and 3D-CNN improve the classification performances through considering the spectral and spatial information of the hyperspectral image simultaneously. DLE combines deep learning and ensemble methods to improve the performance of hyperspectral classification. CasRNN utilizes gated recurrent units (GRUs) to explore the redundant and complementary information of hyperspectral images.

4.5. Experimental results

This section reports the experimental results. In order to validate the effectiveness of each proposed relation learning module and the superiority of the proposed method, we design the ablation experiments and contrasting experiments, the details are described in the followings.

4.5.1. Ablation experiments

This subsubsection reports the ablation experimental results, the baseline network architecture is a 3-layer 3D-CNN, which is defined in Table 1. Additionally, the ablation experiment is performed on the IP dataset.

The ablation experimental results are shown in Table 2, where SpaRL, SpeRL and CatRL represents the proposed spatial relation learning module, spectral relation learning module and category relation learning module respectively. From Table 2, we can see that, each of the proposed relation learning module contribute improvement of the classification accuracy. Specifically, based on the traditional 3D-CNN architecture, spatial relation learning module obtains 0.44% and 0.74% increments in terms of OA and AA respectively. The reason is that, the proposed spatial relation learning module takes full consideration for the significant dependence of pixels in the flat region and the strong independence of pixels in the boundary and near-boundary regions simultaneously. Additionally, when the proposed spatial relation learning module and spectral relation learning module are incorporated into the traditional 3D-CNN, the classification performance is improved further. Specifically, compared with 3D-CNN + SpaRL method, 3D-CNN + SpaRL + SpeRL method gains 0.81% OA increment and 0.43% AA increment. The reason is that, the spectral relation learning module filters the noisy and redundant bands through giving each bands of the hyperspectral image a specific importance weight, which contributes to the robustness and effectiveness for the learning of the subsequent representation and classifier and further improves the classification performance. Finally, when the category relation learning module is incorporated into the 3D-CNN + SpaRL + SpeRL architecture, it achieves 0.19% and 2.78% improvements in terms of OA and AA respectively. Comparatively speaking, the AA improvement is more significant. This is because that, the proposed category relation learning module takes full consideration of the data imbalance problem in hyperspectral image classification task, and addresses the ill-conditioned problem through giving different penalty weights to different categories according to their sample distribution. To demonstrate this point furthermore,

Table 1
The architecture of the baseline 3D-CNN.

Layer	Input	Kernel	Stride	Padding Mode	Output
Conv1	$9 \times 9 \times BN$	$3 \times 3 \times 3 \times 128$	$1 \times 1 \times 1$	SAME	$9 \times 9 \times 128$
Pool1	$9 \times 9 \times 128$	/	$2 \times 2 \times 1$	None	$4 \times 4 \times 128$
Conv2	$4 \times 4 \times 128$	$3 \times 3 \times 3 \times 256$	$1 \times 1 \times 1$	SAME	$4 \times 4 \times 256$
Conv3	$4 \times 4 \times 256$	$3 \times 3 \times 3 \times 512$	$1 \times 1 \times 1$	SAME	$4 \times 4 \times 512$
Pool2	$4 \times 4 \times 512$	/	$4 \times 4 \times 1$	None	1×512
SoftMax	1×512	/	/	/	$1 \times N_c$

Table 2
Ablation experimental results.

/	SpaRL	SpeRL	CatRL	OA ± std$_{OA}$ (%)	AA ± std$_{AA}$ (%)	Time (s)
Ablation	–	–	–	96.81 ± 0.73	94.10 ± 0.88	15.96
	√	–	–	97.25 ± 0.69	94.84 ± 0.86	19.35
	√	√	–	98.06 ± 0.75	95.27 ± 0.93	23.84
	√	√	√	98.27 ± 0.60	98.05 ± 0.74	26.71

Table 3
The influence of CategoryRL for each class on IP dataset.

No.	Category	Sample Number	Without CatRL (%)	with CatRL (%)
1	Asphalt	46	92.68	97.56
2	Corn-notill	1428	97.98	98.13
3	Corn-min	830	98.13	97.86
4	Corn	237	94.95	97.25
5	Grass-pasture	483	97.01	98.16
6	Grass-trees	730	98.02	97.72
7	Grass-pasture-mowed	28	80.00	100.00
8	Hay-windrowed	478	98.14	98.37
9	Oats	20	88.89	94.44
10	Soybean-notill	972	98.17	98.40
11	Soybean-mintill	2455	99.64	99.55
12	Soybean-clean	593	97.94	98.31
13	Wheat	205	95.75	97.83
14	Woods	1265	99.03	98.86
15	Buildings-Grass-Trees	386	98.27	97.98
16	Stone-Steel-Towers	93	90.48	97.62

we have counted the classification performance of each category, which is shown in Table 3. From which we can see that, the incorporation of category relation learning module significantly improve the classification accuracy of category with small sample scale such as grass-pasture-mowed, oats, and stone-steel-towers.

Besides the classification accuracy, we have analyzed the time-consuming situation when each of the proposed relation learning modules is incorporated into the original baseline network, which is counted in the last column of Table 2. From which we can see that, the incremental time consumption of spatial relation learning module, spectral relation learning module, and category relation learning module are 3.57s, 4.49s, and 2.87s respectively. Compared with the performance increments, the incremental time consumptions are within acceptable limits.

4.5.2. Contrasting experiments

This subsection reports the contrasting experimental results on four testing datasets, and our proposed three relation learning modules are applied to conventional 2D-CNN and 3D-CNN models and we named them Ours1 and Ours2 respectively. All the experimental results are presented in Table 4. Next, we will demonstrate the contrasting experimental results on four testing datasets successively.

Contrasting experimental results on IP dataset are shown in the first block in Table 4 and the visualized results are shown in Fig. 5. From which we can see that, 1D-CNN achieves better performances than MLP on IP dataset, 1.21% OA increment and 0.42% AA increment. Meanwhile, the standard deviations of OA and AA become larger, in other words, the robustness of 1D-CNN is weaker than MLP. The reason is that, MLP is actually a fully connected network, which can not depict the similarity information of adjacent bands while 1D-CNN considers this point through some kernels with different fixed sizes. In this case, 1D-CNN achieves better classification performance. However, 1D-CNN utilizes same descriptors to depict the relationships of bands in different intervals, which makes the trained model sensitive to the characters of data, and further results in the testing results with relatively large randomness. In addition, DLE [41] achieves better performance than 1D-CNN, which is because of the applications of multi-branch feature representation and decision-level integration.

Contrasting experimental results on UP dataset are shown in the second block in Table 4 and the visualized results are shown in Fig. 6. From which we can see that, 2D-CNN achieves better classification performances than 1D-CNN on UP dataset, which achieves 0.83% increment in terms of OA and 1.47% increment in terms of AA. Investigating its reason, the spatial context information plays a key role for this phenomenon. Specifically, inherent local similarity on spatial domains of hyperspectral image makes

Table 4
Experimental results on four datasets

Dataset	Method	OA ± std_{OA} (%)	AA ± std_{AA} (%)
IP	MLP	84.38 ± 0.52	81.63 ± 1.33
	1D-CNN	85.59 ± 1.03	82.05 ± 1.48
	2D-CNN	86.42 ± 0.94	83.52 ± 1.27
	DLE [41]	91.20 ± 1.15	89.79 ± 1.34
	CasRNN [42]	91.79 ± 1.07	95.94 ± 1.19
	2D-CNN + MRL (Ours1)	90.55 ± 0.88	89.43 ± 1.06
	3D-CNN	96.81 ± 0.73	94.10 ± 0.88
	3D-CNN + MRL (Ours2)	98.27 ± 0.60	98.05 ± 0.74
UP	MLP	92.23 ± 0.19	89.45 ± 0.33
	1D-CNN	94.78 ± 0.26	93.05 ± 0.67
	2D-CNN	95.43 ± 1.02	94.18 ± 1.25
	DLE [41]	91.23 ± 1.38	90.45 ± 1.42
	CasRNN [42]	90.30 ± 1.45	87.97 ± 1.36
	2D-CNN + MRL (Ours1)	96.68 ± 0.93	96.39 ± 1.08
	3D-CNN	98.25 ± 0.44	97.48 ± 0.64
	3D-CNN + MRL (Ours2)	99.16 ± 0.40	98.96 ± 0.52
SV	MLP	90.76 ± 0.34	93.84 ± 0.26
	1D-CNN	92.85 ± 0.71	96.09 ± 0.48
	2D-CNN	94.01 ± 1.62	96.18 ± 1.85
	DLE [41]	93.35 ± 1.75	91.86 ± 1.53
	CasRNN [42]	94.21 ± 1.58	91.39 ± 1.71
	2D-CNN + MRL (Ours1)	96.23 ± 1.49	96.42 ± 1.37
	3D-CNN	97.34 ± 1.28	98.01 ± 1.04
	3D-CNN + MRL (Ours2)	98.82 ± 0.97	99.14 ± 0.49
KSC	MLP	90.87 ± 1.65	89.54 ± 1.73
	1D-CNN	93.05 ± 1.71	91.43 ± 1.88
	2D-CNN	94.28 ± 1.69	93.01 ± 1.76
	DLE [41]	94.42 ± 1.58	93.25 ± 1.63
	CasRNN [42]	95.08 ± 1.36	93.94 ± 1.40
	2D-CNN + MRL (Ours1)	95.34 ± 1.57	95.06 ± 1.64
	3D-CNN	98.79 ± 0.94	97.04 ± 1.05
	3D-CNN + MRL (Ours2)	99.04 ± 0.59	98.92 ± 0.62

that, context information can contribute to the discriminative descriptor of a given pixel, and 2D-CNN has natural advantages for representing the 2D spatial information. In this case, 2D-CNN shows better classification performances, compared to 1D-CNN.

Contrasting experimental results on SV dataset are shown in the third block in Table 4. From which we can see that, CasRNN [42] achieves better performance than DLE [41] and 2D-CNN on SV dataset. In addition, 3D-CNN obtains significant improvement compared to 2D-CNN on SV dataset, which achieves 3.23% OA increment and 1.83% AA increment. The reason is that, compared to 2D-CNN and DLE [41], 3D-CNN and CasRNN [42] take full consideration of spatial and spectral information of each pixel in the hyperspectral image simultaneously. Specifically, the spatial–temporal descriptor, 3D convolution kernel, is transformed to depict

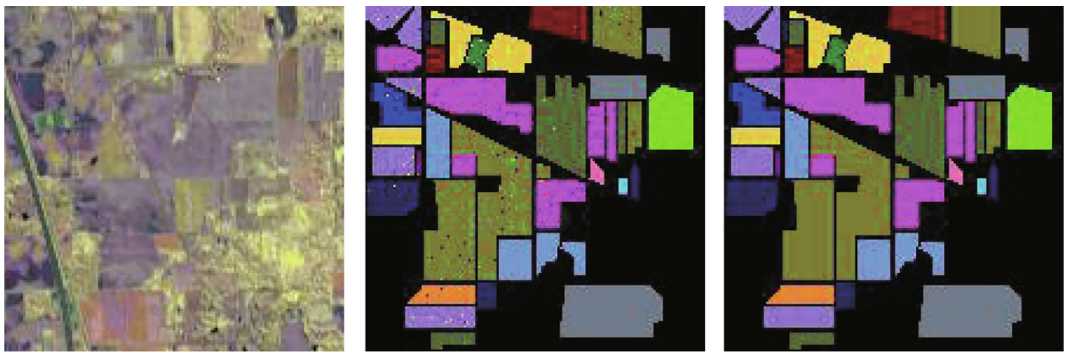

Fig. 5. The visualized results of our 3D-CNN + MRL method on IP dataset. The left, middle, and right image denote pseudocolor image, predicted label mask, and groundtruth of IP dataset respectively.

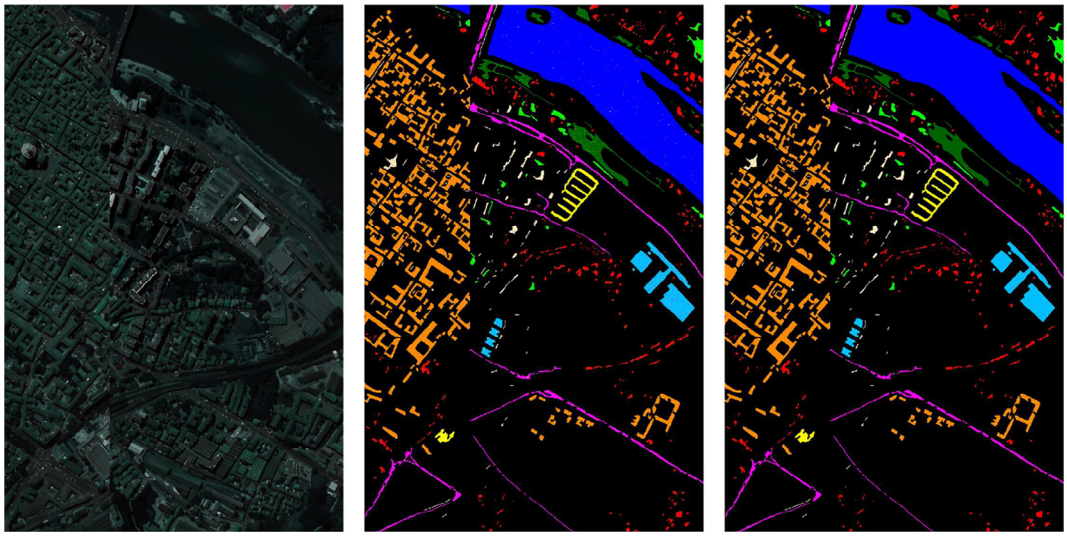

Fig. 6. The visualized results of our 3D-CNN + MRL method on UP dataset. The left, middle, and right image denote pseudocolor image, predicted label mask, and groundtruth of UP dataset respectively.

the spatial-spectral information of the hyperspectral image, which is appropriate and reasonable. CasRNN [42] utilizes gated recurrent units to represent the relationships of different spectral bands creatively. In these cases, 3D-CNN and CasRNN [42] achieve competitive performances.

Contrasting experimental results on KSC dataset are shown in the fourth block in Table 4. From which we can see that, the proposed three relation learning modules can significantly improve the performances for hyperspectral image classification on KSC dataset, no matter the based line network is 2D-CNN or 3D-CNN. For instance, when the proposed three relation learning module are incorporated into 2D-CNN architecture, it achieves 1.06% OA increment and 2.21% AA increment. The reason is that, the spatial relation learning module alleviates the spatial diffusion issue, the spectral relation learning module alleviate the interferences from the noisy and redundant spectral bands, and the category relation learning module alleviates the ill-conditioned classifier problem from imbalanced samples.

Generally speaking, through the ablation experiments and contrasting experiments, the effectiveness and robustness of the proposed method are validated.

5. Conclusion and future work

In this paper, we propose a multidimensional relation learning based method for hyperspectral image classification. Specifically, we propose a weighted gaussian mask based spatial relation learning module to alleviate the spatial diffusion issue. Besides, we propose a channel attention alike spectral relation learning module to alleviate the interferences from noisy and redundant spectral bands. Additionally, we propose a integrated decision based category relation learning module to alleviate the ill-conditioned classifying issue from imbalanced samples. Finally, we test the proposed method on four public and challenging datasets, and the experimental results validate the effectiveness and robustness of our method.

Even though the proposed multidimensional relation learning method improves the hyperspectral image classification performances, there is still an important issue to be addressed: how to ensure the sparsity of the weight coefficient in spectral relation learning module to achieve the purpose of band selection? In the future, we intend to apply 0-norm (or 1-norm) constraint to the spectral relation learning module to improve the band selection performance and improve the classification results further.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Jie Fang: Conceptualization, Methodology, Writing - original draft. **Xiaoqian Cao:** Writing - review & editing.

References

[1] D. Landgrebe, Hyperspectral image data analysis, IEEE Signal Process. Mag. 19 (1) (2002) 17–28.
[2] J.M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N.M. Nasrabadi, J. Chanussot, Hyperspectral remote sensing data analysis and future challenges, IEEE Geosci. Remote Sensing Mag. 1 (2) (2013) 6–36.
[3] B. Uzkent, A. Rangnekar, M.J. Hoffman, Aerial vehicle tracking by adaptive fusion of hyperspectral likelihood maps, IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017, pp. 233–242.
[4] Z. Xia, Y. Sun, K. Shang, L. Zhang, S. Wang, Crop classification based on feature band set construction and object-oriented approach using hyperspectral images, IEEE J. Selected Top. Appl. Earth Observations Remote Sensing 9 (9) (2016) 4117–4128.
[5] W. Li, G. Wu, F. Zhang, Q. Du, Hyperspectral image classification using deep pixel-pair features, IEEE Trans. Geosci. Remote Sens. 55 (2) (2016) 844–853.
[6] G. Cheng, Z. Li, J. Han, X. Yao, L. Guo, Exploring hierarchical convolutional features for hyperspectral image classification, IEEE Trans. Geosci. Remote Sens. 99 (2018) 1–11.
[7] M. Paoletti, J. Haut, J. Plaza, A. Plaza, A new deep convolutional neural network for fast hyperspectral image classification, ISPRS J. Photogrammetry Remote Sensing 145 (2018) 120–147.
[8] B. Pan, Z. Shi, X. Xu, Mugnet: Deep learning for hyperspectral image classification using limited samples, ISPRS J. Photogrammetry Remote Sensing 145 (2018) 108–119.
[9] Q. Wang, X. He, X. Li, Locality and structure regularized low rank representation for hyperspectral image classification, IEEE Trans. Geosci. Remote Sensing 99 (2018) 1–13.
[10] P. Zhou, J. Han, G. Cheng, B. Zhang, Learning compact and discriminative stacked autoencoder for hyperspectral image classification, IEEE Trans. Geosci. Remote Sensing..
[11] W. Song, S. Li, L. Fang, T. Lu, Hyperspectral image classification with deep feature fusion network, IEEE Trans. Geosci. Remote Sens. 56 (6) (2018) 3173–3184.
[12] Q. Wang, J. Lin, Y. Yuan, Salient band selection for hyperspectral image classification via manifold ranking, IEEE Trans. Neural Networks Learn. Syst. 27 (6) (2016) 1279–1289.
[13] J. Yue, W. Zhao, S. Mao, H. Liu, Spectral–spatial classification of hyperspectral images using deep convolutional neural networks, Remote Sensing Lett. 6 (6) (2015) 468–477.
[14] 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 Sensing 54 (10) (2016) 6232–6251.
[15] H. Lee, H. Kwon, Going deeper with contextual cnn for hyperspectral image classification, IEEE Trans. Image Process. 26 (10) (2017) 4843–4855.
[16] Z. Zhong, J. Li, Z. Luo, M. Chapman, Spectral–spatial residual network for hyperspectral image classification: A 3-d deep learning framework, IEEE Trans. Geosci. Remote Sens. 56 (2) (2017) 847–858.
[17] J. Haut, M. Paoletti, A. Paz-Gallardo, J. Plaza, A. Plaza, Cloud implementation of logistic regression for hyperspectral image classification, in: Proceedings of the 17th International Conference on Computational and Mathematical Methods in Science and Engineering, CMMSE, 2017, pp. 1063–2321.
[18] Z. Yang, Y. Yuan, N. Feiping, W. Qi, Spectral clustering based on iterative optimization for large-scale and high-dimensional data, Neurocomputing 318 (2018) 227–235.
[19] A. Dong, J. Li, Z. Bei, M. Liang, S.O. Science, C. University, Hyperspectral image classification algorithm based on spectral clusteringand sparse representation, Acta Optica Sinica 37 (8) (2017) 0828005..
[20] A. Marinoni, P. Gamba, Unsupervised data driven feature extraction by means of mutual information maximization, IEEE Trans. Comput. Imaging 3 (2) (2017) 243–253.
[21] E.G. Njoku, T.G. Farr, Encyclopedia of remote sensing, Encyclopedia Earth Sci. (4).
[22] G. Campsvalls, D. Tuia, L. Bruzzone, J.A. Benediktsson, Advances in hyperspectral image classification: Earth monitoring with statistical learning methods, IEEE Signal Process. Mag. 31 (1) (2014) 45–54.
[23] G. Camps-Valls, L. Bruzzone, Kernel-based methods for hyperspectral image classification, IEEE Trans. Geosci. Remote Sensing 43 (6) (2005) 1351–1362.
[24] Y. Li, Z. Wu, W. Jie, A. Plaza, J. Li, Z. Wei, Fast principal component analysis for hyperspectral imaging based on cloud computing, Geoscience & Remote Sensing Symposium, 2015.
[25] J. Fang, Y. Yuan, X. Lu, Y. Feng, Robust space-frequency joint representation for remote sensing image scene classification, IEEE Trans. Geosci. Remote Sens. (2019) 1–11, https://doi.org/10.1109/TGRS.2019.2913816.
[26] Y. Yuan, J. Fang, X. Lu, Y. Feng, Remote sensing image scene classification using rearranged local features, IEEE Trans. Geosci. Remote Sens. 57 (3) (2019) 1779–1792.
[27] X. Lu, Y. Yuan, J. Fang, et al., Jm-net and cluster-svm for aerial scene classification., in: IJCAI, 2017, pp. 2386–2392.
[28] C. Wang, X. Bai, S. Wang, J. Zhou, P. Ren, Multiscale visual attention networks for object detection in vhr remote sensing images, IEEE Geosci. Remote Sens. Lett. 16 (2) (2019) 310–314.
[29] J. Fang, X. Cao, Gan and dcn based multi-step supervised learning for image semantic segmentation, in: Chinese Conference on Pattern Recognition and Computer Vision (PRCV), Springer, 2018, pp. 28–40.
[30] S. Mei, J. Ji, Q. Bi, J. Hou, D. Qian, L. Wei, Integrating spectral and spatial information into deep convolutional neural networks for hyperspectral classification, Geoscience & Remote Sensing Symposium, 2016.
[31] W. Li, G. Wu, F. Zhang, Q. Du, Hyperspectral image classification using deep pixel-pair features, IEEE Trans. Geosci. Remote Sens. 55 (2) (2017) 844–853, https://doi.org/10.1109/TGRS.2016.2616355.
[32] H. Zhang, L. Ying, Y. Zhang, S. Qiang, Spectral-spatial classification of hyperspectral imagery using a dual-channel convolutional neural network, Remote Sensing Lett. 8 (5) (2017) 438–447.
[33] 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) (2016) 4544–4554, https://doi.org/10.1109/TGRS.2016.2543748.
[34] Z. Zhong, J. Li, Z. Luo, M. Chapman, Spectral-spatial residual network for hyperspectral image classification: a 3-d deep learning framework, IEEE Trans. Geosci. Remote Sens. 56 (2) (2018) 847–858.
[35] R.K. Srivastava, K. Greff, J. Schmidhuber, Training very deep networks, in: C. Cortes, N.D. Lawrence, D.D. Lee, M. Sugiyama, R. Garnett (Eds.), Advances in Neural Information Processing Systems, 28, Curran Associates Inc, 2015, pp. 2377–2385.
[36] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.
[37] S. Xie, R. Girshick, P. Dollár, Z. Tu, K. He, Aggregated residual transformations for deep neural networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 1492–1500.
[38] M.E. Paoletti, J.M. Haut, R. Fernandez-Beltran, J. Plaza, A.J. Plaza, F. Pla, Deep pyramidal residual networks for spectral-spatial hyperspectral image classification, IEEE Trans. Geosci. Remote Sens. 57 (2) (2019) 740–754.
[39] R.O. Green, M.L. Eastwood, C.M. Sarture, T.G. Chrien, M. Aronsson, B.J. Chippendale, J.A. Faust, B.E. Pavri, C.J. Chovit, M. Solis, Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (aviris), Remote Sens. Environ. 65 (3) (1998) 227–248.
[40] B. Kunkel, F. Blechinger, R. Lutz, R. Doerffer, H.V.D. Piepen, Rosis (reflective optics system imaging spectrometer) – a candidate instrument for polar platform missions, in: Optoelectronic Technologies for Remote Sensing from Space, 1988.
[41] Y. Chen, Y. Wang, Y. Gu, X. He, P. Ghamisi, X. Jia, Deep learning ensemble for hyperspectral image classification, IEEE J. Selected Topics Appl. Earth Observations Remote Sensing.
[42] R. Hang, Q. Liu, D. Hong, P. Ghamisi, Cascaded recurrent neural networks for hyperspectral image classification, IEEE Trans. Geosci. Remote Sensing.

Jie Fang is with the Key Laboratory of Spectral Imaging Technology CAS, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 7101-19, Shaanxi, PR China, and with the University of Chinese Academy of Sciences, Beijing 100049, PR China. His research interests include Artificial intelligence, Machine Learning and Video/Audio Analysis.

Xiaoqian Cao is with the College of Electric and Information Engineering, Shaanxi University of Science and Technology, Xi'an 710021, Shaanxi, PR China.

Her research interests include Pattern Recognition, Visual Understanding, and Stereo Matching.