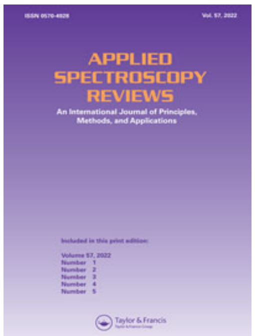

Taylor & Francis
Taylor & Francis Group

Applied Spectroscopy Reviews

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/laps20

An overview of hyperspectral image feature extraction, classification methods and the methods based on small samples

Xueying Li, Zongmin Li, Huimin Qiu, Guangli Hou & Pingping Fan

APPLIED SPECTROSCOPY REVIEWS
https://doi.org/10.1080/05704928.2021.1999252

An overview of hyperspectral image feature extraction, classification methods and the methods based on small samples

Xueying Li[a,b], Zongmin Li[c], Huimin Qiu[a], Guangli Hou[a], and Pingping Fan[a]

[a]Institute of oceanographic Instrumentation, Qilu University of Technology (Shandong Academy of Sciences), Qingdao, China; [b]School of Geosciences, China University of Petroleum (East China), Qingdao, China; [c]College of computer science and technology, China University of Petroleum (East China), Qingdao, China

ABSTRACT

Hyperspectral image (HSI) contains rich spatial and spectral information, which has been widely used in resource exploration, ecological environment monitoring, land cover classification and target recognition. However, the nonlinearity of HSI data and the strong correlation between bands also bring difficulties and challenges to HSI application. In particular, the limited available hyperspectral training samples will lead to the classification accuracy cannot be improved. Therefore, making full use of the advantages of HSI data, through algorithms and strategies to solve the limited training samples, high-dimensional HSI data and effective classification method, so as to improve the classification accuracy. This paper reviews the research results of the feature extraction methods and classification methods of HSI classification in recent years. In addition, this paper expounds five kinds of small sample strategies, and solves the problem of small sample in HSI classification from different angles. Small sample strategy will be the focus of HSI classification research in the future. To solve the problem of small sample classification can greatly promote the application of HSI.

KEYWORDS
Hyperspectral image; small samples; feature extraction; classification methods

1. Introduction

Hyperspectral imaging technology began to rise in the field of earth observation in the 1980s. Imaging spectrometer is a hardware device for collecting hyperspectral images (HSI). With the rapid development of imaging spectrometer in recent decades, hyperspectral imaging technology has been widely used in resource exploration, ecological environment monitoring, land cover classification and target recognition.[1–2]

Hyperspectral remote sensing technology integrates spectrum and image, and can obtain image and spectral data at the same time.[3] Hyperspectral image sensor provides HSI of hundreds of spectral bands in the same region. There are many spectral bands in spectral information, including visible spectrum, near infrared spectrum, mid infrared spectrum, etc., and each pixel has a continuous and high resolution spectrum. HSI can

CONTACT Pingping Fan fanpp_sdioi@126.com Institute of oceanographic Instrumentation, Qilu University of Technology (Shandong Academy of Sciences), Qingdao, 266061, China

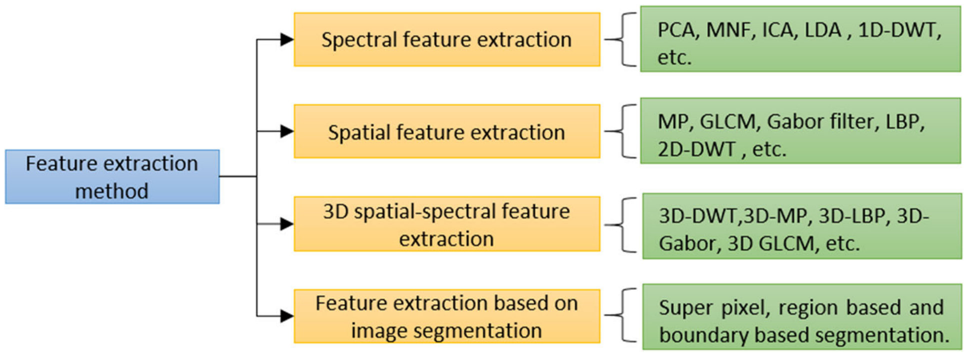

Figure 1. Categorization of the feature extraction methods.

be regarded as a cube containing two-dimensional spatial information and the third dimensional is spectral information. These characteristics of HSI improve the accuracy of the application in the qualitative and quantitative analysis of surface features and targets. At the same time, there are some challenges in the application of HSI. Firstly, hyperspectral data is nonlinear, and due to the interference of a variety of uncontrollable external factors, including environment, atmosphere, illumination, data transmission, etc., and spectral information will be lost or affected by noise.[4] Secondly, hyperspectral data contains a lot of data and the correlation between adjacent bands is strong, it is easy to cause information redundancy and overlap. Thirdly, it is difficult to obtain the available training samples, which will lead to the ill posed problem of some methods, reduce the generalization ability of the classifier, and limit the wide application of hyperspectral.[5]

HSI contains rich spatial and spectral information, which brings further development of land cover classification. However, the nonlinearity of HSI data and the strong correlation between bands also bring difficulties and challenges to HSI application in object recognition. In particular, the limited available hyperspectral training samples will lead to the classification accuracy cannot be improved. Therefore, making full use of the advantages of HSI data, through algorithms and strategies to solve the limited training samples and high-dimensional HSI data, so as to improve the classification accuracy, it is the research hotspot of current researchers.

2. Feature extraction methods

The feature extraction methods in HSI classification mainly include spectral feature extraction, spatial feature extraction, 3D spatial-spectral feature extraction and feature extraction based on image segmentation. The spectral feature extraction method is based on the spectral data of each pixel and extracts the effective data according to the correlation rules. Spatial feature extraction is mainly aimed at the spatial features of HSI, which is often combined with spectral feature extraction. 3D spatial-spectral feature extraction is to extract the three-dimensional data of HSI directly, and retain the information of the original data to the maximum extent. According to the image segmentation theory, HSI is segmented according to the similarity between pixels, and then classified. The categorization of the feature extraction methods is shown in Figure

Table 1. The advantages and disadvantages of feature extraction methods.

Feature extraction method	Advantages	Disadvantages
Spectral feature extraction	Model determination, simple algorithm, easy to understand and process.	It ignores the spatial distribution information of HSI, and cannot reveal the internal structure of the data.
Spatial feature extraction	Combined the spatial and spectral information, the classification accuracy can be improved to a certain extent.	Due to the limitation of dimension and training samples, the obtained long feature vector contains redundant information, which leads to over fitting.
3D spatial-spectral feature extraction	It cannot change the original shape, and retain the original spectral-spatial correlation to the greatest extent.	There are many data, large amount of calculation, long running time and easy to introduce interference information.
Feature extraction based on image segmentation	The probability that the adjacent pixels of HSI are of the same kind is very high, which is conducive to feature extraction.	The coefficient selection between spatial data and spectral data needs to be tested and this method is vulnerable to noise.

1, and the advantages and disadvantages of feature extraction methods is shown in Table 1.

2.1. Spectral feature extraction

In HSI, each pixel contains hundreds of spectral bands. Compared with the traditional imaging system, rich spectral information is helpful to better identify the surface features and objects. However, due to the small spectral band spacing, these bands are highly correlated and contain redundant information. Some spectral bands can reveal important information in specific applications, while these bands may not provide useful information in other applications. At the same time, redundant bands and noise bands not only increase the amount of unnecessary calculation, but also affect the classification accuracy. Therefore, it is very important to extract the effective spectral feature for the full use of HSI and improve the classification accuracy.

The principle of spectral feature extraction algorithm is to project the data into a new feature space while retaining its recognition information. The new feature space is usually a low dimensional space, but sometimes it is the same or higher dimensional space. The feature extraction algorithm combines the original bands to generate new spectral features, and the new features retain most of the important information. Traditional spectral feature extraction methods include principal component analysis (PCA), independent component analysis (ICA), linear discriminant analysis (LDA), etc.

PCA is a very common method to extract the feature information of hyperspectral spectral information. It is a method to get the minimum root mean square error of data, and generates features that minimize the correlation between principal components. Through PCA algorithm, the useful information in the original hyperspectral data can be concentrated into as few feature images as possible, and the images of different bands are not correlated with each other. In order to reduce the amount of calculation, the segmented PCA is proposed which divides the spectral information into multi segment to analysis, and the local PCA is proposed which establishes the local relationship according to the surrounding information. However, when PCA is used, sometimes the image quality does not decline steadily with the decrease of principal component. In order to solve this problem, the minimum noise fraction (MNF) is proposed. MNF introduces signal variance and noise variance to extract feature information and separate noise effectively. Based on MNF, a segmented maximum noise fraction (SMNF) transform is proposed for HSI feature extraction.[6] PCA and MNF are easy to ignore some important information in characterizing fine substance due to abandoning the latter several principal components. ICA is to linearly decompose the original data into statistically independent potential variables to find a projection direction that can make the data most independent, which is conducive to the effective retention of small target and small category information.

PCA, MNF and ICA are all unsupervised feature extraction methods. Supervised feature extraction depends on the prior knowledge provided by labeled samples. LDA is a supervised feature extraction method widely used in classification. The essence of LDA is to project high-dimensional feature data into the best discriminant coordinate space, so as to reduce the complexity of feature calculation and extract effective classification feature. However, when the number of samples is far less than the feature dimension, the distance between samples increases, which leads to the failure of distance measurement. It makes the intra-class and inter-class dispersion matrix singular, and is unable to get the optimal projection direction. According to the shortcomings of LDA, local Fisher's discriminant analysis (LFDA) is proposed. LFDA has a good effect in feature extraction of different types of images.[7] However, for high-dimensional images, because the intra-class scattering matrix is usually singular, the effect of extracting feature information is often limited. Non-parametric Weighted Feature Extraction (NWFE) overcomes the shortcomings of LFDA, defines new intra-class and inter-class scatter matrices, and calculates the weighted average. NWFE is a successful HSI processing method, but it faces the problem of computing time

PCA, MNF, ICA, LDA and other algorithms are based on the principle of statistics. According to a certain optimization criterion of statistics, an optimal model is constructed, and the feature information is extracted by data linear method. The advantages of these methods are model determination, simple algorithm, easy to understand and process, but it ignores the spatial distribution information of hyperspectral data, and can not reveal the internal structure of the data. In addition, spectral feature extraction methods include one-dimensional discrete wavelet transform (1 D-DWT), deep learning method, etc.

2.2. Spatial feature extraction

Although the spectral information in HSI has a lot of material information, the analysis results only using the spectral feature of HSI need to be improved. When two targets samples have the same spectral information, they can be classified or recognized by shape and texture. Only using spectral information or spatial information can not achieve effective results. Spatial neighborhood information also provides important information for the classification. If the spatial and spectral information can be combined, the classification accuracy can be improved to a certain extent.[8] The commonly used spatial feature methods include morphological profiles (MP), gray level cooccurrence matrix (GLCM), Gabor filter, Local binary pattern (LBP), two-dimensional wavelet transform (2D-DWT), etc.

Mathematical morphology (MP) are constructed by repeatedly using open and close operations on the image and increasing the size of structural elements (SE). It uses morphological operations to generate spatial structure features. MP can well characterize the multi-scale variability, but it has some shortcomings for simulating other geometric attributes of SE, so attribute filters (AF) are proposed.[9] AF operates on pixels based on different attributes such as area, standard deviation, volume, etc. Attribute profiles (AP) are constructed with AF. Due to successful performance of MP, it produced an improved version. MP and AP are created for each image after dimensionality reduction to obtain extended morphological profiles (EMP), and extended attribute profiles (EAP).[10] If the EAP of multiple attributes is superimposed together, it will form the extended multi attribute profile (EMAP). In order to solve the difficulty of EMAP in selecting filter parameters for constructing contour, all filter parameters sampled in a very small interval are used, that is, entire extended multi attribute profile (EEMAP).[11] But it has its own limitations in very large dimension and high redundancy. The methods of MP combined with spectral information to reduce the redundant information of images often use PCA, and PCA can maximize the retention of useful information.[11]

GLCM is a method to calculate the image texture characteristics. GLCM includes variance, median, energy, homogeneity, correlation, contrast, extraction, difference, contrast, correlation, second-order distance and other texture feature parameters. In HSI feature extraction, GLCM combined with spectral information is a relatively simple feature extraction method combining spatial and spectral information.

Gabor filteris a technique of texture analysis, especially suitable for texture representation and discrimination. Gabor transform is evolved from Fourier transform. When local information of signal is extracted by Fourier transform, Gabor transform is obtained by introducing window function of time localization. The frequency features of Fourier transform in different positions are often mixed together, but Gabor filter can extract spatial local frequency features. The two-dimensional Gabor filter can generate different scale filters according to different spatial frequencies, and has certain selectivity for spatial position and direction. Gabor filter combined with spectral feature extraction is a common method for spatial-spectral feature extraction of HSI.

Local binary pattern (LBP) is an operator used to describe the local texture features of an image, which reflects the relationship between each pixel and its surrounding pixels. The principle of the algorithm is to compare the gray value of a pixel in the image with that of the neighboring pixel. The conventional LBP method generates LBP code image for each band in the input HSI. In order to describe the spatial characteristics of pixels, the LBP histogram of each interest pixel and its corresponding neighborhood are calculated. Li, et al. noted that the texture features of HSI were extracted by LBP and fused with spectral spatial features. The HSI were classified by kernel extreme learning machine.[12]

Gabor, MP, GLCM and LBP provide multiple features of HSI, including directivity, shape and size, which are complementary. In addition, many types of wavelet transform is also a method of spatial information extraction, for such 2D-DWT. Wavelet can separate high frequency and low frequency information, and retain the energy and spatial geometric information at different scales. Other spatial information is often stacked with spectral information. Spectral information is extracted first, and then spatial information is extracted, so as to obtain better classification results. But on the other hand, due to the limitation of dimension and training samples, the obtained long feature vector contains redundant information, which also leads to over fitting.

2.3. 3D spatial-spectral feature extraction

Spectral and spatial information extraction separately ignores the spectral spatial correlation. HSI is a three-dimensional cube, using three-dimensional spatial-spectral information extraction can not change the original shape, and retain the original spectral-spatial correlation to the greatest extent.

On the basis of 2D-DWT, 3D-DWT can analyze the horizontal, vertical and spectral values of HSI at the same time, and realize the extraction of spatial-spectral feature of HSI. Ghasemzadeh et al.[13] proposed that Spatial-spectral feature were extracted from HIS based on 3D-DWT, the classification results were better than raw HIS data. Other wavelet methods include 3D scattering wavelet transform uses cascaded wavelet decomposition, complex modulus and local weighted average to filter HSI cube data. Scattering features can capture spectral-spatial information for classification. Its advantage is that the cascade of wavelet decomposition in multiple directions and scales provides a complex structure description for HSI classification. The local weighted average method is used to reduce the variability of features and makes the pixel labels in the neighborhood have local consistency. 3D-shearlet transform made full use of spatial information in the process of preprocessing and classification, highlighted the inherent characteristics of HSI·[14]

The traditional morphological filter is two-dimensional and ignores the spectral-spatial correlation in the three-dimensional structure of HSI. Compared the traditional 2D-MP, 3D morphological profiles (3D-MP) extract three-dimensional features of HSI directly, which have better classification accuracy. He et al.[15] proposed that the joint sparse low rank multi task learning method based on 3D-MP was used to classify HIS. Compared with 2D-Gabor filter, which only contains texture information of image, 3D-Gabor filter can capture specific scale, direction and wavelength correlation characteristics. On the basis of extracting texture information, it can better save spectral information, and keep the close correlation between spectral and spatial information to the greatest extent. Jia et al.[16] proposed that 3D-Gabor phase coding was introduced, which was used in HSI classification together with matching based on Hamming distance. It used Gabor phase feature instead of amplitude feature to overcome the disadvantage of large Gabor feature. Although two-dimensional LBP can be used for local feature extraction of HSI, HSI had three-dimensional features, and its spatial features could not be fully extracted. Jia et al.[17] proposed that 3D-LBP was proposed, which used three-dimensional regular octahedron frame to represent the spectral and spatial relationship instead of two-dimensional pixel, which fully mined the context information hidden in the spectral and spatial structure. Since HSI was a cube data, GLCM was extended to three-dimensional form to classify HSI.[18] It was found that the texture calculated by 3D-GLCM had better classification effect than 2D-GLCM. Zhu et al.[19] proposed that three kinds of 3D feature extraction methods, 3D-MP, 3D-LBP and 3D-Gabor, were used to extract HSI feature from morphology, local dependence and shape smoothness respectively. They were fused in the framework of multi task sparse representation, and multiple 3D features were fully used for classification.

2.4. Feature extraction based on image segmentation

The purpose of image segmentation is to extract the object of interest from the image according to different regions. Traditional image segmentation algorithms include region based segmentation algorithm and boundary based segmentation algorithm. The former is based on the similarity of regional features, while the latter is based on some similarity search criteria to find the edge points. The HSI classification methods based on image segmentation are mainly improved based on these two methods. In unsupervised image segmentation, clustering algorithm is widely used. It is a spectral domain segmentation method, which is according to the spectral information of each pixel to achieve HSI classification.

At present, the most commonly used image segmentation method in HSI is super pixel (SP). SP has the similar region block composed of adjacent pixels of color, brightness, texture and other characteristics. In hyperspectral remote sensing image, the probability of adjacent pixels being of the same kind is very high, so the image is segmented into sub regions with homogeneity as much as possible, which is widely used in feature extraction of spectral and spatial information. SP algorithm can use spatial information to improve the classification effect, which obtains redundant information, reduces the complexity of HSI processing, and improves the classification efficiency. SP algorithm can be divided into based on graph theory algorithm and based on gradient descent algorithm. The based on graph theory algorithm is that each pixel is regarded as a vertex of the graph, and the similarity between the vertices is regarded as the weight of the edge of the graph. According to the objective function, the image is divided into several subgraphs, and each subgraph is a super pixel. The based on graph theory algorithm includes graph-based method, normalized cuts method (Ncut), superpixel lattice method, based entropy rate method. Graph-based realizes clustering based on minimum spanning tree. Ncut method realizes segmentation by constructing optimization function on image texture and contour. Superpixel lattice method finds the optimal path to realize segmentation by adding the constraint of image topology information. The objective cost function of based entropy rate method adopts balance term function and entropy rate function of random walk. Based gradient descent method includes watersheds method, MeanShift method, Turbopixels method and simple linear iterative clustering (SLIC) method. The disadvantage of watersheds method is easy to over segment. The MeanShift method is based on the density function and classifies the pixels with similar density into a super pixel. The Turbopixels method uses the idea of grid to divide. SLIC method realizes SP segmentation through simple iterative clustering, which becomes the most popular super-pixel segmentation method because of its simple algorithm and fast speed.

SP algorithm is introduced into HSI, which reduces the complexity of hyperspectral data and improves the classification effect. Jia et al.[20] proposed that the original HSI was segmented into disjoint super pixels by using the classic SP segmentation method, and then classified. Compared with single-scale SP segmentation, multi-scale SP segmentation can get different structure information at different scales.[21] Large-scale segmentation can obtain large super pixels and provide more powerful and recognizable features, while the super pixels generated by small-scale segmentation have high consistency in intensity and texture features. Therefore, multi-scale super-pixel can solve the problem that single scale SP is difficult to adapt to different object sizes in HSI.[22] Jin et al.[23] proposed that two different scale SP segmentation images were established, that is, a thinner SP segmentation image had higher segmentation accuracy, and a thicker SP segmentation image retained the objectivity of the original image. Li et al.[24] proposed a multi-scale SP fusion method for HSI classification, which avoided the problem of defining the optimal SP scale and combined all SP scales to guide HSI classification. Multi-scale SP are combined with other algorithms to obtain a variety of multi-scale SP extension algorithms. Zhan et al.[22] proposed that PCA was used to reduce the spectral dimension, then different number of super pixels were set, and the entropy rate segmentation method was used for multi-scale SP segmentation. Cui et al.[25] proposed that a multi-scale image segmentation algorithm was used to generate super pixels, and a weighted image was established with super pixels as nodes. SVM classifier was used to get the pixel level probability map, which was used as the prior probability distribution of the SP, so as to measure the probability that the super-pixel belonged to a certain class. Jia et al.[26] proposed a super pixel based weighted linear programming method for HSI classification. The HSI was over segmented by using the entropy rate segmentation method, and the similarity matrix based on SP was established. Lu et al.[27] proposed a fusion framework of HSI classification based on complementary information of sub-pixel, pixel and SP, which combined multiple features into composite core for subsequent classification. Liu et al.[28] proposed a new method of multi-form SP (MMSP) to extract spectral and spatial features. Morphological features were extracted from the original HSI, and each morphological feature was segmented by SLIC. The MMSP belonging to the same class were merged and extracted by using consistency constraint and mean filter, and then integrated into SVM classification. Multi-scale SP algorithm can effectively extract HSI features, which lays a foundation for the correct classification of HIS.

3. Classification methods

HSI are classified after feature extraction. Classification algorithms are mainly divided into classical classification algorithm, kernel-based classification algorithm, representation-based classification algorithm, decision fusion based classification algorithm and deep learning classification algorithm. The categorization of the classification method is shown in Figure 2, and the advantages and disadvantages of classification methods is shown in Table 2.

3.1. Classical classification methods

The classical classification methods based on spectral features include spectral angle mapping (SAM) and spectral information divergence (SID). These two methods are

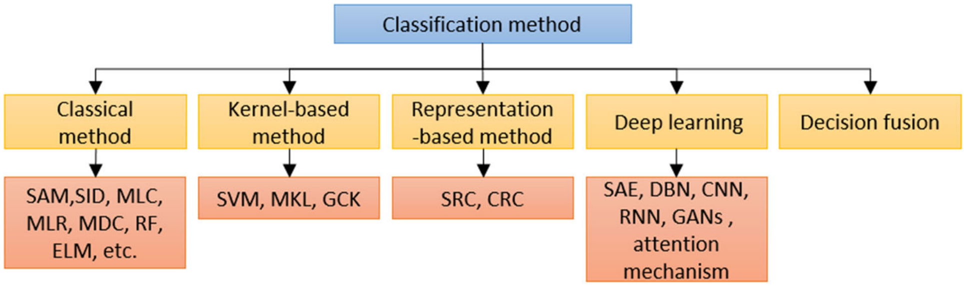

Figure 2. Categorization of the classification methods.

Table 2. The advantages and disadvantages of classification methods.

Classification methods	Advantages	Disadvantages
Classical method	The algorithm is simple and easy to implement	The classification accuracy of a large amount of data is not very high.
Kernel-based method	It can calculate the inner product of high-dimensional space without knowing the mapping function, and solve the nonlinear problem by linear algorithm.	It is easy to over fit, takes a long time to calculate, and needs to select more parameters.
Representation-based method	It is a nonparametric method without any knowledge of data distribution, which not need enough training samples.	When the sample dimension increases, the calculation cost increases.
Deep learning	It can extract not only simple feature information, but also more complex feature information. The learning process is fully automated, and different networks can extract different feature types.	The network parameters of deep learning are difficult to determine, and the calculation time of the model is long,
Decision fusion	Due to the limitations of a single classifier, using multiple classifiers for decision fusion can also improve the classification accuracy to a certain extent.	Multiple classification models need to be calculated, and the amount of calculation is large

based on the similarity to compare the spectral information of the test sample with that of the standard sample to find the closest sample and realize the classification. The SAM method is based on the mapping projection to calculate the similarity, while the SID is based on the information measure to calculate the similarity between the test sample and the standard sample. Since HSI is easy to be introduced interference, only relying on the classification method based on spectral features has great uncertainty in HSI classification.

Classical classification methods based on both spectral and spatial features, include maximum likelihood classification (MLC), which determines the attribution of pixels to be classified according to statistical methods. Multinomial logistic regression (MLR) uses dependent variables based on multiple independent variables to predict category or membership probability, and uses maximum likelihood estimation to evaluate. Minimum distance classification (MDC) is a classification method, in which the distance between the pixel to be classified and the center vector of each class is calculated, and then the vector to be classified belongs to the class with the smallest distance. Decision tree classification (DTC) uses inductive algorithm to generate readable rules and decision tree to process data, and then uses decision tree to analyze new data. Random forest (RF) is a classifier with multiple decision trees. Extreme learning machine (ELM) can initialize the input weight and bias randomly, and get the corresponding output weight, which is fast and accurate.[29]

3.2. Kernel-based classification method

Kernel method is to map the original feature space (usually low dimensional space) to an available feature space (usually high dimensional space) by using nonlinear mapping function, so as to transform the data that can't be linearly segmented into data that can be linearly segmented. Kernel method is an effective method to solve nonlinear classification problem. Its advantage is that it can calculate the inner product of high-dimensional space by kernel function without knowing the mapping function, and solve the nonlinear problem by linear algorithm.

The most representative kernel method is support vector machine (SVM). It makes the distance between the two classes as large as possible by finding the optimal hyperplane in high-dimensional space. SVM can process high-dimensional data with limited training samples and get better classification results, so it is widely used in HSI classification. Cheng et al.[30] proposed that HSI data was transformed into a new spectral space feature through CNN and metric learning, and was trained with SVM to get classification results. SVM also has some improved algorithms, such as bootstrapped SVM. The principle of bootstrapped SVM is to remove the training samples with wrong classification from the training set, redistribute the correct labels, and reintroduce the training set in the subsequent training cycle.[31] Rotating support vector machine produces different SVM classification results through random feature selection and data transformation, which can improve the accuracy and diversity of individuals in the set at the same time.[32] In order to alleviate the problem of information redundancy in HSI space, support tensor machines is proposed, which processes HSI into a data cube, and then recognizes information in tensor space.[33]

SVM uses a single kernel function, which has some limitations. When the sample size is too large and the multidimensional data is not normalized, the single kernel function can not fully express the complex feature space. Multi kernel learning (MKL) method is based on SVM, which combines multiple kernel functions and has higher flexibility in exploring HSI information. MKL is a fusion method, which generates a composite kernel through the linear or nonlinear combination of basic kernels, and searches for the corresponding optimal kernel through multiple kernel functions and different feature spaces. So that the data can be more accurate expression, and the accuracy of classification can be improved. It not only inherits the advantages of SVM in solving small sample nonlinear problems, but also has better classification performance in dealing with complex distributed data. For high-dimensional data of HSI, some improved MKL algorithms are proposed, including kernel scale nonlinear MKL (NMKL) and kernel alignment MKL (KA-MKL). The former is to improve the accuracy by optimizing kernel combination and related classifiers, while the latter is to optimize by reducing kernel scale. Gu et al.[34] proposed multiple structure element nonlinear MKL. NMKL was introduced to learn the optimal combination kernel from the predefined linear basis kernel, which made full use of the similar information generated by the nonlinear interaction between different kernels, and improved the accuracy. Wang et al.[35] proposed that a discriminant MKL (DMKL) method. By maximizing the separability of reproducing kernel Hilbert space, DMKL learned the optimal combination kernel from the predefined basic kernel.

The convex combination of kernel based on SVM and MKL has some limitations in practical application. In order to solve this problem, a generalized composite kernel is proposed. Generalized composite kernel (GCK) introduces composite kernel function, which combines spectral and spatial information. It has great flexibility in the contribution of spectral and spatial information, and is not limited by convexity. Li et al.[36] proposed that GCK and MLR were used to classify HSI to get better classification results. The kernel based classification method maps the low dimensional space to the high dimensional space, which can solve the nonlinear classification problem well. The classification accuracy is high and the running speed of the model is fast, so it is widely used in HSI classification.

3.3. Representation-based classification method

Representation-based classifier is a nonparametric method without any knowledge of data distribution. Therefore, it has some advantages when there are not enough training samples. The principle of this method is that the test samples are approximated linearly by the trained dictionary. In HSI classification, sparse representation classification (SRC) and collaborative representation classification (CRC) are the most widely used as the representation-based classification method.

In SRC method, test samples are sparse approximated by a small number of dictionary atoms through L1 minimization problem, and a large number of original data can be represented by linear combination of a small number of sample. In the classical SRC, the test sample is represented as a sparse linear combination of all training samples, and the class with the least reconstruction error is regarded as the test sample classification. In HSI, adjacent pixels are usually highly correlated, while traditional SRC only considers spectral information. Therefore, joint sparse representation classification (JSRC) is proposed. On the basis of SRC, the neighborhood pixel centered on the test pixel is selected to combine the spatial information. Based on JSRC, nonlocal weighted joint sparse representation classification (NLW-JSRC) algorithm is proposed.[37] In the classification process of the center test pixel, the weights of the neighboring pixels are added. Kernel sparse representation based classification (kernel-SRC) is a nonlinear extension of SRC.[38] By mapping nonlinear separable samples to high-dimensional feature space, it shows good performance in HSI classification. Sun et al.[39] proposed that a sparse representation framework based on super-pixel features. HSI was divided into different spatial regions, and each region was suitable for shape and size, which was considered as a super-pixel. Combined with spatial information, SRC was used to classify. In order to solve the problem that SRC is not robust to outliers in practical application, robust SRC (RSRC) is proposed to deal with this problems.[40] In practical applications, HSI are often destroyed by different types of noise. Huang et al.[41] proposed that the mixed noise model was combined with the prior knowledge of input data representation coefficients. Three methods of SRC, JSRC and super pixel joint SRC were proposed as the robust classification methods.

CRC method is derived from SRC, and the difference is that test samples use L2 minimization problem, which is represented by all atoms in dictionary. SRC means that each atom has a choice to participate in the representation process of a given pixel, while CRC means that each atom has an equal opportunity to participate in the representation process. According to the characteristics of spatial-spectral information in HSI, a HSI classification method based on probabilistic cooperative representation was proposed.[42] The performance of HSI was evaluated on different types of spatial features, which reduced the amount of computation while maintaining the classification accuracy. Jia et al.[43] proposed a multi-scale super pixel fusion method based on cooperative representation was proposed for HSI classification. 3D Gabor and EMAP features were convoluted firstly, and EMAP-Gabor features were extracted by CRC, then multi-scale super-pixel mapping was generated by EMAP features, and the classification mapping obtained by CRC was regularized. Du et al.[44] proposed that a new classifier based on multi-core cooperative representation. The MKL model was embedded in CRC, which was better than other based-representation classifier.

Due to the different characteristics of SRC and CRC, they are conducive to the good characteristics of HSI in different classification scenarios. Li et al.[45] proposed a hyperspectral image classification method based on SRC and CRC. It solved the problem that SRC might select too few samples to reflect the changes of within-class, while CRC might contain interference of inter class atoms. Combining the advantages of CRC and SRC, the classification accuracy was improved. Compared with traditional SRC and CRC, kernel based SRC and kernel based CRC can significantly improve the classification accuracy of complex data. The combination of kernel based SRC and CRC can realize the dual advantages of sparse representation and cooperative representation in kernel space.

3.4. Classification method based on decision fusion

When only using a single feature set or data set to classify HSI, some important information and details will be ignored. In order to improve the classification accuracy, multiple feature subsets can be extracted and fused to get the global decision results. Due to the limitations of a single classifier, using multiple classifiers for decision fusion can also improve the classification accuracy to a certain extent. The most commonly used decision fusion rule is the majority voting (MV) rule. Classification methods based on decision fusion are mainly divided into feature level fusion, data level fusion and classification level fusion.

Feature level fusion is to extract or select different feature sets, then fuze and assign a classifier. Different feature extraction methods are used to extract different feature subsets from HSI. Each subset is assigned to the same classifier, and the final classification result is obtained by decision fusion. Imani et al.[46] proposed a method of spectral-spatial feature extraction based on morphological-based feature space discriminant analysis

(MBFSDA). After PCA transformation, each principal component generated MP, which contained spatial features such as the size and shape information of context structure, and used MV rule to provide final classification map. Priya et al.[47] proposed that the HSI was segmented into SP. Each SP was classified independently by using statistical Bayesian classification method, and the decision was combined to get the unique class label of each SP, and then classified.

In data level fusion, different data values are assigned to the same classifier. Guo et al.[48] proposed that considering the global and local features of HSI, the global feature was the hyperspectral reflection curve, and the local feature was the absorption feature. The local feature set was used to modify the result of multi label classification. Since the reflection feature had more accurate classification results than the absorption feature, the reflection feature and SVM classifier were used for classification. If the result was satisfactory, it would be reported as the final result. Otherwise, the classification result obtained by absorbing features by using multi label classification method would be the final result. Guo et al.[49] proposed that different algorithms were used to process the reflection feature and absorption feature of HSI, and the decision of reflection feature was modified by the result of absorption feature through the fusion of decision layer algorithm.

Classification level fusion uses the same feature set and gives different classifiers. HSI uses the same feature extraction method to obtain a feature subset, and then different classifiers are used for decision fusion. Bo et al.[50] proposed that a new HSI classification method and decision fusion based on joint collaborative representation (JCR) and SVM model. The classification results were obtained by combining JCR model and SVM model with multiplicative fusion rule. Ye et al.[51] proposed that the windowed 3D discrete wavelet transform was combined with the correlation matrix of characteristic group wavelet coefficients, and the multi classifier decision fusion method of the maximum likelihood estimation, Gaussian mixture model and SVM were used for final classification.

There are also some researches that combine several fusion methods. Li et al.[52] proposed that the LBP features extracted from HSI, global Gabor features and original spectral features were fused at feature level fusion and classification level fusion. Feature level fusion was to concatenate multiple features before pattern classification, while classification level fusion was to fuze the probability output of each classification, and used soft decision fusion rules to fuze the results of classifier integration. Liu et al.[53] proposed a new framework of joint classification fusion and feature fusion, which combined the classification results obtained from multi-scale features of different scales of weighted extended multi-attribute profiles and extreme learning machine into the final classification results for HSI classification.

3.5. Deep learning classification algorithm

Deep learning has made great progress in image classification and target detection. In recent years, some achievements have been made by introducing deep learning into HSI. Compared with traditional methods, deep learning extracts useful feature information from hyperspectral raw data through multiple hierarchies. Deep learning can extract not only simple feature information, such as texture and edge information, but also more complex feature information. Another advantages of deep learning are that the learning process is fully automated, and different networks can extract different feature types.

Deep learning networks include stacked auto encoders (SAE), deep belief networks (DBN), convolutional neural networks (CNN), recurrent neural networks (RNN) and generative adversarial networks (GAN). SAE has more powerful expression ability and all the advantages of deep network, which tends to learn the characteristic representation of data. DBN is based on Bayesian idea, the high-level information hidden in the data is automatically obtained by finding out the joint probability distribution of the data. The advantages of DBN is to extract the deep features of the training data through learning. The feature of CNN is that the original signal is directly used as the input of the network, which avoids the complex process of feature extraction and image reconstruction in traditional recognition algorithm. RNN can be used to identify data with sequence property. By using a recursive hidden layer to identify the patterns and dynamic time characteristics in the data sequence. When traditional RNN processes long-term sequence data, its performance will be degraded due to gradient disappearance or gradient explosion. So long short-term memory (LSTM) is proposed, which is a variant of RNN. RNN saves all information, while LSTM allows information to be retained or forgotten, which is more suitable for classification and prediction. GAN includes generation model G and discriminant model D. Model G and model D are trained in a confrontational way. Model G generates false samples as real as possible, while model D tries to distinguish real samples from false samples. Through the confrontation and competition between the two modes, the training process will continue to be effective until they reach a balanced and harmonious state.

Traditional spectral feature extraction methods have limitations in dealing with complex hyperspectral features, and can not extract effective information well. The spectral feature extraction method based on deep learning can extract the deep HSI features, which is more conducive to improve the accuracy of HSI classification. Hu et al.[54] proposed that the deep CNN was used to classify HSI directly in the spectral domain, which had better classification performance than the traditional SVM. One dimensional CNN was also used to extract HSI features for classification.[55–56] According to the frequency band selection of different application scenarios, two different versions of BSNet-FC and BSNet-Conv were adopted, which adopted fully connected network and convolution network, respectively.[57] It could accurately select the information band set with less redundancy, and had great advantages in classification accuracy and time.

HSI includes not only spectral information, but also spatial information. Compared with the traditional feature extraction methods, the spatial-spectral feature extraction method based on deep learning can automatically extract depth features from complex HSI data. Using deep learning to extract features from spatial and spectral features can improve the accuracy of classification. Spatial-spectral features based on deep learning are widely used in HSI classification. CNN is a commonly used deep learning method, which combines spectral features with spatial features. 2D- CNN and several 2D- CNN combined algorithms have some applications in HSI classification. Zhang et al.[58] proposed a dual channel CNN framework. One dimensional CNN was used to extract hierarchical spectral features automatically, and two dimensional CNN was used to extract hierarchical spatial correlation features. Softmax regression classifier combined spectral features with spatial features to predict the classification results. Li et al.[59] proposed that two 2D-CNN networks were used to extract spectral features, local spatial features and global spatial features at the same time, and the most informative was identified by using channel correlation. Compared with 2D-CNN, 3D-CNN can extract more feature information from HIS. Li et al.[60] proposed that the CNN was a three-dimensional network using both spectral and spatial information, the network had high classification accuracy and efficiency. Only 2D-CNN can not extract the feature map with good discrimination ability from the spectral dimension, and the calculation of 3D-CNN is often more complex. In order to solve this problem, a hybrid spectral CNN for HSI classification was proposed.[61] The combination of 3D-CNN layer and 2D-CNN layer made full use of spectral and spatial feature map to achieve maximum accuracy. Since CNN has a good classification effect in HSI classification, some improved methods based on CNN are developed. Zhang et al.[62] proposed that CNN based on different regions was proposed, which represented spatial-spectral context information to obtain effective HSI classification features.

In addition to CNN, other deep learning networks also show good classification accuracy in HSI spatial-spectral feature extraction. Song et al.[63] proposed a deep feature fusion network for HSI classification. Residual learning was introduced to optimize multiple convolution layers, which simplified the training of deep network and improved the classification accuracy. Pan et al.[64] proposed a spectral and spatial classification model of HSI based on single gate recursive unit (GRU). Spectral and spatial features were calculated and expanded in one GRU. Sun et al.[65] proposed a deep extraction of localized spectral features and multi-scale spatial features convolution was proposed. Considering the correlation among spectral bands, the local spectral information of grouped sub cubes was mined by fuzing spectral and spatial information, and the multi-scale spatial feature extraction strategy was realized. Most of the existing HSI classification methods extract spatial-spectral feature by combining pixels in small neighborhood or aggregating statistical features and morphological features.

With the great success of deep learning in HIS classification, attention mechanism is introduced to further improve the application of deep learning in HSI classification. Inspired by human visual attention process, the attention mechanism is designed to focus more on the informative areas and takes less account of non-essential areas. In HSI classification, attention mechanism is mainly divided into channel attention, spatial attention, hybrid attention and self-attention. Channel attention aims to show the correlation between different channels. The importance of each feature channel is automatically obtained through network learning, and then different weight coefficients are given to each channel to strengthen the important features and suppress the non-important features. Qing et al.[66] proposed a multi-scale residual convolutional neural network model fused with an efficient channel attention network. The efficient channel attention network reduced the model complexity, manages output feature channels with different weights, and achieved the extraction of important features within the image. Spatial attention aims to improve the feature expression of key regions. It transforms the spatial information in the original image into another space and retains the key information through the spatial conversion module. It generates a mask for each position and outputs it with weight, so as to enhance the specific target regions of interest and weaken the irrelevant background regions. Because HSI contains spectral information, spectral attention is often combined with spatial attention, and spectral attention recalibrates the importance of different spectral bands. Huang et al.[67] proposed an attention aided CNN model, which included a spectral attention sub-network and a spatial attention sub-network for spectral and spatial HIS classification, respectively. Spatial attention ignores the information interaction between channels because it processes the features in each channel equally. While channel attention is to process the information in a channel directly, which is easy to ignore the information interaction in space. Hybrid attention combines the advantages of the two and forms a more comprehensive feature attention method. Qu et al.[68] proposed that a channel-spectral-spatial-attention module to optimize the information transmission between different subnetworks. The attention mechanism filters the feature mapped of any subnetwork to obtain stronger spectral–spatial information and more important feature channels as input for the succeeding subnetwork. Self-attention is a variant of attention mechanism. Its purpose is to reduce the dependence on external information and make full use of the inherent information of features for attention interaction. Qing et al.[55] proposed that the spectral attention and the self-attention mechanism was used to extract the spectral–spatial features of the HSI, respectively. Several multiple multi-head self-attention modules was employed to extract the image features and a residual network structure was constructed, which was to solve the gradient dispersion and over-fitting problems

The advantage of deep learning in HSI classification is that it can fully extract the deep information of HSI, which is more conducive to the analysis of the inherent characteristics of HSI and improve the accuracy of classification. However, due to the difficulty in determining the network parameters of deep learning and the long calculation time of the model, deep learning has some limitations in HIS classification.

4. Classification methods based on small sample strategy

In HSI classification, the most effective classification method is supervised classification, which needs a large number of labeled samples. In practical application, only a small number of labeled samples are available, and the acquisition of labeled samples needs to be based on experience or field investigation, which is time-consuming, laborious and expensive. Deep learning and other classification methods need a large number of samples for training and learning in order to get a good classification model. A small number of samples for direct training may lead to poor classification performance. Therefore, under the condition of limited samples, it is of great significance to use the small sample strategy for HSI classification. The categorization of the methods based on the small sample is shown in Figure 3, and the advantages and disadvantages of classification methods based on small sample strategy is shown in Table 3.

4.1. Data augmentation

Data augmentation is an effective method to solve the problem of the small samples, which is by creating new training samples from the given known samples. The main

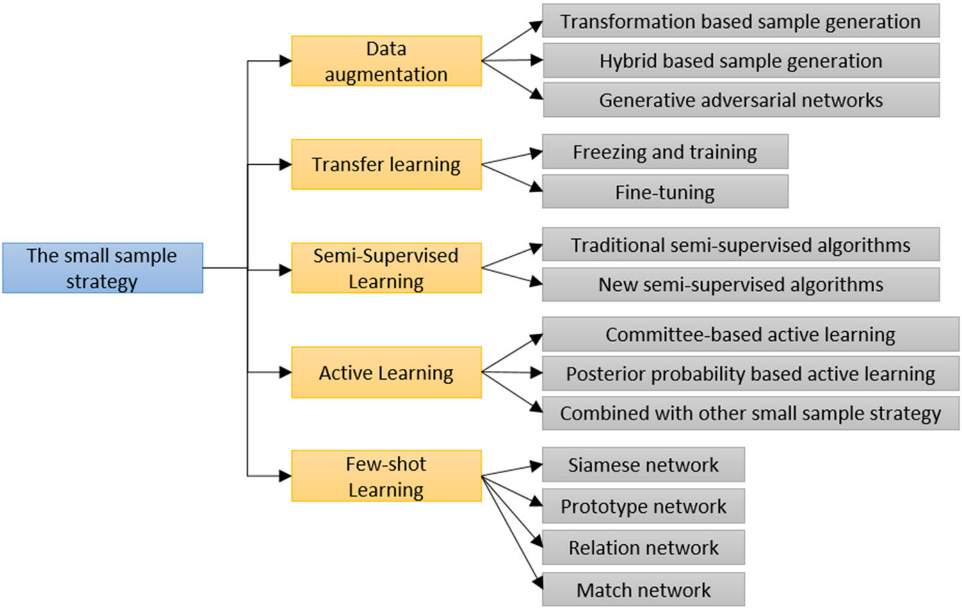

Figure 3. Categorization of the methods based on the small sample.

Table 3. The advantages and disadvantages of classification methods based on small sample strategy.

Small sample strategy	Advantages	Disadvantages
Data augmentation	It can create a large number of new training samples.	It is easy to produce invalid samples, affect the classification effect and increase the calculation time.
Transfer learning	It can significantly reduce the demand for training samples.	It may lead to negative migration, prone to over adaptation, and the data distribution of the source domain and the target domain is often different.
Semi-Supervised Learning	It uses a large number of unlabeled data combined with a small number of labeled samples to improve learning performance without manual intervention and external interaction.	The model cannot correct its own mistakes.
Active Learning	It selects high-value samples for manual and accurate labeling, and will not introduce wrong class labels.	Sometimes human-computer interaction is too frequent, and sometimes synthetic samples have no meaning and cannot be labeled.
Few-shot Learning	It makes the model learn to learn. The model learned through this learning mechanism can better classify. new and unprecedented meta tasks	High dimensional parameter space training needs to be improved, which may lead to additional space and computational cost.

strategies of generating new samples are transformation based sample generation, hybrid based sample generation and generative adversarial networks.

The transformation based samples generation transform the current known samples to generate virtual samples, and adopt data augmentation strategies such as translation, clipping, flipping, rotation and noise to increase the number and diversity of samples.

On the basis of the strategies of generating new samples, combined with other spatial-spectral feature extraction algorithms, HSI classification of small sample is realized. Attribute profile can obtain the geometric and spectral features of HSI. Aptoula et al.[69] proposed that the image after attribute filtering was superimposed as the input of CNN to realize data augmentation, which improved the classification performance. Accion et al.[70] proposed a dual window super-pixel data augmentation framework based on the combination of super-pixel segmentation and geometric transformation, which was called dual window super-pixel (DWS). Four kinds of rotation and flip transformations were carried out on DWS framework, including inner rotation, double rotation, inner flip and double flip to realize data augmentation.

The hybrid based sample generation is to generate new samples from two given samples of the same class according to the similar spectral characteristics of the same class in a certain range. Wang et al.[71] proposed that a data mixing model was established to expand the labeled samples twice, and the DCNN classifier was used to train the samples. Through random sampling the coefficients in the data mixing model, multiple independent classifiers were obtained. The voting strategy was used for fusion to obtain the final HSI classification results. Li et al.[72] proposed that pixel pair features (PPF) method was used to improve the number of training samples, and CNN was used to realize HSI classification. Firstly, the training samples were paired with any two selected samples by a certain standard. In this standard, a pair of samples from the same class was marked as unchanged, while samples from different classes were marked as 0. On this basis, the data augmentation method of pixel block pair (PBP) was proposed.[73] PBP could significantly improve the disadvantages of the PPF, which ignored rich spatial information. PBP also used the spectral and spatial information in HSI to study the influence of training sample size on classification accuracy. A method of expanding multi label training samples was proposed.[74] By giving a small number of pixels a label (called a single label sample) to accurately label them, and then a large number of pixels in a specific region were annotated by using a single label sample to multiple labels (called multi label samples). Based on the method of super-pixel segmentation and recursive filtering, multi label training samples were made full use of to achieve HSI classification. Another data augmentation strategies was according to the clustering distribution characteristics of HSI, similar samples were selected from the number and similarity.[75] In the neighborhood of a specific sample, the most likely sample was selected first. Then, according to the similarity between the candidate sample and the current sample, the contribution weight of the sample to the clustering center was calculated to realize the data augmentation.

GAN is another data augmentation strategies in HSI classification, which can generate new samples to simulate the distribution of real data through the competition between generator and discriminator. On the basis of GAN, a symmetric convolution GAN based on cooperative learning and attention mechanism was proposed.[76] Wang et al.[77] proposed that a new discriminator was designed by exploiting capsule network and convolutional long short-term memory, which formed the high-level contextual features with local space sequence information and low-level features. Cooperative learning and competitive learning were used to generate high-quality samples. A clustering based conditional GAN was proposed,[78] which could improve the scale and quality of training samples. The strategy of subtractive clustering was used to automatically select the most representative initial samples, which maintained the diversity of sample generation. A self-attention generation adversary adaptive network was proposed,[79] which aimed to increase the number and quality of training samples and avoided the influence of over fitting. The method introduced domain adaptive term to make the generated samples closer to the original samples, and used the self-attention mechanism to capture the correlation among spectral bands. It could further improve the quality of the generated samples.

4.2. Transfer learning

Deep learning methods needs the massive parameters and lots of labeled samples to achieve the final classification results, which needs a long time to train the model. Transfer learning is a technology that introduces the useful information learned from the source data into the target data, which can significantly reduce the demand for training samples. It can improve training process by transferring well trained parameters from the source data to the target data. In HSI classification, transfer learning is used to set the initial value of network parameters copied from other training depth networks. Compared with the random initialization method, it has better classification performance. In the deep network, the shallow layer of the network mainly obtains the shallow layer features. It uses the relevant data set with enough tags to pre train the network, and directly transfers the network parameters of the lower layer and the middle layer to the new network with the same structure as the original network. The top layer of the network extracts deep features, and the subsequent classification is divided into two type. They are freezing and training, fine-tuning.

Freezing and training is to train the classifier by using the deep features extracted from the transfer network. It includes freezing all layers except the last layer and training the last one. It can also freeze the first few layers and fine tune the rest. The hyperspectral data processing method based on transfer learning and deep learning used data sets similar to the target data to pre train the deep learning network, and used transfer learning method to find out the common features of the source domain data and the target domain data. Yang et al.[80] proposed that the transfer learning combined with two branches of the deep convolution neural network, which respectively studied the characteristics of spectral and spatial domain. It pre-trained the bottom and middle layer of the network, and only the top lay was trained. The deep model was solved in the case of limited training samples. Jiang et al.[81] proposed a transfer learning method based on Bayesian framework combined with spectral and spatial information. The Markov characteristics of images were used to distinguish and separate the images with class labels, which could make further use of spatial information. Existing remote sensing image processing methods require all images to have the same dimension. ImageNet datasets had three channels, while HSI datasets had hundreds. Therefore, three channels of HSI were randomly selected to form transferring CNN, several transferring CNN were combined to build an ensemble classification system with diversity.[82] To solve the problem of having different dimension, a new Iterative reweighting heterogeneous transfer learning framework was proposed.[83] It iteratively learned the common space of source data and target data, and used a new iterative weighting strategy to weight the source samples. Since the sample is limited to one source domain, it can only partially alleviate the shortage of labeled samples. In order to further alleviate the problem of limited samples, a multi-source deep transfer learning framework was proposed, which consisted of a multi-source compatible model and a user-defined loss function.[84]

Fine-tuning is to fine tune the network by using a small number of training samples of target data, which called fine-tuning. It includes using the pre training network on the source data and training all layers in the target data. Although the classification effect of deep learning is better than that of shallow classifier, it needs a certain number of labeled samples to build the model. Li et al.[85] proposed a hyperspectral data processing method based on transfer learning and deep learning. The hyperspectral data sets similar to the target data were used to pre train the deep learning network, found the common features and established the classification model. Liu et al.[86] proposed that a deep convolutional recursive neural network for HSI classification was constructed by Siamese network, which was composed of two CNN. The whole network was adjusted by a small number of label samples combined with transfer learning. Transfer learning combined with other algorithms can further improve the classification effect in small sample. Xie et al.[87] proposed a spectral-spatial HSI classification scheme based on SP pooling convolutional neural network with transfer learning of fine tune, results proved that the proposed algorithm could effectively classify the HSI with limited training labels. An improved method based on 3D CNN was proposed,[88] which combined parameter optimization, transfer learning and virtual samples. Using CNN model structure and weights trained on well labeled datasets such as ImageNet datasets, the model and weights of HSI classification network were initialized, and the well-designed neural network was used to complete the task of HSI classification.[89–90] In order to eliminate the difference between the ImageNet datasets and HIS datasets, attention mechanism was used to reweight the feature map, which further improved the classification accuracy.[84] Based on the fine-tuning, two transfer learning strategies were proposed: cross-sensor strategy and cross-mode strategy.[91] In the cross-sensor strategy, a 3D model was pre trained in the source HSI data set with a large number of labeled samples, and then transferred to the target HSI data set. The cross-mode strategy was to pre train a 3D model in the target HSI dataset, which included two dimensional RGB image data set with a large number of samples. Then it was transferred to the target HSI data set, which had good performance in HSI classification. Jiang et al.[92] proposed that through collaborative 3-D separable ResNet and the transfer learning of cross-sensor, the pre trained model was converted to the target HSI data set for fine tuning, and the classification task was completed.

4.3. Semi-Supervised learning

Supervised classification and unsupervised classification are two traditional framework of pattern recognition classification model. Supervised classification uses the prior knowledge of experts or measured data to mark the unknown state in the data, which is called training sample. Unsupervised classification is to divide the input data into different categories according to the similarity of specific model comparison. The disadvantage of supervised classification is that the cost of label training is very high, resulting in a limited number of training samples. While the disadvantage of unsupervised classification is that the classification accuracy is low. So the semi-supervised classification is proposed.[93] The main purpose of semi-supervised feature learning is to extract useful features from a large number of unlabeled data. Hyperspectral labeled samples are difficult to obtain, and HSI classification research is devoted to designing robust and effective semi-supervised learning framework.

Traditional semi-supervised algorithms include generative model semi-supervised algorithm, self-training semi-supervised algorithm, collaborative training semi-supervised algorithm and graph based semi-supervised algorithm. The generative model estimates the model parameters according to the labeled and unlabeled samples, which has the advantage of simple implementation. Traditional generative model semi-supervised algorithm has the problem that the model can not accurately adapt to the basic distribution. The unbiased estimation generative model semi supervised is proposed to overcome the shortcomings of traditional methods and improve the convergence of the model.[94] Self-training semi-supervised framework is one of the most commonly used for semi-supervised learning. It uses a small number of labeled training tags to predict the categories of unlabeled data, so as to improve the final accuracy. In the framework of self-training semi-supervised learning, a new unlabeled data trainer was implemented by using minimum trust and maximum uncertainty to estimate the fusion entropy of unlabeled samples.[93] Cooperative training semi-supervised classification method trains multiple models, each model selects the sample with the highest learning confidence, learns from each other and iterates until the set conditions are met. Li et al.[95] proposed that the weak supervised information between labeled samples and unlabeled samples was obtained by collaborative training, and a projection matrix was found to maximize the classification. The representation coefficient of each unlabeled sample was obtained. The separability of classification was improved by using the information of all labeled samples. The graph based semi-supervised model constructs a graph, in which the nodes represent all samples and the edge weights represent the similarity between data points. Then label information of labeled data is propagated to unlabeled data through graph. Shao et al.[96] proposed a method of probabilistic class structure regularized sparse representation graph for semi-supervised learning. By adding class structure information to the sparse representation model, discriminant graph could be learned from data. Taking full account of the spatial information, the semi-supervised spatial and class structure regularized sparse representation was adopted, and the spatial information was introduced into the sparse model through the graph Laplace regularization to improve the discernibility of the graph.[97] Aydemir et al.[98] proposed that the method of subtractive clustering was used to select the initial labeled training samples, which provided the most useful samples for graph based self-training.

At present, the HSI classification algorithm based on semi-supervised learning has made some progress. Wu et al.[99] proposed that the constrained Dirichlet process hybrid model was used to generate high-quality pseudo label data. A deep CRNN was pre trained with a large number of unlabeled data and pseudo labels, and then fine-tuned with limited available label data. Wu et al.[100] proposed a semi supervised local Fisher discriminant analysis (SFDA) based on pseudo label to discriminate labeled samples and unlabeled samples. By learning from the clustering algorithm of Dirichlet process hybrid model, the SFDA was applied to unlabeled data. The proposed semi supervised deep learning framework based on residual network (RESNET) was a new dual strategy sample selection joint training algorithm, which made full use of the complementary clues of spectral and spatial features to guide RESNET to learn from unlabeled data and improve the classification performance.[101] In order to improve the effectiveness of the pseudo labeled samples, the learning super-pixel image and the initial classification image were used to select the pseudo labeled samples·[102] Cui et al.[103] adopted a semi-supervised classification method based on extended label propagation (ELP) and rolling guide filter (RGF). ELP first implemented a graph based label propagation algorithm, and then used super-pixel to correct the labels with error marks. RGF was used to remove noise and small texture structure, and optimize the features of the initial HSI. The initial labeled samples and high confidence pseudo labeled samples were used as training samples.

4.4. Active learning

Active learning is to select the candidate samples with the most information in the unlabeled sample set according to the sampling strategy. Then the candidate samples are added to the training classifier through manual marking. It reduces the cost of obtaining a large number of labeled training samples and the number of training samples. Active learning probability function is applied to extract samples with more uncertainty, so as to provide more information for the model.[104] Active learning is different from semi-supervised learning. Semi-supervised learning algorithm does not need human intervention and uses unlabeled data based on itself. But active learning selects the most useful unlabeled data for classification according to the specific query strategy. According to the query rules, active learning in HIS classification applications is mainly divided into two families: committee-based active learning and posterior probability based active learning.[105]

The committee-based active learning uses entropy other indicators to measure the amount of information in unlabeled samples. Firstly, different kinds of learners are used to label the samples, and then the labeling personnel make the final judgment on the controversial labeling results. The divergence between different learners is caused by the difference of their prediction of sample annotation results. Through the voting mode of multiple models, we can select the sample data which is difficult to distinguish. Committee-based active learning can be adopted by any classifier. Active learning and deep learning have good effect in small sample HSI classification. According to the spatial and spectral similarity of HSI, the active learning method combined with CNN was used to label the unlabeled samples, and the training set was added to improve the classification accuracy of the classifier.[106,107] Mario et al.[104] proposed that a HSI classification algorithm combining Bayesian CNN and active learning was proposed to avoid the over fitting of small data sets and improve its generalization ability. Liu et al.[108] proposed that an additional network was added to the designed deep CNN to predict the loss of input samples, which was different from the traditional active learning method. The additional network could be used to imply that unlabeled samples. Shi et al.[109] proposed a multi-channel image classification framework based on active learning and deep learning. Three active learning algorithms were used as the selection criteria, including minimum confidence, marginal sampling and entropy, and the image pool was introduced to generate the image. Ahmad et al.[110] proposed a fuzzy active learning framework. By estimating the boundary of each class, the fuzzy distance between each sample and the estimated class boundary was calculated to select the optimal sample. Aiming at different types of hyperspectral features, a regularized multi-scale active learning framework was proposed to avoid over fitting when the size of training data was very small.[111] A batch pattern classification strategy combining uncertainty and diversity was used, and k-nearest neighbor classification was combined to enrich the labeled sample set. Ahmad et al.[112] proposed an active learning algorithm based on weighted incremental dictionary learning was proposed. The training samples were selected to maximize the representativeness and uncertainty of the two selection criteria. By actively selecting training samples in each iteration, the deep network was trained effectively.

In posterior probability based active learning, a specific model is used to estimate the posterior probability of all samples in the candidate pool. Then the posterior probability is used to input the following formula to generate the measurement of sample uncertainty. The posterior probability based active learning requires classifiers that can provide posterior probabilities, such as Maximum-Likelihood classifier, linear discriminant analysis classifiers, probabilistic SVM classifier, and multinomial logistic regression.[113] Sun et al.[114] proposed a new framework based on MRF model. In this framework, the unlabeled samples whose predicted results vary before and after the MRF processing step are regarded as uncertain samples. Li et al.[115] proposed that a new supervised Bayesian HSI classification method based on active learning. We use the multinomial logistic regression (MLR) model to learn the class posterior probability distribution. Active learning based on MLR posterior probability can reduce the cost of obtaining large training set. Considering the randomness between the existing samples and the new samples, the concept of spatial prior fuzziness was used.[116] Combining multiple logistic regression, split augmented Lagrange classifier and double stop criterion, active learning method was used to solve the problems.

Active learning combined with other small sample strategy methods can also improve the accuracy of small sample HSI classification. Both active learning and transfer learning promote the training process by selecting unlabeled data or using the knowledge obtained from relevant data. Active learning is an iterative process, which selects a small number of unlabeled samples with the most information through a query function to train a robust classifier. Transfer learning aims to transfer useful knowledge from the source domain to the target domain. Combining the advantages of the two methods, the network can be trained effectively by using only a limited number of labeled samples.[117] Active learning solves problems by improving the quality of training samples, while semi-supervised learning solves problems by increasing the number of training samples. Using semi-supervised active learning to find representativeness and discriminability has the advantages of both active learning and semi supervised learning. Through morphological component analysis, the original hyperspectral data was decomposed into morphological components.[118] In each feature domain, active learning and semi supervised learning were combined to expand the training data set based on super pixel.

4.5. Few-shot learning

Few-shot learning is that uses a small number of labeled samples to train the model to recognize objects, and its purpose is to study the differences between samples. Unlike most other learning methods, it does not directly learn what the sample is, but makes the model learn to learn. In the training stage, C categories and K samples of each category (a total of C*K data) will be randomly selected from the training set to construct a meta task as the support set. Then a batch of samples are selected from the remaining data of the C classes as the batch set. The model is required to learn how to distinguish these C categories from C * K data. Such a task is called C-way K-shot problem. In the process of training, different meta tasks are sampled in each training session. This mechanism enables the model to learn the common parts of different meta tasks. The model learned through this learning mechanism can better classify new and unprecedented meta tasks. Few-shot learning models can be roughly divided into three categories: model based, metric based and optimization based. The model based method aims to update parameters on a small number of samples quickly through the design of model structure, and directly establish the mapping function of input and predictive value. Metric based method measures the distance between the samples in batch and the support set, and classifies them by the idea of nearest neighbor. Optimization based method completes the task of small sample classification by adjusting the optimization method. In HIS classification, metric based method is the most commonly used method.

Metric based method includes siamese network, prototype network, relation network and match network. Siamese Network is that trains a two-way neural network in a supervised way to learn, and then the features extracted from the network are reused for few-shot learning. Huang et al.[119] proposed a dual-path siamese CNN for HSI classification, which was a combination of extended MP, CNN, siamese network, and spectral–spatial feature fusion. It improved the classification performance with limited training samples. The basic idea of prototype network is: there is a prototype expression for each class, and the prototype of this class is the mean value of support set in the embedding space. Then, the classification problem becomes the nearest neighbor in the embedding space. Zhang et al.[120] proposed a vector called global prototypical representation for each class was proposed. The similarity between the unclassified samples and the global prototype representation of each class was evaluated, and the nearest neighbor classifier was used to complete the classification. Chen et al.[121] proposed that the convolutional block attention module was embedded in the convolution blocks of prototypic networks, which improved the feature extraction efficiency. Relation network has a relationship module, which is used to calculate the similarity of samples. Computing similarity is not satisfied with a single and fixed distance measurement, but training a network to learn distance measurement. Gao et al.[122] proposed that a new deep classification model based on relational network to learn the relationship by comparing the similarity between samples. Task based learning strategy could make the model continuously improve the learning ability. It had good generalization ability, and only needed a small number of labeled samples to get satisfactory classification results. Match network uses an attention mechanism, and constructs different encoders for support set and batch set. The final output of classifier is the weighted sum of predicted values between the labeled set of samples (the support set) and the unlabeled samples (the query set). Inspired by the prototype network and match network, a deep few-shot learning method was proposed.[123] A deep residual 3-D convolutional neural network was trained by episodes to learn a metric space where samples from the same class are close and those from different classes are far, the testing samples were classified by a nearest neighbor classifier in the learned metric space.

5. Future research direction

This article summarizes the current mainstream feature extraction methods, classification methods and small sample strategies. The lack of available training samples has always been a difficult and hot point in studying HSI classification problems. To solve the problem of small sample hyperspectral classification, feature extraction methods, classification methods and small sample strategies are complementary to each other. Choosing the appropriate feature extraction method and classification method plays an important role in the effective realization of the small sample strategy.

Since HSI have spatial and spectral information, purely spectral feature extraction or spatial feature extraction can no longer meet the demand. 3D spatial-spectral feature extraction and Feature extraction based on image segmentation combine spatial and spectral information, which can extract useful information. And it also can reduce redundancy and increase the calculation rate, which is the mainstream of feature extraction. In the future, it is still necessary to propose more effective spatial-spectral feature extraction methods for the characteristics of HSI, which will lay the foundation for the classification of small samples of hyperspectral spectrum.

The classification method is the key to the HIS classification, and effective classification methods play a vital role in the results of HSI classification. Classical classification algorithm, kernel-based classification algorithm, representation-based classification algorithm have always occupied an important position in the previous HIS classification. Researchers optimize and improve these algorithms to achieve better classification results. The emergence of deep learning has led to the further development of hyperspectral classification. Deep learning is particularly suitable for large data sets and high-dimensional data sets, and shows high prediction accuracy, so that HSI classification based on deep learning has always been studied Hot spot. However, due to the need to learn a large number of parameters, a large amount of training data is needed, and the effect in small sample classification needs to be improved. At the same time, the problem of deep learning parameter setting and the problem of large amount of calculation also need to be solved in the future.

The emergence of the small sample strategy has alleviated the problem of limited training sample acquisition. Data augmentation is the mainstream of small sample strategy. Most algorithms are improved and optimized based on data augmentation, which is also the focus of future small sample strategy research. Although the current Data augmentation algorithm has increased the number of training samples to a certain extent, it has also greatly increased the training time. Whether the increased training samples can accurately establish a classification model remains to be studied. Therefore, the future research direction is to increase the most effective training sample information as much as possible, while improving the classification accuracy, ensure that the training time is not too long. Other small sample strategies can also improve the classification results under the condition of insufficient training samples, and the combination of multiple methods also improves the classification accuracy to a certain extent. However, the classification results of these small sample strategies in very small samples still need to be improved, and new algorithms are still needed to solve the problem of very small samples in the future.

6. Conclusion

HSI is rich in spectral and spatial information, which has been widely used in resource exploration, ecological environment monitoring, land cover classification and target recognition. This paper introduces the feature extraction methods, classification methods and classification strategy based on small sample of HSI. In the feature extraction method, the spectral feature extraction method, spatial feature extraction method, 3D spatial feature extraction method and graph based segmentation feature extraction method are described in detail. In the classification method, the classical classification method, kernel based classification method, representation based classification method, decision fusion based classification method and deep learning classification method are given a full explanation. In practical hyperspectral applications, supervised classification is the best method to classify, which requires a large number of labeled samples. It is difficult to obtain labeled samples, which limits the wide application of HSI. Therefore, it is of great significance to use small samples for HSI classification. Aiming at the problem of small sample strategy, this paper mainly expounds four kinds of small sample strategies, which are data augmentation, transfer learning, semi-supervised learning and active learning. They solve the problem of small sample in HSI classification from different angles. Small sample strategy will be the focus of HSI classification research in the future. To solve the problem of small sample classification can greatly promote the application of HSI.

Disclosure statement

No potential conflict of interest was reported by the authors

Nomenclature

AE	Auto encoders
AF	Attribute filters
AP	Attribute profiles
CNN	Convolutional neural networks
CRC	Collaborative representation classification
DBN	Deep belief networks
DCNN	Deep convolutional neural networks

DTC	Decision tree classification
DWS	Dual window super-pixel
DWT	Discrete wavelet transform
EAP	Extended attribute profiles
EEMAP	Entire extended multi attribute profile
ELM	Extreme learning machine
EMAP	Extended multi attribute profile
EMP	Extended morphological profiles
GAN	Generative adversarial networks
GCK	Generalized composite kernel
GLCM	Gray level co-occurrence matrix
GRU	Gate recursive unit
HIS	Hyperspectral image
ICA	Independent component analysis
JCR	Joint collaborative representation
KPCA	Kernel principal component analysis
LBP	Local binary pattern
LDA	Linear discriminant analysis
LDE	Local discriminant embedding
LFDA	Local Fisher's discriminant analysis
LSTM	Long short-term memory
MDC	Minimum distance classification
MKL	Multi kernel learning
MLC	Maximum likelihood classification
MLR	Multinomial logistic regression
MNF	Minimum noise fraction
MP	Morphological profiles
MV	Majority voting
Ncut	Normalized cuts method
NWFE	Non-parametric Weighted Feature Extraction
PCA	Principal component analysis
RBF	Radial basis function
RBM	Restricted Boltzmann machine
RF	Random forest
RNN	Recurrent neural networks
SAE	Stacked auto encoders
SAM	Spectral angle mapping
SE	Structural elements
SID	Spectral information divergence
SLIC	Simple linear iterative clustering
SMNF	Segmented maximum noise fraction
SP	Super pixel
SPCA	Segmented principal component analysis
SRC	Sparse representation classification
SVM	Support vector machine

Funding

This work has received funding from the National Natural Science Foundation of China, grant number is 32171578 and U2006209, and the Natural Science Foundation of Shandong Province, grant number is ZR2018LD007 and ZR2017BB037.

References

1. Ng, W.; Malone, B. P.; Minasny, B. Rapid Assessment of Petroleum-Contaminated Soils with Infrared Spectroscopy. *Geoderma* 2017, *289*, 150–160. doi:10.1016/j.geoderma.2016.11.030
2. Jia, S.; Li, H. Y.; Wang, Y. J.; Tong, R. Y.; Li, Q. Recursive Variable Selection to Update near-Infrared Spectroscopy Model for the Determination of Soil Nitrogen and Organic Carbon. *Geoderma* 2016, *268*, 92–99. doi:10.1016/j.geoderma.2016.01.018
3. Jia, S.; Li, H. Y.; Wang, Y. J.; Tong, R. Y.; Li, Q. Hyperspectral Imaging Analysis for the Classification of Soil Types and the Determination of Soil Total Nitrogen. *Sensors* 2017, *17*, 2252. doi:10.3390/s17102252
4. Ghamisi, P.; Yokoya, N.; Li, J.; Liao, W.; Liu, S.; Plaza, J.; Rasti, B.; Plaza, A. Advances in Hyperspectral Image and Signal Processing: A Comprehensive Overview of the State of the Art. *IEEE Geosci. Remote Sens. Mag.* 2017, *5*, 37–78. doi:10.1109/MGRS.2017.2762087
5. Li, S.; Song, W.; Fang, L.; Chen, Y.; Ghamisi, P.; Benediktsson, J. A. Deep Learning for Hyperspectral Image Classification: An Overview. *IEEE Trans. Geosci. Remote Sens.* 2019, *57*, 6690–6709. doi:10.1109/TGRS.2019.2907932
6. Guan, L.; Xie, W.; Pei, J. Segmented Minimum Noise Fraction Transformation for Efficient Feature Extraction of Hyperspectral Images. *Pattern Recogn.* 2015, *48*, 3216–3226.
7. Wang, Z.; Ruan, Q. Q.; An, G. Y. Facial Expression Recognition Using Sparse Local Fisher Discriminant Analysis. *Neurocomputing* 2016, *174*, 756–766. doi:10.1016/j.neucom.2015.09.083
8. Zhao, J.; Zhong, Y. F.; Jia, T. Y.; Wang, X. Y.; Xu, Y.; Shu, H.; Zhang, L. P. Spectral-Spatial Classification of Hyperspectral Imagery with Cooperative Game. *ISPRS J. Photogram.* 2018, *135*, 31–42. doi:10.1016/j.isprsjprs.2017.10.006
9. Dalla Mura, M.; Benediktsson, J. A.; Waske, B.; Bruzzone, L. Morphological Attribute Profiles for the Analysis of Very High Resolution Images. *IEEE Trans. Geosci. Remote Sens.* 2010, *48*, 3747–3762. doi:10.1109/TGRS.2010.2048116
10. Dalla Mura, M.; Benediktsson, J. A.; Waske, B.; Bruzzone, L. Extended Profiles with Morphological Attribute Filters for the Analysis of Hyperspectral Data. *Int. J. Remote Sens.* 2010, *31*, 5975–5991. doi:10.1080/01431161.2010.512425
11. Bhardwaj, K.; Patra, S. An Unsupervised Technique for Optimal Feature Selection in Attribute Profiles for Spectral-Spatial Classification of Hyperspectral Images. *ISPRS J. Photogram.* 2018, *138*, 139–150. doi:10.1016/j.isprsjprs.2018.02.005
12. Li, F.; Wang, J.; Lan, R. S.; Liu, Z. B.; Luo, X. N. Hyperspectral Image Classification Using Multi-Feature Fusion. *Opt. Laser Technol.* 2019, *110*, 176–183. doi:10.1016/j.optlastec.2018.08.044
13. Ghasemzadeh, A.; Demirel, H. 3D Discrete Wavelet Transform-Based Feature Extraction for Hyperspectral Face Recognition. *IET Biom.* 2018, *7*, 49–55. doi:10.1049/iet-bmt.2017.0082
14. Zaouali, M.; Bouzidi, S.; Zagrouba, E. 3-D Shearlet Transform Based Feature Extraction for Improved Joint Sparse Representation HSI Classification. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* 2018, *11*, 1306–1314. doi:10.1109/JSTARS.2018.2792532
15. He, Z.; Wang, Y.; Hu, J. Joint Sparse and Low-Rank Multitask Learning with Laplacian-Like Regularization for Hyperspectral Classification. *Remote Sens.* 2018, *10*, 322. doi:10.3390/rs10020322
16. Jia, S.; Shen, L. L.; Zhu, J. S.; Li, Q. Q. A 3-D Gabor Phase-Based Coding and Matching Framework for Hyperspectral Imagery Classification. *IEEE Trans. Cybern.* 2018, *48*, 1176–1188.
17. Jia, S.; Jie, H.; Zhu, J.; Jia, X.; Li, Q. Three-Dimensional Local Binary Patterns for Hyperspectral Imagery Classification. *IEEE Trans. Geosci. Remote Sens.* 2017, *55*, 2399–2413. doi:10.1109/TGRS.2016.2642951

18. Tsai, F.; Lai, J. S. Feature Extraction of Hyperspectral Image Cubes Using Three-Dimensional Gray-Level Cooccurrence. *IEEE Trans. Geosci. Remote Sens.* 2013, *51*, 3504–3513. doi:10.1109/TGRS.2012.2223704
19. Zhu, J.; Jie, H.; Jia, S.; Jia, X.; Li, Q. Multiple 3-D Feature Fusion Framework for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* 2018, *56*, 1873–1886. doi:10.1109/TGRS.2017.2769113
20. Jia, S.; Wu, K. L.; Zhu, J. S.; Jia, X. P. Spectral-Spatial Gabor Surface Feature Fusion Approach for Hyperspectral Imagery Classification. *IEEE Trans. Geosci. Remote Sens.* 2019, *57*, 1142–1154. doi:10.1109/TGRS.2018.2864983
21. Zhang, S. Z.; Li, S. T.; Fu, W.; Fang, L. Y. Multiscale Superpixel-Based Sparse Representation for Hyperspectral Image Classification. *Remote Sens.* 2017, *9*, 139. doi:10.3390/rs9020139
22. Zhan, T. M.; Lu, Z. Y.; Wan, M. H.; Yang, G. W. Multiscale Superpixel Kernel-Based Low-Rank Representation for Hyperspectral Image Classification. *IEEE Geosci. Remote Sens. Lett.* 2020, *17*, 1642–1646. doi:10.1109/LGRS.2019.2949893
23. Jin, X.; Gu, Y. Superpixel-Based Intrinsic Image Decomposition of Hyperspectral Images. *IEEE Trans. Geosci. Remote Sens.* 2017, *55*, 4285–4295. doi:10.1109/TGRS.2017.2690445
24. Li, S. S.; Ni, L.; Jia, X. P.; Gao, L. R.; Zhang, B.; Peng, M. Multi-Scale Superpixel Spectral-Spatial Classification of Hyperspectral Images. *Int. J. Remote Sens.* 2016, *37*, 4905–4922. doi:10.1080/01431161.2016.1225175
25. Cui, B.; Xie, X.; Ma, X.; Ren, G.; Ma, Y. Superpixel-Based Extended Random Walker for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* 2018, *56*, 3233–3243. doi:10.1109/TGRS.2018.2796069
26. Jia, S.; Deng, X.; Xu, M.; Zhou, J.; Jia, X. Superpixel-Level Weighted Label Propagation for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* 2020, *58*, 5077–5091. doi:10.1109/TGRS.2020.2972294
27. Lu, T.; Li, S.; Fang, L.; Jia, X.; Benediktsson, J. A. From Subpixel to Superpixel: A Novel Fusion Framework for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* 2017, *55*, 4398–4411. doi:10.1109/TGRS.2017.2691906
28. Liu, T.; Gu, Y.; Chanussot, J.; Mura, M. D. Multimorphological Superpixel Model for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* 2017, *55*, 6950–6963. doi:10.1109/TGRS.2017.2737037
29. Kayabol, K. Approximate Sparse Multinomial Logistic Regression for Classification. *IEEE Trans. Pattern Anal. Mach Intell.* 2020, *42*, 490–493. doi:10.1109/TPAMI.2019.2904062
30. Cheng, G.; Li, Z.; Han, J.; Yao, X.; Guo, L. Exploring Hierarchical Convolutional Features for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* 2018, *56*, 5712–6722.
31. Castillo, C.; Chollett, I.; Klein, E. Enhanced Duckweed Detection Using Bootstrapped SVM Classification on Medium Resolution RGB MODIS Imagery. *Int. J. Remote Sens.* 2008, *29*, 5595–5604. doi:10.1080/01431160801961375
32. Xia, J. S.; Chanussot, J.; Du, P. J.; He, X. Y. Rotation-Based Support Vector Machine Ensemble in Classification of Hyperspectral Data with Limited Training Samples. *IEEE Trans. Geosci. Remote Sens.* 2016, *54*, 1519–1531. doi:10.1109/TGRS.2015.2481938
33. Xian, G.; Xin, H.; Zhang, L. F.; Zhang, L. P.; Plaza, A.; Benediktsson, J. A. Support Tensor Machines for Classification of Hyperspectral Remote Sensing Imagery. *IEEE Trans. Geosci. Remote Sens.* 2016, *54*, 3248–3264.
34. Gu, Y.; Liu, T.; Jia, X.; Benediktsson, J. A.; Chanussot, J. Nonlinear Multiple Kernel Learning with Multiple-Structure-Element Extended Morphological Profiles for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* 2016, *54*, 3235–3247. doi:10.1109/TGRS.2015.2514161
35. Wang, Q. W.; Gu, Y. F.; Tuia, D. Discriminative Multiple Kernel Learning for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* 2016, *54*, 3912–3927. doi:10.1109/TGRS.2016.2530807

36. Li, J.; Marpu, P. R.; Plaza, A.; Bioucas-Dias, J. M.; Benediktsson, J. A. Generalized Composite Kernel Framework for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* 2013, *51*, 4816–4829. doi:10.1109/TGRS.2012.2230268
37. Zhang, H.; Li, J.; Huang, Y.; Zhang, L. A Nonlocal Weighted Joint Sparse Representation Classification Method for Hyperspectral Imagery. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sens.* 2014, *7*, 2056–2065. doi:10.1109/JSTARS.2013.2264720
38. Gan, L.; Xia, J.; Du, P.; Chanussot, J. Class-Oriented Weighted Kernel Sparse Representation with Region-Level Kernel for Hyperspectral Imagery Classification. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* 2018, *11*, 1118–1130. doi:10.1109/JSTARS.2017.2757475
39. Sun, H.; Ren, J.; Zhao, H.; Yan, Y.; Zabalza, J.; Marshall, S. Superpixel Based Feature Specific Sparse Representation for Spectral-Spatial Classification of Hyperspectral Images. *Remote Sens.* 2019, *11*, 536. doi:10.3390/rs11050536
40. Li, C.; Ma, Y.; Mei, X. G.; Liu, C. Y.; Ma, J. Y. Hyperspectral Image Classification with Robust Sparse Representation. *IEEE Geosci. Remote Sensing Lett.* 2016, *13*, 641–645. doi:10.1109/LGRS.2016.2532380
41. Huang, S.; Zhang, H.; Aleksandra, P. A Robust Sparse Representation Model for Hyperspectral Image Classification. *Sensors* 2017, *17*, 2087. doi:10.3390/s17092087
42. Xu, Y.; Du, Q.; Li, W.; Younan, N. H. Efficient Probabilistic Collaborative Representation-Based Classifier for Hyperspectral Image Classification. *IEEE Geosci. Remote Sens. Lett.* 2019, *16*, 1746–1750. doi:10.1109/LGRS.2019.2906839
43. Jia, S.; Deng, X.; Zhu, J.; Xu, M.; Zhou, J.; Jia, X. Collaborative Representation-Based Multiscale Superpixel Fusion for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* 2019, *57*, 7770–7784. doi:10.1109/TGRS.2019.2916329
44. Du, P.; Le, G.; Xia, J.; Wang, D. Multikernel Adaptive Collaborative Representation for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* 2018, *56*, 4664–4677. doi:10.1109/TGRS.2018.2833882
45. Li, W.; Du, Q.; Zhang, F.; Hu, W. Hyperspectral Image Classification by Fusing Collaborative and Sparse Representations. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sens.* 2016, *9*, 4178–4187. doi:10.1109/JSTARS.2016.2542113
46. Imani, M.; Ghassemian, H. Discriminant Analysis in Morphological Feature Space for High-Dimensional Image Spatial–Spectral Classification. *J. Appl. Rem. Sens.* 2018, *12*, 1. doi:10.1117/1.JRS.12.016024
47.. Priya, T.; Prasad, S.; Wu, H. Superpixels for Spatially Reinforced Bayesian Classification of Hyperspectral Images. *IEEE Geosci. Remote. Sens. Lett.* 2015, *12*, 1071–1075.
48. Guo, B. F.; Shen, H. H.; Yang, M. Y. Improving Hyperspectral Image Classification by Fusing Spectra and Absorption Features. *IEEE Geosci. Remote Sens. Lett.* 2017, *14*, 1363–1367. doi:10.1109/LGRS.2017.2712805
49. Guo, B. Entropy-Mediated Decision Fusion for Remotely Sensed Image Classification. *Remote Sens.* 2019, *11*, 352. doi:10.3390/rs11030352
50. Bo, C.; Lu, H.; Wang, D. Hyperspectral Image Classification via JCR and SVM Models with Decision Fusion. *IEEE Geosci. Remote. Sens. Lett.* 2016, *13*, 177–181.
51. Ye, Z.; Prasad, S.; Li, W.; Fowler, J. E.; He, M. Classification Based on 3-D DWT and Decision Fusion for Hyperspectral Image Analysis. *IEEE Geosci. Remote Sens. Lett.* 2014, *11*, 173–177. doi:10.1109/LGRS.2013.2251316
52. Li, W.; Chen, C.; Su, H. J.; Du, Q. Local Binary Patterns and Extreme Learning Machine for Hyperspectral Imagery Classification. *IEEE Trans. Geosci. Remote Sens.* 2015, *53*, 3681–3693. doi:10.1109/TGRS.2014.2381602
53. Liu, M. Z.; Cao, F. X.; Yang, Z. J.; Hong, X. B.; Huang, Y. Z. Hyperspectral Image Denoising and Classification Using Multi-Scale Weighted EMAPs and Extreme Learning Machine. *Electronics-Switz* 2020, *9*, 2137. doi:10.3390/electronics9122137
54. Hu, W.; Huang, Y. Y.; Wei, L.; Zhang, F.; Li, H. C. Deep Convolutional Neural Networks for Hyperspectral Image Classification. *J. Sens.* 2015, *2015*, 1–12. doi:10.1155/2015/258619

55. Qing, Y. H.; Liu, W. Y.; Feng, L. Y.; Gao, W. J. Improved Transformer Net for Hyperspectral Image Classification. *Remote Sens.* 2021, *13*, 2216. doi:10.3390/rs13112216
56. Yang, X.; Ye, Y.; Li, X.; Lau, R. Y. K.; Zhang, X.; Huang, X. Hyperspectral Image Classification with Deep Learning Models. *IEEE Trans. Geosci. Remote Sens.* 2018, *56*, 5408–5423. doi:10.1109/TGRS.2018.2815613
57. Cai, Y.; Liu, X.; Cai, Z. BS-Nets: An End-to-End Framework for Band Selection of Hyperspectral Image. *IEEE Trans. Geosci. Remote Sens.* 2020, *58*, 1969–1984. doi:10.1109/TGRS.2019.2951433
58. Zhang, H. K.; Li, Y.; Zhang, Y. Z.; Shen, Q. Spectral-Spatial Classification of Hyperspectral Imagery Using a Dual-Channel Convolutional Neural Network. *Remote Sens. Lett.* 2017, *8*, 438–447. doi:10.1080/2150704X.2017.1280200
59. Li, X.; Ding, M.; Pizurica, A. Deep Feature Fusion via Two-Stream Convolutional Neural Network for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* 2020, *58*, 2615–2629. doi:10.1109/TGRS.2019.2952758
60. Li, Y.; Zhang, H. K.; Shen, Q. Spectral–Spatial Classification of Hyperspectral Imagery with 3D Convolutional Neural Network. *Remote Sens.* 2017, *9*, 67. doi:10.3390/rs9010067
61. Roy, S. K.; Krishna, G.; Dubey, S. R.; Chaudhuri, B. B. HybridSN: Exploring 3D-2D CNN Feature Hierarchy for Hyperspectral Image Classification. *IEEE Geosci. Remote Sens. Lett.* 2020, *17*, 277–281. doi:10.1109/LGRS.2019.2918719
62. Zhang, M. M.; Li, W.; Du, Q. Diverse Region-Based CNN for Hyperspectral Image Classification. *IEEE Trans. Image Process.* 2018, *27*, 2623–2634. doi:10.1109/TIP.2018.2809606
63. Song, W.; Li, S.; Fang, L.; Lu, T. Hyperspectral Image Classification with Deep Feature Fusion Network. *IEEE Trans. Geosci. Remote Sens.* 2018, *56*, 3173–3184. doi:10.1109/TGRS.2018.2794326
64. Pan, E. T.; Mei, X. G.; Wang, Q. D.; Ma, Y.; Ma, J. Y. Spectral-Spatial Classification for Hyperspectral Image Based on a Single GRU. *Neurocomputing* 2020, *387*, 150–160. doi:10.1016/j.neucom.2020.01.029
65. Sun, G. Y.; Zhang, X. M.; Jia, X. P.; Ren, J. C.; Zhang, A. Z.; Yao, Y. J.; Zhao, H. M. Deep Fusion of Localized Spectral Features and Multi-Scale Spatial Features for Effective Classification of Hyperspectral Images. *Int. J. Appl. Earth Obs.* 2020, *91*, 102157. doi:10.1016/j.jag.2020.102157
66. Qing, Y. H.; Liu, W. Y. Hyperspectral Image Classification Based on Multi-Scale Residual Network with Attention Mechanism. *Remote Sens.* 2021, *13*, 335. doi:10.3390/rs13030335
67. Hang, R. L.; Li, Z.; Liu, Q. S.; Ghamisi, P.; Bhattacharyya, S. S. Hyperspectral Image Classification with Attention Aided CNNs. *IEEE Trans. Geosci. Remote Sens.* 2021, *59*, 2281–2293. doi:10.1109/TGRS.2020.3007921
68. Qu, L.; Zhu, X. L.; Zheng, J. N.; Zou, L. Triple-Attention-Based Parallel Network for Hyperspectral Image Classification. *Remote Sens.* 2021, *13*, 324. doi:10.3390/rs13020324
69. Aptoula, E.; Ozdemir, M. C.; Yanikoglu, B. Deep Learning with Attribute Profiles for Hyperspectral Image Classification. *IEEE Geosci. Remote Sens. Lett.* 2016, *13*, 1970–1974. doi:10.1109/LGRS.2016.2619354
70. Accion, A.; Arguello, F.; Heras, D. B. Dual-Window Superpixel Data Augmentation for Hyperspectral Image Classification. *Appl. Sci-Basel* 2020, *10*, 8833. doi:10.3390/app10248833
71. Wang, C.; Zhang, L.; Wei, W.; Zhang, Y. Hyperspectral Image Classification with Data Augmentation and Classifier Fusion. *IEEE Geosci. Remote Sens. Lett.* 2020, *17*, 1420–1424. doi:10.1109/LGRS.2019.2945848
72. Li, W.; Wu, G. D.; Zhang, F.; Du, Q. Hyperspectral Image Classification Using Deep Pixel-Pair Features. *IEEE Trans. Geosci. Remote Sens.* 2017, *55*, 844–853. doi:10.1109/TGRS.2016.2616355
73. Li, W.; Chen, C.; Zhang, M. M.; Li, H. C.; Du, Q. Data Augmentation for Hyperspectral Image Classification with Deep CNN. *IEEE Geosci. Remote Sens. Lett.* 2019, *16*, 593–597. doi:10.1109/LGRS.2018.2878773

74. Hao, Q. B.; Li, S. T.; Kang, X. D. Multilabel Sample Augmentation-Based Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* 2020, *58*, 4263–4278. doi:10.1109/TGRS.2019.2962014
75. Wang, W. N.; Liu, X. B.; Mou, X. Q. Data Augmentation and Spectral Structure Features for Limited Samples HyperspectralClassification. *Remote Sens.* 2021, *13*, 547. doi:10.3390/rs13040547
76. Feng, J.; Feng, X. L.; Chen, J. T.; Cao, X. H.; Zhang, X. R.; Jiao, L. C.; Yu, T. Generative Adversarial Networks Based on Collaborative Learning and Attention Mechanism for Hyperspectral Image Classification. *Remote Sens.* 2020, *12*, 1149.
77. Wang, W.-Y.; Li, H.-C.; Deng, Y.-J.; Shao, L.-Y.; Lu, X.-Q.; Du, Q. Generative Adversarial Capsule Network with Convlstm for Hyperspectral Image Classification. *IEEE Geosci. Remote Sens. Lett.* 2021, *18*, 523–527. doi:10.1109/LGRS.2020.2976482
78. Zhao, W. Z.; Chen, X. H.; Bo, Y. C.; Chen, J. G. Semisupervised Hyperspectral Image Classification with Cluster-Based Conditional Generative Adversarial Net. *IEEE Geosci. Remote Sens. Lett.* 2020, *17*, 539–543. doi:10.1109/LGRS.2019.2924059
79. Zhao, W. Z.; Chen, X.; Chen, J. G.; Qu, Y. Sample Generation with Self-Attention Generative Adversarial Adaptation Network (SaGAAN) for Hyperspectral Image Classification. *Remote Sens.* 2020, *12*, 843. doi:10.3390/rs12050843
80. Yang, J.; Zhao, Y. Q.; Chan, C. W. Learning and Transferring Deep Joint Spectral-Spatial Features for Hyperspectral Classification. *IEEE Trans. Geosci. Remote Sens.* 2017, *55*, 4729–4742. doi:10.1109/TGRS.2017.2698503
81. Jiang, X. F.; Zhang, Y.; Li, Y.; Li, S. Y.; Zhang, Y. N. Hyperspectral Image Classification with Transfer Learning and Markov Random Fields. *IEEE Geosci. Remote Sensing Lett.* 2020, *17*, 544–548. doi:10.1109/LGRS.2019.2923647
82. He, X.; Chen, Y. S. Transferring CNN Ensemble for Hyperspectral Image Classification. *IEEE Geosci. Remote Sens. Lett.* 2021, *18*, 876–880. doi:10.1109/LGRS.2020.2988494
83. Li, X.; Zhang, L. P.; Du, B.; Zhang, L. F.; Shi, Q. Iterative Reweighting Heterogeneous Transfer Learning Framework for Supervised Remote Sensing Image Classification. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* 2017, *10*, 2022–2035. doi:10.1109/JSTARS.2016.2646138
84. Zhao, X.; Liang, Y.; Guo, A. J. X.; Zhu, F. Classification of Small-Scale Hyperspectral Images with Multi-Source Deep Transfer Learning. *Remote Sens. Lett.* 2020, *11*, 303–312. doi:10.1080/2150704X.2020.1714772
85. Li, K.; Wang, M. J.; Liu, Y. X.; Yu, N.; Lan, W. A Novel Method of Hyperspectral Data Classification Based on Transfer Learning and Deep Belief Network. *Appl. Sci* 2019, *9*, 1379. doi:10.3390/app9071379
86. Liu, B.; Yu, X. C.; Yu, A. Z.; Wan, G. Deep Convolutional Recurrent Neural Network with Transfer Learning for Hyperspectral Image Classification. *J. Appl. Rem. Sens.* 2018, *12*, 1. doi:10.1117/1.JRS.12.026028
87. Xie, F. D.; Gao, Q. S.; Jin, C.; Zhao, F. X. Hyperspectral Image Classification Based on Superpixel Pooling Convolutional Neural Network with Transfer Learning. *Remote Sens.* 2021, *13*, 930. doi:10.3390/rs13050930
88. Liu, X. F.; Sun, Q. Q.; Meng, Y.; Fu, M.; Bourennane, S. Hyperspectral Image Classification Based on Parameter-Optimized 3D-CNNs Combined with Transfer Learning and Virtual Samples. *Remote Sens.* 2018, *10*, 1425. doi:10.3390/rs10091425
89. He, X.; Chen, Y. S.; Ghamisi, P. Heterogeneous Transfer Learning for Hyperspectral Image Classification Based on Convolutional Neural Network. *IEEE Trans. Geosci. Remote Sens.* 2020, *58*, 3246–3263. doi:10.1109/TGRS.2019.2951445
90. Windrim, L.; Melkumyan, A.; Murphy, R. J.; Chlingaryan, A.; Ramakrishnan, R. Pretraining for Hyperspectral Convolutional Neural Network Classification. *IEEE Trans. Geosci. Remote Sens.* 2018, *56*, 2798–2810. doi:10.1109/TGRS.2017.2783886
91. Zhang, H. K.; Li, Y.; Jiang, Y. A.; Wang, P.; Shen, Q.; Shen, C. H. Hyperspectral Classification Based on Lightweight 3-D-CNN with Transfer Learning. *IEEE Trans. Geosci. Remote Sens.* 2019, *57*, 5813–5828.

92. Jiang, Y. N.; Li, Y.; Zhang, H. K. Hyperspectral Image Classification Based on 3-D Separable ResNet and Transfer Learning. *IEEE Geosci. Remote Sens. Lett.* 2019, *16*, 1949–1953. doi:10.1109/LGRS.2019.2913011
93. Wang, C. Y.; Xu, Z. F.; Wang, S. T.; Zhang, H. B. Semi-Supervised Classification Framework of Hyperspectral Images Based on the Fusion Evidence Entropy. *Multimed. Tools Appl.* 2018, *77*, 10615–10633. doi:10.1007/s11042-017-4686-x
94. Fox-Roberts, P.; Rosten, E. Unbiased Generative Semi-Supervised Learning. *J. Mach. Learn. Res.* 2014, *15*, 367–443.
95. Li, F. D.; Lv, M.; Jing, L. Collaborative Representation-Based Semisupervised Feature Extraction of Hyperspectral Images Using Attraction Points. *J. Appl. Remote Sens.* 2020, *14*, 026505.
96. Shao, Y. J.; Sang, N.; Gao, C. X.; Ma, L. Probabilistic Class Structure Regularized Sparse Representation Graph for Semi-Supervised Hyperspectral Image Classification. *Pattern Recognit.* 2017, *63*, 102–114. doi:10.1016/j.patcog.2016.09.011
97. Shao, Y. J.; Sang, N.; Gao, C. X.; Ma, L. Spatial and Class Structure Regularized Sparse Representation Graph for Semi-Supervised Hyperspectral Image Classification. *Pattern Recognit.* 2018, *81*, 81–94. doi:10.1016/j.patcog.2018.03.027
98. Aydemir, M. S.; Bilgin, G. Semisupervised Hyperspectral Image Classification Using Small Sample Sizes. *IEEE Geosci. Remote Sens. Lett.* 2017, *14*, 621–625. doi:10.1109/LGRS.2017.2665679
99. Wu, H.; Prasad, S. Semi-Supervised Deep Learning Using Pseudo Labels for Hyperspectral Image Classification. *IEEE Trans Image Process* 2018, *27*, 1259–1270. doi:10.1109/TIP.2017.2772836
100. Wu, H.; Prasad, S. Semi-Supervised Dimensionality Reduction of Hyperspectral Imagery Using Pseudo-Labels. *Pattern Recognit.* 2018, *74*, 212–224. doi:10.1016/j.patcog.2017.09.003
101. Fang, B.; Li, Y.; Zhang, H. K.; Chan, J. C. W. Semi-Supervised Deep Learning Classification for Hyperspectral Image Based on Dual-Strategy Sample Selection. *Remote Sens.* 2018, *10*, 574. doi:10.3390/rs10040574
102. Zhang, Y. X.; Liu, K.; Dong, Y. N.; Wu, K.; Hu, X. Y. Semisupervised Classification Based on SLIC Segmentation for Hyperspectral Image. *IEEE Geosci. Remote Sens. Lett.* 2020, *17*, 1440–1444. doi:10.1109/LGRS.2019.2945546
103. Cui, B. E.; Xie, X. Y.; Hao, S. Y.; Cui, J. D.; Lu, Y. Semi-Supervised Classification of Hyperspectral Images Based on Extended Label Propagation and Rolling Guidance Filtering. *Remote Sens.* 2018, *10*, 515. doi:10.3390/rs10040515
104. Mario, H. J.; Paoletti, M. E.; Plaza, J.; Li, J.; Plaza, A. Active Learning with Convolutional Neural Networks for Hyperspectral Image Classification Using a New Bayesian Approach. *IEEE Trans. Geosci. Remote Sens.* 2018, *56*, 6440–6461.
105. Jia, S.; Jiang, S. G.; Lin, Z. J.; Li, N. Y.; Xu, M.; Yu, S. Q. A Survey: deep Learning for Hyperspectral Image Classification with Few Labeled Samples. *Neurocomputing* 2021, *448*, 179–204. doi:10.1016/j.neucom.2021.03.035
106. Cao, X. Y.; Yao, J.; Xu, Z. B.; Meng, D. Y. Hyperspectral Image Classification with Convolutional Neural Network and Active Learning. *IEEE Trans. Geosci. Remote Sens.* 2020, *58*, 4604–4616. doi:10.1109/TGRS.2020.2964627
107. Cao, X. Y.; Yao, J.; Fu, X. Y.; Bi, H. X.; Hong, D. F. An Enhanced 3-d Discrete Wavelet Transform for Hyperspectral Image Classification. *IEEE Geosci. Remote Sens. Lett.* 2021, *18*, 1104–1108. doi:10.1109/LGRS.2020.2990407
108. Liu, B.; Yu, A. Z.; Zhang, P. Q.; Ding, L.; Guo, W. Y.; Gao, K. L.; Zou, X. B. Active Deep Densely Connected Convolutional Network for Hyperspectral Image Classification. *Int. J. Remote Sens.* 2021, *42*, 5905–5924.
109. Shi, F. Y.; Wang, Z. D.; Hu, M. H.; Zhai, G. T. Active Learning plus Deep Learning Can Establish Cost-Effective and Robust Model for Multichannel Image: A Case on Hyperspectral Image Classification. *Sensors* 2020, *20*, 4975. doi:10.3390/s20174975
110. Ahmad, M.; Protasov, S.; Khan, A. M.; Hussain, R.; Khattak, A. M.; Khan, W. A. Fuzziness-Based Active Learning Framework to Enhance Hyperspectral Image

Classification Performance for Discriminative and Generative Classifiers. *Plos One*. 2018, *13*, e0188996. doi:10.1371/journal.pone.0188996
111. Zhang, Z.; Crawford, M. M. A Batch-Mode Regularized Multimetric Active Learning Framework for Classification of Hyperspectral Images. *IEEE Trans. Geosci. Remote Sens.* 2017, *55*, 6594–6609. doi:10.1109/TGRS.2017.2730583
112. Ahmad, M.; Mazzara, M.; Raza, R. A.; Distefano, S.; Asif, M.; Sarfraz, M. S.; Khan, A. M.; Sohaib, A. Multiclass Non-Randomized Spectral–Spatial Active Learning for Hyperspectral Image Classification. *Appl. Sci.* 2020, *10*, 4739. doi:10.3390/app10144739
113. Saboori, A.; Ghassemian, H.; Razzazi, F. Active Multiple Kernel Fredholm Learning for Hyperspectral Images Classification. *IEEE Geosci. Remote Sens. Lett.* 2021, *18*, 356–360. doi:10.1109/LGRS.2020.2969970
114. Sun, S. J.; Zhong, P.; Xiao, H. T.; Wang, R. S. An MRF Model-Based Active Learning Framework for the Spectral–Spatial Classification of Hyperspectral Imagery. *IEEE J. Sel. Top. Signal Process.* 2015, *9*, 1074–1088. doi:10.1109/JSTSP.2015.2414401
115. Li, J.; Bioucas-Dias, J. M.; Plaza, A. Hyperspectral Image Segmentation Using a New Bayesian Approach with Active Learning. *IEEE Trans. Geosci. Remote Sens.* 2011, *49*, 3947–3960. doi:10.1109/TGRS.2011.2128330
116. Liu, P.; Zhang, H.; Eom, K. B. Active Deep Learning for Classification of Hyperspectral Images. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* 2017, *10*, 712–724. doi:10.1109/JSTARS.2016.2598859
117. Deng, C.; Xue, Y. M.; Liu, X. L.; Li, C.; Tao, D. C. Active Transfer Learning Network: A Unified Deep Joint Spectral-Spatial Feature Learning Model for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* 2019, *57*, 1741–1754. doi:10.1109/TGRS.2018.2868851
118. Wang, Z. M.; Du, B.; Zhang, L. F.; Zhang, L. P.; Jia, X. P. A Novel Semisupervised Active-Learning Algorithm for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* 2017, *55*, 3071–3083. doi:10.1109/TGRS.2017.2650938
119. Huang, L. B.; Chen, Y. S. Dual-Path Siamese CNN for Hyperspectral Image Classification with Limited Training Samples. *IEEE Geosci. Remote Sens. Lett.* 2021, *18*, 518–522. doi:10.1109/LGRS.2020.2979604
120. Zhang, C. Y.; Yue, J.; Qin, Q. Global Prototypical Network for Few-Shot Hyperspectral Image Classification. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sens.* 2020, *13*, 4748–4759. doi:10.1109/JSTARS.2020.3017544
121. Chen, L.; Tian, X. M.; Chai, G. Q.; Zhang, X. L.; Chen, E. R. A New CBAM-P-Net Model for Few-Shot Forest Species Classification Using Airborne HyperspectralImages. *Remote Sens* 2021, *13*, 1269. doi:10.3390/rs13071269
122. Gao, K. L.; Liu, B.; Yu, X. C.; Qin, J. C.; Zhang, P. Q.; Tan, X. Deep Relation Network for Hyperspectral Image Few-Shot Classification. *Remote Sens.* 2020, *12*, 923. doi:10.3390/rs12060923
123. Liu, B.; Yu, X. C.; Yu, A. Z.; Zhang, P. Q.; Wan, G.; Wang, R. R. Deep Few-Shot Learning for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* 2019, *57*, 2290–2304. doi:10.1109/TGRS.2018.2872830